Б. САЛЕХ, М. ТЕЙХ

ОПТИКА И ФОТОНИКА
ПРИНЦИПЫ И ПРИМЕНЕНИЯ

Том 2

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. 2 / Б. Салех, М. Тейх – Долгопрудный: Издательский Дом «Интеллект», 2012. – 784 с.: цв. вкл.

ISBN 978-5-91559-135-5

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.

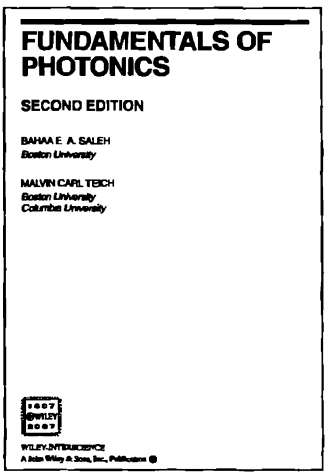
Этот термин отражает квантовую (фотонную) природу света и включает широкий круг физических явлений, методов и устройств, используемых для генерации света, управления его свойствами, передачи, регистрации, воздействия светом на вещество и оптической диагностики материальных сред. В учебной литературе 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 978-5-91559-135-5
ISBN 978-0-4713-5832-9 (англ.)

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

ОГЛАВЛЕНИЕ

Глава 15
ЛАЗЕРЫ 9

15.1. Теория лазерной генерации 11
15.1.1. Оптическое усиление и обратная связь 11
15.1.2. Условия лазерной генерации 15
15.2. Выходные характеристики лазера 19
15.2.1. Мощность 19
15.2.2. Спектральное распределение 25
15.2.3. Пространственное распределение и поляризация 31
15.2.4. Селекция мод 34
15.3. Распространенные лазеры 37
15.3.1. Твердотельные лазеры 38
15.3.2. Газовые лазеры 51
15.3.3. Другие лазеры 53
15.3.4. Таблица характеристик 58
15.4. Импульсные лазеры 58
15.4.1. Методы получения импульсной генерации 61
*15.4.2. Анализ переходных эффектов 64
*15.4.3. Модуляция добротности 66
15.4.4. Синхронизация мод 72

Глава 16
ОПТИКА ПОЛУПРОВОДНИКОВ 87

16.1. Полупроводники 88
16.1.1. Энергетические зоны и носители заряда 89
16.1.2. Полупроводниковые материалы 94
16.1.3. Концентрации электронов и дырок 101
16.1.4. Генерация, рекомбинация и инжекция 110
16.1.5. Переходы 115
16.1.6. Гетеропереходы 120
16.1.7. Квантово-размерные структуры 122

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

16.2. Взаимодействие фотонов с носителями заряда 130
16.2.1. Взаимодействие фотонов с объемными полупроводниками 130
16.2.2. Межзонные переходы в объемных полупроводниках 131
16.2.3. Поглощение, испускание и усиление в объемных полупроводниках 136
16.2.4. Взаимодействие фотонов с квантово-размерными структурами 143
16.2.5. Показатель преломления 145

Глава 17
ПОЛУПРОВОДНИКОВЫЕ ИСТОЧНИКИ ФОТОНОВ 152

17.1. Светоизлучающие диоды 154
17.1.1. Инжекционная электролюминесценция 154
17.1.2. Характеристики СИД 161
17.1.3. Материалы и структуры устройств 172
17.2. Полупроводниковые оптические усилители 180
17.2.1. Усиление и ширина полосы 181
17.2.2. Накачка 187
17.2.3. Гетероструктуры 190
17.2.4. Структуры с квантовыми ямами 192
17.2.5. Сверхлюминесцентные диоды 197
17.3. Лазерные диоды 198
17.3.1. Усиление, обратная связь и генерация 198
17.3.2. Мощность и коэффициент преобразования 203
17.3.3. Спектральные и пространственные характеристики 209
17.4. Квантово-размерные лазеры и лазеры с микрорезонаторами 214
17.4.1. Квантово-размерные лазеры 214
17.4.2. Лазеры с микрорезонаторами 222
17.4.3. Материалы и структуры устройств 225

Глава 18
ПОЛУПРОВОДНИКОВЫЕ ДЕТЕКТОРЫ ФОТОНОВ 241

18.1. Фотоприемники 242
18.1.1. Внешний и внутренний фотоэффект 242
18.1.2. Общие свойства 246
18.2. Фотопроводники 254
18.2.1. Собственные полупроводники 255
18.2.2. Примесные материалы 258
18.2.3. Гетероструктуры 259
18.3. Фотодиоды 260
18.3.1. *p*—*n*-Фотодиод 260
18.3.2. *p*—*i*—*n*-Фотодиод 263
18.3.3. Гетероструктуры 264
18.4. Лавинные фотодиоды 267

18.4.1. Принципы действия ... 267
18.4.2. Усиление и токовая чувствительность ... 270
18.4.3. Время отклика ... 274
18.4.4. Лавинные диоды для регистрации одиночных фотонов (SPAD) ... 276
18.5. Матричные детекторы ... 277
18.6. Шум в фотодетекторах ... 279
18.6.1. Фотоэлектронный шум ... 281
18.6.2. Шум усиления ... 286
18.6.3. Шум схемы ... 292
18.6.4. Отношение сигнал—шум и обнаружительная способность приемника 296
18.6.5. Частота появления ошибочных битов и обнаружительная способность приемника ... 302

Глава 19
АКУСТООПТИКА ... 316

19.1. Взаимодействие света и звука ... 318
19.1.1. Дифракция Брэгга ... 319
*19.1.2. Теория связанных волн ... 328
19.1.3. Брэгговская дифракция пучков ... 331
19.2. Акустооптические устройства ... 335
19.2.1. Модуляторы ... 335
19.2.2. Сканеры ... 338
19.2.3. Пространственные переключатели ... 341
19.2.4. Фильтры, преобразователи частоты и вентили ... 345
*19.3. Акустооптика анизотропных сред ... 346

Глава 20
ЭЛЕКТРООПТИКА ... 355

20.1. Принципы электрооптики ... 357
20.1.1. Эффекты Поккельса и Керра ... 357
20.1.2. Электрооптические модуляторы и переключатели ... 359
20.1.3. Сканеры ... 366
20.1.4. Направленные ответвители ... 367
20.1.5. Пространственные модуляторы света ... 371
*20.2. Электрооптика анизотропных сред ... 374
20.2.1. Эффекты Поккельса и Керра ... 375
20.2.2. Модуляторы ... 383
20.3. Электрооптика жидких кристаллов ... 385
20.3.1. Фазовые пластинки и модуляторы ... 385
20.3.2. Пространственные модуляторы света ... 392
*20.4. Фоторефрактивность ... 395
20.5. Электропоглощение ... 401

Глава 21
НЕЛИНЕЙНАЯ ОПТИКА 409

21.1. Нелинейные оптические среды 411
21.2. Нелинейная оптика второго порядка 416
21.2.1. Генерация второй гармоники (ГВГ) и оптическое выпрямление 416
21.2.2. Электрооптический эффект 419
21.2.3. Трехволновое смешение 421
21.2.4. Фазовый синхронизм и кривые настройки 426
21.2.5. Квазисинхронизм 432
21.3. Нелинейная оптика третьего порядка 435
21.3.1. Генерация третьей гармоники (ГТГ) и оптический эффект Керра 436
21.3.2. Самомодуляция фазы (СМФ), самофокусировка и пространственные солитоны 437
21.3.3. Фазовая кросс-модуляция (ФКМ) 442
21.3.4. Четырехволновое смешение (ЧВС) 443
21.3.5. Обращение волнового фронта (ОВФ) 446
*21.4. Нелинейная оптика второго порядка: теория связанных волн 450
21.4.1. Генерация второй гармоники (ГВГ) 455
21.4.2. Преобразование оптической частоты (ПОЧ) 459
21.4.3. Параметрическое усиление (ПУ) и параметрическая генерация (ПГ) света 460
*21.5. Нелинейная оптика третьего порядка: теория связанных волн 464
21.5.1. Четырехволновое смешение (ЧВС) 464
21.5.2. Трехволновое смешение и генерация третьей гармоники (ГТГ) 467
21.5.3. Обращение волнового фронта (ОВФ) 470
*21.6. Анизотропные нелинейные среды 472
*21.7. Нелинейные среды с дисперсией 477

Глава 22
ОПТИКА СВЕРХБЫСТРЫХ ПРОЦЕССОВ 489

22.1. Характеристики импульсов 490
22.1.1. Временны́е и спектральные характеристики 490
22.1.2. Гауссовы импульсы и гауссовы импульсы с чирпом 497
22.1.3. Пространственные характеристики 500
22.2. Формирование и компрессия импульсов 504
22.2.1. Фильтры с чирпом 505
22.2.2. Осуществление фильтрации с изменением чирпа 515
22.2.3. Сжатие импульсов 521
22.2.4. Формирование импульсов 523

22.3. Распространение импульсов в оптических волноводах ... 526
22.3.1. Оптическое волокно как фильтр с чирпом ... 526
22.3.2. Распространение гауссова импульса в оптическом волокне ... 530
*22.3.3. Уравнение диффузии для медленной огибающей ... 538
*22.3.4. Аналогия между дисперсией и дифракцией ... 541
22.4. Линейная оптика ультракоротких импульсов ... 546
22.4.1. Оптика лучей ... 546
*22.4.2. Волновая и Фурье-оптика ... 548
*22.4.3. Оптика пучков ... 552
22.5. Нелинейная оптика ультракоротких импульсов ... 561
22.5.1. Импульсные параметрические процессы ... 562
22.5.2. Оптические солитоны ... 569
*22.5.3. Суперконтинуум ... 582
22.6. Детектирование импульсов ... 584
22.6.1. Измерение интенсивности ... 585
22.6.2. Измерение спектральной интенсивности 591
22.6.3. Измерение фазы ... 593
*22.6.4. Измерение спектрограмм ... 597

Глава 23
ОПТИЧЕСКИЕ МЕЖСОЕДИНЕНИЯ И КОММУТАТОРЫ ... 606

23.1. Оптические межсоединения ... 608
23.1.1. Межсоединения в свободном пространстве на основе рефракции и дифракции ... 611
23.1.2. Волноводные межсоединения ... 616
23.1.3. Невзаимные оптические межсоединения ... 618
23.1.4. Оптические межсоединения в микроэлектронике ... 618
23.2. Пассивные оптические маршрутизаторы ... 625
23.2.1. Маршрутизаторы с разделением по длине волны ... 626
23.2.2. Маршрутизаторы с разделением по поляризации, фазе и интенсивности ... 631
23.3. Фотонные коммутаторы ... 635
23.3.1. Архитектуры пространственных коммутаторов ... 635
23.3.2. Конструкции оптических пространственных коммутаторов ... 637
23.3.3. Полностью оптические пространственные коммутаторы ... 648
23.3.4. Коммутаторы с разделением по длине волны 654
23.3.5. Коммутаторы с разделением по времени ... 658
23.3.6. Коммутаторы пакетов ... 662
23.4. Оптические логические элементы ... 665
23.4.1. Бистабильные системы ... 665
23.4.2. Основы оптической бистабильности 667
23.4.3. Бистабильные оптические устройства ... 670

Глава 24
ВОЛОКОННО-ОПТИЧЕСКИЕ СИСТЕМЫ СВЯЗИ 684

24.1. Волоконно-оптические компоненты 686
24.1.1. Оптические волокна 686
24.1.2. Источники для оптических передатчиков 693
24.1.3. Оптические усилители 695
24.1.4. Детекторы для оптических приемников 698
24.2. Волоконно-оптические системы связи 700
24.2.1. Эволюция волоконно-оптических систем связи 701
24.2.2. Эксплуатационные показатели оптических волоконных систем 705
24.2.3. Системы, ограниченные по ослаблению и дисперсии 709
24.2.4. Компенсация ослабления и дисперсии и управление ими 718
24.2.5. Солитонная оптическая связь 721
24.3. Модуляция и мультиплексирование 723
24.3.1. Модуляция 723
24.3.2. Мультиплексирование 725
24.3.3. Мультиплексирование с разделением по длине волны (WDM) 727
24.4. Волоконно-оптические сети 729
24.4.1. Топологии сетей и коллективный доступ 730
24.4.2. Сети, использующие мультиплексирование с разделением по длине волны (WDM) 733
24.5. Когерентная оптическая связь 738

Приложение А
ПРЕОБРАЗОВАНИЕ ФУРЬЕ 752

А.1. Одномерное преобразование Фурье 752
А.2. Длительность и спектральная ширина 754
А.3. Двумерное преобразование Фурье 762

Приложение Б
ЛИНЕЙНЫЕ СИСТЕМЫ 766

Б.1. Одномерные линейные системы 766
Б.2. Двумерные линейные системы 770

Приложение В
МОДЫ ЛИНЕЙНЫХ СИСТЕМ 773

В.1. Моды дискретной линейной системы 775
В.2. Моды непрерывной системы, описываемой интегральным оператором 776
В.3. Моды системы, описываемой обыкновенными дифференциальными уравнениями 778
В.4. Моды системы, описываемой дифференциальным уравнением в частных производных 778

ГЛАВА 15 ЛАЗЕРЫ



Артур Шавлов (1921—1999)



Теодор Мейман (род. 1927)

В 1958 г. Артур Шавлов вместе с Чарльзом Таунсом показали, как обобщить принцип мазера на оптический диапазон спектра. Шавлов разделил Нобелевскую премию 1981 г. с Николасом Бломбергеном. Теодор Мейман впервые получил генерацию в рубиновом лазере в 1960 г.

Лазер — оптический генератор. Он состоит из резонансного оптического усилителя, выходной сигнал которого подается на вход с совпадением по фазе (рис. 15.1). Процесс генерации может быть инициирован при наличии на входе усилителя даже малого шума, содержащего спектральные компоненты, попадающие в полосу усиления. Этот входной сигнал усиливается и снова подается на вход, где претерпевает дальнейшее усиление. Процесс продолжается неограниченно, пока выходной сигнал не станет большим. Рост сигнала в конечном счете ограничивается насыщением усиления, и система приходит к стационарному состоянию, в котором выходной сигнал создается на частоте резонансного усилителя.

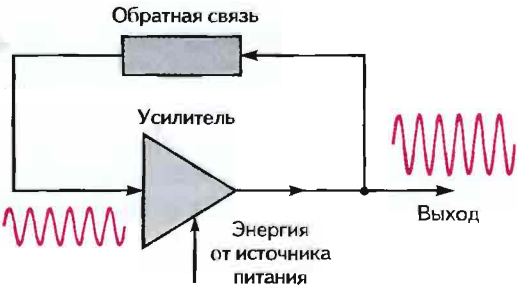


Рис. 15.1. Генератор как усилитель с положительной обратной связью

Для осуществления генерации необходимо выполнение двух условий:

• усиление в усилителе должно превосходить потери в системе обратной связи, так чтобы при полном обходе петли обратной связи происходило нарастание сигнала;

• полный фазовый сдвиг за один полный обход должен быть кратным, чтобы фаза сигнала обратной связи совпадала с фазой первоначального входного сигнала.

Если эти условия выполнены, система становится неустойчивой и начинается генерация. С увеличением мощности усиление насыщается и становится ниже своего первоначального значения. Устойчивость достигается, когда усиление уменьшается до уровня потерь (рис. 15.2). В этом состоянии усиление лишь компенсирует потери, так что в ходе цикла «усиление—обратная связь» сигнал не меняется, что и обеспечивает стационарность генерации.

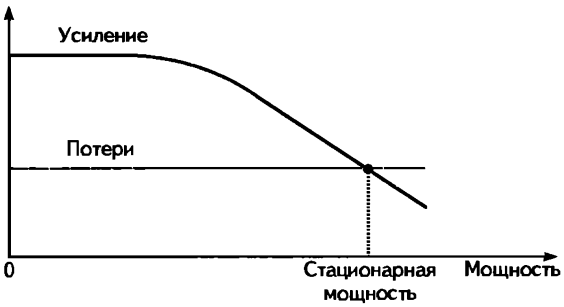


Рис. 15.2. Если первоначальное усиление превосходит потери, может начаться генерация. С ростом мощности генерации усилитель насыщается, что приводит к снижению усиления. Стационарное состояние достигается, когда усиление равно потерям

Поскольку усиление и фазовый сдвиг являются функциями частоты, оба условия генерации удовлетворяются только на одной (или нескольких) частотах, которые являются резонансными частотами генератора. Полезный выходной сигнал получается путем вывода части мощности из генератора.

Таким образом, генератор включает в себя:

• усилитель с насыщением;

• систему обратной связи;

• механизм селекции частоты;

• схему вывода полезного сигнала.

Лазер — оптический генератор (рис. 15.3), в котором усилителем служит активная среда с накачкой, рассмотренная в разд. 14.1 и 14.2. Насыщение усиления является основным свойством лазерных усилителей, что обсуждалось в разд. 14.4. Обратная связь обеспечивается путем помещения активной среды в оптический резонатор, отражающий свет туда и обратно между зеркалами, как описано в гл. 10. Селекция частот достигается совместным действием усилителя и резонатора, поддерживающего лишь определенные моды. Для вывода излучения одно из зеркал резонатора делается частично пропускающим.

Лазеры имеют невероятное разнообразие форм и приложений в науке и технике, включая интерферометрию, спектроскопию, формирование изображений, литографию, метрологию, связь, оптическое зондирование, охлажде-

ние атомов и обработку материалов. Излишне говорить о неоценимой роли лазеров в фундаментальных исследованиях по фотонике, как и во всех областях науки, техники и медицины. Предшественником лазера был **мазер** (англ. maser — *m*icrowave *a*mplification by *s*timulated *e*mission of *r*adiation). Принцип генерации в мазерах и лазерах можно применить к волнам, отличным от электромагнитного излучения. Например, **сазер** — акустическая версия лазера, излучающая пучок фононов за счет усиления звука посредством вынужденного испускания (*s*ound *a*mplification by *s*timulated *e*mission of *r*adiation)

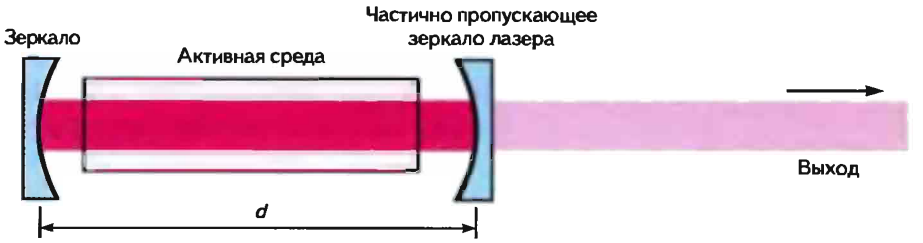


Рис. 15.3. Лазер состоит из оптического усилителя (включающего активную среду), помещенного внутрь оптического резонатора. Выходной сигнал извлекается через частично пропускающее зеркало

О данной главе

Глава содержит введение в принципы действия лазеров. В разд. 15.1 описано поведение лазерного резонатора и лазерного усилителя и выведены условия генерации. Свойства света, излучаемого лазерами, таких как мощность, спектральное и пространственное распределение, поляризация, рассмотрены в разд. 15.2. Распространенные типы лазеров обсуждаются в разд. 15.3, а работа импульсных лазеров — в разд. 15.4.

15.1. ТЕОРИЯ ЛАЗЕРНОЙ ГЕНЕРАЦИИ

В начале раздела дадим сводку сведений о двух основных компонентах лазера — усилителе и резонаторе. Хотя они детально обсуждались в гл. 14 и 10, соответственно, их краткий обзор приводится для удобства чтения.

15.1.1. Оптическое усиление и обратная связь

Лазерное усиление

Лазерный усилитель — узкополосный когерентный усилитель света. Усиление достигается посредством вынужденного излучения света атомной или молекулярной системой с инверсией заселенностей в одном из переходов (заселенность верхнего энергетического уровня больше, чем нижнего). Ширине

на полосы усиления определяется шириной линии атомного перехода либо одним из механизмов неоднородного уширения, таким как эффект Доплера в газовых лазерах

Лазерный усилитель — устройство с распределенным усилением, которое локально характеризуется коэффициентом усиления (усиление на единицу длины) $\gamma(\nu)$. Этот коэффициент определяет, насколько быстро нарастает плотность потока фотонов ϕ (или оптическая интенсивность $I = h\nu\phi$). При малых плотностях потока фотонов ϕ коэффициент усиления выражается как

$$\gamma_0(\nu) = N_0\sigma(\nu) = N_0\frac{\lambda^2}{8\pi t_{\text{сп}}}g(\nu), \tag{15.1}$$

Коэффициент усиления слабого сигнала

где N_0 — равновесная разность плотностей заселенности (плотность числа атомов в верхнем энергетическом состоянии минус плотность числа атомов в нижнем состоянии); N_0 растет с увеличением скорости накачки; $\sigma(\nu)$ — сечение перехода:

$$\sigma(\nu) = \frac{\lambda^2}{8\pi t_{\text{сп}}}g(\nu);$$

$t_{\text{сп}}$ — спонтанное время жизни; $g(\nu)$ — форма линии перехода; $\lambda = \lambda_0/n$ — длина волны в среде, где n — показатель преломления.

С ростом плотности потока фотонов действие усилителя становится нелинейным. За счет насыщения усиление снижается. За счет процесса усиления исходная разность заселенностей N_0 снижается до значения

$$N = \frac{N_0}{1 + \phi/\phi_s(\nu)}$$

для среды с однородным уширением линии, где $\phi_s(\nu) = 1/\tau_s\sigma(\nu)$ — насыщающая плотность потока фотонов; τ_s — постоянная времени насыщения, которая зависит от времен распада энергетических уровней; в идеальной четырехуровневой схеме накачки $\tau_s \approx t_{\text{сп}}$, а в идеальной трехуровневой схеме $\tau_s \approx 2t_{\text{сп}}$.

Таким образом, коэффициент усиления для насыщенного усилителя снижается до $\gamma(\nu) = N\sigma(\nu)$, поэтому для однородного уширения

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \phi/\phi_s(\nu)}. \tag{15.2}$$

Насыщенный коэффициент усиления

Процесс лазерного усиления включает также фазовый сдвиг. Когда форма линии лоренцева с шириной $\Delta\nu$,

$$g(\nu) = \frac{\Delta\nu/2\pi}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2},$$

сдвиг фазы при усилении, приходящийся на единицу длины, равен

$$\boxed{\varphi(\nu) = \frac{\nu - \nu_0}{\Delta\nu}\gamma(\nu).} \qquad (15.3)$$

Коэффициент фазового сдвига (лоренцева форма линии)

Этот фазовый сдвиг добавляется к сдвигу, вносимому средой, в которой расположены активные атомы. Графики коэффициентов усиления и фазового сдвига для усилителя с лоренцевой функцией формы линии показаны на рис. 15.4.

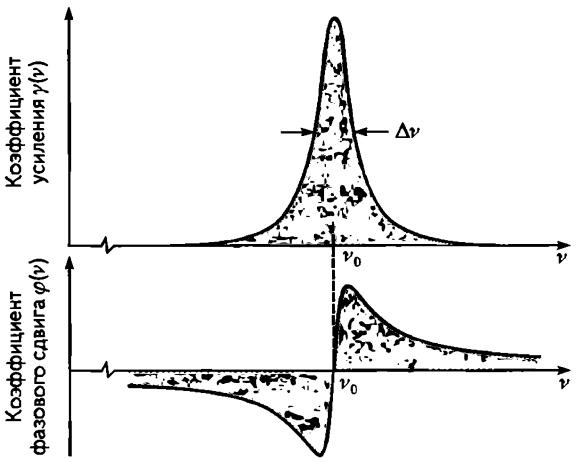


Рис. 15.4. Спектральная зависимость коэффициентов усиления и фазового сдвига для лазерного усилителя с лоренцевой функцией формы линии

Обратная связь и потери: оптический резонатор

Оптическая обратная связь достигается помещением активной среды в оптический резонатор. Резонатор Фабри—Перо, состоящий из двух зеркал, разделенных расстоянием d, содержит среду с показателем преломления n, в которой находятся активные атомы усилителя. Прохождение единицы длины среды вносит фазовый сдвиг, равный волновому числу

$$\boxed{k = \frac{2\pi\nu}{c}.} \qquad (15.4)$$

Коэффициент фазового сдвига

Резонатор также вносит потери в систему. Поглощение и рассеяние света в среде вносят распределенные потери, характеризуемые коэффициентом ослабления α_s (потери на единицу длины). При полном обходе резонатора длины d плотность потока фотонов уменьшается на множитель $\mathcal{R}_1\mathcal{R}_2 \exp(-2\alpha_s d)$, где $\mathcal{R}_1$ и $\mathcal{R}_2$ — коэффициенты отражения зеркал. Следовательно, полные потери

за один полный обход можно описать с помощью эффективного коэффициента распределенных потерь α_r, такого что

$$\exp(-2\alpha_r d) = \mathcal{R}_1\mathcal{R}_2 \exp(-2\alpha_s d), \tag{15.5}$$

так что

$$\boxed{\begin{aligned} \alpha_r &= \alpha_s + \alpha_{m_1} + \alpha_{m_2}; \\ \alpha_{m_1} &= \frac{1}{2d}\ln\frac{1}{\mathcal{R}_1}; \\ \alpha_{m_2} &= \frac{1}{2d}\ln\frac{1}{\mathcal{R}_2}, \end{aligned}} \tag{15.6}$$

Коэффициент потерь

где α_{m_1} и α_{m_2} представляют вклады зеркал 1 и 2 соответственно. Вклад обоих зеркал:

$$\alpha_m = \alpha_{m_1} + \alpha_{m_2} = \frac{1}{2d}\ln\frac{1}{\mathcal{R}_1\mathcal{R}_2}. \tag{15.7}$$

Поскольку α_r описывает полные потери энергии (или числа фотонов) на единице длины, то $\alpha_r c$ — потери числа фотонов в секунду. Таким образом,

$$\tau_\Phi = \frac{1}{c\alpha_r} \tag{15.8}$$

представляет **время жизни фотона**.

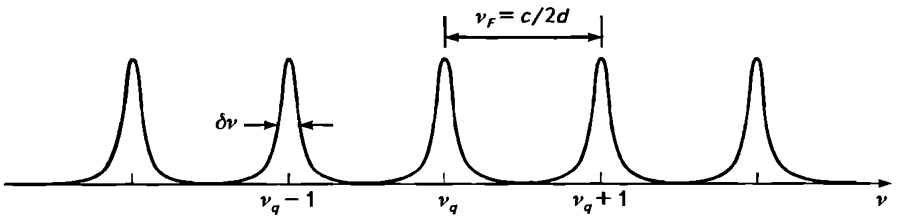


Рис. 15.5. Моды резонатора разделены частотным интервалом $\nu_F = c/2d$ и имеют ширину линии $\delta\nu = \nu_F/\mathcal{F} = 1/2\pi\tau_\Phi$

Резонатор поддерживает только те частоты, которым соответствует сдвиг фазы, кратный 2π, за полный обход резонатора. Для резонатора без активных атомов («холодного» резонатора) набег фазы за полный обход:

$$k2d = \frac{2\pi\nu d}{c} = q2\pi,$$

что соответствует модам с частотами

$$\nu_q = q\nu_F; \quad q = 1,\ 2, \tag{15.9}$$

где $\nu_F = c/2d$ — межмодовое расстояние резонатора; c — скорость света в среде (рис. 15.5). Спектральная ширина этих мод по половине высоты максимума

$$\delta\nu \approx \frac{\nu_F}{\mathcal{F}}, \tag{15.10}$$

где $\mathcal{F}$ — параметр резкости резонатора (см. подразд. 10.1.1). При малых потерях резонатора параметр резкости велик и приближенно равен

$$\mathcal{F} \approx \frac{\pi}{\alpha_r d} = 2\pi\tau_\Phi \nu_F. \tag{15.11}$$

15.1.2. Условия лазерной генерации

Два условия необходимо выполнить, чтобы лазер мог генерировать. *Условие усиления* определяет минимальную разность заселенностей и, следовательно, пороговое значение накачки, необходимое для генерации. *Фазовое условие* определяет частоту (или частоты), на которых происходит генерация.

Условие усиления: порог генерации

Чтобы инициировать лазерную генерацию, необходимо, чтобы коэффициент усиления слабого сигнала превышал коэффициент потерь:

$$\boxed{\gamma_0(\nu) > \alpha_r,} \tag{15.12}$$

Пороговое условие усиления

или, что равносильно, усиление должно превышать потери. В соответствии с (15.1) коэффициент усиления слабого сигнала $\gamma_0(\nu)$ пропорционален равновесной разности заселенностей N_0, которая, в свою очередь, как известно из гл. 14, увеличивается с ростом скорости накачки R. Поэтому с помощью (15.1) можно преобразовать (15.12) в условие, накладываемое на разность заселенностей:

$$N_0 = \frac{\gamma_0(\nu)}{\sigma(\nu)} > \frac{\alpha_r}{\sigma(\nu)}.$$

Итак,

$$N_0 > N_\text{п}, \tag{15.13}$$

где величина

$$N_\text{п} = \frac{\alpha_r}{\sigma(\nu)} \tag{15.14}$$

называется **пороговой разностью заселенностей**. Величина $N_\text{п}$, пропорциональная α_r, определяет минимальную скорость накачки $R_\text{п}$, достаточную для возбуждения лазерной генерации.

С помощью (15.8) величину α_r можно выразить через время жизни фотона, $\alpha_r = 1/c\tau_ф$, тогда (15.14) принимает вид

$$N_п = \frac{1}{c\tau_ф \sigma(\nu)}. \tag{15.15}$$

Следовательно, пороговая плотность разности заселенностей прямо пропорциональна α_r и обратно пропорциональна $\tau_ф$. Большие потери (более короткие времена жизни фотона) требуют более сильной накачки для достижения генерации.

Наконец, использование стандартной формулы для сечения перехода приводит к еще одному выражению для пороговой разности заселенностей:

$$\boxed{N_п = \frac{8\pi}{\lambda^2 c}\frac{t_{сп}}{\tau_ф}\frac{1}{g(\nu)}}, \tag{15.16}$$

Пороговая разность заселенностей

из которого ясно, что $N_п$ — функция частоты ν. Порог самый низкий и, следовательно, лазерный эффект легче всего достигается на той частоте, где функция формы линии наибольшая, т. е. на центральной частоте линии $\nu = \nu_0$. Для лоренцевой функции формы линии $g(\nu_0) = 2/\pi\Delta\nu$, так что минимальная разность заселенностей для генерации на центральной частоте ν_0 оказывается равной

$$N_п = \frac{2\pi}{\lambda^2}\frac{2\pi\Delta\nu t_{сп}}{\tau_ф}. \tag{15.17}$$

Она прямо пропорциональна ширине линии $\Delta\nu$.

Далее, если уширение перехода обусловлено временем жизни $t_{сп}$, то $\Delta\nu$ принимает значение $1/2\pi t_{сп}$ (см. подразд. 13.3.4), в результате чего (15.17) упрощается до

$$N_п = \frac{2\pi}{\lambda^2 c\tau_ф} = \frac{2\pi\alpha_r}{\lambda^2}. \tag{15.18}$$

Эта формула показывает, что минимальная пороговая разность заселенностей, требуемая для достижения генерации, является простой функцией длины волны λ и времени жизни фотона $\tau_ф$. Ясно, что лазерная генерация труднее достигается при уменьшении длины волны. Приведем численный пример: если $\lambda_0 = 1$ мкм, $\tau_ф = 1$ нс, а показатель преломления $n = 1$, получаем $N_п \approx 2{,}1 \ \ 10^7$ см^3.

Упражнение 15.1

Порог рубинового лазера

а. В центре линии перехода с $\lambda_0 = 693{,}4$ нм коэффициент поглощения рубина при тепловом равновесии (без накачки) при $T = 300$ °K равен $\alpha(\nu_0) \equiv -\gamma(\nu_0) \approx 0{,}2$ см$^{-1}$. При концентрации ионов Cr^{3+}, ответственных за данный переход, равной $N_a = 1{,}58 \cdot 10^{19}$ см$^{-3}$, определите сечение перехода $\sigma_0 = \sigma(\nu_0)$.

б. Стержень рубинового лазера имеет длину 10 см (показатель преломления $n = 1{,}76$) и площадь сечения 1 см^2. Рабочая длина волны $\lambda_0 = 693{,}4$ нм. Оба торца стержня отполированы и покрыты отражающим слоем с коэффициентом отражения 80 %. Предполагая, что отсутствуют рассеяние и другие дополнительные источники потерь, определите коэффициент потерь резонатора α_r и время жизни фотона $\tau_ф$.

в. С увеличением накачки лазера $\gamma(\nu_0)$ растет от своего начального значения $-0{,}2$ см$^{-1}$ в состоянии теплового равновесия, меняет знак и, следовательно, обеспечивает усиление. Определите пороговое значение разности заселенностей $N_п$ для лазерной генерации.

Фазовое условие: лазерные частоты

Второе условие генерации требует, чтобы сдвиг фазы, приобретаемый светом за полный обход резонатора, был целым кратным 2π, т. е.

$$2kd + 2\varphi(\nu)d = 2\pi q; \quad q = 1,\ 2, \tag{15.19}$$

Если вклад от атомов активной среды $2\varphi(\nu)d$ мал, то деление (15.19) на $2d$ приводит к ранее полученному результату для холодного резонатора, $\nu = \nu_q = q(c/2d)$.

В присутствии активной среды, когда вклад $2\varphi(\nu)d$ требует учета, решение уравнения (15.19) порождает набор частот генерации ν_q', которые слегка сдвинуты относительно собственных частот холодного резонатора ν_q. Оказывается, что активная среда слегка сдвигает все частоты мод холодного резонатора в сторону центральной частоты атомного перехода, как показано ниже.

*Затягивание частоты

С использованием соотношения $k = 2\pi\nu/c$ и коэффициента фазового сдвига для лоренцевой формы линии (15.3) условие фазового сдвига (15.19) дает

$$\nu + \frac{c}{2\pi}\frac{\nu - \nu_0}{\Delta\nu}\gamma(\nu) = \nu_q. \tag{15.20}$$

Это уравнение можно решить относительно частот генерации $\nu = \nu_q'$, соответствующих каждой моде холодного резонатора ν_q. Поскольку уравнение нелинейное, полезно провести графическое решение. Обозначим левую часть уравнения (15.20) через $\psi(\nu)$ и построим ее графически (рис. 15.6) (это сумма линейной функции и лоренцева коэффициента фазового сдвига, схематически показанного на рис. 15.1). Значение $\nu = \nu_q'$, при котором $\psi(\nu) = \nu_q$, определяется графически. Из рисунка видно, что частоты мод холодного резонатора ν_q всегда затягиваются в сторону центральной частоты перехода резонансной среды ν_0.

Приближенное аналитическое решение уравнения (15.20) также можно получить. Запишем (15.20) в виде

$$\nu = \nu_q - \frac{c}{2\pi}\frac{\nu - \nu_0}{\Delta\nu}\gamma(\nu). \tag{15.21}$$

Когда $\nu = \nu_q' \approx \nu_q$, второй член в (15.21) мал, вследствие чего ν можно заменить на ν_q без большой потери точности. Тогда

$$\nu_q' = \nu_q - \frac{c}{2\pi}\frac{\nu_q - \nu_0}{\Delta\nu}\gamma(\nu_q), \tag{15.22}$$

что представляет собой явное выражение для частоты генерации ν_q' как функции частоты моды холодного резонатора ν_q. Более того, в стационарном режиме усиление равно потерям, так что

$$\gamma(\nu_q) = \alpha_r \approx \frac{\pi}{\mathcal{F}d} = \frac{2\pi}{c}\delta\nu,$$

где $\delta\nu$ — спектральная ширина мод холодного резонатора. Поставляя это соотношение в (15.22), находим

$$\boxed{\nu_q' \approx \nu_q - (\nu_q - \nu_0)\frac{\delta\nu}{\Delta\nu}.} \tag{15.23}$$

Лазерные частоты

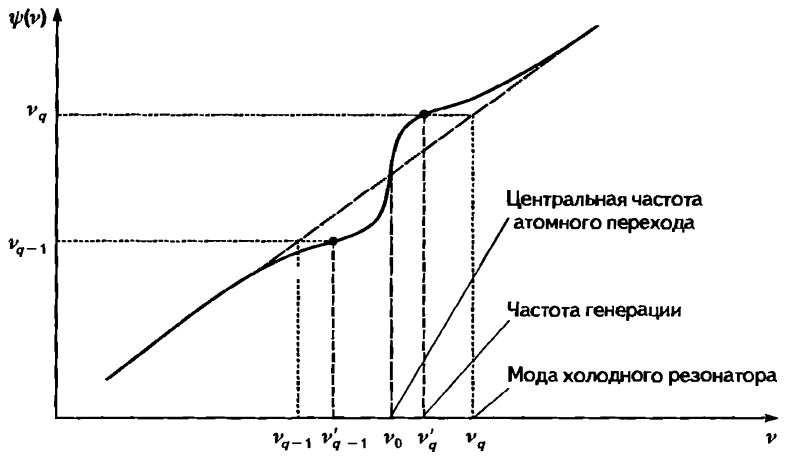


Рис. 15.6. Левая часть уравнения (15.20) $\psi(\nu)$ как функция ν. Частота ν, при которой $\psi(\nu) = \nu_q$, есть решение уравнения (15.20). Частота каждой моды холодного резонатора ν_q соответствует своей частоте «горячего» резонатора ν_q', которая сдвинута в направлении центральной частоты атомного резонанса ν_0

Следовательно, частота моды холодного резонатора ν_q затягиваются по направлению к частоте атомного резонанса ν_0 на долю $\delta\nu/\Delta\nu$ своего исходного расстояния от центральной частоты ($\nu_q - \nu_0$), как показано на рис. 15.7. Чем острее мода резонатора (т. е. чем меньше $\delta\nu$), тем меньше эффект затягивания частоты. Напротив, чем уже атомный резонанс (меньше $\Delta\nu$), тем более эффективно затягивание.

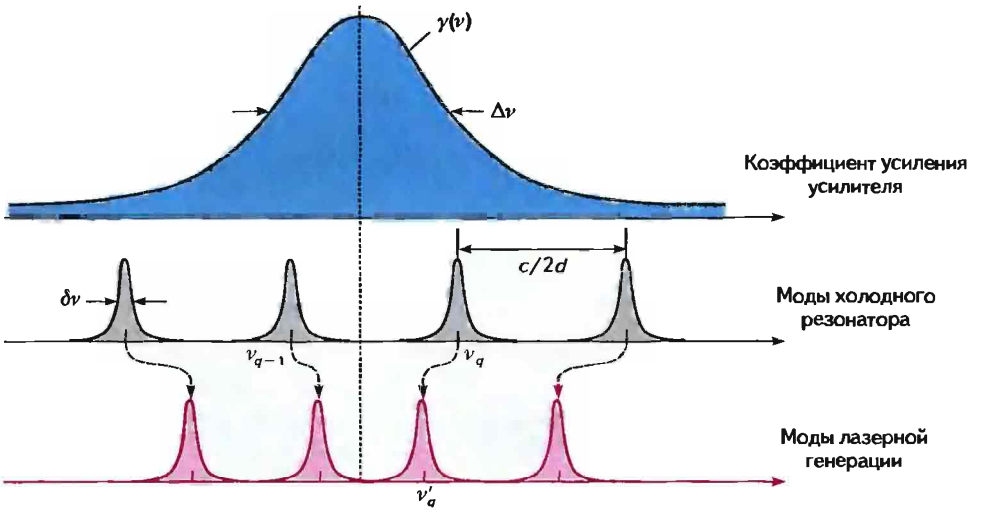


Рис. 15.7. Частоты лазерной генерации находятся вблизи мод холодного резонатора. Они слегка затянуты к центральной частоте атомного резонанса ν_0. Рисунок схематический и выполнен без соблюдения масштабов

15.2. ВЫХОДНЫЕ ХАРАКТЕРИСТИКИ ЛАЗЕРА

15.2.1. Мощность

Внутренняя плотность потока фотонов

Лазер, накачиваемый с превышением порога ($N_0 > N_п$), имеет коэффициент усиления слабого сигнала $\gamma(\nu_0)$, который больше коэффициента потерь α_r, как показывает (15.12). Тогда может начаться лазерная генерация, если удовлетворяется фазовое условие (15.19). С ростом плотности потока фотонов ϕ внутри резонатора (рис. 15.8) коэффициент усиления начинает убывать в соответствии с (15.2) для сред с однородным уширением. До тех пор, пока коэффициент усиления продолжает оставаться больше коэффициента потерь, поток фотонов продолжает расти.

В итоге, когда насыщенный коэффициент усиления становится равным коэффициенту потерь (или, что равносильно, $N = N_п$), поток фотонов прекращает свой рост и генерация достигает стационарного режима. Происходит **фиксация усиления** на уровне потерь. Стационарная плотность потока фотонов внутри лазера определяется приравниванием коэффициента усиления сильного сигнала (с насыщением) к коэффициенту потерь,

$$\frac{\gamma_0(\nu)}{1+\phi/\phi_s(\nu)} = \alpha_r,$$

что дает

$$\phi = \begin{cases} \phi_s(\nu)\left(\dfrac{\gamma_0(\nu)}{\alpha_r} - 1\right), & \gamma_0(\nu) > \alpha_r; \\ 0, & \gamma_0(\nu) \le \alpha_r. \end{cases} \tag{15.24}$$

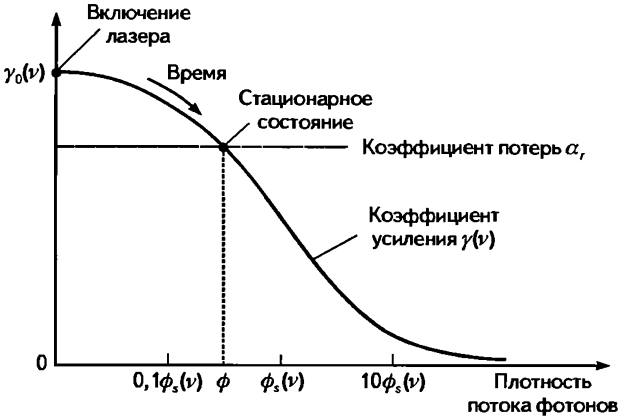


Рис. 15.8. Определение стационарного значения плотности потока фотонов ϕ в лазере. В момент включения лазера $\phi = 0$, так что $\gamma(\nu) = \gamma(\nu_0)$. По мере развития генерации во времени увеличение ϕ вызывает уменьшение $\gamma(\nu)$ вследствие насыщения усиления. Когда γ достигает α_r, плотность потока фотонов прекращает свой рост и устанавливается стационарный режим генерации. Чем меньше потери, тем больше значение ϕ

Формула (15.24) выражает стационарную плотность потока фотонов, возникающую при лазерной генерации. Это среднее число фотонов, пересекающих в секунду единичную площадь в обоих направлениях, поскольку фотоны, распространяющиеся в обоих направлениях, вносят вклад в процесс насыщения. Следовательно, плотность потока фотонов в одном направлении равна $\phi/2$. В данном упрощенном рассмотрении мы пренебрегли спонтанным излучением. Конечно, (15.24) дает среднее значение плотности потока фотонов, вокруг которого происходят случайные флуктуации, как обсуждалось в разд. 12.2.

Поскольку $\gamma_0(\nu) = N_0\sigma(\nu)$ и $\alpha_r = N_{\text{п}}\sigma(\nu)$, (15.24) можно записать в виде

$$\boxed{\phi = \begin{cases} \phi_s(\nu)\left(\dfrac{N_0}{N_{\text{п}}} - 1\right), & N_0 > N_{\text{п}}; \\ 0, & N_0 \le N_{\text{п}}. \end{cases}} \tag{15.25}$$

Стационарная внутренняя плотность потока фотонов

Ниже порога лазерная плотность потока фотонов равна нулю; любое увеличение скорости накачки проявляется в увеличении потока спонтанно испущенных фотонов, но генерация не возникает. Выше порога стационарная внутренняя плотность потока фотонов прямо пропорциональна начальной разности заселенностей N_0 и, следовательно, растет с увеличением скорости накачки R [см. (14.27) и (14.41)]. Если N_0 в два раза превышает пороговое значение $N_{\text{п}}$, то плотность потока фотонов точно равна своему насыщающему значению $\phi_s(\nu)$, при котором коэффициент усиления снижается в два раза по сравнению со своим максимальным значением. Разность заселенностей N и плотность потока фотонов ϕ показаны на рис. 15.9 в зависимости от N_0.

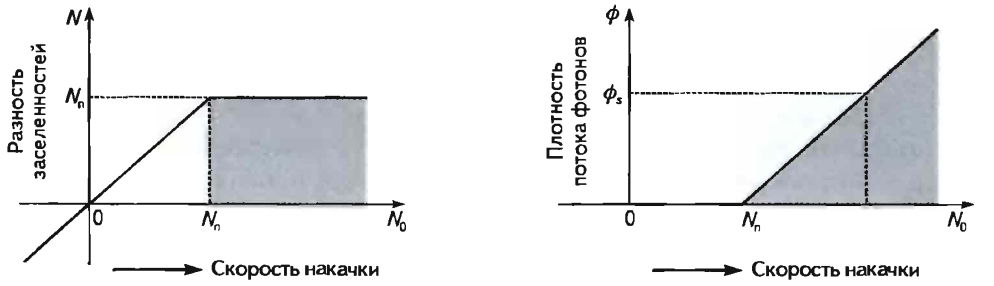


Рис. 15.9. Стационарные значения разности заселенностей N и плотности потока фотонов внутри лазера ϕ как функции N_0 (разности заселенностей в отсутствие излучения; N_0 растет с увеличением скорости накачки R). Лазерная генерация происходит, когда N_0 превышает $N_п$; стационарное значение N при этом насыщается и фиксируется на уровне $N_п$ [так же как $\gamma_0(\nu)$ фиксируется на уровне α_r]. Выше порога ϕ пропорционально $N_0 - N_п$

Плотность потока фотонов на выходе

Только часть стационарной внутренней плотности потока фотонов, определяемой выражением (15.25), покидает резонатор в виде полезного света. Выходная плотность потока фотонов ϕ_0 — это та часть внутренней плотности потока фотонов ϕ, которая проходит сквозь зеркало 1. Если коэффициент пропускания зеркала равен $\mathcal{T}$, то

$$\phi_0 = \frac{\mathcal{T}\phi}{2}. \tag{15.26}$$

Соответствующая оптическая интенсивность на выходе лазера

$$I_0 = \frac{h\nu\mathcal{T}\phi}{2}, \tag{15.27}$$

а выходная мощность лазера $P_0 = I_0 A$, где A — площадь сечения лазерного пучка. Эти соотношения вместе с (15.25) позволяют явно выразить выходную мощность лазера через $\phi_s(\nu)$, N_0, $N_п$, $\mathcal{T}$ и A.

Оптимизация выходной плотности потока фотонов

Полезная плотность потока фотонов на выходе лазера уменьшает внутреннюю плотность потока фотонов и, следовательно, вносит вклад в потери лазерного генератора. Любая попытка увеличить долю фотонов, которым позволяется покинуть резонатор (в ожидании увеличения полезного выхода света) приводит к увеличению потерь, так что стационарная плотность потока фотонов внутри резонатора уменьшается. Итоговый результат может быть отрицательным: выходная мощность лазера не вырастет, а упадет.

Покажем, что существует оптимальный коэффициент пропускания $\mathcal{T}(0 < \mathcal{T} < 1)$, при котором выходная интенсивность лазера максимальна. Плотность потока

фотонов на выходе лазера $\phi_0 = \mathcal{T}\phi/2$ есть произведение коэффициента пропускания зеркала $\mathcal{T}$ и внутренней плотности потока фотонов $\phi/2$. По мере роста $\mathcal{T}$ величина ϕ уменьшается в результате роста потерь. В одном предельном случае $\mathcal{T} = 0$ генератор имеет наименьшие потери (ϕ максимально), однако вообще нет полезного выхода фотонов ($\phi_0 = 0$). В другом предельном случае, когда зеркало убирают, так что $\mathcal{T} = 1$, увеличение потерь приводит к неравенству $\alpha_r > \gamma_0(\nu)$ ($N_n > N_0$), делая невозможной лазерную генерацию. В этом случае $\phi = 0$ и, следовательно, снова $\phi_0 = 0$. Оптимальное значение $\mathcal{T}$ лежит где-то между этими двумя пределами.

Для его нахождения мы должны получить явную связь между ϕ_0 и $\mathcal{T}$. Предположим, что зеркало 1 с коэффициентом отражения $\mathcal{R}_1$ и коэффициентом пропускания $\mathcal{T} = 1 - \mathcal{R}_1$ служит для вывода полезного света. Коэффициент потерь α_r выражается как функция $\mathcal{T}$ путем подстановки в (15.6) коэффициента потерь, вносимых зеркалом 1:

$$\alpha_{m_1} = \frac{1}{2d}\ln\frac{1}{\mathcal{R}_1} = -\frac{1}{2d}\ln(1-\mathcal{T}), \tag{15.28}$$

что дает

$$\alpha_r = \alpha_s + \alpha_{m_2} - \frac{1}{2d}\ln(1-\mathcal{T}), \tag{15.29}$$

где коэффициент потерь, вносимых зеркалом 2, равен

$$\alpha_{m_2} = \frac{1}{2d}\ln\frac{1}{\mathcal{R}_2}. \tag{15.30}$$

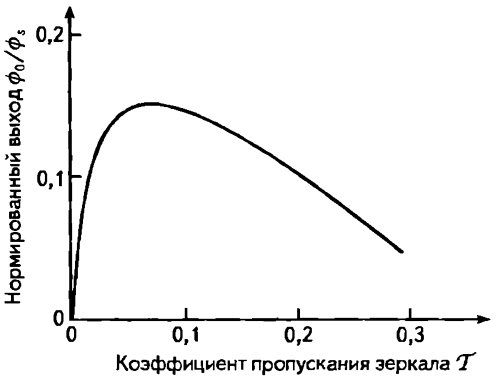


Рис. 15.10. Зависимость стационарной выходной плотности потока фотонов от коэффициента пропускания зеркала. В данном примере фактор усиления $\gamma_0 = 2\gamma_0 d = 0{,}5$, а фактор потерь $L = 2(\alpha_s + \alpha_{m_2})d = 0{,}02$ (2 %). Оптимальный коэффициент пропускания $\mathcal{T}_{\text{опт}}$ получается равным 0,08

Теперь используем (15.24), (15.26) и (15.29) для получения зависимости плотности выходящего потока фотонов ϕ_0 от коэффициента пропускания зеркала

$$\phi_0 = \frac{1}{2}\phi_s\mathcal{T}\left[\frac{g_0}{L - \ln(1-\mathcal{T})} - 1\right]; \quad g_0 = 2\gamma_0(\nu)d; \quad L = 2\left(\alpha_s + \alpha_{m_2}\right)d, \tag{15.31}$$

график которой представлен на рис. 15.10. Заметим, что выходная плотность потока фотонов прямо связана с коэффициентом усиления слабого сигнала. Оптимальный коэффициент пропускания $\mathcal{T}_{\text{опт}}$ получается приравниваем к нулю производной от ϕ_0 по $\mathcal{T}$. При $\mathcal{T} \ll 1$ можно воспользоваться приближением $\ln(1 - \mathcal{T}) \approx -\mathcal{T}$ и получить

$$\mathcal{T}_{\text{опт}} \approx \sqrt{g_0 L} - L. \tag{15.32}$$

Внутренняя плотность числа фотонов

Стационарное число фотонов на единицу объема внутри резонатора n связано со стационарной плотностью потока фотонов ϕ (для фотонов, распространяющихся в обоих направлениях) простым соотношением

$$\boxed{\mathrm{n} = \frac{\phi}{c}.} \tag{15.33}$$

Его смысл легко становится понятным, если рассмотреть цилиндр с площадью основания A, длиной c и объемом cA. (c — скорость света в среде), ось которого параллельна оси резонатора. Если резонатор содержит n фотонов в единице объема, то в цилиндре содержится $cA\mathrm{n}$ фотонов. Эти фотоны распространяются в обоих направлениях параллельно оси резонатора, и половина из них пересекает основание цилиндра каждую секунду. Поскольку на основание цилиндра падает такое же число фотонов с другой стороны, плотность потока фотонов (число фотонов в секунду на единицу площади в обоих направлениях) есть $\phi = 2(cA\mathrm{n}/2) = \mathrm{n}c$, откуда следует (15.33).

Плотность числа фотонов, соответствующая стационарной внутренней плотности потока фотонов (15.25) равна

$$\boxed{\mathrm{n} = \mathrm{n}_s\left(\frac{N_0}{N_{\text{п}}} - 1\right), \quad N_0 > N_{\text{п}},} \tag{15.34}$$

Стационарная плотность числа фотонов

где $\mathrm{n}_s = \phi_s(\nu)/c$ — насыщающее значение плотности числа фотонов. С помощью соотношений $\phi_s(\nu) = [\tau_s \sigma(\nu)]^{-1}$, $\alpha_r = \gamma(\nu)$, $\alpha_r = 1/c\tau_{\text{ф}}$ и $\gamma(\nu) = N\sigma(\nu) = N_{\text{п}}\sigma(\nu)$ формулу (15.34) можно переписать в виде

$$\boxed{\mathrm{n} = (N_0 - N_{\text{п}})\frac{\tau_{\text{ф}}}{\tau_s}, \quad N_0 > N_{\text{п}},} \tag{15.35}$$

Стационарная плотность числа фотонов

Это соотношение допускает простую интерпретацию: $(N_0 - N_{\text{п}})$ есть надпороговое превышение разности заселенностей (на единицу объема), а $(N_0 - N_{\text{п}})\tau_s$ — скорость, с которой генерируются фотоны, которая в стационарном состоянии равна скорости, с которой фотоны теряются, а именно $\mathrm{n}/\tau_{\text{ф}}$. Дробь $\tau_{\text{ф}}/\tau_s$ есть

отношение скорости, с которой фотоны испускаются к скорости, с которой они теряются.

При идеальных условиях накачки в четырехуровневой схеме из формул (14.27) и (14.28) следует, что $\tau_s \approx t_{сп}$ и $N_0 \approx Rt_{сп}$, где R — скорость накачки атомов ($с^{-1} \cdot см^{-3}$). Тогда (15.35) можно записать как

$$\frac{n}{\tau_ф} = R - R_п, \quad R > R_п, \tag{15.36}$$

где $R_п = N_п/t_{сп}$ — пороговое значение скорости накачки. Следовательно, в стационарном режиме полная скорость потерь фотонной плотности $n/\tau_ф$ в точности равна превышению скорости накачки над порогом $R - R_п$.

Выходной поток фотонов и эффективность вывода

Если выход полезного излучения через зеркало резонатора является единственным источником потерь (который учитывается в $\tau_ф$), а V — объем активной среды, то (15.36) дает полный поток выходящих фотонов $Ф_0$ (число фотонов в секунду)

$$Ф_0 = (R - R_п)V, \quad R > R_п. \tag{15.37}$$

Если существуют другие механизмы потерь, то выходной поток фотонов имеет вид

$$\boxed{Ф_0 = \eta_{эфф}(R - R_п)V,} \tag{15.38}$$

Поток фотонов на выходе лазера

где эффективность вывода $\eta_{эфф}$ представляет собой отношение потерь за счет полезного вывода пучка к полной величине потерь в резонаторе α_r.

Если полезный свет выводится только через зеркало 1, то можно использовать выражения (15.8) и (15.28) для α_r и α_{m_1}, чтобы записать $\eta_{эфф}$ как

$$\eta_{эфф} = \frac{\alpha_{m_1}}{\alpha_r} = \frac{c}{2d}\tau_ф \ln\frac{1}{\mathcal{R}_1}. \tag{15.39}$$

Далее, если $\mathcal{T} = 1 - \mathcal{R}_1 \ll 1$, то (15.39) дает

$$\boxed{\eta_{эфф} \approx \frac{\tau_ф}{T_F}\mathcal{T},} \tag{15.40}$$

Эффективность вывода

где мы ввели определение $1/T_F = c/2d$, отражающее тот факт, что эффективность вывода $\eta_{эфф}$ можно выразить через отношение времени жизни фотона к времени полного обхода резонатора, умноженное на коэффициент пропускания зеркала. Тогда для выходной мощности лазера имеем

$$P_0 = h\nu Ф_0 = \eta_{эфф} h\nu (R - R_п)V \tag{15.41}$$

С помощью простых алгебраических преобразований можно показать, что это выражение согласуется с полученным из (15.27).

Потери вносятся также другими источниками, такими как неэффективность процесса накачки. Дополнительные функции, такие как охлаждение и мониторинг, также потребляют энергию. **Коэффициент преобразования мощности** или **коэффициент полезного действия** всей установки определяется как отношение выходной оптической мощности P_0 к мощности P_p, затрачиваемой на накачку:

$$\eta_c = \frac{P_0}{P_p}. \qquad (15.42)$$

Коэффициент преобразования мощности

Примеры значений для различных типов лазеров приведены в табл. 15.1. Поскольку выходная мощность лазера линейно растет с увеличением превышения мощности накачки над порогом в соответствии с (15.41), еще одной часто используемой мерой эффективности преобразования является **дифференциальный коэффициент преобразования мощности**

$$\eta_s = \frac{\mathrm{d}P_0}{\mathrm{d}P_p}. \qquad (15.43)$$

Дифференциальный коэффициент преобразования мощности

В общем случае η_s больше коэффициента полезного действия η_c.

15.2.2. Спектральное распределение

Спектральное распределение света, генерируемого лазером, определяется как формой линии атомов активной среды (включая однородный или неоднородный тип уширения), так и модами резонатора. Это можно проиллюстрировать на примере двух условий лазерной генерации.

1. Условие усиления, требующее, чтобы начальный коэффициент усиления был больше коэффициента потерь $[\gamma_0(\nu) > \alpha_r]$, удовлетворяется для всех частот генерации, лежащих в непрерывной спектральной полосе шириной B с центром на частоте атомного перехода ν_0, как показано на рис. 15.11, *а*. Ширина B растет с увеличением атомной ширины линии $\Delta\nu$ и отношения $\gamma_0(\nu_0)/\alpha_r$; точное их соотношение зависит от конкретного вида функции $\gamma_0(\nu)$.

2. Фазовое условие требует, чтобы частота генерации была одной из модовых частот резонатора ν_q (для простоты предполагаем, что затягивание частоты пренебрежимо мало). Ширина на уровне половины максимума для каждой моды есть $\delta\nu \approx \nu_F/\mathcal{F}$ (рис. 15.11, *б*).

Отсюда следует, что генерация возможна лишь на конечном числе частот (ν_1, ν_2, ν_M). Число генерируемых мод, следовательно, равно

$$M \approx \frac{B}{\nu_F}, \qquad (15.44)$$

Число возможных мод лазера

где ν_F — приближенное значение межмодового расстояния. Однако, сколько из этих M возможных мод будет в действительности нести оптическую мощность, зависит от механизма уширения спектральной линии активной среды. Ниже будет показано, что при неоднородном уширении на всех M модах происходит генерация (хотя и с различной мощностью), в то время как при однородном уширении возникает конкуренция мод, мешающая их одновременной генерации.

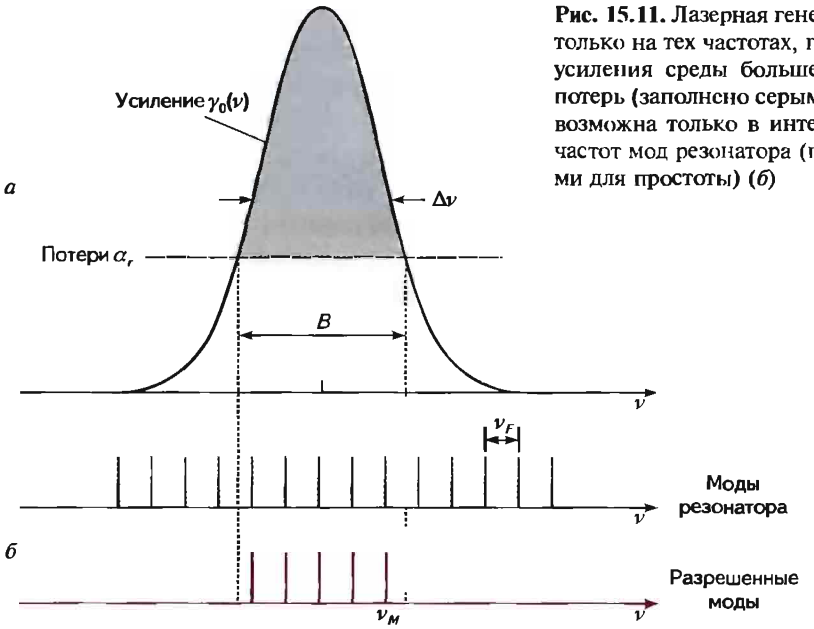


Рис. 15.11. Лазерная генерация возможна только на тех частотах, где коэффициент усиления среды больше коэффициента потерь (заполнено серым) (*а*). Генерация возможна только в интервале $\delta\nu$ вблизи частот мод резонатора (показаны линиями для простоты) (*б*)

Можно ожидать, что ширина каждой лазерной моды приблизительно равна $\delta\nu$, однако она оказывается намного меньше. Она ограничена так называемой шириной Шавлова—Таунса, которая убывает обратно пропорционально оптической мощности. Почти все лазеры имеют ширину линии намного больше ширины Шавлова—Таунса из-за внешних воздействий, таких как акустические и тепловые флуктуации зеркал резонатора, однако к этому пределу можно приблизиться за счет тщательного контроля параметров эксперимента.

Упражнение 15.2

Число мод в газовом лазере

Газовый лазер с доплеровским уширением имеет гауссов профиль коэффициента усиления (см. подразд. 13.3.2) вида

$$\gamma_0(\nu) = \gamma_0(\nu_0)\exp\left[-\frac{(\nu-\nu_0)^2}{2\sigma_D^2}\right],$$

где $\Delta\nu_D = (8\ln 2)^{1/2}\sigma_D$ — ширина на уровне половины максимума.

а. Выразите ширину полосы, в которой возможна генерация, через $\Delta\nu_D$ и отношение $\gamma_0(\nu_0)/\alpha_r$, где α_r — коэффициент потерь резонатора.

б. У He—Ne-лазера доплеровская ширина линии $\Delta\nu_D = 1{,}5$ ГГц, а коэффициент усиления в центре линии $\gamma_0(\nu_0) = 2 \cdot 10^{-3}$ см$^{-1}$. Длина резонатора лазера $d = 100$ см, коэффициенты отражения зеркал 97 и 100 % (остальные потери пренебрежимо малы). Считая, что $n = 1$, определите число лазерных мод M.

Среда с однородным уширением

После включения лазера все моды, для которых начальный коэффициент усиления превышает коэффициент потерь, начинают расти (рис. 15.12, *а*). Плотности потока фотонов ϕ_1, ϕ_2, ..., ϕ_M создаются в M модах. Моды, частоты которых лежат ближе к центральной частоте полосы усиления ν_0, растут наиболее быстро и приобретают наибольшие значения плотности потока фотонов. Эти фотоны взаимодействуют со средой и уменьшают усиление за счет уменьшения разности заселенностей. Насыщенный коэффициент усиления равен

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \sum_{j=1}^{M} \phi_j / \phi_s(\nu_j)}, \tag{15.45}$$

где $\phi_s(\nu_j)$ — насыщающая плотность потока фотонов для моды j. Справедливость (15.45) может быть проверена с помощью анализа, аналогичного тому, который использовался при выводе (14.48). Насыщенный коэффициент усиления показан на рис. 15.12, *б*.

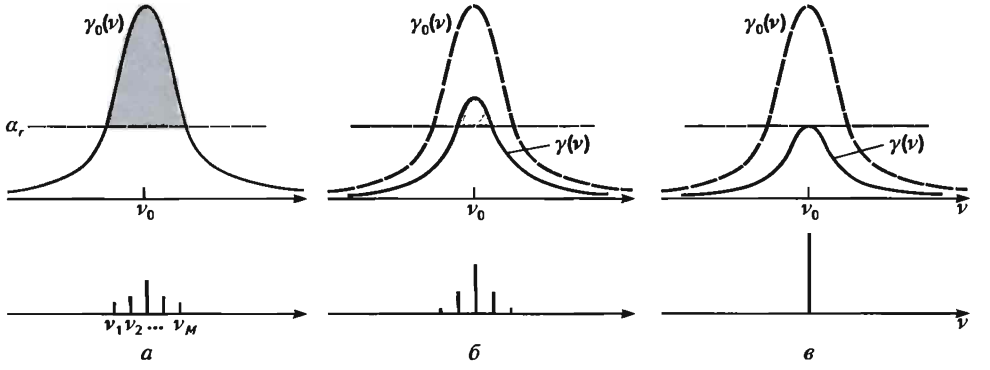


Рис. 15.12. Развитие процесса генерации в идеальной среде с однородным уширением:

а — непосредственно после включения лазера все моды с частотами ν_1, ν_2, ..., ν_M, для которых коэффициент усиления больше коэффициента потерь, начинают расти, причем центральные моды растут с наибольшей скоростью; *б* — через небольшое время усиление насыщается, так что только центральные моды продолжают расти, а периферические моды, для которых усиление становится ниже потерь, ослабляются и в конечном счете исчезают; *в* — в отсутствие выжигания пространственного провала остается только одна мода

Поскольку коэффициент усиления насыщается однородно, для мод, достаточно удаленных от центра линии, потери начинают превышать усиление. Эти моды теряют мощность, в то время как более близкие к центру моды продолжают расти, хотя и с меньшей скоростью. В конце концов остается только одна мода (или две в симметричном случае), у которой усиление поддерживается на уровне потерь, а для остальных мод потери превышают усиление. В идеальных стационарных условиях мощность этой предпочтительной моды остается постоянной, тогда как генерация на остальных модах прекращается (рис. 15.12, *в*). Оставшаяся мода имеет частоту, наиболее близкую к ν_0; значения усиления для ее конкурентов лежат ниже уровня потерь. При известной частоте оставшейся моды плотность потока фотонов в ней можно определить с помощью (15.25).

На практике, однако, лазеры с однородным уширением генерируют на нескольких модах, поскольку различные моды занимают различные пространственные области активной среды. Когда генерация на частоте, наиболее близкой к центру полосы усиления, установилась, коэффициент усиления все еще может превышать коэффициент потерь в тех местах активной среды, где электрическое поле стоячей волны для этой центральной моды мало. Это явление называется **выжиганием пространственного провала.** Оно дает возможность генерации других мод, пространственные максимумы которых лежат вблизи нулей центральной моды.

Среда с неоднородным уширением

В среде с неоднородным уширением общий коэффициент усиления $\bar{\gamma}_0(\nu)$ образуется наложением коэффициентов усиления различных групп атомов (см. подразд. 13.3.4), как показано на рис. 15.13.

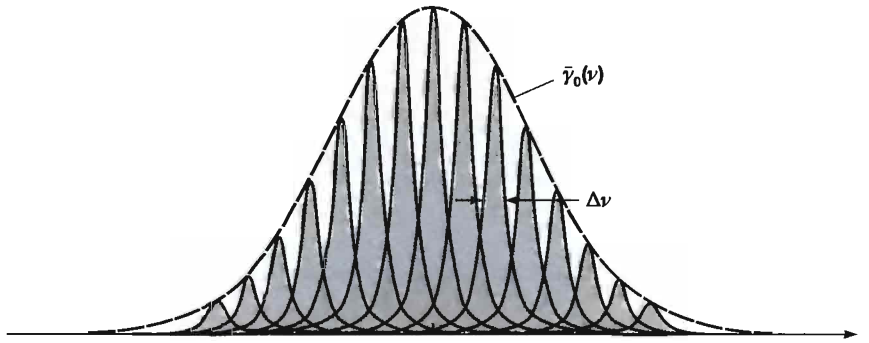
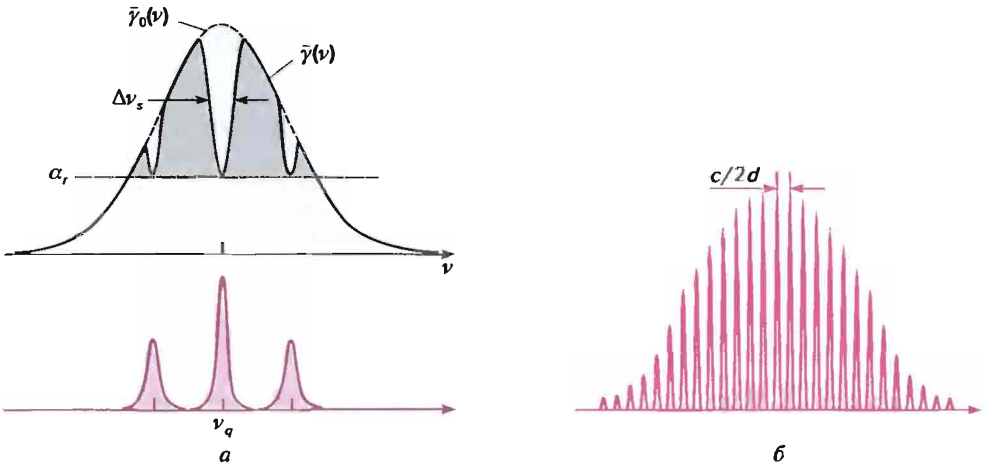


Рис. 15.13. Форма линии усиления среды с неоднородным уширением складывается из множества атомных линий, связанных с различными свойствами или различным окружением атомов

Ситуация сразу после включение лазера такая же, как в среде с однородным уширением. Моды, для которых усиление больше потерь, начинают рас-

ти и усиление снижается. Если межмодовое расстояние больше однородной ширины составляющих атомных линий, то разные моды взаимодействуют с разными атомами. Атомы, линии которых не совпадают ни с одной из мод, не замечают присутствия фотонов в резонаторе. Поэтому для них разность заселенностей не меняется, и их вклад в усиление остается ненасыщенным. В атомах, линии которых совпадают с модами, инверсия заселенностей уменьшается, усиление насыщается, приводя к появлению провалов в спектральном контуре усиления среды (рис 15.14, *а*). Этот процесс называется **выжиганием спектральных провалов**. Ширина спектрального провала растет с плотностью потока фотонов по закону квадратного корня $\Delta\nu_s = \Delta\nu(1 + \phi/\phi_s)^{1/2}$ [см. формулу (14.61)].



а *б*

Рис. 15.14. В среде с неоднородным уширением лазерная генерация происходит независимо для каждой моды с выжиганием провалов в общем спектральном профиле усиления. Усиление, сообщаемое средой одной моде, не влияет на усиление других мод. Центральные моды собирают вклады от большего числа атомов и поэтому несут больше фотонов, чем периферические моды (*а*). Спектр типичного многомодового газового лазера с неоднородным уширением (*б*)

Данный процесс насыщения с выжиганием провалов развивается независимо для разных мод, пока усиление не станет равно потерям для каждой моды в стационарном режиме. Моды не конкурируют, так как они черпают мощность от разных, а не от одних и тех же атомов. Много мод генерируется независимо, причем центральные моды выжигают более глубокие провалы и быстрее нарастают, как показано на рис. 15.14, *а*. Спектр типичного многомодового газового лазера с неоднородным уширением показан на рис. 15.14, *б*. Число мод в типичном случае больше, чем в лазерах с однородным уширением, поскольку, как правило, выжигание пространственных провалов позволяет поддерживать меньшее число мод, чем выжигание спектральных провалов.

*Выжигание спектральных провалов в среде с доплеровским уширением

Форма линии газа при температуре T возникает из наложения смещенных эффектом Доплера линий излучения отдельных атомов, движущихся с различными скоростями (см. подразд. 13.3.4 и упражнение 13.13). Покоящийся атом взаимодействует с излучением на частоте ν_0. Атом, движущийся со скоростью v к источнику света, взаимодействует с излучением на частоте $\nu_0(1 + v/c)$, а от источника — на частоте $\nu_0(1 - v/c)$. Поскольку мода частоты ν_q содержит волны обоих направлений, отражаясь от зеркал туда и обратно, она взаимодействует с двумя группами атомов, движущимися со скоростями $\pm v$, такими что $\nu_q - \nu_0 = \pm\nu_0 v/c$. Отсюда следует, что мода ν_q насыщает разность заселенностей в атомах по обе стороны от центральной частоты и выжигает два провала в профиле усиления, как показано на рис. 15.15. Конечно, при $\nu_q = \nu_0$ выжигается всего один провал в центре профиля.

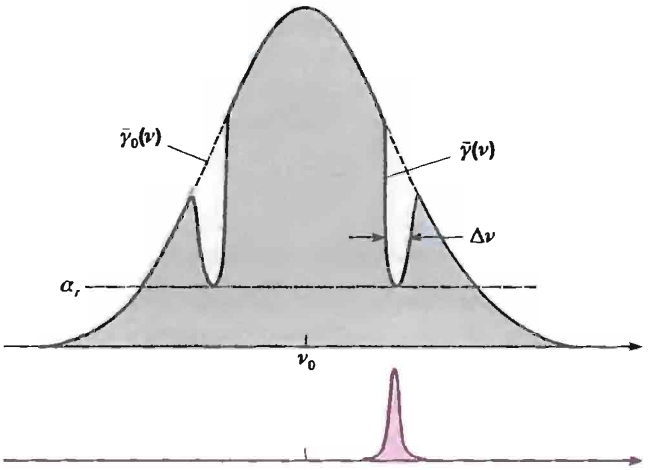


Рис. 15.15. Выжигание провалов в среде с доплеровским уширением. Пробная волна на частоте ν_q вызывает насыщение в атомах, движущихся со скоростями $v = \pm c(\nu_q/\nu_0 - 1)$, что приводит к выжиганию двух провалов в спектральном контуре усиления, симметричных относительно центральной частоты

Стационарная мощность моды растет с глубиной провалов в контуре усиления. При приближении частоты ν_q к ν_0 с любой стороны глубина провалов возрастает, как и мощность моды. Однако в непосредственной близости от ν_0 мода начинает взаимодействовать лишь с одной группой атомов вместо двух, так что два провала сливаются в один. Снижение числа активных атомов, доступных взаимодействию при $\nu_q = \nu_0$, вызывает небольшое уменьшение мощности моды. Таким образом, график зависимости мощности моды от ее частоты принимает вид колоколообразной кривой с понижением в центре, которое называется **провалом Лэмба** (рис. 15.16).

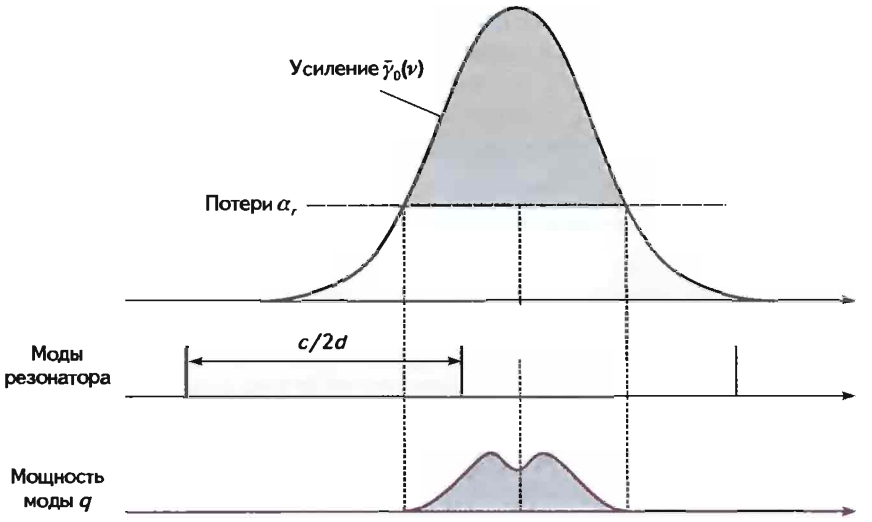


Рис. 15.16. Мощность одной лазерной моды с частотой ν_q в среде с доплеровским уширением профиля коэффициента усиления с центром на частоте ν_0. При $\nu_q = \nu_0$ вместо максимальной мощности наблюдается провал Лэмба

15.2.3. Пространственное распределение и поляризация

Пространственное распределение

Пространственное распределение лазерного излучения зависит от геометрии резонатора и формы активной среды. В теории лазера, рассмотренной до сих пор, мы игнорировали пространственные поперечные эффекты, предполагая, что резонатор образован двумя параллельными плоскими зеркалами бесконечной протяженности, и все пространство между ними заполнено активной средой. В этой идеализированной геометрии на выходе лазера получается плоская волна, бегущая вдоль оси резонатора. Однако из гл. 10 известно, что такой резонатор крайне чувствителен к юстировке.

Лазерные резонаторы обычно имеют сферические зеркала. Как отмечалось в разд. 10.2, моды резонатора со сферическими зеркалами представляют собой гауссовы пучки, подробно изученные в гл. 3. Следовательно, выходное излучение лазера со сферическими зеркалами может иметь вид гауссова пучка.

Мы также показали в подразд. 10.2.4, что резонатор со сферическими зеркалами поддерживает иерархию поперечных электрических и магнитных мод, обозначаемых $\text{TEM}_{l,m,q}$. Каждая пара индексов (l, m) определяет поперечную моду, имеющую свое пространственное распределение. Поперечная мода (0, 0) — гауссов пучок (рис. 15.17). Моды с более высокими l и m — пучки Гаусса—Эрмита (см. разд. 3.3 и рис 3.23). При заданных l и m индекс q определяет номера

продольных (аксиальных) мод с одинаковым пространственным распределением, но различными частотами (которые всегда разделены продольным межмодовым расстоянием $\nu_F = c/2d$, независимо от l и m). Резонансные частоты двух наборов продольных мод, принадлежащих разным поперечным модам, в общем случае смещены друг относительно друга на некоторую долю межмодового расстояния ν_F [см. (10.71)].

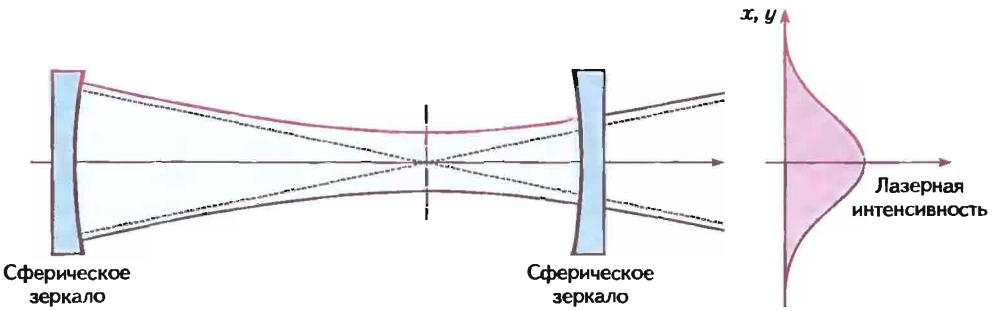


Рис. 15.17. Выходное излучение лазера на поперечной моде (0, 0) резонатора со сферическими зеркалами имеет вид гауссова пучка

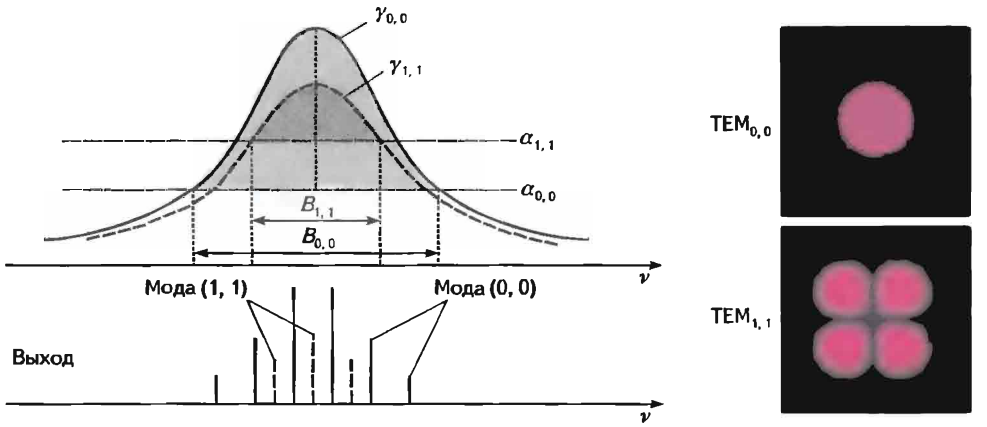


Рис. 15.18. Усиление и потери для двух поперечных мод, скажем, (0, 0) и (1, 1), обычно различны из-за различного пространственного распределения интенсивности. Мода может вносить вклад в выходной сигнал, если она лежит в полосе частот B, где усиление преобладает над потерями. Показаны разрешенные продольные моды, ассоциированные с каждой из поперечных мод

Из-за различных пространственных распределений различные поперечные моды испытывают различное усиление и потери. Например, гауссова мода (0, 0) сосредоточена вблизи оси резонатора и, следовательно, имеет наименьшие дифракционные потери на границах зеркал. Мода (1, 1) исчезает в точках оптической оси (см. рис. 3.23), поэтому если заблокировать центр зеркала небольшим непрозрачным экраном, то на моду (1, 1) это практически не повли-

яет, тогда как потери моды (0, 0) резко увеличатся. Моды высших порядков занимают больший объем и поэтому могут испытывать большее усиление. Отсутствие соответствия между усилением и потерями различных поперечных мод при различных геометриях определяет их конкуренцию и вклад в общее излучение лазера, как показано на рис. 15.18.

В лазерах с однородным уширением наиболее сильная мода стремится подавить усиление остальных мод. Однако выжигание пространственных провалов создает возможность генерации на нескольких модах. Поперечные моды могут иметь существенно различные пространственные распределения, поэтому легко могут генерироваться одновременно. Мода, энергия которой сконцентрирована в данной пространственной области, насыщает усиление среды в этой области, выжигая в ней пространственный провал. Две поперечные моды, которые пространственно не перекрываются, могут сосуществовать без конкуренции, поскольку они черпают энергию от разных атомов. Частичное перекрывание между различными поперечными модами, а также миграция атомов (как в газах) вызывают конкуренцию мод.

Лазеры часто проектируются для работы на одной поперечной моде; обычно это гауссова мода (0, 0), поскольку она имеет наименьший диаметр пучка и может фокусироваться в пятно наименьших размеров (см. гл. 3). С другой стороны, может быть желательна генерация высших мод, например, для получения большей оптической мощности.

Поляризация

Каждая мода (l, m, q) имеет две степени свободы, соответствующие двум независимым ортогональным поляризациям. Одни рассматриваются как две независимые моды. В силу осевой симметрии резонатора со сферическими зеркалами две поляризационные моды с данными l и m имеют одинаковые пространственные распределения. Если резонатор и активная среда обеспечивают одинаковое усиление и потери для обеих поляризаций, то лазер будет генерировать на обеих модах одновременно и с одинаковой интенсивностью. В этом случае излучение лазера не поляризовано (см. разд. 11.4).

Неустойчивые резонаторы

Хотя наше обсуждение сосредоточено на конфигурациях лазеров, использующих устойчивые резонаторы (см. рис. 10.12), необходимо заметить, что для работы лазеров высокой мощности ряд преимуществ имеют **неустойчивые резонаторы**.

1. В выходную мощность лазера дает вклад большая часть активной среды, поскольку достигается больший объем моды.

2. Более высокие мощности достигаются при работе на поперечной моде низшего порядка, а не высших порядков, как в устойчивых резонаторах.

3. Высокая выходная мощность достигается с минимальным оптическим повреждением зеркал в результате применения полностью отражающей опти-

ки, позволяющей излучению выходить из резонатора, огибая края зеркал (такая конструкция позволяет также организовать водяное охлаждение оптики и обеспечить ее устойчивость по отношению к более высоким мощностям).

15.2.4. Селекция мод

Многомодовый лазер можно заставить генерировать одну моду с помощью внутрирезонаторных элементов, вносящих дополнительные потери для подавления нежелательных мод.

Выбор лазерной линии

Активная среда с множеством переходов (атомных линий) с инверсией заселенностей, обеспечиваемой механизмом накачки, будет давать на выходе многочастотное излучение. Отдельную линию для генерации можно выбрать путем помещения призмы внутрь резонатора, как схематически показано на рис. 15.19. Призма устанавливается так, чтобы только свет с нужной длиной волны падал на зеркало резонатора с высоким коэффициентом отражения нормально и полностью отражался назад, обеспечивая обратную связь. Путем поворота призмы каждый раз можно выбирать одну линию. Например, ионные аргоновые лазеры часто имеют внутри резонатора поворотные призмы, осуществляющие выбор одной из шести лазерных линий в области от 488 нм в синей части спектра до 514,5 нм в сине-зеленой. Призма годится только для выбора линий, достаточно удаленных от других линий. Например, с помощью призмы нельзя выбрать одну продольную моду: соседние моды расположены так близко, что дисперсия преломления призмы не позволяет их различить.



Рис. 15.19. Отдельная лазерная линия может быть выбрана с помощью призмы, помещенной внутрь резонатора. Поперечная мода может быть выбрана с помощью диафрагмы тщательно подобранного размера и формы

Выбор поперечной моды

Различные поперечные моды имеют различные пространственные распределения, поэтому для селективного ослабления нежелательных мод можно использовать помещенную внутрь резонатора диафрагму специально подобранной формы (см. рис. 15.19). Зеркала лазера также можно спроектировать так, чтобы создать преимущества для одной из поперечных мод.

Выбор поляризации

Для преобразования неполяризованного света в поляризованный можно использовать поляризатор. Более выгодно поместить его не снаружи, а внутри резонатора. При использовании внешнего поляризатора напрасно теряется половина мощности, генерируемой лазером. Свет, прошедший через внешний поляризатор, может также приобретать шумовую составляющую из-за флуктуаций мощности между двумя поляризационными модами (перескок моды). Внутренний поляризатор создает высокие потери для одной поляризации, так что генерация соответствующей моды вообще не происходит. Атомное усиление в этом случае полностью приходится на оставшуюся поляризационную моду. Внутренний поляризатор обычно реализуется с помощью брюстеровских окон (см. разд. 6.2 и упражнение 6.5), как показано на рис. 15.20.

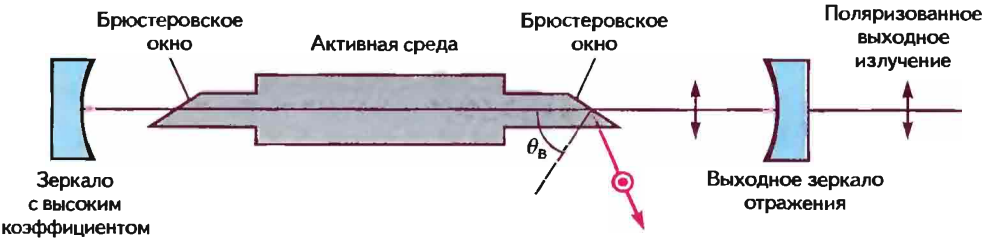


Рис. 15.20. Использование брюстеровских окон в газовом лазере позволяет получить линейно поляризованный лазерный пучок. Свет, поляризованный в плоскости падения (ТМ-волна), проходит без отражения через окно, расположенное под углом Брюстера. Ортогонально поляризованная (ТЕ) мода испытывает потери при отражении и не генерируется

Выбор продольной моды

Возможен также выбор одной продольной моды. Число продольных мод в лазере с неоднородным уширением (например, в газовом лазере с доплеровским уширением) равно числу мод резонатора, попадающих в полосу частот B, в которой усиление за счет активной среды превосходит потери (см. рис. 15.11). Существуют два альтернативных способа добиться генерации на одной продольной моде.

1. Увеличить потери настолько, чтобы условие генерации выполнялось только для одной моды с наибольшим коэффициентом усиления. Это, однако, означает, что и на оставшейся моде генерация будет слабой.

2. Увеличить межмодовое расстояние $\nu_F = c/2d$ путем уменьшения длины резонатора. Однако при этом уменьшатся длина и объем активной среды, а следовательно, и достижимая мощность лазера. В некоторых случаях такой способ практически непригоден. Например, у ионного аргонового лазера $\Delta\nu_D = 3{,}5$ ГГц. Поэтому если $B = \Delta\nu_D$ и $n = 1$, то

$$M = \frac{\Delta\nu_D}{c/2d},$$

так что резонатор должен быть короче 4,3 см для обеспечения генерации на одной продольной моде.

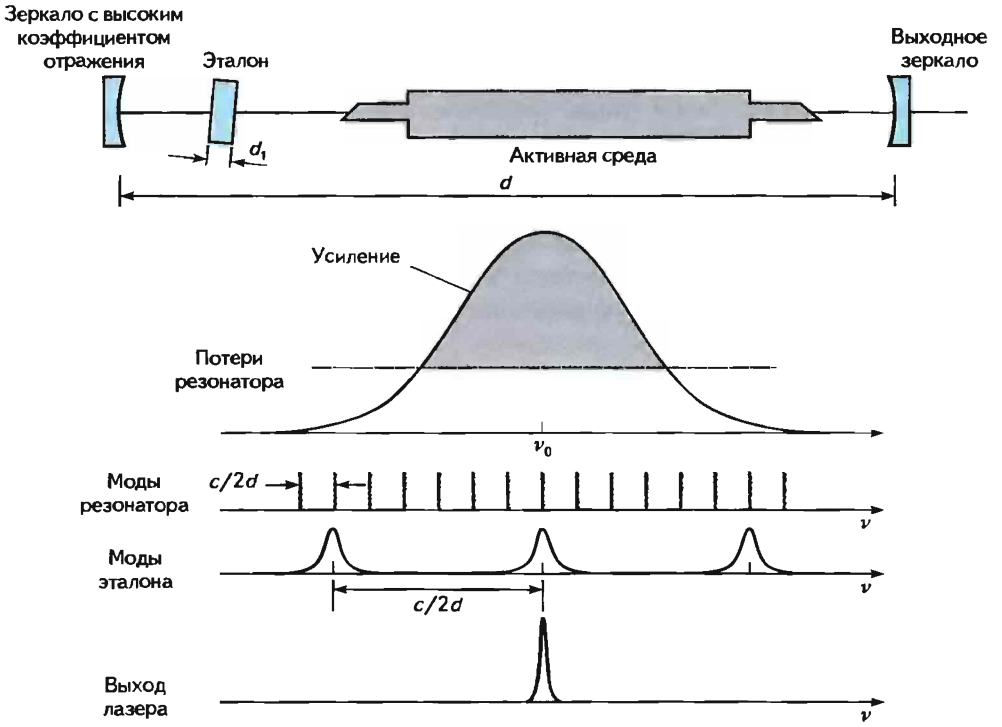


Рис. 15.21. Селекция продольных мод с помощью внутрирезонаторного эталона. Генерация происходит на тех частотах, на которых мода резонатора совпадает с модой эталона. Конечно, обе они должны попадать в полосу частот, где усиление превышает потери

Для изменения межмодового расстояния предложен ряд методов, использующих внутрирезонаторные частотно-селективные элементы.

• *Внутрирезонаторный наклонный эталон* (резонатор Фабри—Перо) с расстоянием между зеркалами d_1 намного меньше длины лазерного резонатора можно использовать для селекции мод (рис. 15.21). Моды эталона разделены большим межмодовым расстоянием $c/2d_1 > B$, так что внутрь полосы лазерного усилителя попадает только одна мода эталона. Эталон проектируется таким образом, чтобы одна из его мод совпадала по частоте с продольной модой лазерного резонатора, на которой усиление максимально (или с любой другой заданной модой). Тонкая настройка эталона обеспечивается за счет его поворота, изменения температуры или небольшого изменения толщины d_1 с помощью пьезоэлектрического (или иного) преобразователя. Эталон слегка наклонен по отношению к оси резонатора, чтобы предотвратить попадание излучения, отраженного от его поверхностей, на зеркала лазерного резонатора и

избежать появления паразитных резонансов. Чтобы гарантировать стабильность частоты, эталон обычно стабилизируют по температуре.

• *Многозеркальные резонаторы* также можно использовать для селекции мод. Некоторые их конфигурации показаны на рис. 15.22. Селекция мод может быть достигнута с помощью двух связанных резонаторов различной длины (рис. 15.22, *а*). Резонатор на рис. 15.22, *б* состоит из двух связанных резонаторов, каждый со своим усилением — по существу, это два связанных лазера. Именно такова конфигурация полупроводникового лазера со связанными резонаторами, образованными сколом (C^3-лазера), обсуждаемого в гл. 17. В другом методе используется резонатор, связанный с интерферометром (рис. 15.22, *в*). Теория связанных резонаторов и резонаторов, связанных с интерферометрами, выходит за рамки нашего рассмотрения.

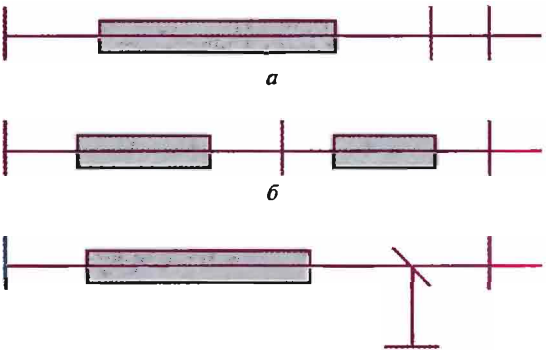


Рис. 15.22. Селекция продольных мод с помощью двух связанных резонаторов (пассивного и активного) (*а*); двух связанных активных резонаторов (*б*); резонатора, связанного с интерферометром (*в*)

15.3. РАСПРОСТРАНЕННЫЕ ЛАЗЕРЫ

Лазерное усиление и генерация распространены повсеместно и осуществляются во множестве различных сред, включая твердые тела (кристаллы, стекла, волокна, порошки), газы (атомные, ионные, молекулярные, эксимерные) и жидкости (растворы органических красителей). В плазме возможна лазерная генерация вакуумного ультрафиолетового и рентгеновского излучения. Энергетические уровни электронов в магнитном поле служат активной средой лазера на свободных электронах. Мы приведем несколько примеров лазеров, относящихся к этим категориям.

Размеры лазеров меняются от нанометров до размеров футбольного поля и даже до размеров астрономических тел. Принцип мазера охватывает огромный диапазон электромагнитных частот, более 18 порядков от 1 ГГц в микроволновой области до 100 ПГц в рентгеновском диапазоне. Разброс спектральных ширин лазерных линий достигает 12 порядков, от герц до терагерц. Лазеры генерируют импульсы с длительностью от фемтосекунд и до непрерывного действия, а диапазон их пиковых мощностей составляет примерно 27 порядков, от пиковатт до петаватт.

15.3.1. Твердотельные лазеры

Схемы энергетических уровней некоторых твердых лазерных материалов (рубин, александрит, Nd^{3+}:YAG и стекло с Nd^{3+}) приведены в подразд. 13.1.3 (см. рис. 13.8 и 13.9), а действие нескольких твердотельных усилителей (рубин, стекло с Nd^{3+} и кварцевое волокно с Er^{3+}) обсуждалось в разд. 14.3 (см. рис. 14.14, 14.16 и 14.19 соответственно). Характеристики основных лазерных переходов в этих и других активных средах сведены в табл. 14.1.

При помещении в резонатор, обеспечивающий обратную связь, все эти твердотельные материалы дают лазерную генерацию. Существует много разновидностей твердотельных лазеров, поскольку десятки прозрачных сред используются в качестве базовых веществ для введения различных активных примесных ионов. Кристаллические базовые вещества включают окислы, гранаты, фториды и ванадаты; наиболее типичными примерами являются Al_2O_3 (сапфир), $Y_3Al_5O_{12}$ (иттрий-алюминиевый гранат, YAG), $Gd_3Ga_5O_{12}$ (гадолиний-галлиевый гранат, GGG), $YLiF_4$ (иттрий-литиевый фторид) и YVO_4 (ванадат иттрия, также известный как ортованадат иттрия). В качестве базовых веществ широко используется также множество видов стекла, включающее композиции на основе силикатов (например, некристаллический SiO_2 — плавленый кварц) и фосфатов, которые предпочтительны для приложений в импульсных лазерах высокой мощности (см., например, подразд. 14.3.2).

Сравнение характеристик лазеров на основе кристаллических и аморфных базовых сред показывает, что первые из них имеют меньшие ширины линий (и, соответственно, более низкий порог генерации), более низкий уровень легирования, повышенную устойчивость к соляризации (потемнению, вызванному ультрафиолетовой составляющей света лампы-вспышки) и более высокую теплопроводность. С другой стороны, стекла в качестве базовых сред имеют ряд очевидных преимуществ: они изотропны, легко изготавливаются с оптическим качеством, однородно легируются, сохраняют оптическую полировку и легко выращиваются больших размеров (см. подразд. 14.3.2). Будучи плохими проводниками тепла, лазерные среды на базе стекол, в основном, применяются в системах, работающих при очень высоких мощностях и больших скважностях импульсов. Уширение линий и времена жизни уровней активных центров часто управляются колебательными характеристиками базовой среды; кристаллические среды в типичном случае создают однородное, а стекла — неоднородное уширение

Львиная доля легирующих ионов, используемых как активные лазерные центры в базовых кристаллах, — это переходные металлы и редкоземельные металлы (лантаниды), однако иногда используются и ионы актинидов (см. рис. 13.3). Легирующие ионы обычно равномерно распределены по объему базовой среды и действуют как независимые излучатели, примерно так же, как ведут себя ионы органических красителей в растворах. Концентрации легирующей примеси, как правило, лежат в окрестности 1 %, однако они могут быть от 0,01 до 50 % в зависимости от типа примеси, базового материала и примене-

ния. Для минимизации напряжений базовый материал обычно выбирается так, чтобы ион активной добавки был совместим по атомным размерам с замещаемым атомом.

Среди этого огромного множества комбинаций наиболее часто встречающиеся твердотельные лазеры — это $Nd^{3+}:YVO_4$, $Nd^{3+}:YAG$, $Yb^{3+}:YAG$, титан-сапфировый, на кварцевом волокне, легированном Er^{3+} и Yb^{3+}; мы рассмотрим их по очереди. Многие другие важные твердотельные лазерные среды также относятся к семейству диэлектриков, легированных редкоземельными ионами. Они включают $Er^{3+}:YAG$, $Ho^{3+}:YAG$, $Tm^{3+}:YAG$, а также оптическое волокно, легированное тулием. Как говорилось в подразд. 13.1.3, энергетические уровни редкоземельных ионов (без учета тонкой структуры) существенно не зависят от базового материала, поскольку $4f$-электроны хорошо экранированы от решетки заполненными подоболочками $5s$ и $5p$ (см. табл. 13.1). Несмотря на то, что он был первым материалом для лазера, рубин сейчас используется редко. Лазер на александрите находит специальное применение в дерматологии.

Лазеры на твердом теле с оптической накачкой от лазерных диодов (**твердотельные лазеры с диодной накачкой, ТЛДН**) преобразуют сравнительно широкополосное многомодовое излучение лазерных диодов в узкополосное одномодовое выходное излучение твердотельного лазера. Они компактны, высокоэффективны и позволяют получать множество различных длин волн. Нелинейное увеличение частоты в два, три или четыре раза (см. гл. 21) часто используется для преобразования излучения в видимый или ультрафиолетовый свет с еще большим разнообразием длин волн. Лазеры на твердом теле находят широкое применение в промышленности, медицине и науке.

Ванадат иттрия, легированный неодимом

$Nd^{3+}:YVO_4$ — диэлектрический материал с показателем преломления $n \approx 2{,}0$. Базовый материал прозрачен в широкой области длин волн от 0,3 до 2,5 мкм. Энергетические уровни, отвечающие за лазерную генерацию, показаны на рис. 15.23. Оптическая накачка от полупроводникового лазерного диода на длине волны $\lambda_0 = 808$ нм заселяет уровень ${}^4F_{5/2}$ с энергией 1,53 эВ. Наборы лазерных диодов обеспечивают высокую мощность накачки, как схематически показано на рис. 14.13, *г*.

Переход ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ отвечает за лазерную генерацию на 1,064 мкм. Это основной лазерный переход для данной среды. Однако генерация возможна и на переходах ${}^4F_{3/2} \rightarrow {}^4I_{13/2}$ и ${}^4F_{3/2} \rightarrow {}^4I_{9/2}$ с длинами волн 1,34 мкм и 914 нм соответственно. В последнем случае схема является квазитрехуровневой. Данный материал отличается от неодимового стекла (см. рис. 13.9) более высоким показателем преломления, однородным уширением и меньшей шириной линии перехода (см. табл. 14.1). Из-за четырехуровневой схемы накачки порог генерации намного ниже, чем у рубина. Вторая гармоника излучения $Nd^{3+}:YVO_4$-лазера на 532 нм часто используется для накачки титан-сапфирового лазера (см. рис. 15.26). Внутрирезонаторное удвоение частоты излучения, генерируемого на переходе ${}^4F_{3/2} \rightarrow {}^4I_{9/2}$, даст синий свет с длиной волны 457 нм.

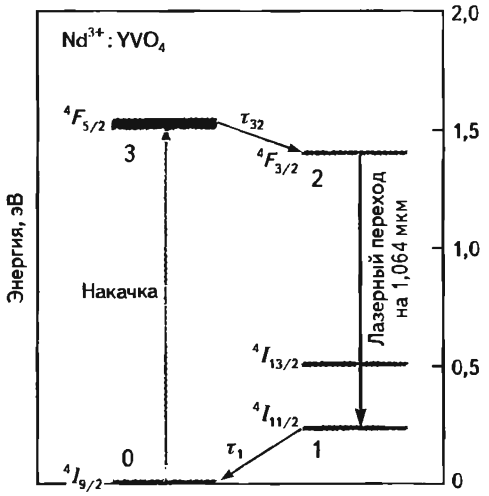


Рис. 15.23. Схема рабочих уровней $Nd^{3+}:YVO_4$. Красная стрелка — основной лазерный переход с длиной волны 1,064 мкм в ближней инфракрасной области. Четырс взаимодействующих уровня энергии отмечены цифрами (*а*). Конфигурация лазера на $Nd^{3+}:YVO_4$ с кристаллом трибората лития (LBO) внутри резонатора, служащим для нелинейного удвоения частоты и генерации света на $\lambda_0/2 = 532$ нм (см. подразд. 21.2.1) (*б*)

Иттрий-алюминиевый гранат с неодимом

Разработанный в 1960-е гг. $Nd^{3+}:YAG$, схема энергетических уровней которого показана на рис. 13.9, является одним из самых широко применяемых твердотельных лазерных материалов. Поскольку оптически активные 4*f*-электроны экранированы от базового вещества, схема его энергетических уровней такая же, как у стекла и ванадата иттрия с неодимом (см. рис. 13.9 и 15.23, соответственно). Nd^{3+} YAG-лазеры часто включают внутрирезонаторные нелинейные кристаллы для удвоения частоты, как показано на рис. 15.23 для $Nd^{3+}:YVO_4$. Хотя накачку $Nd^{3+}:YAG$ можно осуществлять лампой-вспышкой, наиболее удобной накачкой является лазерный диод на 808 нм, что позволяет создать компактный источник ближнего инфракрасного или зеленого света с питанием от батареи. Кристаллы длиной в несколько сотен мкм могут служить эффективными одночастотными лазерами в виде тонких лисков.

Чаще всего используется линия Nd^{3+} YAG лазера на $\lambda_0 = 1{,}06415$ мкм в ближней ИК-области. Уровни тонкой структуры трех состояний, связанных с этим лазерным переходом, показаны на рис. 15.24. Данная лазерная линия соответствует переходу между верхним уровнем тонкой структуры состояния $^4F_{3/2}$ с энергией 1,4269 эВ и третьим снизу уровнем тонкой структуры состояния $^4I_{11/2}$ с энергией 0,2616 эВ. При удвоении частоты этот переход дает известное зеленое излучение на 532 нм.

Число подуровней тонкой структуры каждого состояния равно $g/2$, где $g = 2J + 1$. Величина g — **параметр вырождения**, а J — квантовое число полного углового момента, которое содержится в символе терма $^{2S+1}\mathfrak{L}_J$, введенном в подразд. 13.1.1. Из рис. 15.24 видно, что переходы между различными подуровнями тонкой структуры верхнего и нижнего состояний порождают множество

лазерных линий, длины волн которых лежат в интервале от 1,052 до 1,122 мкм. В частности, генерация может достигаться на $\lambda_0 = 1{,}12238$ мкм, что соответствует переходу между нижним из двух подуровней состояния $^4F_{3/2}$ с энергией 1,4165 эВ и верхним подуровнем состояния $^4I_{11/2}$ с энергией 0,3117 эВ. Это наибольшая длина волны для данной группы переходов. При удвоении частоты получается желто-зеленый свет с $\lambda_0 = 561$ нм. Nd^{3+} : YAG может генерировать также по квазитрехуровневой схеме с рабочим переходом $^4F_{3/2} \rightarrow {}^4I_{9/2}$. При этом получается излучение на длине волны 946 нм, которое после внутрирезонаторного удвоения частоты дает синий свет на 473 нм. Специально разработанные фотонные кристаллы могут использоваться как фильтры, подавляющие генерацию на преобладающем переходе. Имеется большое число других возможностей, поскольку генерация может происходить на многих из указанных переходов.

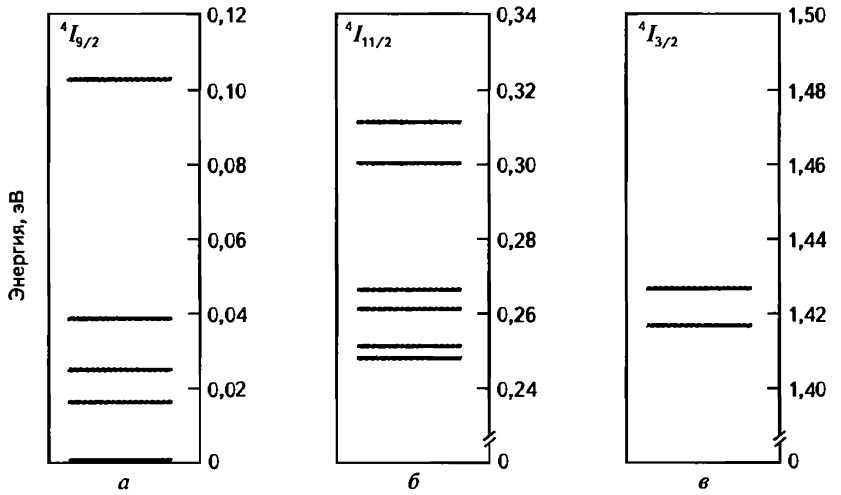


Рис. 15.24. Тонкая структура энергетических уровней трех состояний, связанных с лазерными переходами Nd^{3+} : YAG вблизи 1,06 мкм в ближней ИК-области (см. рис. 13.9):

а — основное состояние $^4I_{9/2}$; *б* — нижний лазерный уровень $^4I_{11/2}$; *в* — верхний лазерный уровень $^4F_{3/2}$. Число отдельных подуровней для указанных состояний равно $(2J + 1)/2 = 5$, 6 и 2 соответственно. Конкретные значения энергий зависят от базового материала — в стеклах уровни существенно расплываются

Принципиальными недостатками Nd^{3+} : YAG-лазера по сравнению с Nd^{3+} : YVO_4 является более узкая полоса поглощения с переходом в состояние $^4F_{5/2}$ (что повышает чувствительность к изменениям длины волны лазерного диода накачки), более высокий порог генерации, низкий дифференциальный коэффициент преобразования мощности и неполяризованное выходное излучение. Тем не менее Nd^{3+} : YAG продолжает оставаться основной «рабочей лошадью» из всех твердотельных лазеров с диодной накачкой.

Иттрий-алюминиевый гранат, легированный иттербием

Yb^{3+} YAG-лазеры в виде тонких дисков используют накачку на 940 нм от лазерного диода (рис. 15.25). Высокая эффективность поглощения накачки достигается многократным пропусканием ее через активную среду с помощью соответствующей оптики. Большое усиление достигается высоким содержанием Yb^{3+} в базовом веществе. Поскольку длина волны накачки, равная 940 нм, близка к лазерной длине волны, равной 1030 нм, тепловая нагрузка, приходящаяся на один фотон накачки, мала, поэтому в кристалле выделяется мало теплоты. Более того, геометрия в виде тонкого диска дает возможность эффективного теплоотвода, что позволяет поддерживать пространственную моду TEM_{00}. Несмотря на тот факт, что $Yb^{3+}:YAG$ — квазитрехуровневая система, лазеры в виде тонких дисков из этого материала способны генерировать сотни ватт в непрерывном режиме на длине волны 1,030 мкм. При удвоении частоты получается мощный источник зеленого света на 515 нм, который может заменить во многих приложениях значительно более сложный аргоновый лазер.

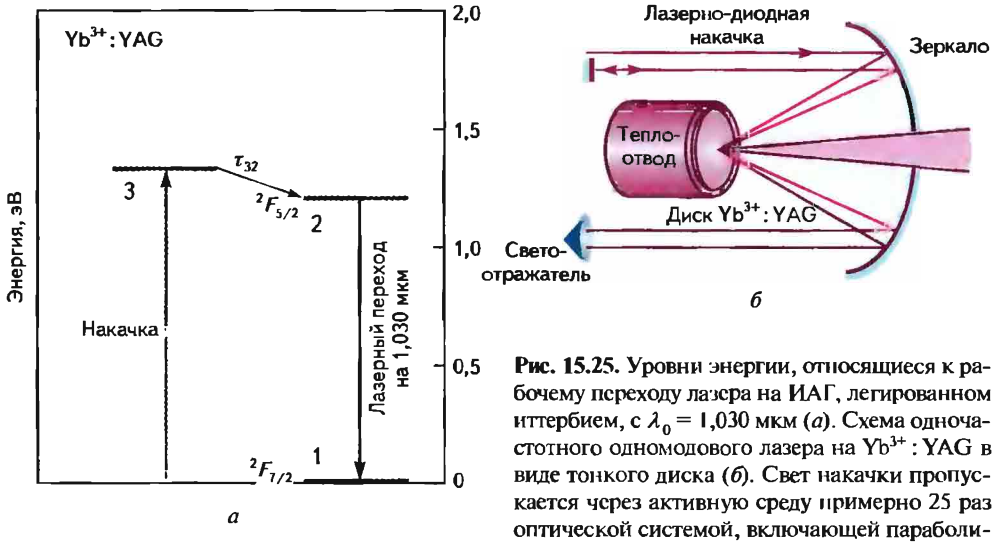


Рис. 15.25. Уровни энергии, относящиеся к рабочему переходу лазера на ИАГ, легированном иттербием, с $\lambda_0 = 1{,}030$ мкм (*а*). Схема одночастотного одномодового лазера на $Yb^{3+}:YAG$ в виде тонкого диска (*б*). Свет накачки пропускается через активную среду примерно 25 раз оптической системой, включающей параболическое зеркало и световозвращающий отражатель. Высокое усиление около 25 % достигается применением примесных уровней Yb^{3+}

Титан—сапфир

Лазер на сапфире, легированном ионами Ti^{3+}, широко применяется, поскольку его длина волны перестраивается в широком диапазоне. Другой его особенностью является возможность синхронизации мод для генерации ультракоротких импульсов (см. подразд. 15.4.4). По мере роста кристалла сапфира небольшая часть ионов Al (~1 %) замещается ионами Ti. Как и рубин, данный материал в принципе является сапфиром и, следовательно, имеет пока-

затель преломления $n \approx 1{,}76$. Оптическая накачка обычно осуществляется второй гармоникой $Nd^{3+}:YVO_4$- или $Nd^{3+}:YAG$-лазера на 532 нм (см. рис. 15.23), ионным аргоновым лазером или второй гармоникой $Yb^{3+}:YAG$-лазера на 515 нм (см. рис. 15.25), либо непосредственно зеленым светом диодного лазера.

Каждый ион титана, имеющий один активный электрон $3d^1$ (см. табл. 13.1), окружен шестью атомами кислорода в позиции октаэдрической решетки. Поэтому этот ион испытывает значительное действие кристаллического поля. Как и для других ионов переходных металлов в диэлектрических базовых средах, уровни энергии для ионов титана в сапфире, показанные на рис. 15.26, обозначаются символами представлений группы симметрии, а не символами термов (см. подразд. 13.1.3). Электронные уровни энергии сильно связаны с колебаниями решетки, что приводит к появлению широких энергетических зон вибронных состояний. Поэтому вынужденное излучение сопровождается одновременным испусканием одного или нескольких фононов. Заселенности вибронных уровней состояния $^2T_2$ подчиняются распределению Больцмана, поэтому верхние уровни почти не заселены и система ведет себя как четырехуровневый лазер, что иллюстрирует рис. 15.26, *а*.

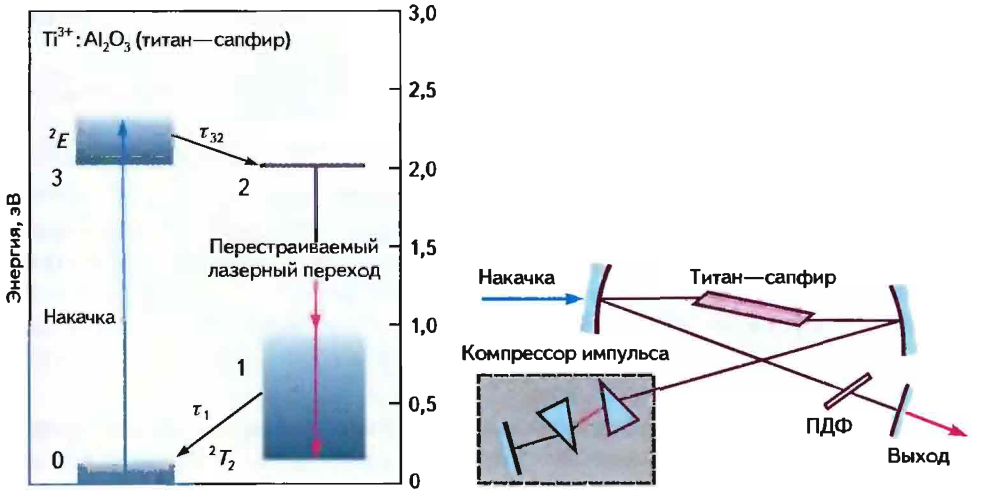


Рис. 15.26. Избранные энергетические зоны $Ti^{3+}:Al_2O_3$ (*а*). Красная стрелка — основной лазерный переход в этой вибронной системе, перестраиваемый между 700 и 1050 нм. Изменение окраски внутри каждой зоны от темной к светлой отражает уменьшение относительной заселенности уровней с повышением энергии. Схема $Ti^{3+}:Al_2O_3$-лазера с синхронизацией мод (*б*). Две призмы внутри блока, очерченного штриховой линией, обеспечивают внутрирезонаторную компенсацию дисперсии. Перестройка длины волны в пределах десятков нанометров достигается с помощью поворотного двулучепреломляющего фильтра (ПДФ), который действует как полосовой фильтр для поляризованного пучка внутри резонатора. Грубая перестройка осуществляется установкой одной из призм. Зеленый свет для накачки часто получают в виде второй гармоники излучения лазера на $Nd^{3+}:YVO_4$, схема которого показана на рис. 15.23

Частоту лазерного перехода, показанного красной стрелкой на рис. 15.26, *а*, можно перестраивать в пределах нескольких десятков нанометров с помощью поворотного установленного под углом Брюстера внутри резонатора двулучепреломляющего фильтра (рис. 15.26, *б*), который действует как полосовой фильтр для поляризованного пучка. Большие изменения длины волны достигаются настройкой внутренней оптики, поскольку дисперсия групповой скорости в резонаторе меняется с длиной волны. Сочетая эти методы, можно получать генерацию в широком диапазоне длин волн от 700 нм в красной области до 1050 нм в ближней инфракрасной. Лазер на $Ti^{3+}:Al_2O_3$ может обеспечивать оптическую мощность ≈5Вт в непрерывном режиме, а при синхронизации мод генерировать последовательность импульсов длительностью 10 фс с энергией 50 нДж, частотой повторения ≈80 МГц и пиковой мощностью ≈1 МВт.

По причине важности колебаний решетки для перестройки частоты лазера на $Nd^{3+}:YVO_4$ он называется **лазером на фононно-ограниченном переходе**, или **вибронным лазером**. Лазер на александрите (см. рис. 13.8) тоже относится к этому классу, как и лазеры на красителях (см. подразд. 15.3.3), поскольку в них молекулярные колебания играют ту же роль, что и колебания решетки. В общем случае вибронный переход сопровождается одновременным изменением электронного и колебательного состояния системы.

Волоконные лазеры

При наличии подходящей обратной связи волокна, легированные редкоземельными ионами, действуют как высокоэффективные волоконные лазеры в диапазоне от видимого до среднего инфракрасного. Упрощенная схема, иллюстрирующая использование накачки от диодного лазера и волоконных отражателей в виде решеток Брэгга, приведена на рис. 15.27, *а*. Чтобы избежать нелинейных эффектов, сопутствующих высокой концентрации мощности накачки в малом объеме сердцевины волокна, широко используются волокна с двойной оболочкой (рис. 15.27, *б*).

Особенно хорошими характеристиками обладают лазеры на основе кварцевого волокна с многослойной оболочкой, легированного иттербием. Накачка от лазерного диода вводится в активную среду через многомодовые волокна, сращенные с бухтой волокна с многослойной оболочкой. Обратная связь обеспечивается **волоконными решетками Брэгга**. Выходная мощность намного больше 1 кВт в непрерывном режиме достигается в диапазоне длин волн 1070—1080 нм с шириной линии ≈1 ТГц. Выход излучения осуществляется через одномодовое волокно с диаметром сердцевины в несколько микрометров. Качество пучка превосходное ($\mathbb{M}^2 < 1{,}1$), как и полный коэффициент преобразования мощности ($\eta_c > 25\,\%$). Линейная поляризация, уменьшенная ширина линии и более широкий диапазон длин волн могут быть достигнуты при меньших уровнях мощности. С другой стороны, намного большие мощности могут быть достигнуты в многомодовом режиме; мощности до 50 кВт можно получать для таких приложений, как плавление, резка и сверление.

Лазеры на кварцевом волокне, легированном эрбием, обеспечивают мощности в сотни ватт в непрерывном режиме в диапазоне длин волн 1550—1570 нм при ширине линии <400 ГГц. Качество пучка приближается к дифракционному пределу ($M^2 < 1{,}1$), а полный коэффициент преобразования мощности $\eta_c > 10\,\%$. Эти лазеры с диодной накачкой компактны и охлаждаются воздухом, они могут работать со случайной, линейной или круговой поляризацией. Более широкий диапазон перестройки длины волны от 1530 до 1620 нм доступен при снижении уровня мощности до десятков ватт.

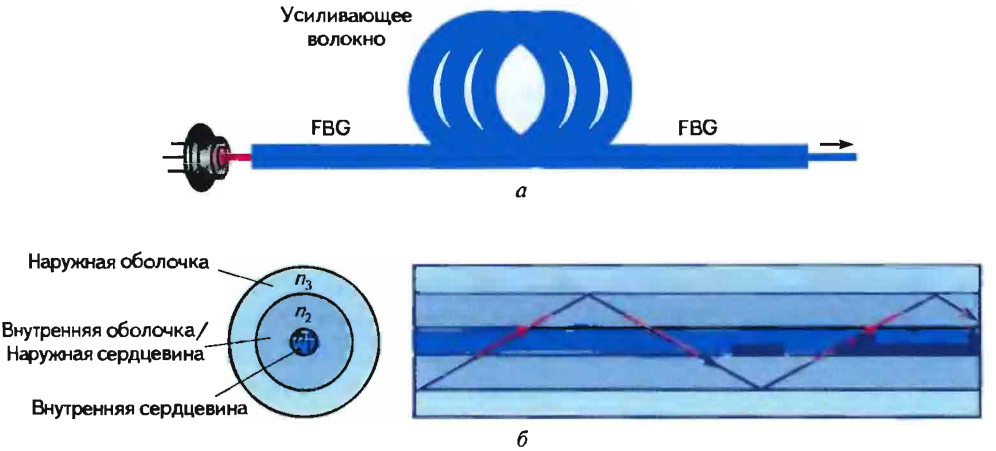


Рис. 15.27. Упрощенная схема волоконного лазера с накачкой от диодного лазера и с волоконными решетками Брэгга (FBG) в качестве отражателей (*а*). Для накачки часто используется набор из большого числа многомодовых лазерных диодов большой площади, излучение которых вводится во внешнюю оболочку волокна через многомодовые соединители как в прямом, так и в обратном направлении. Одномодовая внутренняя сердцевина способствует генерации одной поперечной моды. Существует множество других конфигураций волоконных лазеров. Концентрическая конфигурация волокна с двойной оболочкой (*б*). Проектируются и другие конфигурации волокон с двойной оболочкой с целью обеспечить лучшее перекрывание между внутренней оболочкой и косыми лучами внешней оболочки (см. рис. 9.2). Например, внутренняя сердцевина может быть сдвинута от оси к краю наружной оболочки, либо внешняя оболочка может быть прямоугольной, шестиугольной, восьмиугольной или иметь профиль буквы D

На еще больших длинах волн 1,8—2,1 мкм лазеры на волокне, легированном туллием, дают оптическую мощность 150 Вт при ширине линии ≈75 ГГц. Типичный коэффициент преобразования составляет $\eta_c \approx 5\,\%$, а качество пучка превосходно ($M^2 < 1{,}05$). И здесь линейная поляризация достигается ценой уменьшения мощности. На длине волны 1,94 мкм излучение туллиевого лазера попадает в область поглощения воды, находящейся в мягких тканях, поэтому этот лазер представляет интерес для клинической медицины.

Лазеры на легированных волокнах имеют широкое применение, область которого простирается от обработки материалов и хирургии до затравки усилителей в Национальном комплексе зажигания (см. подразд. 14.3.2). Волоконные лазеры могут работать при больших средних мощностях в режимах модуляции добротности и синхронизации мод (см. разд. 15.4). Более того, **фотонно-кристаллический волоконный лазер** можно сделать так, чтобы свет излучался радиально во все стороны, а не вдоль оси, что открывает новые возможности формирования и демонстрации изображений[1]. Сравнение волоконных лазеров и лазерных диодов проводится в подразд. 17.3.3.

Волоконные ВКР-лазеры

Волоконные ВКР-лазеры работают на основе вынужденного комбинационного рассеяния (ВКР) — процесса, первоначально рассмотренного в подразд. 13.5.3, а затем в подразд. 14.3.4 в связи с работой волоконных ВКР-усилителей. Вынужденное комбинационное рассеяние схематически показано на рис. 14.20. Сигнальный фотон с энергией $h\nu_s$ стимулирует испускание такого же точно фотона, который получается за счет стоксова сдвига частоты фотона накачки на величину колебательной энергии $h\nu_R$, так что энергия испускаемого сигнального фотона точно равна энергии первичного. Оптическое усиление зависит от коэффициента усиления ВКР g_R [см. (21.55)], а ширина полосы усиления определяется колебательным спектром базового стекла (см. подразд. 14.3.4 и рис. 14.20).

Так же как волоконный усилитель с редкоземельными активными ионами превращается в волоконный лазер при введении обратной связи, как показано на рис 15.27, так и волоконный ВКР-усилитель превращается в волоконный ВКР-лазер. Волоконные решетки Брэгга служат отражателями и образуют резонатор, способствующий генерации на тех частотах, на которых коэффициент отражения велик (см. подразд. 7.1.3). Частота генерации ν сдвинута относительно частоты накачки ν_p на частоту стоксова сдвига ν_R, которая может принимать значения только в пределах колебательного спектра базового стекла, как показано на рис. 14.20.

Уникальная особенность комбинационного рассеяния состоит в том, что генерируемая стоксова частота привязана к частоте накачки. Вследствие этого излучение на стоксовой частоте, генерируемое в резонаторе, образованном парой волоконных решеток Брэгга FBG1, можно снова использовать для накачки того же волокна. Как показано на рис. 15.28, *а*, эта вторая накачка с пониженной частотой $\nu_{p1} = \nu_p - \nu_R$ может вызвать генерацию стоксовой компоненты второго порядка, которая устанавливается на частоте максимального отражения ν_{p2} второй пары волоконных решеток Брэгга FBG2. Каскад можно продолжать, используя вложенные пары брэгговских решеток, пока он не прервется

---

[1] См. *Shapira O., Kuriki K., Orf N.D., Abouraddy A.F., Benoit G., Viens J.F., Rodriguez A., Ibanescu M., Joannopoulos J.D., Fink Y., Brewster M.M.*. Surface-Emitting Fiber Lasers. Optics Express. Vol. 14, 2006. P. 3929—3935.

выходным соединением, направляющим свет из волокна на желаемой частоте. Волоконные ВКР-лазеры с большим числом стоксовых сдвигов, известные как **каскадные ВКР-лазеры**, значительно расширяют доступный диапазон длин волн.

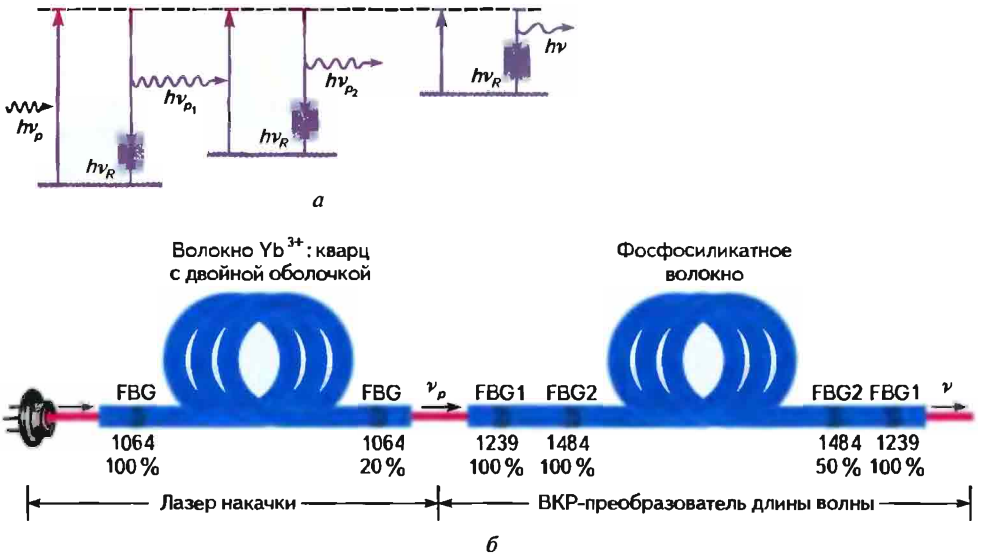


Рис. 15.28. Каскад стоксовых сдвигов (*а*). Схема ВКР-лазера на фосфосиликатном волокне (*б*). Накачка осуществляется лазером на волокне Yb^{3+} : кварц с двойной оболочкой, который, в свою очередь, накачивается матрицей диодных лазеров. Числа под FBG — коэффициенты отражения по интенсивности на соответствующих длинах волн (нм)

Волоконный ВКР-лазер — одномодовый источник света высокой мощности, пригодный для множества приложений, включая обработку материалов, клиническую медицину и оптоволоконную связь. Эти лазеры особенно хорошо подходят для накачки волоконных ВКР-усилителей в многоканальных системах связи с высокой степенью спектрального уплотнения (DWDM-системах), а также для дистанционной накачки волоконных усилителей, легированных эрбием. Наиболее привлекательная черта этих лазеров заключается в том, что генерацию можно получать в широком диапазоне длин волн путем подходящего выбора длины волны накачки, материала волокна и волоконных решеток Брэгга. При соответствующих условиях волоконный ВКР-лазер может непосредственно служить усилителем, если в него вводить сигнал, подлежащий усилению.

Пример 15.1

ВКР-лазер на фосфосиликатном волокне

ВКР-лазер можно построить с использованием лазера на волокне с двойной оболочкой, легированном иттербием. Этот лазер излучает в области 1050—1120 нм и служит накачкой для ВКР-преобразователя длины волны из фосфосиликат-

ного волокна длиной 1 км, как показано на рис. 15.28, *б*. Лазер накачки на кварцевом волокне с Yb^{3+}, накачка которого, в свою очередь, обеспечивается матрицей лазерных диодов, работает в окрестности 960 нм и излучает одномодовый свет, который можно ввести в одномодовый ВКР-преобразователь длины волны намного более эффективно, чем многомодовый свет от матрицы лазерных диодов непосредственно. В качестве частного примера предположим, что мы хотим преобразовать свет накачки на 1064 нм от иттербиевого лазера в свет большей длиной волны, например, 1484 нм, так что волоконный ВКР-лазер подходит для накачки волоконного ВКР-усилителя (см. подразд. 14.3.4). Поскольку мы хотим сдвинуть частоту волны на довольно большую величину 80 ТГц, используем фосфосиликатное волокно с большим стоксовым сдвигом $\nu_R \approx 40$ ТГц, позволяющее осуществить требуемое преобразование всего в две ступени. В качестве альтернативы мы могли бы использовать кварцевое волокно, легированное германием, но тогда потребовалось бы шесть ступеней, поскольку стоксов сдвиг этого материала намного меньше и составляет $\nu_R \approx 13$ ТГц, как видно из рис. 14.20. Как показано на рис. 15.28, *б*, первая пара волоконных решеток Брэгга, FBG1, используется для преобразования частоты накачки с длиной волны 1064 нм вниз на 40 ТГц, что дает 1239 нм, а вторая пара, FBG2, сдвигает частоту вниз еще на 40 ТГц и получается искомая длина волны 1484 нм. Решетки пары FBG1 имеют коэффициент отражения 100 %, а у одной из решеток пары FBG2 он понижен до 50 % с целью вывода излучения на 1484 нм. Компактный волоконный ВКР-лазер, подобный изображенному на рис. 15.28, *б*, имеет выходную мощность в десятки ватт в непрерывном режиме при ширине линии в несколько нм на любой желаемой длине волны в диапазоне 1200—1700 нм. Полный коэффициент преобразования мощности у лазера накачки на волокне с иттербием составляет ≈25 %, а грубая оценка дифференциального коэффициента преобразования мощности для ВКР-преобразователя длины волны дает 0,25 Вт/Вт.

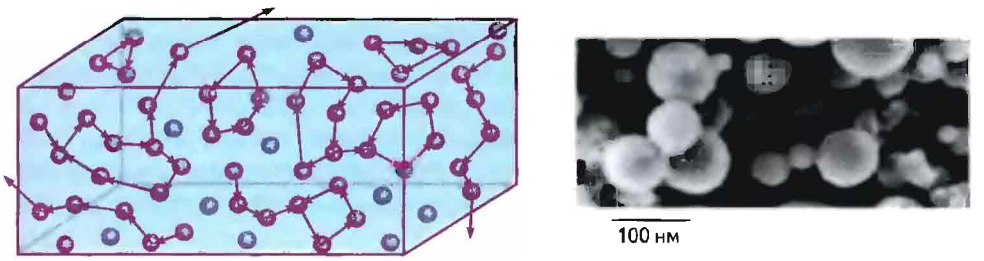
Хотя объемные ВКР-лазеры известны давно, именно в волоконном исполнении ВКР-методика добилась успеха. Причин для этого несколько. Волокна имеют большую длину, обеспечивая высокое усиление, выдерживают высокую интенсивность в малом сечении одномодовой сердцевины, допускают эффективную накачку твердотельными лазерами с накачкой от лазерных диодов, в них легко размещаются многочисленные брэгговские отражатели. Вынужденное рассеяние Мандельштама—Бриллюэна можно использовать аналогичным образом для создания волоконных ВМБР-лазеров.

Случайные лазеры

Как обсуждалось в разд. 15.1 и 15.2, частоты генерации обычных лазеров определяются модами резонатора Фабри—Перо и профилем усиления резонансных переходов активной среды. Свет, выходящий через частично отражающее зеркало, обычно имеет узкий спектр, сильную направленность и

высокую степень пространственной и временно́й когерентности. Рассеяние в активной среде вносит потери и его старательно избегают.

Однако, когда рассеяние в активной среде очень велико, оно само может служить источником обратной связи. Случайные лазеры работают на основе такой обратной связи, обеспечиваемой многократным рассеянием внутри неупорядоченной усиливающей среды, которая служит закрытым трехмерным резонатором. Можно считать, что фотоны внутри среды совершают случайные блуждания в трех измерениях (рис. 15.29, *а*). Поскольку сильное рассеяние связано с неупорядоченностью среды, лазеры такого типа называют **случайными лазерами**. Они называются также **порошковыми лазерами** или **плазерами**. В отличие от обычных лазеров излучение, отраженное назад в каждую точку активной среды, имеет случайную фазу. Таким образом, обратная связь некогерентна и осуществляется через интенсивность, в то время как для обычного лазера она когерентна и осуществляется через амплитуду поля. Так как в таких лазерах резонансная обратная связь отсутствует, центральная частота генерации определяется спектральным профилем усиления активной среды. По причине большой полной длины пробега фотона при многократном рассеянии в активной среде может достигаться значительное усиление. Чем сильнее рассеяние, тем больше обратная связь и, следовательно, ниже порог генерации. Обратная связь через рассеяние играет важную роль в наблюдаемой лазерной генерации астрономических объектов, например,молекулярных облаков H_2O, OH и SiO. Рассеяние претендует на обеспечение обратной связи в рентгеновском диапазоне, где зеркальное отражение достигается с трудом.



а *б*

Рис. 15.29. Случайный лазер основан на некогерентной и нерезонансной обратной связи за счет многократного рассеяния с большой длиной пробега фотона внутри активной среды (*а*). При повторяющемся рассеянии (схематически изображенном замкнутыми петлями) поле многократно проходит один или несколько локальных путей, каждый из которых служит локальным резонатором, обеспечивающим когерентную и резонансную обратную связь. В случайном микролазере плотно упакованные нанокристаллиты ZnO одновременно служат активной средой и рассеивающими элементами, обеспечивающими обратную связь (*б*)

Отличительными признаками случайного лазера является отсутствие направленности и пространственной когерентности испускаемого света. Пространственная картина лазерного излучения с торца кюветы, содержащей порошко-

вую активную среду, часто напоминает картину от светодиода с поверхностным излучателем (см. рис. 17.14, *а*). Однако по многим свойствам случайные лазеры похожи на обычные лазеры, в том числе по своему разнообразию. Их накачка может быть оптической, электрической или осуществляться электронным пучком. Генерация может происходить в широком диапазоне длин волн от инфракрасной до ультрафиолетовой области спектра. Размеры активных областей меняются от микрорезонаторов объемом 1 мкм^3 до макроскопических устройств с объемами в несколько кубических сантиметров.

Обычные твердотельные лазеры с такими активными средами, как рубин, Nd^{3+} : YAG, Nd^{3+} : стекло, Ti^{3+} : сапфир и GaAs, функционируют как случайные лазеры при измельчении активной среды в порошок. Во многих порошках усиливающая и рассеивающая среда одна и та же, но это не обязательно так. Например, усиление может обеспечиваться родамином 6Ж, а рассеяние — микрочастицами Al_2O_3 при помещении обоих в раствор метанола. Активные среды случайных лазеров включают неорганические диэлектрики, полимеры, жидкости, растворы красителей, жидкие кристаллы с примесью красителей, неупорядоченные полупроводниковые наноструктуры и даже биологические ткани.

Если частицы активной среды достаточно велики и имеют правильную форму, позволяющую поддерживать резонаторные моды, то они ведут себя не как случайный лазер, а как набор отдельных микролазеров, каждый со своим направлением излучения. В противном случае резонансы поддерживаются локальным расположением рассеивателей. Если рассеяние имеет возвратный характер (см. рис. 15.29, *а*) и оптическое усиление превосходит потери на замкнутом пути, последний может служить резонатором для генерации. Возникающее при этом лазерное излучение имеет острый пик на соответствующей случайной частоте моды. Поскольку в разных областях случайной среды наборы замкнутых путей различны, частоты генерации зависят от конкретной области материала, подвергаемой накачке. Такие лазеры называются **когерентными случайными лазерами**, так как обратная связь у них когерентна, а пространственная конфигурация резонаторов случайна. Такие лазеры аналогичны наборам микролазеров со случайными направлениями излучения.

Кластеры рассеивающих частиц можно использовать для изготовления **случайных микролазеров**, в которых свет ограничен объемом порядка куба длины волны за счет сильного рассеяния, а не отражения. Другие виды лазеров с микрорезонаторами рассмотрены в подразд. 17.4.2.

Пример 15.2

Случайный микролазер на ZnO

Плотно упакованная совокупность нескольких тысяч нанокристаллитов ZnO (см. рис. 15.29, *б*), каждый диаметром порядка десятков нанометров, может образовать микрокластер диаметром ≈1 мкм. Нанокристаллиты служат как усиливающей средой, так и рассеивающими элементами обратной связи случайного микролазера. Длина волны излучения $\lambda_0 \approx 380$ нм лежит вблизи длины

волны запрещенной зоны ZnO. Поскольку удержание света происходит за счет рассеяния, а не отражения от поверхности микрокластера, такие микролазеры не нуждаются в правильной форме и гладкой поверхности.

15.3.2. Газовые лазеры

Атомные и ионные лазеры

Атомные и ионные **газовые лазеры**, такие как He—Ne, Ar^+ и Kr^+, дают красивые разноцветные пучки, которые в течение десятков лет были основой оптических лабораторий (см. табл. 15.1). В частности, криптоновый ионный лазер обеспечивает оптическую мощность в сотни милливатт на длинах волн от $\lambda_0 = 350$ нм в ближнем ультрафиолете до 676 нм в красной области спектра. Он может одновременно генерировать на множестве линий и давать «белый лазерный свет». Множество других одноатомных веществ и их ионов также служат активными средами и генерируют на бесчисленном множестве длин волн в инфракрасном и видимом диапазоне. Тем не менее газовые лазеры в настоящее время используются только для специализированных приложений, так как твердотельные лазеры с диодной накачкой и диодные лазеры обладают лучшими рабочими характеристиками, легче перестраиваются и физически более надежны.

Молекулярные лазеры

Молекулярные газовые лазеры, такие как CO_2 лазер (см. табл. 15.1 и рис. 13.4), который генерирует вблизи $\lambda_0 = 9{,}6$ и 10,6 мкм в ближнем инфракрасном диапазоне, могут давать тысячи ватт в непрерывном режиме с высоким коэффициентом преобразования. Их приложения — резка, плавление, разметка, гравировка и маркировка. В дальнем ИК-диапазоне фаворитом является метанол, генерирующий на $\lambda_0 = 9{,}6$ и 10,6 мкм, а также громадном множестве других длин волн. Большинство молекулярных переходом в инфракрасной области могут служить для генерации; даже простой водяной пар (H_2O) генерирует на многих длинах волн в дальнем ИК-диапазоне, как показано в табл. 15.1.

Эксимерные лазеры

Эксимерные лазеры важны в ультрафиолетовой области спектра. Термин «эксимер», являющийся сокращением выражения «excited dimer» (возбужденный димер), означает короткоживущую молекулу из двух атомов в возбужденном электронном состоянии. Если атомы не одинаковы, то вместо термина «эксимер» часто употребляется «эксиплекс». Галогениды инертных газов, такие как XeCl, существуют в виде эксиплексов, так как химические свойства возбужденного атома инертного газа аналогичны свойствам щелочных металлов, которые легко реагируют с галогенами (см. рис. 13.3). При возвращении

эксиплекса в основное состояние он диссоциирует на отдельные атомы, которые часто отталкивают друг друга. Поэтому нижний лазерный уровень незаселен, что обеспечивает исконно присущую данной среде инверсию заселенностей. Примерами эксимеров и эксимерных лазеров с основными рабочими длинами волн являются F_2 (153 нм), ArF (193 нм), KrF (248 нм), XeCl (308 нм), XeF (351 нм) [см. табл. 15.1]. Например, лазерные импульсы от XeCl можно генерировать при электрическом разряде в смеси газов Xe и Cl_2. Коротковолновый свет не поглощается глубоко в большинстве материалов, что делает эксимерные лазеры полезными для острой резки без заметного нагрева материала. Это, в сочетании со значительной энергией на один импульс, обеспечивает приложения таких лазеров в микролитографии, механической микрообработке, фотохимии и глазной хирургии. Стандартным применением является изготовление полупроводниковых микросхем с использованием дальнего ультрафиолетового излучения с $\lambda_0 = 193$ нм.

Химические лазеры

Хемилюминесценция — излучение света при химической реакции — наблюдается, когда реакция между двумя или несколькими веществами сопровождается выделением энергии, достаточной для заселения возбужденного состояния продукта реакции (см. подразд. 13.5.1). Химические лазеры на основе смеси газов являются самонакачивающимися в том смысле, что источником энергии накачки является химическая реакция в самой активной среде. Лазер на HF, генерирующий в области длин волн 2,7—3,1 мкм, является, вероятно, наиболее известным из данного класса лазеров. Смесь газов H_2 и F_2 подвергается электрическому разряду, в результате чего образуются молекулы HF в возбужденных колебательных состояниях, обозначаемые HF^* Такая молекула излучает инфракрасный фотон и диссоциирует. Получающиеся атомы снова реагируют с газами H_2 и F_2 с образованием новых колебательно-возбужденных молекул, создавая цепную реакцию. Химические лазеры могут генерировать высокую мощность и представляют интерес главным образом для военных приложений.

Пример 15.3

Химический лазер на фториде дейтерия в Уайт Сэндз

Вероятно, наиболее известным химическим лазером является лазер армии США «Mid-Infrared Advanced Chemical Laser» (MIRACL), расположенный на ракетной базе Уайт Сэндз. Этот чудовищный агрегат сжигает этилен (C_2H_4) с трифторидом азота (NF_3). Образующиеся свободные атомы фтора реагируют с инжектируемым газообразным дейтерием с образованием возбужденных молекул фторида дейтерия DF^* Испускание фотона, молекулярная диссоциация и образование новых колебательно-возбужденных молекул происходит так же, как в лазере на HF. Однако DF-лазер генерирует излучение на множестве линий в области длин волн 3,5—4,0 мкм, которое значительно меньше поглоща-

ется в атмосфере, чем излучение лазера на HF. Выход DF-лазера может иметь мощность на уровне мегаватт в непрерывном режиме, примерно в течение минуты, при этом диаметр пучка составляет около 4 см.

15.3.3. Другие лазеры

Лазеры на красителях

Лазеры на органических красителях в прошлом играли центральную роль в фотонике по причине их способности к перестройке частоты в широком диапазоне. Активной средой этих лазеров обычно является раствор органического красителя в спирте или воде с концентрацией $\approx 10^{-4}$ М, хотя молекулы красителя могут находиться и в другой базовой среде — полимерной, стеклообразной или кристаллической (**твердотельные лазеры на красителях**). Лазеры на красителях, как правило, работают по четырехуровневой схеме (см. рис. 13.5). Полиметиновые красители генерируют в красной и ближней инфракрасной областях спектра (0,7—1,5 мкм), ксантеновые красители — в видимой области (500—700 нм), кумариновые — в сине-зеленой области (400—500 нм), а сцинтилляторные — в ультрафиолете (<400 нм). Наиболее типичный пример — родамин 6Ж допускает перестройку длины волны в диапазоне от 560 до 640 нм.

К сожалению, лазеры на красителях сложны в эксплуатации, в немалой степени это связано с тем, что химическая жизнь красителя в растворе сравнительно коротка. В результате твердотельные лазеры с диодной накачкой (см. подразд. 15.3.1) в общем и целом вытеснили лазеры на красителях из большинства приложений, за исключением специальных. Например, твердотельный лазер на $Ti^{3+}:Al^2O^3$ с диодной накачкой обеспечивает более широкий диапазон перестройки длины волны вблизи $\lambda_0 = 800$ нм, чем типичный лазер на красителе, и при этом намного проще в эксплуатации. Вторая гармоника от титан-сапфирового лазера дает перестраиваемое излучение в окрестности 400 нм. Перестройка от 400 до 800 нм может быть достигнута путем удвоения частоты выходного излучения оптического параметрического генератора, работающего в области 1—2 мкм (см. подразд. 21.2.3 и 21.4.3).

Лазеры вакуумно-ультрафиолетового и рентгеновского диапазонов

Достижение лазерной генерации в вакуумно-ультрафиолетовом (ВУФ) и рентгеновском диапазонах — заманчивое и сложное предприятие, ввиду сложности создания инверсии заселенностей на таких коротких длинах волн. Согласно (15.16), при фиксированном значении $t_{сп}$ пороговая разность заселенностей $N_п \propto 1/\tau_ф \lambda^2 g(\nu)$, так что пороговая плотность мощности накачки $\propto 1/\tau_ф \lambda^3 g(\nu)$. Если преобладает доплеровское уширение, как обычно бывает в активных лазерных средах ВУФ-диапазона, то формула (13.53) дает $g(\nu) \propto \lambda$, и пороговая плотность мощности накачки становится пропорциональной $1/\tau_ф \lambda^4$.

Поэтому по мере уменьшения λ с продвижением в ВУФ и рентгеновскую область достичь порога генерации становится все труднее.

Другая проблема связана с оптическими элементами, которые в данной спектральной области традиционно трудно сконструировать. Это и потому, что коэффициент поглощения велик (что уменьшает $\tau_ф$ и, следовательно, повышает $N_п$), и потому, что показатель преломления большинства веществ в этой области близок к единице. Тем не менее два подхода оказались успешными.

• *Оптика скользящего падения.* Поскольку частоты ВУФ-излучения намного выше плазменной частоты, металлы имеют показатель преломления, лежащий чуть ниже единицы (см. рис. 5.23 и подразд. 5.5.4). Таким образом, возможно полное внутреннее отражение и металлы могут служить зеркалами. Однако это возможно только при скользящем падении, так как низкий контраст показателей преломления требует большого угла падения (аналогичная ситуация имеет место на границе между оболочкой и сердцевиной оптического волокна).

• *Многослойная оптика.* Создание многослойных оптических элементов не так просто, как в видимом диапазоне, так как показатель преломления в ВУФ-диапазоне близок к единице и мало меняется от одного материала к другому. Тем не менее за счет увеличения числа слоев удается изготовить многослойные зеркала с высоким коэффициентом отражения и приемлемыми потерями (см. разд. 7.1). Например, многослойные зеркала из десятков чередующихся слоев Si и Mo обеспечивают коэффициент отражения >70 % в ВУФ-диапазоне. Многослойная оптика может также встраиваться в структуры типа зонных пластин.

Лазерная генерация рентгеновского излучения впервые наблюдалась в ходе драматического эксперимента, проводившегося в Ливерморской национальной лаборатории им. Лоуренса в 1980 г. Подземный ядерный взрыв служил для создания потока рентгеновских лучей, которые, в свою очередь, служили накачкой для атомов ансамбля металлических стержней. Прежде чем ядерный взрыв испарил установку, был сгенерирован лазерный импульс рентгеновского излучения.

В наши дни ВУФ- и рентгеновское излучения, как правило, генерируются посредством рекомбинации в горячей плазме ионизированных атомов. Электронный переход сверху вниз в сильно ионизированном атоме рождает фотон с высокой энергией, который, в свою очередь, индуцирует рождение своей точной копии соседним ионом вследствие вынужденного испускания. Накачка такой системы, как правило, достигается путем фокусировки лазерного пучка высокой интенсивности на твердотельную мишень. Для достижения высоких интенсивностей обычно используются короткие импульсы накачки. Большинство **лазеров вакуумпо-ультрафиолетового и рентгеновского диапазонов** работает на основе этого принципа. Такие устройства генерируют **усиленное спонтанное излучение** (УСИ) [см. разд. 14.5], а не излучение, возникающее из резонатора, поскольку на таких малых длинах волн трудно организовать обратную связь. Пространственная когерентность может вноситься в процессе распространения (см. подразд. 11.3.3).

В качестве иллюстрации рассмотрим пример ВУФ-лазера на основе плазмы ионизированного углерода. В эксперименте, выполненном в середине 1980-х гг., импульс CO_2 лазера с длиной волны 10,6 мкм, длительностью 50 нс и энергией 300 Дж фокусировался на твердый углеродный диск. Инфракрасный лазерный импульс накачки нагревал мишень настолько, что некоторые атомы углерода полностью лишались электронов в результате ионизации. Образовавшаяся плазма ионизированного углерода удерживалась в радиальном направлении с помощью магнитного поля. Охлаждение плазмы по окончании импульса накачки приводило к захвату электронов на оболочке с $\nu = 3$ и одновременно к недостатку электронов на оболочках с $\nu = 2$ из-за быстрого радиационного распада с переходом в основное состояние. В результате получался ансамбль водородоподобных ионов C^{5+} с инверсией заселенностей (см. рис. 13.1).

Как следует из (13.4) спонтанный переход электрона с оболочки $\nu = 3$ на оболочку $\nu = 2$ (наибольшее сечение имеет переход $3d \rightarrow 2p$) будет сопровождаться излучением фотона с энергией

$$E = \frac{M_r Z^2 e^4}{(4\pi\varepsilon_0)^2 2\hbar^2}\left(\frac{1}{2^2} - \frac{1}{3^2}\right). \qquad (15.46)$$

При $Z = 6$ это соответствует ВУФ-фотону с энергией 68 эВ и длиной волны $\lambda_0 = 18{,}2$ нм. В эксперименте с ионизированным углеродом фотон спонтанного излучения ($t_{сп} \approx 12$ пс) инициировал вынужденное испускание ВУФ-фотонов другими ионами, в результате чего получалось усиленное спонтанное излучение. Произведение коэффициента усиления на длину γd составляло около 6, поэтому, согласно (14.7), полное усиление получалось равным $G \approx e^6$. Выход имел вид импульса ВУФ- (или мягкого рентгеновского) излучения с длительностью 20 нс, мощностью 100 кВт, энергией 2 мДж и расходимостью 5 мрад. Аналогичные результаты были получены при использовании накачки от лазера на стекле с неодимом на 1,06 мкм.

Активные среды ВУФ-лазеров, привлекающие интерес в настоящее время, — это, по большей части, атомные ионы типа Ne, Ni и Pd. Лазеры созданы на десятках подобных ионов. Обычная конфигурация накачки включает цилиндрическую линзу, фокусирующую излучение накачки на мишень для создания вытянутого облака плазмы, которая служит активной средой. Накачка обычно производится титан-сапфировым лазером либо первой, второй или третьей гармоникой излучения лазера на стекле с неодимом (см. подразд. 15.3.1 и 14.3.2). Применение последовательности импульсов накачки увеличивает инверсию заселенностей, что повышает эффективность и делает возможной лазерную генерацию в режиме насыщенного усиления. Доставка основного импульса накачки в режиме скользящего падения увеличивает поглощение и уменьшает требуемую энергию накачки. Примером действия такого лазера является рассмотренный в примере 15.3 **ВУФ-лазер на ионах Ag^{19+}**, подобных атому Ni.

Пример 15.4

ВУФ-лазер на никелеподобных ионах серебра

ВУФ-лазер на ионах Ag^{19+}, подобных атому Ni, работает на переходе $4d^1S_0 \rightarrow 4p^1P_1$. Накачка предварительным импульсом длительностью 4 пс с энергией 1,5 Дж, за которым через 1,2 нс следует основной импульс накачки с энергией 10,5 Дж и длительностью 4 пс, приводит к генерации ВУФ-импульса УСИ с энергией 25 мкДж на длине волны $\lambda_0 = 13{,}9$ нм. Коэффициент усиления равен 35 $см^{-1}$, произведение коэффициента усиления на длину усилителя $\gamma d \approx 13{,}6$. Расходимость пучка 6 мрад.

Фокусировка высокоинтенсивного пучка лазерной накачки на твердую мишень — не единственный способ достижения лазерной генерации в ВУФ-диапазоне. Инверсия заселенностей может быть достигнута также прямым возбуждением активной среды с помощью короткого сильного электрического импульса, создающего горячую плазму. В **рентгеновском лазере с капиллярным разрядом** используется активная среда, протекающая по капилляру (или заключенная в нем) диаметром несколько миллиметров и длиной десятки сантиметров. Такие устройства с накачкой электрическим током, как правило, обеспечивают более высокие скорости повторения импульсов и большую пространственную когерентность, чем устройства с лазерной накачкой. Однако их главным достоинством является сравнительно малый размер, за что они получили неформальное название «настольных» рентгеновских лазеров.

Ионизацию активной среды можно обеспечить непосредственно через полевые эффекты и многофотонные процессы. Тогда инверсия заселенностей получается за счет возбуждения ионов ударами испущенных электронов, а не термализованными частицами горячей плазмы. Такие **лазеры с ионизацией оптически полем** производят холодную плотную совокупность ионизированных атомов, окруженную облаком горячих электронов. В принципе можно построить также **лазеры с фотонакачкой внутренних оболочек**, в которых ионизация происходит за счет вырывания электронов внутренних оболочек у нейтрального атома. Хотя время жизни вакансий на внутренних оболочках очень мало из-за быстрого радиационного распада и Оже-переходов, методы оптики сверхбыстрых процессов вполне могут оказаться полезными для реализации такой накачки (см. гл. 22).

Значительный прогресс был достигнут также в генерации когерентного рентгеновского излучения с помощью **генерации высших гармоник**. Гармоники высоких порядков можно генерировать при экстремально нелинейном взаимодействии интенсивных импульсов фемтосекундного лазера с молекулами газа. Если лазерное поле достаточно сильное, то оно может ионизировать молекулы газа и ускорять освободившиеся валентные электроны, способствуя их вылету из образовавшегося иона. Гармоники высоких порядков генерируются, когда поле обращается, и осциллирующие электроны возвращаются в ион. Этот подход обычно порождает ультракороткие импульсы с длиной волны 5—30 нм и

значительной средней мощностью. Другим средством генерации когерентного ВУФ- и рентгеновского излучения являются лазеры на свободных электронах (см. далее), однако их можно строить только на больших синхротронных установках.

Применение ВУФ- и рентгеновских лазеров включает нанолитографию для полупроводниковых интегральных схем, наномоделирование и нановизуализацию, диагностику плазмы, динамическую визуализацию и голографию биологических и других структур.

Лазеры на свободных электронах

В лазерах на свободных электронах (ЛСЭ) используется периодическое магнитное **поле вигглера** — периодического ансамбля магнитов чередующейся полярности, которое заставляет электрон двигаться по извилистой траектории. При малой амплитуде колебаний устройства такого типа известны как **ондуляторы**. Релятивистские электроны, движущиеся в поле вигглера, представляют собой активную среду (рис. 15.30). При прохождении через вигглер электроны испытывают колебания и когерентно излучают. Несмотря на название «лазер на свободных электронах», электроны на самом деле не являются свободными, так как на их движение влияет поле вигглера.

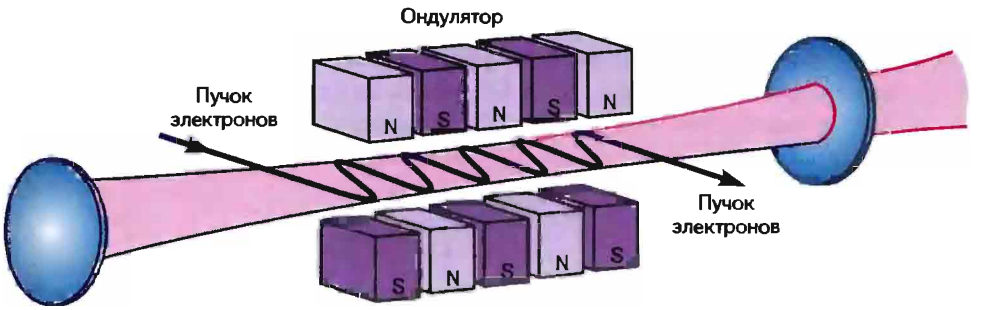


Рис. 15.30. Схема типичного лазера на свободных электронах в оптическом диапазоне. Ондулятор создает периодическое поперечное поле с периодом Λ в несколько сантиметров, 100 периодов укладывается на полной длине в несколько метров. Ондулятор помещен между зеркалами резонатора, расстояние между которыми d примерно вдвое больше длины ондулятора. Максимальное магнитное поле в ондуляторе составляет обычно несколько килогаусс. Электронный пучок направляется в ондулятор с помощью отклоняющих магнитов. Ток электронного пучка составляет от нескольких ампер до нескольких килоампер, а энергия электронов может меняться от нескольких мега- до нескольких гигаэлектронвольт; типичное значение — 50 МэВ. Радиус электронного пучка составляет примерно 1 мм, а перетяжка оптического пучка — примерно 3 мм. Длительность импульса электронного пучка может меняться от пико- до микросекунд. Временная структура излучения следует временной структуре электронного пучка

Излучение лазера на свободных электронах можно перестраивать в широком диапазоне длин волн путем изменения энергии электронов, напряженнос-

ти магнитного поля и периода ондулятора. В зависимости от конструкции, длина волны излучения ЛСЭ может меняться от миллиметрового до ВУФ-диапазона. Поскольку ЛСЭ работают в вакууме, высокие мощности излучения можно получать без повреждения материала и эффектов тепловой линзы. Разрабатываемое поколение ЛСЭ будет предназначено для генерации когерентного жесткого рентгеновского излучения с длиной волны до $\lambda_0 \approx 1$ Å $= 0{,}1$ нм. Первый ЛСЭ этого класса — когерентный источник света Linac (Linac Coherent Light Source, LCLS) длиной 1 км в Стэнфордском центре линейных ускорителей — будет ускорять сгруппированные сгустки примерно из 10^{10} электронов для генерации сгустков из $\approx 10^{12}$ фотонов в импульсах длительностью 100 фс на $\lambda_0 = 1{,}5$ Å. Несколько ЛСЭ, работающих на более коротких волнах, намечается запустить к 2015 г.

15.3.4. Таблица характеристик

В табл. 15.1 приведен список типичных характеристик некоторых хорошо известных лазеров в порядке возрастания длины волны. Отметим широкий диапазон длин волн переходов, полных коэффициентов преобразования и выходных мощностей для различных лазеров.

Сечения переходов, спонтанные времена жизни и ширины атомных линий для многих лазерных переходов приведены в табл. 14.1. Ширина линии излучения лазера, как правило, на много порядков меньше атомной ширины линии перехода, приведенной в табл. 14.1, благодаря дополнительной частотной селективности, навязанной оптическим резонатором. Некоторые лазерные системы не могут поддерживать инверсию заселенностей непрерывно и, следовательно, работают только в импульсном режиме.

15.4. ИМПУЛЬСНЫЕ ЛАЗЕРЫ

Иногда желательно, чтобы лазер работал в импульсном режиме, поскольку оптическая мощность может быть сильно увеличена, когда выходной импульс имеет ограниченную длительность. Длительность импульсов, генерируемых современными лазерами, достигает нескольких фемтосекунд; дальнейшая компрессия импульсов до аттосекундных масштабов возможна с помощью нелинейно-оптических методов. Максимальная частота повторения импульсов меняется от одного импульса в течение нескольких часов для некоторых ВУФ-лазеров до более чем 100 ГГц. Максимальная энергия импульса меняется от фемтоджоулей до мегаджоулей, т. е. более чем на 20 порядков величины, при этом пиковая мощность превосходит 10 МВт, а интенсивность — 10 ТВт/см^2. Некоторые лазеры могут работать только в импульсном режиме, поскольку непрерывный режим в них не поддерживается, что видно из табл. 15.1.

Таблица 15.1. **Типичные характеристики и параметры ряда хорошо известных лазеров на основе различных форм вещества* в порядке увеличения длины волны**

Лазерная среда	Длина волны перехода λ_0	Одномодовый (О) или многомодовый (М)	Непрерывный или импульсный**	Примерный полный коэффициент преобразования мощности η_c, %***	Выходная мощность или энергия****	Схема уровней энергии
Ar^{19+} (п)	13,9 нм	М	Имп.	0,0002	25 мкДж	
C^{5+} (п)	18,2 нм	М	Имп.	0,0005	2 мкДж	Рис. 13.1
Эксимер ArF (г)	193 нм	М	Имп.	1,0	200 мДж	
Эксимер KrF (г)	248 нм	М	Имп.	1,0	500 мдж	
Ht—Cd (г)	442 нм	О/М	Непр.	0,1	100 мВт	
Ar^+ (г)	515 нм	О/М	Непр.	0,05	10Вт	
Родамин 6Ж (ж)	560—640 нм	О/М	Непр.	0,005	100 мВт	Рис. 13.5
He—Ne (г)	633 нм	О/М	Непр.	0,05	10 мВт	Рис. 13.2
Kr^+ (г)	647 нм	О/М	Непр.	0,01	1 Вт	
Рубин (т)	694 нм	М	Непр.	0,1	5 Вт	Рис. 14.14
Александрит (т)	700—820 нм	М	Непр.	0,1	1 Вт	Рис. 13.8
Ti : сапфир (т)	700—1050 нм	О/М	Непр.	0,01	5 Вт	Рис. 15.26
Yb^{3+} : YAG (т)	1030 нм	О/М	Непр.	5,0	100 Вт	Рис. 15.25
Nd^{3+} : стекло (т)	1053 нм	М	Имп.	1,0	50 Дж	Рис. 14.16
Nd^{3+} : YAG (т)	1064 нм	О/М	Непр.	5,0	50 Вт	Рис. 13.9
Nd^{3+} : YVO_4 (т)	1064 нм	О/М	Непр.	10,0	30 Вт	Рис. 15.23
Yb^{3+} : кварцевое волокно (т)	1075 нм	О/М	Непр.	20,0	1500 Вт	

Окончание табл. 15.1

Лазерная среда	Длина волны перехода λ_0	Одномодовый (О) или многомодовый (М)	Непрерывный или импульсный**	Примерный полный коэффициент преобразования мощности η_c, %***	Выходная мощность или энергия****	Схема уровней энергии
Er^{3+}: кварцевое волокно (т)	1550 нм	О/М	Непр.	10,0	100 Вт	Рис. 14.19
Tm^{3+}: фторидное волокно (т)	1,8—2,1 мкм	О/М	Непр.	5,0	150 Вт	
He—Ne (г)	3,39 мкм	О/М	Непр.	0,05	20 мВт	Рис. 13.2
CO_2 (г)	10,6 мкм	О/М	Непр.	10,0	500 Вт	Рис. 13.4
H_2O (г)	28 мкм	О/М	Непр.	0,02	100 мВт	
ЛСЭ в Калифорнийском университете в Санта-Барбаре	60 мкм—2,5 мм	М	Имп.	0,5	5 мДж	
H_2O (г)	118,7 мкм	О/М	Непр.	0,01	50 мВт	
CH_3OH (г)	118,9 мкм	О/М	Непр.	0,02	100 мВт	
HCN (г)	336,8 мкм	О/М	Непр.	0,01	20 мВт	

* Газ (г), жидкость (ж), твердое тело (т), плазма (п).

** Лазеры, отмеченные «непр.», конечно, могут работать и в импульсном режиме. Для лазеров, отмеченных «имп.», импульсный режим является типичным

*** Полный коэффициент преобразования мощности η_c или коэффициент полезного действия есть отношение полезной мощности света на выходе к электрической мощности, потребляемой на входе (для импульсных лазеров — соответствующее отношение энергий). Приведенные значения имеют значительную неопределенность, так как в некоторых случаях в затраченную мощность включались расходы на охлаждение и мониторинг. Лазерные диоды имеют наибольший коэффициент преобразования, легко достигающий 50 %, что обсуждается в подразд. 17.4.4.

**** Выходная мощность для непрерывного режима и энергия импульса меняются на значительную величину, частично из-за большой разницы в длительностях импульса; приведены типичные значения.

15.4.1. Методы получения импульсной генерации

Наиболее прямой метод получения импульсов от лазера заключается в использовании лазера непрерывного действия в сочетании с внешним переключателем или модулятором, который пропускает свет только в течение избранных коротких интервалов времени. Этот простой способ, однако, страдает двумя явно выраженными недостатками. Во-первых, схема неэффективна, поскольку она блокирует (и, следовательно, расходует впустую) энергию лазерного пучка в промежутках между импульсами. Во-вторых, пиковая мощность импульсов не превышает стационарной мощности непрерывного источника, как показано на рис. 15.31, *а*.

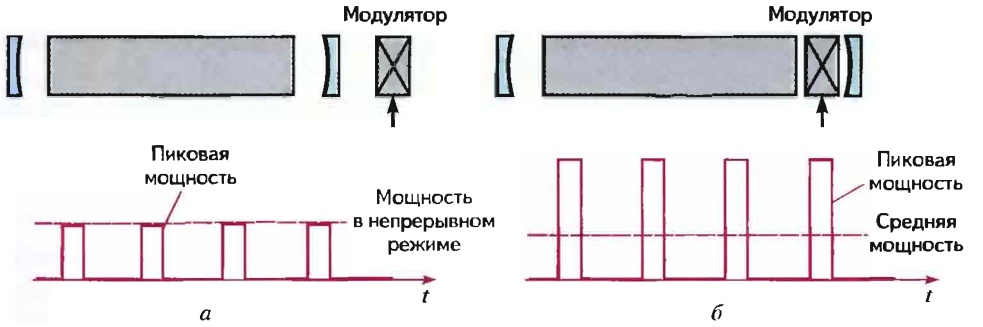


Рис. 15.31. Сравнение выходной мощности импульсного лазера при внешней (*а*) и внутренней модуляциях (*б*)

Более эффективные схемы импульсной генерации основаны на включении и выключении самого лазера и разработаны с таким расчетом, чтобы энергия запасалась в промежутках между импульсами и высвобождалась в течение импульса. Энергия может запасаться либо в резонаторе в виде света, которому периодически позволяют выходить наружу, либо в атомах в виде инверсии заселенностей, которой периодически дают возможность возбуждать генерацию лазера. Такие схемы позволяют генерировать короткие импульсы, максимальная мощность которых намного превышает постоянную мощность лазера в непрерывном режиме, как показано на рис 15.31, *б*.

Существует четыре метода внутренней модуляции лазерного света: модуляция усиления, модуляция добротности, режим открытия резонатора и синхронизация мод.

Модуляция усиления

Модуляция усиления — прямой метод, в котором управление усилением осуществляется путем включения и выключения накачки (рис. 15.32). Например, в рубиновом лазере с накачкой от лампы-вспышки накачка (лампа-вспышка) периодически включается на короткое время с помощью после-

довательности коротких электрических импульсов. В течение импульса коэффициент усиления превышает коэффициент потерь и происходит генерация. Большинство импульсных полупроводниковых лазеров работают в режиме модуляции усиления, поскольку электрический ток, используемый для накачки, легко модулировать (см. гл. 17). Времена нарастания и спада мощности в импульсе, достижимые при модуляции усиления, определены в подразд. 15.4.2.

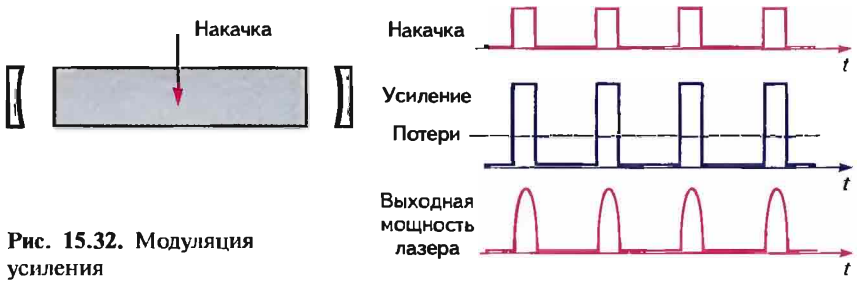


Рис. 15.32. Модуляция усиления

Модуляция добротности

При модуляции добротности лазер выключается посредством периодического увеличения потерь резонатора (ухудшения добротности резонатора Q) с помощью модулируемого поглотителя внутри резонатора (рис. 15.33). Таким образом, модуляция добротности — это модуляция потерь. Поскольку накачка продолжает постоянно снабжать активную среду энергией, она накапливается в атомах среды в виде инверсии заселенностей в течение промежутков между импульсами, когда потери велики. Когда потери резко снижаются, возбуждается генерация и накопленная энергия высвобождается в виде интенсивного (обычно короткого) светового импульса. Анализ этого метода приведен в подразд. 15.4.3.

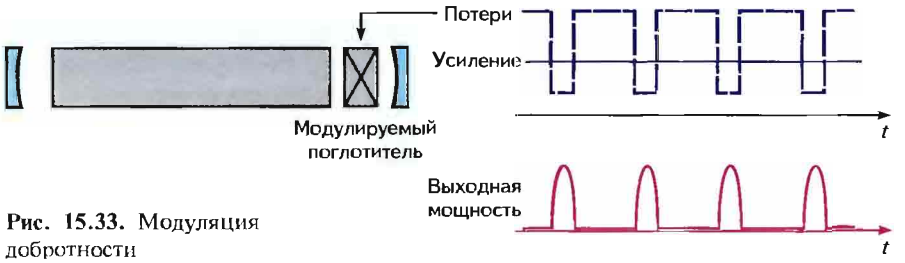


Рис. 15.33. Модуляция добротности

Режим открытия резонатора

Режим открытия резонатора — это метод, основанный на накоплении фотонов (а не разности заселенностей) в резонаторе во время выключения и их высвобождении во время включения. Он отличается от модуляции тем,

что потери резонатора модулируются за счет изменения коэффициента пропускания зеркал (рис. 15.34). Система работает подобно ведру, в которое вода наливается из шланга с постоянной скоростью. По истечении времени накопления воды дно ведра внезапно убирают, и ведро опорожняется. Затем дно возвращается на место и процесс повторяется. Так постоянный поток воды превращается в импульсный поток. Для лазера в режиме открытия резонатора ведро — это резонатор, шланг — постоянная накачка, а дно ведра — выходное зеркало резонатора. Во время между импульсами утечка света, *в том числе и полезного*, из резонатора запрещена. Потери резонатора пренебрежимо малы, и мощность света внутри резонатора нарастает. Фотоны накапливаются в резонаторе и не могут выйти из него. Внезапно зеркало полностью убирают (например, смещая его с оси резонатора путем поворота) и коэффициент пропускания становится равным 100 %. Когда накопленные фотоны покинут резонатор, внезапный рост потерь блокирует генерацию. В результате получается мощный импульс лазерного света. Анализ режима открытия резонатора здесь не проводится, поскольку он тесно связан с режимом модуляции добротности: изменения усиления и потерь со временем одинаковы, как можно увидеть из сравнения рис. 15.34 и 15.33.

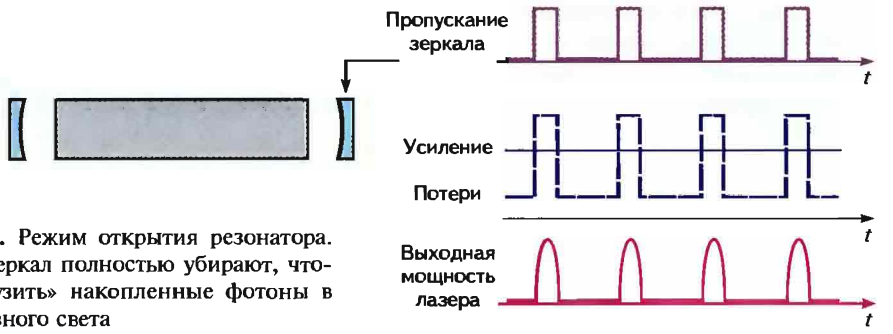


Рис. 15.34. Режим открытия резонатора. Одно из зеркал полностью убирают, чтобы «выгрузить» накопленные фотоны в виде полезного света

Синхронизация мод

Три метода получения импульсной генерации, рассмотренные выше, основаны на нестационарной динамике лазерной среды. **Синхронизация мод** отличается от этих подходов тем, что это динамический стационарный процесс. Это наиболее важный метод для получения цугов ультракоротких лазерных импульсов. Импульсное действие лазера достигается за счет связывания мод и синхронизации их фаз между собой. Примером являются продольные моды в многомодовом лазере, которые генерируют на равноотстоящих частотах с межмодовым расстоянием $c/2d$. Когда фазы таких компонент синхронизованы, они ведут себя как Фурье-компоненты периодической функции и, следовательно, формируют периодический цуг импульсов. Связывание мод осуществляется за счет периодической модуляции потерь в резонаторе. Синхронизация мод исследована в подразд. 15.4.4.

*15.4.2. Анализ переходных эффектов

Аналитическое описание действия импульсных лазеров требует понимания динамики процесса лазерной генерации, т. е. процесса включения генерации и ее прекращения. Стационарные решения, представленные ранее в этой главе, непригодны для этой цели. Процесс лазерной генерации описывается двумя переменными: числом фотонов в единице объема резонатора $n(t)$ и разностью заселенностей атомных уровней энергии на единицу объема среды $N(t) = N_2(t) - N_1(t)$. Обе величины являются функциями времени t.

Скоростное уравнение для плотности числа фотонов

Плотность числа фотонов $\nu(t)$ удовлетворяет скоростному уравнению

$$\frac{dn}{dt} = -\frac{n}{\tau_ф} + NW_i. \tag{15.47}$$

Первый член описывает потери фотонов из-за утечки из резонатора со скоростью, определяемой обратным временем жизни фотона $1/\tau_ф$. Второй член описывает полное нарастание числа фотонов со скоростью NW_i в результате процессов вынужденного излучения и поглощения. $W_i = \phi\sigma(\nu) = cn\sigma(\nu)$ — плотность вероятности вынужденного излучения/поглощения. Спонтанное испускание предполагается малым. С помощью соотношения $N_п = \alpha_r/\sigma(\nu) = 1/c\tau_ф\sigma(\nu)$, где $N_п$ — пороговая разность заселенностей [см. (15.15)], запишем $\sigma(\nu) = 1/c\tau_ф N_п$, откуда

$$W_i = \frac{n}{N_п\tau_ф}. \tag{15.48}$$

Подставляя это выражение в (15.47), получаем простое дифференциальное уравнение для плотности числа фотонов n:

$$\boxed{\frac{dn}{dt} = -\frac{n}{\tau_ф} + \frac{N}{N_п}\frac{n}{\tau_ф}.} \tag{15.49}$$

Скоростное уравнение для числа фотонов

Пока $N > N_п$, dn/dt положительно и n растет. После достижения стационарного ($dn/dt = 0$) состояния $N = N_п$.

Скоростное уравнение для разности заселенностей

Динамика разности заселенностей $N(t)$ определяется конфигурацией накачки. Здесь будет проанализирована трехуровневая схема (см. подразд. 14.2.2). Скоростное уравнение для заселенности верхнего уровня энергии перехода, согласно (14.22), имеет вид

$$\frac{dN_2}{dt} = R - \frac{N_2}{t_{сп}} - W_i(N_2 - N_1), \tag{15.50}$$

где предполагается, что $\tau_2 = t_{сп}$. Скорость накачки R предполагается не зависящей от разности заселенностей N. Обозначая полную плотность числа атомов $N_2 + N_1$ через N_a, так что $N_1 = (N_a - N)/2$ и $N_2 = (N_a + N)/2$, получаем дифференциальное уравнение для разности заселенностей $N = N_2 - N_1$:

$$\frac{dN}{dt} = \frac{N_0}{t_{сп}} - \frac{N}{t_{сп}} - 2W_i N, \tag{15.51}$$

где разность заселенностей при слабом сигнале $N_0 = 2Rt_{сп} - N_a$ [см. (14.41)]. Подставляя соотношение $W_i = n/N_п\tau_ф$, полученное выше, в (15.51), находим

$$\boxed{\frac{dN}{dt} = \frac{N_0}{t_{сп}} - \frac{N}{t_{сп}} - 2\frac{N}{N_п}\frac{n}{\tau_ф}.} \tag{15.52}$$

Скоростное уравнение для разности заселенностей

Третий член в правой части уравнения (15.52) равен удвоенному второму члену в правой части уравнения (15.49) с противоположным знаком. Это отражает тот факт, что генерация одного фотона посредством вынужденного перехода понижает заселенность энергетического уровня 2 на один атом и настолько же повышает заселенность уровня 1, в итоге уменьшая их разность на 2.

Уравнения (15.49) и (15.52) — связанные нелинейные дифференциальные уравнения, решение которых определяет нестационарное поведение плотности числа фотонов $n(t)$ и разности заселенностей $N(t)$. Полагая $dN/dt = 0$ и $dn/dt = 0$, получим $N = N_п$ и $n = (N_0 - N_п)(\tau_ф/2t_{сп})$. Это совпадает с ранее полученными стационарными значениями N и n, что очевидно из (15.35) при $\tau_ф = 2t_{сп}$, согласно (14.42) для трехуровневой схемы накачки.

Упражнение 15.3

Скоростное уравнение для разности заселенностей в четырехуровневой схеме

Получите скоростное уравнение для разности заселенностей в четырехуровневой системе при $\tau_1 = t_{сп}$. Объясните отсутствие множителя 2, присутствующего в (15.52).

Модуляция усиления

Модуляция усиления осуществляется включением и выключением накачки R; в свою очередь, это эквивалентно модуляции разности заселенностей при слабом сигнале $N_0 = 2Rt_{сп} - N_a$. Схематическая иллюстрация типичной временной эволюции разности заселенностей $N(t)$ и плотности числа фотонов $n(t)$ при ступенчатом изменении N_0 приведена на рис. 15.35. В этом процессе, очевидно, можно выделить четыре режима.

- При $t < 0$ разность заселенностей $N(t) = N_{0a}$ лежит ниже порога $N_п$ и генерация не происходит.

• Накачка включается при $t = 0$, и значение N_0 скачком увеличивается с N_{0a} ниже порога до N_{0b} выше порога. Разность заселенностей $N(t)$ начинает нарастать. Однако, пока $N(t) < N_п$, плотность числа фотонов $n = 0$. В этой области уравнение (15.52) превращается в $dN/dt = (N_0 - N)/t_{сп}$, что отражает рост и экспоненциальное приближение $N(t)$ к равновесному значению N_{0b} с постоянной времени $t_{сп}$.

• Как только $N(t)$ пересекает пороговый уровень $N_п$ при некотором $t = t_1$, начинается лазерная генерация и $n(t)$ начинает расти. Затем разность заселенностей начинает обедняться, так что скорость роста $N(t)$ снижается. По мере роста $n(t)$ обеднение идет все более эффективно, так что $N(t)$ начинает затухать, стремясь к значению $N_п$. В конце концов $N(t)$ достигает $N_п$; в это время $n(t)$ выходит на свой стационарный уровень.

• В момент $t = t_2$ накачка выключается и N_0 снижается до своего начального значения N_{0a}. $N(t)$ и $n(t)$ затухают к значениям N_{0a} и 0 соответственно.

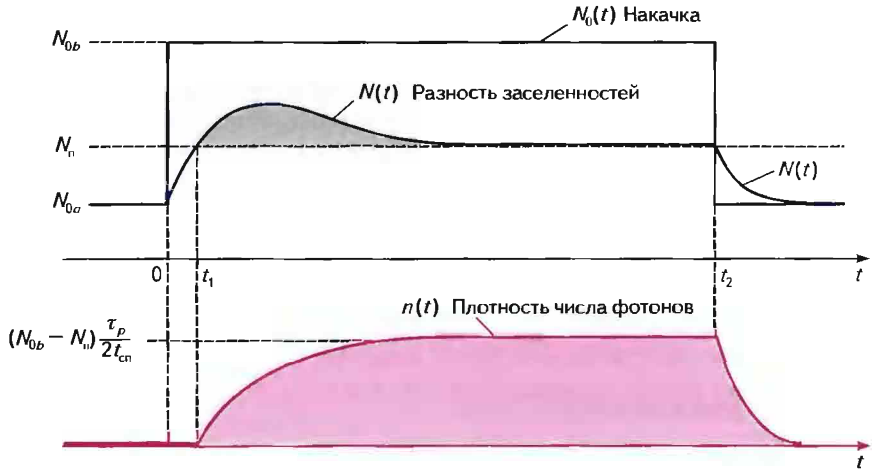


Рис. 15.35. Изменение разности заселенностей $N(t)$ и плотности числа фотонов $n(t)$ со временем при ступенчатом изменении накачки, вызывающем скачок N_0 от низкого значения N_{0a} до высокого значения N_{0b} и обратно к N_{0a}

Реальный профиль нарастания и спада $n(t)$ получается путем численного решения (15.49) и (15.52). Точный вид решения зависит от $t_{сп}$, $\tau_ф$, $N_п$, а также от N_{0a} и N_{0b} (см. задачу 2 к разд. 14.4).

*15.4.3. Модуляция добротности

В лазере с модуляцией добротности импульсный режим достигается посредством переключения коэффициента потерь резонатора α_r с большого значения в промежутках между импульсами на малое значение во время каждого импульса. Это можно сделать разными путями, например, помещая внутрь резонатора модулятор, периодически вносящий большие потери. Поскольку

пороговое значение разности заселенностей $N_п$ пропорционально коэффициенту потерь резонатора α_r [см. (15.14) и (15.6)], результатом модуляции α_r является скачкообразное изменение $N_п$ от высокого значения $N_{па}$ до низкого $N_{пb}$ и обратно, как показано на рис. 15.36. Таким образом, при модуляции добротности меняется $N_п$ при постоянном N_0, тогда как при модуляции усиления меняется N_0 при постоянном $N_п$ (см. рис. 15.36). Разность заселенностей и плотность числа фотонов ведут себя следующим образом:

• При $t = 0$ включается накачка, так что N_0 — ступенчатая функция. Потери поддерживаются на достаточно высоком уровне ($N_п = N_{па} > N_0$), и лазерная генерация не начинается. Разность заселенностей $N(t)$ растет с постоянной времени $t_{сп}$. Хотя среда является хорошим усилителем, потери достаточно велики для предотвращения генерации.

• При $t = t_1$ потери внезапно уменьшаются, так что $N_п$ падает до значения $N_{пb} < N_0$. Начинается генерация, и число фотонов резко возрастает. Присутствие излучения приводит к обеднению инверсии заселенностей (насыщению усиления), так что $N(t)$ начинает убывать. Когда $N(t)$ спадает ниже $N_{пb}$, потери снова превосходят усиление, что приводит к быстрому уменьшению плотности числа фотонов (с постоянной времени порядка времени жизни фотона $\tau_ф$).

• При $t = t_2$ потери восстанавливаются, гарантируя достаточное время нарастания разности заселенностей, необходимое чтобы приготовиться к следующему импульсу. Процесс повторяется периодически, что приводит к генерации цуга импульсов.

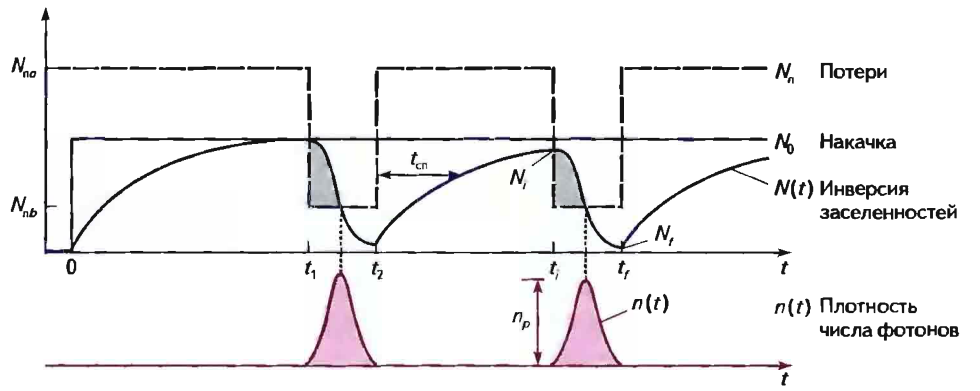


Рис. 15.36. Действие лазера с модуляцией добротности. Изменения порогового значения разности заселенностей $N_п$ (которая пропорциональна потерям резонатора), параметра накачки N_0, разности заселенностей $N(t)$ и плотности числа фотонов $n(t)$

Теперь проведем анализ, необходимый для определения пиковой мощности, энергии, длительности и формы оптического импульса, генерируемого лазером с модуляцией добротности в установившемся импульсном режиме. Будем опираться на два основных скоростных уравнения (15.49) и (15.52) для $n(t)$ и $N(t)$ соответственно. Решим эти уравнения на отрезке времени между моментом включения t_i и моментом выключения t_f, показанном на рис. 15.36. Задача,

конечно, может быть решена численно. Однако упрощение, достаточное для аналитического решения, возможно, если допустить, что первые два члена в правой части (15.52) пренебрежимо малы. Это предположение справедливо, если на коротком отрезке времени от t_i до t_f накачкой и спонтанным излучением можно пренебречь по сравнению с действием вынужденного излучения. Такое приближение оказывается разумным, если длительность генерируемого оптического импульса намного меньше $t_{сп}$. Если это так, то (15.49) и (15.52) принимают вид

$$\boxed{\frac{dn}{dt} = \left(\frac{N}{N_п} - 1\right)\frac{n}{\tau_ф};} \tag{15.53}$$

$$\boxed{\frac{dN}{dt} = -2\frac{N}{N_п}\frac{n}{\tau_ф}.} \tag{15.54}$$

Это два связанных дифференциальных уравнения относительно $n(t)$ и $N(t)$ с начальными условиями $n = 0$ и $N = N_i$ при $t = t_i$. На протяжении интервала времени от t_i до t_f $N_п$ фиксировано на низком уровне $N_{пb}$.

Деля (15.53) на (15.54), получим одно дифференциальное уравнение, связывающее n и N:

$$\frac{dn}{dN} \approx \frac{1}{2}\left(\frac{N}{N_п} - 1\right), \tag{15.55}$$

интегрирование которого дает

$$n \approx \frac{1}{2}N_п \ln(N) - \frac{1}{2}N + \text{const}. \tag{15.56}$$

Использование начального условия $n = 0$ при $N = N_i$ дает окончательно

$$\boxed{n \approx \frac{1}{2}N_п \ln\frac{N}{N_i} - \frac{1}{2}(N - N_i).} \tag{15.57}$$

Пример 15.5

Миниатюрный Nd^{3+} : YAG-лазер с модуляцией добротности

Тонкий срез Nd^{3+} : YAG, слой насыщаемого поглотителя и внутрирезонаторный кристалл для удвоения частоты, сложенные вместе, образуют резонатор длиной 1 мм. При накачке светом мощностью 1 Вт через оптоволокно от диодного лазера этот миниатюрный лазер в виде чипа генерирует импульсы в режиме модулированной добротности на 532 нм. Каждый импульс имеет энергию 30 мкДж и длительность 250 пс. Частота повторения примерно 10 кГц, средняя мощность 300 Вт.

Мощность импульса

Согласно (15.33) и (15.26), внутренняя плотность потока фотонов (включая оба направления) равна $\phi = \mathrm{n}c$, в то время как внешняя плотность потока фотонов, выходящего из зеркала 1 (с коэффициентом пропускания $\mathcal{T}$), равна $\phi_0 = \mathcal{T}\mathrm{n}c/2$. Считая, что плотность потока фотонов равномерно распределена по поперечному сечению A выходящего пучка, найдем соответствующую выходную оптическую мощность:

$$P_0 = h\nu A\phi_0 = \frac{1}{2}h\nu c\mathcal{T}A\mathrm{n} = h\nu\mathcal{T}\frac{c}{2d}V\mathrm{n}, \tag{15.58}$$

где $V = Ad$ — объем резонатора. Согласно (15.40), если $\mathcal{T} \ll 1$, то доля потерь резонатора, приходящаяся на полезное излучение на выходе, составляет $\eta_{\text{эфф}} \approx \mathcal{T}(c/2d)\tau_{\text{ф}}$, поэтому получаем

$$P_0 = \eta_{\text{эфф}}h\nu\frac{\mathrm{n}V}{\tau_{\text{ф}}}. \tag{15.59}$$

Выражение (15.59) легко интерпретируется, поскольку множитель $\mathrm{n}V/\tau_{\text{ф}}$ есть число фотонов, теряемых резонатором в единицу времени.

Пиковая мощность импульса

Как обсуждалось выше и показано на рис. 15.36, n достигает своего максимального значения n_p, когда $N = N_{\text{п}} = N_{\text{п}b}$. Это подтверждается, если положить $\mathrm{d}\mathrm{n}/\mathrm{d}t = 0$ в (15.53), что немедленно приводит к $N = N_{\text{п}}$. Подставляя это в (15.57), находим

$$\mathrm{n}_p = \frac{1}{2}N_i\left(1 + \frac{N_{\text{п}}}{N_i}\ln\frac{N_{\text{п}}}{N_i} - \frac{N_{\text{п}}}{N_i}\right). \tag{15.60}$$

Используя этот результат вместе с (15.58), получаем для пиковой мощности

$$P_p = h\nu\mathcal{T}\frac{c}{2d}V\mathrm{n}_p. \tag{15.61}$$

Когда $N_i \gg N_{\text{п}}$, как должно быть в случае импульсов с большой пиковой мощностью, $N_{\text{п}}/N_i \ll 1$, и из (15.60) получается

$$\mathrm{n}_p \approx \frac{1}{2}N_i. \tag{15.62}$$

Пиковая плотность потока фотонов равна половине начальной плотности разности заселенностей. В этом случае выражение для пиковой мощности принимает особенно простой вид

$$\boxed{P_p \approx \frac{1}{2}h\nu\mathcal{T}\frac{c}{2d}VN_i.} \tag{15.63}$$

Пиковая мощность импульса

Энергия импульса

Энергия импульса дается выражением

$$E = \int_{t_1}^{t_2} P_0 \mathrm{d}t, \tag{15.64}$$

которое, в соответствии с (15.58), можно записать как

$$E = h\nu\mathcal{T}\frac{c}{2d}V\int_{t_i}^{t_f} \mathrm{n}(t)\mathrm{d}t = h\nu\mathcal{T}\frac{c}{2d}V\int_{N_i}^{N_f} \mathrm{n}(t)\frac{\mathrm{d}t}{\mathrm{d}N}\mathrm{d}N. \tag{15.65}$$

Подставляя (15.54) в (15.65), получаем

$$E = \frac{1}{2}h\nu\mathcal{T}\frac{c}{2d}VN_{\mathrm{п}}\tau_{\Phi}\int_{N_f}^{N_i}\frac{\mathrm{d}N}{N}. \tag{15.66}$$

В результате интегрирования имеем

$$E = \frac{1}{2}h\nu\mathcal{T}\frac{c}{2d}VN_{\mathrm{п}}\tau_{\Phi}\ln\frac{N_i}{N_f}. \tag{15.67}$$

Конечная разность заселенностей N_f определяется, если положить $\mathrm{n} = 0$ и $N = N_f$ в (15.57), что дает

$$\ln\frac{N_i}{N_f} = \frac{N_i - N_f}{N_{\mathrm{п}}}. \tag{15.68}$$

Подстановка этого выражения в (15.67) дает

$$\boxed{E = \frac{1}{2}h\nu\mathcal{T}\frac{c}{2d}VN_{\mathrm{п}}\tau_{\Phi}\left(N_i - N_f\right).} \tag{15.69}$$

Энергия импульса при модуляции добротности

Когда $N_i \gg N_f$,

$$E \approx \frac{1}{2}h\nu\mathcal{T}\frac{c}{2d}V\tau_{\Phi}N_i,$$

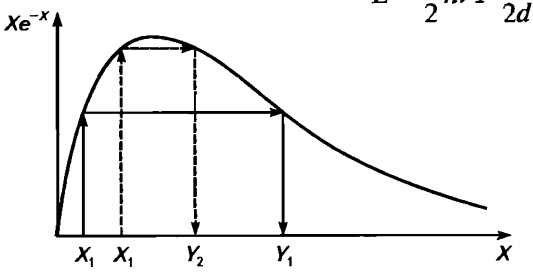


Рис. 15.37. Графическое построение для определения N_f по N_i, где $X = N_i/N_{\mathrm{п}}$ и $Y = N_f/N_{\mathrm{п}}$. Для $X = X_1$ ордината имеет значение $X_1 \exp(-X_1)$. Поскольку соответствующее решение для Y_1 удовлетворяет $Y_1 \exp(-Y_1) = X_1 \exp(-X_1)$, ему соответствует то же самое значение ординаты

как и ожидалось. Остается решить (15.68) относительно N_f. Один из подходов состоит в том, чтобы переписать его в виде $Y\exp(-Y) = X\exp(-X)$, где $X = N_i/N_п$ и $Y = N_f/N_п$. При данном $X = N_i/N_п$ можно легко решить это уравнение относительно Y численно или при помощи графика, приведенного на рис. 15.37.

Длительность импульса

Грубая оценка длительности импульса дается отношением энергии импульса к его пиковой мощности. Используя (15.60), (15.61) и (15.69), получаем

$$\tau_{имп} = \tau_ф \frac{N_i/N_п - N_f/N_п}{N_i/N_п - \ln(N_i/N_п) - 1}. \qquad (15.70)$$

Длительность импульса

Когда $N_i \gg N_п$ и $N_i \gg N_f$, имеем $\tau_{имп} \approx \tau_ф$.

Форма импульса

Форма оптического импульса, как и другие его характеристики, описанные выше, может быть определена путем численного интегрирования (15.53) и (15.54). Примеры получающихся форм импульса показаны на рис. 15.38.

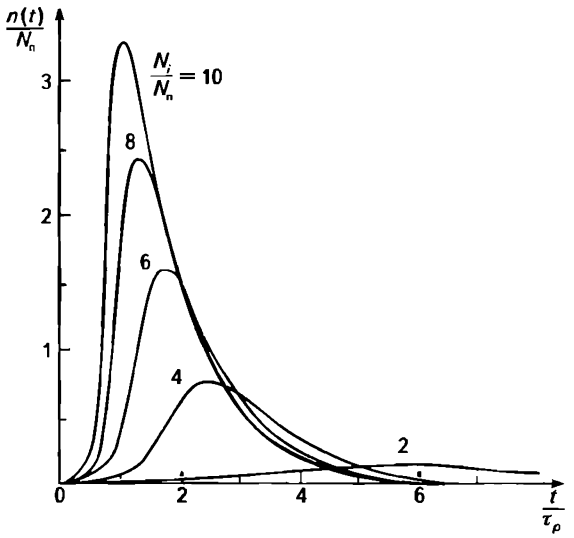


Рис. 15.38. Типичные формы импульсов при модуляции добротности, полученные численным интегрированием приближенных скоростных уравнений. Плотность числа фотонов $n(t)$ нормирована на пороговое значение разности заселенностей $N_п = N_{пb}$, а время t нормировано на время жизни фотона $\tau_ф$. Импульс становится уже и достигает более высоких пиковых значений по мере того, как отношение $N_i/N_п$ растет. В пределе $N_i/N_п \gg 1$ пиковое значение $n(t)$ стремится к $N_i/2$

Упражнение 15.4

Импульсный рубиновый лазер

Рассмотрим рубиновый лазер из упражнения 15.1. При модуляции добротности, осуществляемой так, чтобы в конце цикла накачки (при $t = t_i$ на рис. 15.36) разность заселенностей составляла $N_i = 6N_п$, с помощью рис. 15.38 оцените форму лазерного импульса, его длительность, пиковую мощность и полную энергию.

15.4.4. Синхронизация мод

Лазер может генерировать на множестве продольных мод с частотами, разделенными межмодовым расстоянием $\nu_F = c/2d$ резонатора Фабри—Перо. Хотя обычно эти моды генерируются независимо (в этом случае они называются несинхронизированными модами), с помощью внешних воздействий можно связать их и синхронизировать их фазы. Тогда эти моды можно рассматривать как Фурье-компоненты периодической функции времени с периодом $T_F = 1/\nu_F = 2d/c$, которая представляет собой периодический цуг импульсов. Этот подход мы применяли в подразд. 2.6.2 при рассмотрении интерференции M монохроматических волн с одинаковой интенсивностью и равноотстоящими частотами. Рассмотрим теперь сначала свойства цуга импульсов при синхронизации мод, а потом методы ее достижения. Затем приведем несколько примеров лазеров с синхронизацией мод.

Свойства цуга импульсов при синхронизации мод

Если аппроксимировать каждую из лазерных мод однородной плоской волной, распространяющейся в направлении z со скоростью $c = c_0/n$, то комплексную волновую функцию поля можно записать в виде суммы

$$U(z,\ t) = \sum_q A_q \exp\left[j2\pi\nu_q\left(t - \frac{z}{c}\right)\right], \tag{15.71}$$

где ν_q — частота моды q,

$$\nu_q = \nu_0 + q\nu_F, \quad q = 0,\ \pm 1,\ \pm 2, \tag{15.72}$$

A_q — ее комплексная огибающая. Для удобства мы предполагаем, что частота моды $q = 0$ совпадает с центральной частотой атомной линии ν_0. Модули $|A_q|$ можно определить, зная спектральный профиль усиления и потерь резонатора (см. подразд. 15.2.2). Поскольку моды взаимодействуют с разными группами атомов в неоднородно уширенной среде, их фазы $\arg\{A_q\}$ случайны и статистически независимы.

Подставляя (15.72) в (15.71), находим

$$U(z,\ t) = \mathcal{A}\left(t - \frac{z}{c}\right)\exp\left[j2\pi\nu_0\left(t - \frac{z}{c}\right)\right], \tag{15.73}$$

где комплексная огибающая $\mathcal{A}(t)$ есть функция

$$\mathcal{A}(t) = \sum_q A_q \exp\left(\frac{jq2\pi t}{T_F}\right); \tag{15.74}$$

$$T_F = \frac{1}{\nu_F} = \frac{2d}{c}. \tag{15.75}$$

Комплексная огибающая $\mathcal{A}(t)$ в (15.74) — периодическая функция t с периодом T_F, а $\mathcal{A}(t - z/c)$ — периодическая функция z с периодом $cT_F = 2d$. Если модули и фазы комплексных коэффициентов A_q подобраны должным образом, то функции $\mathcal{A}(t)$ можно придать вид периодической последовательности коротких импульсов.

Рассмотрим, например, M мод ($q = 0, \pm 1, \ldots, \pm S$, так что $M = 2S + 1$), у которых все комплексные коэффициенты равны, $A_q = A$, $q = 0, \pm 1, \ldots, \pm S$. Тогда

$$\mathcal{A}(t) = A \sum_{q=-S} \exp\left(\frac{jq2\pi t}{T_F}\right) = A \sum_{q=-S}^{S} x^q = \\ = A\frac{x^{S+1} - x^{-S}}{x-1} = A\frac{x^{S+1/2} - x^{-S-1/2}}{x^{1/2} - x^{-1/2}}, \tag{15.76}$$

где $x = \exp(j2\pi t/T_F)$ (см. подробности в подразд. 2.6.2). После несложных алгебраических преобразований $\mathcal{A}(t)$ можно представить в виде

$$\mathcal{A}(t) = A\frac{\sin(M\pi t/T_F)}{\sin(\pi t/T_F)}. \tag{15.77}$$

Тогда оптическая интенсивность, определяемая как $I(t, z) = |\mathcal{A}(t - z/c)|^2$, выражается как

$$I(t, z) = |A|^2 \frac{\sin^2\left[\dfrac{M\pi(t - z/c)}{T_F}\right]}{\sin^2\left[\dfrac{\pi(t - z/c)}{T_F}\right]} \tag{15.78}$$

и является периодической функцией времени, график которой показан на рис. 15.39.

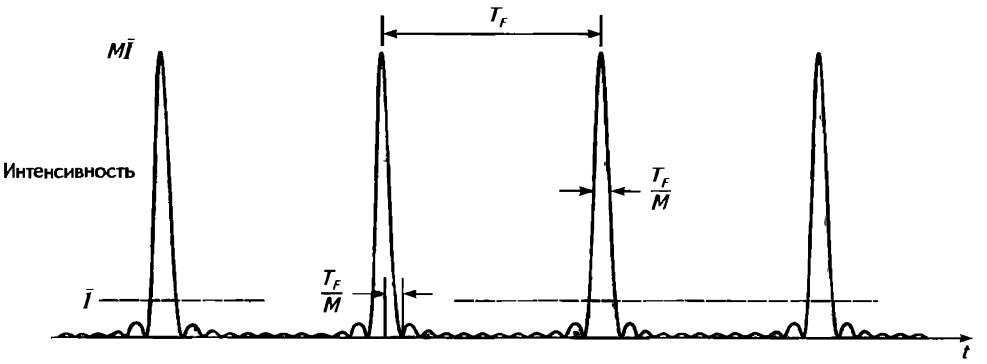


Рис. 15.39. Интенсивность периодического цуга импульсов, получающихся в результате сложения M лазерных мод с равными амплитудами и фазами. Каждый импульс имеет длительность, которая в M раз меньше периода следования T_F, и пиковую интенсивность, которая в M раз больше средней интенсивности

Таким образом, цуг импульсов, получающийся при синхронизации мод, зависит от числа мод M, которое пропорционально ширине атомной линии усиления $\Delta\nu$. Если $M \approx \Delta\nu/\nu_F$, то $\tau_{имп} = T_F/M \approx 1/\Delta\nu$. Длительность импульса $\tau_{имп}$ обратно пропорциональна ширине атомной линии $\Delta\nu$. Поскольку $\Delta\nu$ может быть весьма большой, за счет синхронизации мод можно генерировать очень короткие импульсы. Отношение пиковой интенсивности к средней интенсивности равно числу мод M, которое также может быть весьма большим.

Период следования импульсов равен $T_F = 2d/c$. Он равен времени полного обхода резонатора. Действительно, в лазере с синхронизацией мод свет можно представить как один узкий пакет фотонов, отражающийся туда и обратно между зеркалами (рис. 15.40). При каждом отражении от выходного зеркала часть фотонов проходит через зеркало наружу в виде импульса излучения. Прошедшие импульсы разделены в пространстве расстоянием $c(2d/c) = 2d$ и имеют пространственную протяженность $d_{имп} = c\tau_{имп} = 2d/M$.

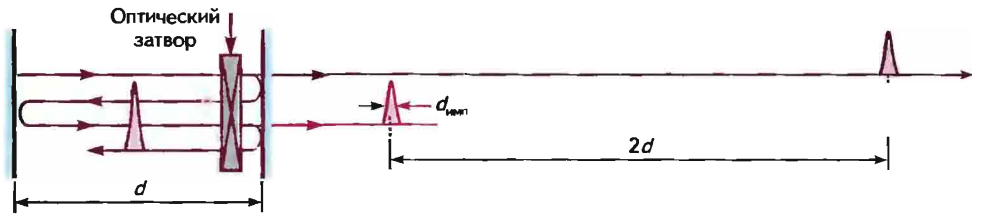


Рис. 15.40. При синхронизации мод лазерный импульс отражается туда и обратно между зеркалами резонатора. Достигая выходного зеркала, он каждый раз частично проходит через него. Прошедшие импульсы разделены расстоянием $2d$ и распространяются со скоростью c. Затвор открывается, только когда импульс достигает его, и открыт только в течение длительности импульса. Таким образом, периодическая последовательность импульсов не подвергается воздействию со стороны затвора. Другие волны, однако, испытывают большие потери и не допускаются к генерации

Сводка свойств цуга импульсов при синхронизации мод приведена в табл. 15.2.

Таблица 15.2. **Характерные свойства цуга импульсов при синхронизации мод**

Временной период	$T_F = \dfrac{2d}{c}$	Длительность импульса	$\tau_{имп} = \dfrac{T_F}{M} = \dfrac{1}{\Delta\nu}$
Пространственный период	$2d$	Длина импульса	$d_{имп} = \dfrac{2d}{M}$
Средняя интенсивность	$\bar{I}$	Пиковая интенсивность	$I_p = M\bar{I}$

В качестве конкретного примера рассмотрим лазер на стекле с неодимом, работающий на $\lambda_0 = 1{,}05$ мкм (см. табл. 14.1). Он имеет показатель преломления $n = 1{,}5$ и ширину линии усиления $\Delta\nu = 7$ ТГц. Таким образом, длительность импульса $\tau_{имп} = 1/\Delta\nu \approx 140$ фс, а его длина в пространстве $d_{имп} \approx 42$ мкм. Если резонатор

имеет длину $d = 15$ см, то межмодовое расстояние равно $\nu_F = c/2d = 1$ ГГц, что дает $M = \Delta\nu/\nu_F = 7000$ мод. Это значит, что пиковая интенсивность в 7000 раз больше средней интенсивности. В средах с широкими спектральными линиями синхронизация мод имеет преимущества перед модуляцией добротности для получения коротких импульсов. Газовые лазеры обычно имеют узкие линии, так что в них ультракороткие импульсы не удается получать с помощью синхронизации мод.

Формулы, приведенные выше, получены для особого случая, когда все амплитуды и фазы мод равны, однако расчеты на основе более реалистических моделей показывают, что сделанные нами выводы не утрачивают силы в общем случае.

Упражнение 15.5

Демонстрация пульсаций при синхронизации мод

Напишите компьютерную программу для построения графика интенсивности волны $I(t, z) = |\mathcal{A}(t)|^2$, огибающая которой $\mathcal{A}(t)$ дается суммой (15.74). Считайте, что число мод $M = 11$ и используйте следующий выбор коэффициентов A_q.

а. Все модули и фазы равны (воспроизведение рассмотренного случая).

б. Модули подчиняются гауссову распределению $|A_q| = \exp[-(1/2)(q/5)^2]$, а фазы равны.

в. Модули равны, а фазы случайны (получите их с помощью генератора случайных чисел с равномерным распределением на интервале от 0 до 2π).

Методы синхронизации мод

Мы показали, что если значительное число мод M синхронизованы по фазе, то они образуют гигантский узкий импульс фотонов, которые отражаются туда и обратно между зеркалами резонатора. Пространственная длина этого импульса в M раз меньше, чем длина резонатора. Остается ответить на вопрос, как синхронизовать моды таким образом, чтобы они имели одинаковую фазу. Эту задачу можно решить с помощью активного или пассивного модулятора (затвора), помещенного внутрь резонатора. Мы по очереди рассмотрим **активную** и **пассивную синхронизацию мод**.

Предположим, что оптический переключатель, управляемый внешним сигналом (например, акустооптический или электрооптический модулятор, которые рассматриваются в гл. 19 и 20), помещен внутрь резонатора и блокирует свет все время, кроме тех промежутков, когда импульс собирается пройти через него. Во время прохождения импульса модулятор остается открытым (см. рис. 15.40). Поскольку импульсу позволяется свободно проходить через модулятор, он не испытывает влияния со стороны последнего, и цуг импульсов не прерывается. В отсутствие синхронизации мод отдельные моды имеют различные фазы, определяемые случайными условиями начала генерации каждой из них. Если фазы случайно окажутся равными, то сумма мод сформирует гигантский импульс, который не будет подвержен влиянию модулятора. Любая дру-

гая комбинация фаз приведет к формированию такого распределения поля, которое будет полностью или частично блокироваться модулятором, что вносит в систему дополнительные потери. Следовательно, в присутствии модулятора лазерная генерация возможна только при условии, что фазы мод одинаковы. Лазер ждет «счастливого случая» такого совпадения фаз, но уж если это произошло и генерация началась, то фазы остаются синхронизованными.

Проблема также может быть исследована математически. Оптическое поле должно удовлетворять волновому уравнению с граничными условиями, наложенными присутствием модулятора. Многомодовое оптическое поле (15.71) действительно удовлетворяет волновому уравнению для любой комбинации фаз. В случае одинаковых оно удовлетворяет также граничным условиям, которые накладывает модулятор, следовательно, это должно быть единственное решение.

Пассивный затвор, такой как **насыщаемый поглотитель**, также может использоваться для достижения синхронизации мод. Насыщаемый поглотитель (см. подразд. 14.4.1) — это среда, коэффициент поглощения которой убывает с ростом интенсивности проходящего через нее света. Такая среда сравнительно мало поглощает мощные импульсы и сильно поглощает слабое излучение. Поэтому генерация может происходить только тогда, когда фазы различных мод связаны друг с другом так, что образуется интенсивный импульс, который в дальнейшем будет проходить через затвор. Широкое применение имеют полупроводниковые зеркала с насыщаемым поглощением (SESAM), которые представляют собой насыщаемые поглотители, работающие на отражение. Чем больше интенсивность света, тем больше коэффициент отражения этих устройств. SESAM работают на длинах волн от 800 до 1600 нм при длительностях импульсов от фс до наносекунд и уровнях мощности от милливатт до ватт. Насыщаемый поглотитель может вызывать также синхронизацию мод за счет модуляции добротности, когда лазер излучает цуг импульсов синхронизации мод в пределах огибающей импульса модуляции добротности.

Пассивную синхронизацию мод можно осуществить также с помощью **керровской линзы**, которая представляет собой нелинейно-оптическое явление, когда показатель преломления материала меняется в зависимости от оптической интенсивности (см. разд. 21.3). Керровская среда, в качестве которой может выступать как сама активная среда лазера, так и материал, дополнительно вводимый в резонатор, действует как линза с фокусным расстоянием, обратно пропорциональным интенсивности света (см. упражнение 21.4). Таким образом, керровская линза уменьшает площадь лазерной моды при высоких интенсивностях. Такой свет проходит через диафрагму, расположенную в нужном месте резонатора. В качестве альтернативы можно использовать уменьшение площади сечения моды в усиливающей среде для увеличения перекрывания мод с сильно сфокусированным пучком накачки и, следовательно, для повышения эффективного усиления. Подход, основанный на керровской линзе, изначально широкополосный, в силу параметрической природы явления. Быстрое восстановление системы, типичное для пассивной синхронизации мод, приводит к генерации более коротких импульсов света, чем можно достичь при активной

синхронизации. Пассивные и активные модуляторы (оптические затворы) применяются для синхронизации мод в средах с неоднородным и однородным уширением сходным образом.

Примеры лазеров с синхронизацией мод

В табл. 15.3 в порядке возрастания приведены длительности импульсов различных лазеров с синхронизацией мод. Диапазон представленных длительностей импульсов весьма широк. Наблюдаемые длительности импульса, которые для данного вещества могут сильно различаться, зависят от метода достижения синхронизации мод и могут быть ограничены нелинейностью и дисперсией среды.

Обладая возможностью перестройки центральной длины волны в пределах от 700 до 1050 нм и позволяя получать отдельные импульсы длительностью до 10 фс, именно титан-сапфировый лазер чаще всего используется для работы в режиме синхронизации мод. Промышленный вариант такого лазера дает импульсы с энергией 50 нДж, длительностью 10 фс, пиковой мощностью 1 МВт и частотой повторения 80 МГц. Существенно лучшие рабочие характеристики можно получить на лабораторных установках. Дальнейшее уменьшение длительности возможно с помощью методов сжатия импульсов (см. подразд. 22.2.1). Следует заметить, что спектральную ширину полосы усиления $\Delta\nu$ этого лазера всегда можно ограничить так, чтобы в режиме синхронизации мод он генерировал пикосекундные импульсы. Помимо исследовательских задач фотоники, лазеры с синхронизацией мод находят широкое применение во многих прикладных областях, включая измерения с разрешением во времени, формирование изображений, метрологию, связь, обработку материалов и клиническую медицину.

Таблица 15.3. **Типичные длительности импульсов для ряда лазеров с однородным (О) и неоднородным (Н) уширением при синхронизации мод**

Активная среда	Тип уширения	Ширина линии перехода* $\Delta\nu$	Рассчитанная длительность импульса $\tau_{имп} = 1/\Delta\nu$	Наблюдаемая длительность импульса
$Ti^{3+}:Al_2O_3$	О	100 ТГц	10 фс	10 фс
Краситель родамин 6Ж	О/НО	40 ТГц	25 фс	27 фс
Nd^{3+}: фосфатное стекло	НО	7 ТГц	140 фс	150 фс
Er^{3+}: кварцевое волокно	О/НО	5 ТГц	200 фс	200 фс
$Nd^{3+}:YAG$	О	150 ГГц	7 пс	7 пс
Ar^+	НО	3,5 ГГц	286 пс	150 пс
He–Ne	НО	1,5 ГГц	667 пс	600 пс
CO_2	НО	60 МГц	16 нс	20 нс

* Ширина линий переходов $\Delta\nu$ взята из табл. 14.1.

Пример 15.6

Синхронизация мод в волоконном лазере, легированном иттербием

Лазер на кварцевом волокне, легированном иттербием, работающий в режиме синхронизации мод на длине волны $\lambda_0 = 1070$ нм, генерирует излучение со средней мощностью 10 Вт в виде импульсов с энергией 200 нДж и пиковой мощностью 40 кВт. Импульсы имеют длительность 5 пс и частоту повторения 50 МГц. Учитывая, что $\Delta\nu = 5$ ТГц, видим, что длительность импульса существенно больше ожидаемого значения $\tau_{имп} = 1/\Delta\nu = 200$ фс. Расхождение возникает из-за дисперсии групповой скорости, которая вносит вклад в уширение и чирп импульса, распространяющегося в оптической среде (см. рис. 5.26). Нормальная дисперсия в кварце вблизи $\lambda_0 = 1$ мкм (см. рис. 5.40) может быть скомпенсирована введением аномальной дисперсии с помощью волоконной решетки Брэгга или фотонно-кристаллического волокна, что уменьшает наблюдаемую длительность импульса до 200 фс.

Уровень мощности титан-сапфирового лазера с синхронизацией мод позволяет использовать генерацию гармоник и другие нелинейно-оптические методы для изменения длины волны (см. гл. 21), что дает источник коротких импульсов в более коротковолновой области спектра. В частности, генерация второй гармоники даст импульсы в области 350—525 нм, а третьей — в области 230—350 нм. Для получения импульсов на длинах волн больше $\lambda_0 = 1$ мкм в ИК-диапазоне можно использовать волоконный ВКР-лазер или нелинейное параметрическое преобразование частоты вниз (см. подразд. 21.4.3). Титан-сапфировый лазер в режиме синхронизации мод служит источником для оптического параметрического генератора с синхронной накачкой, в котором используется нелинейный оптический кристалл, такой как триборат лития (LBO), или кристалл с периодической гребневой структурой. При этом генерируется сигнальная и холостая волна в виде коротких импульсов в диапазоне 1,0—3,3 мкм.

Рекомендуемая литература

КНИГИ ПО ЛАЗЕРАМ

См. также список литературы к гл. 14 и книги по оптоэлектронике к гл. 17.

Koechner W. Solid-State Laser Engineering. Springer-Verlag, 6th ed. 2006.

Quimby R.S. Photonics and Lasers: An Introduction. Wiley, 2006.

Yariv A., Yeh P. Photonics: Optical Electronics in Modern Communications. Oxford University Press, 6th ed. 2006.

Band Y.B. Light and Matter: Electromagnetism, Optics, Spectroscopy and Lasers. Wiley, 2006.

Noginov M.A. Solid-State Random Lasers. Springer-Verlag, 2005.

Rulliere C., ed. Femtosecond Laser Pulses: Principles and Experiments. Springer, 1998; 2nd ed., 2005.

Silfvast W.T. Laser Fundamentals. Cambridge University Press, 2nd ed. 2004.

Svelto O. Principles of Lasers. Springer-Verlag, 4th ed. 2004.

Meschede D. Optics, Light and Lasers: The Practical Approach to Modern Aspects of Photonics and Laser Physics. Wiley-VCH, 2004.

Saldin E.L., Schneidmiller E.A., Yurkov M.V. The Physics of Free Electron Lasers, Springer-Verlag, 2000.

Ciocci F., Dattoli G., Torre A., Renieri A. Insertion Devices for Synchrotron Radiation and Free Electron Laser. World Scientific, 2000.

Weber M.J. Handbook of Laser Wavelengths. CRC Press, 1999.

Kaminskii A.A. Crystalline Lasers: Physical Processes and Operating Schemes. CRC Press, 1996.

Verdeyen J.T. Laser Electronics. Prentice Hall, 3rd ed. 1995.

Duarte F.J., ed. Tunable Lasers Handbook. Academic Press, 1995.

Yokoyama H., Ujihara K., eds., Spontaneous Emission and Laser Oscillation in Microcavities. CRC Press, 1995.

Elitzur M. Astronomical Masers. Kluwer, 1992.

Schäfer F.P., ed. Dye Lasers. Springer-Verlag, 3rd ed. 1990.

Basov N.G., Bashkin A.S., Igoshin V.I., Oraevsky A.N., Shcheglov A.A. Chemical Lasers. Springer-Verlag, 1990.

Luchini P., Motz H. Undulators and Free-electron Lasers. Oxford University Press, 1990.

Milonni P W., Eberly J.H. Lasers. Wiley, 1988.

Cheo P.K., ed. Handbook of Molecular Lasers. Marcel Dekker, 1987.

Siegman A.E. Lasers. University Science, 1986.

Shimoda K. Introduction to Laser Physics. Springer-Verlag, 2nd ed. 1986.

Rhodes C.K., ed. Excimer Lasers. Springer-Verlag, 2nd ed. 1984.

КНИГИ И СТАТЬИ ПО УЛЬТРАФИОЛЕТОВЫМ И РЕНТГЕНОВСКИМ ЛАЗЕРАМ

Jaeglé P. Coherent Sources of XUV Radiation: Soft X-Ray Lasers and High-Order Harmonic Generation. Springer-Verlag, 2006.

Rocca J.J., Kapteyn H.C., Atwood D.T., Murnane M.M., Menoni C.S., Anderson E.H. Tabletop Lasers in the Extreme Ultraviolet. Optics & Photonics News. Vol. 17, № 11, 2006. P. 30—38.

Paganin D.M. Coherent X-Ray Optics. Oxford University Press, 2006.

Feature section on extreme ultraviolet coherent sources and applications. IEEE Journal of Quantum Electronics. Vol. 42, № 1, 2006.

Weith A., Larotonda M.A., Wang Y., Luther B.M., Alessi D., Marconi M.C., Rocca J.J., Dunn J. Continuous High-Repetition-Rate Operation of Collisional Soft-X-Ray Lasers With Solid Targets. Optics Letters. Vol. 31, 2006. P. 1994—1996.

Zhang J., ed. X-Ray Lasers 2004 (Institute of Physics Conference Series). Taylor & Francis, 2005.

Kapteyn H.C., Murnane M.M., Christov I.P. Extreme Nonlinear Optics: Coherent X-Rays from Lasers. Physics Today. Vol. 58, № 3, 2005. P. 39—44.

Shvyd'ko Y. X-Ray Optics: High-Energy-Resolution Applications. Springer-Verlag, 2004.

Issue on short wavelength and EUV lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 10, № 6, 2004.

Rocca J.J., Dunn J., Suckewer S., eds. X-Ray Lasers 2002 (AIP Conference Proceedings 641). American Institute of Physics, 2002.

Attwood D. Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications. Cambridge University Press, 1999.

Turcu I.C.E., Dance J.B. X-Rays From Laser Plasmas: Generation and Applications. Wiley, 1999.

Issue on short wavelength lasers and applications. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 5, № 6, 1999.

Rullhusen P., Artru X., Dhez P. Novel Radiation Sources Using Relativistic Electrons: From Infrared to X-Rays. World Scientific, 1998.

Duley W.W. UV-Lasers: Effects and Applications in Materials Science. Cambridge University Press, 1996.

Waynant R.W., Ediger M.N., eds. Selected Papers on UV-, VUV- and X-Ray Lasers. SPIE Optical Engineering Press (Milestone Series. Vol. 71), 1993.

Elton R.C. X-Ray Lasers. Academic Press, 1990.

Matthews D.L., Rosen M.D. Soft X-Ray Lasers. Scientific American. Vol. 259, № 6, 1988. P. 86—91.

Matthews D.L., Hagelstein P.L., Rosen M.D., Eckart M.J., Ceglio N.M., Hazi A.U., Medecki H., MacGowan B.J., Trebes J.E., Whitten B.L., Campbell E.M., Hatcher C.W., Hawryluk A.M., Kauffman R.L., Pleasance L.D., Rambach G., Scofield J.H., Stone G., Weaver T.A. Demonstration of a Soft X-Ray Amplifier. Physical Review Letters. Vol. 54, 1985. P. 110—113.

Suckewer S., Skinner C.H., Milchberg H., Keane C., Voorhees D. Amplification of Stimulated Soft X-Ray Emission in a Confined Plasma Column. Physical Review Letters. Vol. 55, 1985. P. 1753—1756.

СТАТЬИ

Feature issue on fiber lasers. Journal of the Optical Society of America B. Vol. 24, № 8, 2007.

Kent A.J., Kini R.N., Stanton N.M., Henini M., Glavin B.A., Kochelap V.A., Linnik T.L. Acoustic Phonon Emission from a Weakly Coupled Superlattice Under Vertical Electron Transport: Observation of Phonon Resonance. Physical Review Letters. Vol. 96, 2006. 215504.

Cao H. Review on Latest Developments in Random Lasers with Coherent Feedback. Journal of Physics A: Mathematical and General. Vol. 38, 2005. P. 10497—10535,

Cao H. Random Lasers: Development, Features and Applications. Optics & Photonics News. Vol. 16, № 1, 2005. P. 24—29.

Issue on solid-state lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 11, № 3, 2005.

Claussen M. Astronomical Masers. Science. Vol. 306, 2004. P. 235—236.

Walsworth R.L. The Maser at 50. Science. Vol. 306, 2004. P. 236—237.

Keller U. Recent Developments in Compact Ultrafast Lasers. Nature. Vol. 424, 2003. P. 831—838.

Ter-Mikirtychev V.V., ed. Selected Papers on Tunable Solid-State Lasers. SPIE Optical Engineering Press (Milestone Series. Vol. 173), 2002.

Colson W.B., Johnson E.D., Kelley M.J., Schwettman H.A., Putting Free-Electron Lasers to Work. Physics Today. Vol. 55, № 1, 2002. P. 35—41.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Kim N.S., Prabhu M., Li C., Song J., Ueda K-i. 1239/1484 nm Cascaded Phosphosilicate Raman Fiber Laser with CW Output Power of 1.36 W at 1484 nm Pumped by CW Vb-Doped Double-Clad Fiber Laser at 1064 nm and Spectral Continuum Generation. Optics Communications. Vol. 176, 2000. P. 219—222.

Feld M.S., An K. The Single-Atom Laser. Scientific American. Vol. 279, № 1, 1998. P. 56—63.

Elitzur M. Masers in the Sky. Scientific American. Vol. 272, № 2, 1995. P. 68—74.

Perry M.D., Mourou G. Terawatt to Petawatt Subpicosecond Lasers. Science. Vol. 264, 1994. P. 917—924.

Silfvast W.T., ed. Selected Papers on Fundamentals of Lasers. SPIE Optical Engineering Press (Milestone Series. Vol. 70), 1993.

Mumma M.J., Buhl D., Chin G., Deming D., Espenak F., Kostiuk T. Discovery of Natural Gain Amplification in the 10 μm CO_2 Laser Bands on Mars: A Natural Laser. Science. Vol. 212, 1981. P. 45—49.

ИСТОРИЧЕСКИЙ МАТЕРИАЛ

Hecht J. Beam: The Race to Make the Laser. Oxford University Press, 2005.

Bertolotti M. The History of the Laser. Taylor & Francis, 2004.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Hecht J., ed. Laser Pioneer Interviews. High Tech Publications, 1985.

Kastler A. Birth of the Maser and Laser. Nature. Vol. 316, 1985. P. 307—309.

Special issue: «Twenty-Five Years of the Laser», Optica Acta (Journal of Modern Optics). Vol. 32, № 9/10, 1985.

Centennial issue. IEEE Journal of Quantum Electronics. Vol. QE-20, № 6, 1984.

Townes C.H. Science, Technology, and Invention: Their Progress and Interactions. Proceedings of the National Academy of Sciences (USA). Vol. 80, 1983. P. 7679—7683.

Schawlow A.L. Spectroscopy in a New Light (Nobel lecture). Reviews of Modern Physics. Vol. 54, 1982. P. 697—707.

O'Shea D., Peckham D.C. Lasers: Selected Reprints. American Association of Physics Teachers, 1982.

O'Shea D.C., Peckham D.C. Resource Letter L-1: Lasers. American Journal of Physics. Vol. 49, 1981. P. 915—925.

Schawlow A.L. Maser and Laser. IEEE Transactions on Electron Devices. Vol. ED-23, 1976. P. 773—779.

Schawlow A.L. From Maser to Laser, in Impact of Basic Research on Technology. B. Kursunoglu and A. Perlmutter, eds. Plenum, 1973.

Lamb, Jr. W.E. Physical Concepts in the Development of the Maser and Laser. In Impact of Basic Research on Technology. B. Kursunoglu and A. Perlmutter, eds. Plenum, 1973.

Kastler A. Optical Methods for Studying Hertzian Resonances, in Nobel Lectures in Physics, 1963—1970. Elsevier, 1972.

Townes C.H. Production of Coherent Radiation by Atoms and Molecules, in Nobel Lectures in Physics, 1963—1970. Elsevier, 1972.

Basov N.G. Semiconductor Lasers. In Nobel Lectures in Physics, 1963—1970. Elsevier, 1972.

Prokhorov A.M. Quantum Electronics. In Nobel Lectures in Physics, 1963—1970. Elsevier, 1972.

Barnes F.S., ed. Laser Theory. IEEE Press Reprint Series, IEEE Press, 1972.

Stolen R.H., Ippen E.P., Tynes A.R. Raman Oscillation in Glass Optical Waveguide. Applied Physics Letters. Vol. 20, 1972. P. 62—65.

Schawlow A.L., ed. Lasers and Light-Readings from Scientific American. Freeman, 1969.

Townes C.H. Quantum Electronics and Surprise in the Development of Technology. Science. Vol. 159, 1968. P. 699—703.

Letokhov V.S. Generation of Light by a Scattering Medium with Negative Resonance Absorption. Soviet Physics-JETP. Vol. 26, 1968. P. 835—840, [Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki. Vol. 53, 1967. P. 1442—1452].

Weber J., ed. Lasers: Selected Reprints with Editorial Comment. Gordon and Breach, 1967.

Lengyel B.A. Evolution of Masers and Lasers. American Journal of Physics. Vol. 34, 1966. P. 903—913.

Cohen- Tannoudji C., Kastler A. Optical Pumping., In Progress in Optics. E. Wolf, ed. North-Holland. Vol. 5, 1966.

Lamb, Jr. W.E. Theory of an Optical Maser. Physical Review. Vol. 134, 1964. P. A1429—A1450.

Yariv A., Gordon J.P. The Laser. Proceedings of the IEEE. Vol. 51, 1963. P 4—29.

Maiman T.H. Stimulated Optical Radiation in Ruby. Nature. Vol. 187, 1960. P. 493—494.

Schawlow A.L., Townes C.H. Infrared and Optical Masers. Physical Review. Vol. 112, 1958. P. 1940—1949.

Dicke R.H. Molecular Amplification and Generation Systems and Methods. U.S. Patent 2,851,652. September 9, 1958.

Gordon J.P., Zeiger H.J., Townes C.H. The Maser-New Type of Microwave Amplifier, Frequency Standard, and Spectrometer. Physical Review. Vol. 99, 1955. P. 1264—1274.

Basov N.G., Prokhorov A.M. Possible Methods of Obtaining Active Molecules for a Molecular Oscillator. Soviet Physics-JETP Vol. 1, 1955. P. 184—185 [Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki. Vol. 28, 1955. P. 249—250].

Fabrikant V.A. The Emission Mechanism of Gas Discharges. Trudi Vsyesoyuznogo Elektrotekhnicheskogo Instituta (Reports of the All-Union Electrotechnical Institute, Moscow). Vol. 41, Elektronnie i Ionnie Pribori (Electron and Ion Devices), 1940. P. 236—296.

Задачи

К РАЗДЕЛУ 15.2

1. **Число продольных мод.** Аргоновой ионный лазер имеет резонатор длиной 100 см. Показатель преломления $n = 1$.

а. Определите частотное расстояние ν_F между модами резонатора.

б. Определите число продольных мод, на которых может генерировать лазер, если ширина доплеровского контура линии усиления $\Delta\nu_D = 3{,}5$ ГГц, а коэффициент потерь составляет половину максимума коэффициента усиления при слабом сигнале.

в. Какой должна быть длина резонатора d для достижения генерации на одной продольной моде? Какой должна быть длина у CO_2 лазера с гораздо меньшей доплеровской шириной линии $\Delta\nu_D = 60$ МГц при тех же условиях?

2. **Дрейф частоты лазерных мод.** He—Ne-лазер имеет следующие характеристики:

1) резонатор с коэффициентами отражения зеркал 97 и 100 % и пренебрежимо малыми внутренними потерями;

2) доплеровски-уширенная линия атомного перехода с шириной $\Delta\nu_D = 1{,}5$ ГГц;

3) максимальный коэффициент усиления слабого сигнала $\gamma_0(\nu_0) = 2{,}5 \cdot 10^{-3}$ см$^{-1}$.

В процессе генерации частоты мод дрейфуют со временем в результате тепловых изменений длины резонатора. Найдите допустимые пределы изменений длины резонатора, внутри которых лазер генерирует одну или две пролольные моды (но не более). Показатель преломления $n = 1$.

3. **Селекция мод с помощью эталона.** Газовый лазер с доплеровским уширением работает на 515 нм при расстоянии между зеркалами резонатора 50 см.

Время жизни фотона составляет 0,33 нс. Спектральное окно, в котором возможна генерация, имеет ширину $B = 1{,}5$ ГГц. Показатель преломления $n = 1$. С целью выбора одной моды для генерации свет пропускают через эталон (пассивный резонатор Фабри—Перо), расстояние между зеркалами которого d, а параметр резкости $\mathcal{F}$. Эталон служит фильтром. Предложите подходящие значения d и $\mathcal{F}$. Как лучше расположить эталон: внутри или снаружи резонатора?

4. **Мощности мод в многомодовом лазере.** Не—Ne-лазер, работающий на $\lambda_0 = 632{,}8$ нм, имеет выходную мощность 50 мВт в многомодовом режиме. Он имеет неоднородно уширенный спектральный профиль усиления с лоплеровской шириной $\Delta\nu_D = 1{,}5$ ГГц и показатель преломления $n = 1$. Длина резонатора 30 см.

а. При условии, что коэффициент потерь вдвое меньше максимального коэффициента усиления слабого сигнала, определите число продольных мод лазера.

б. Если зеркала настроены так, чтобы обеспечить максимальную интенсивность наиболее сильной моды, оцените мощность этой моды.

5. **Выходное излучение одномодового газового лазера.** Рассмотрим газовый лазер длиной 10 см, работающий на центре линии перехода с длиной волны 600 нм в режиме одной продольной и одной поперечной моды. Коэффициенты отражения зеркал $\mathcal{R}_1 = 99\,\%$ и $\mathcal{R}_2 = 100\,\%$. Показатель преломления $n = 1$, а площадь поперечного сечения выходящего пучка 1 мм^2. Коэффициент усиления слабого сигнала $\gamma_0(\nu_0) = 0{,}1$ см$^{-1}$, а насыщающая плотность потока фотонов $\phi_s = 1{,}43 \cdot 10^{19}$ фотонов/(см$^2 \cdot$ с).

а. Определите коэффициенты распределенных потерь α_{m_1} и α_{m_2}, связанные с зеркалами 1 и 2. Предполагая, что $\alpha_s = 0$, определите коэффициент потерь резонатора α_r.

б. Найдите время жизни фотона τ_Φ.

в. Определите плотность потока фотонов на выходе ϕ_0 и выходную мощность P_0.

6. **Коэффициент пропускания лазерного резонатора.** Монохроматический свет от перестраиваемого источника пропускают через оптический резонатор газового лазера с выключенной накачкой. Наблюдаемый коэффициент пропускания как функция частоты показан на рис. 15.41.

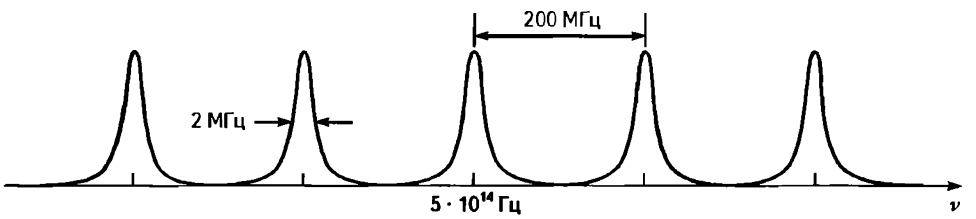


Рис. 15.41. Коэффициент пропускания лазерного резонатора

а. Определите длину резонатора, время жизни фотона и пороговый коэффициент усиления лазера. Показатель преломления считайте равным единице.

б. Полагая центральную частоту лазерного перехода равной $5 \cdot 10^{14}$ Гц, изобразите графически коэффициент пропускания в зависимости от частоты, если включить накачку лазера, на уровне ниже порога, недостаточного для генерации.

7. **Скоростные уравнения для четырехуровневого газового лазера.** Рассмотрим четырехуровневый газовый лазер с объемом активной среды $V = 1$ см^3. Плотности заселенностей нижнего и верхнего лазерных уровней обозначим N_2 и N_1, а их разность $N = N_2 - N_1$. Скорость накачки такова, что стационарное значение разности заселенностей N в отсутствие вынужденного излучения равна N_0. Плотность числа фотонов равна n, а время жизни фотона $\tau_ф$. Запишите скоростные уравнения для переменных N_2, N_1, N и n, в которые в качестве параметров входят N_0, сечение перехода $\sigma(\nu)$ и времена $t_{сп}$, τ_1, τ_2, τ_{21} и $\tau_ф$. Определите стационарные значения N и n.

К РАЗДЕЛУ 15.3

1. **Генерация Yb^{3+} : YAG-лазера.** Yb^{3+} : YAG — диэлектрический материал, легированный редкоземельными ионами, генерирующий на переходе $^2F_{5/2} \rightarrow {}^2F_{7/2}$ излучение с длиной волны $\lambda_0 = 1{,}030$ мкм (см. табл. 13.1, 14.1, 15.1 и рис. 15.25). Этот трехуровневый лазер обычно накачивают лазерным диодом на InGaAs.

а. Энергетическая зона, через которую производится накачка (уровень 3), имеет центральную энергию 1,35915 эВ и ширину 0,02475 эВ. Определите желательную длину волны лазерного диода накачки и ширину полосы поглощения в нм.

б. На центральной частоте лазерного перехода ν_0 пиковое значение сечения перехода $\sigma_0 \equiv \sigma(\nu_0) = 2 \cdot 10^{-20}$ см^2. При плотности ионов Yb^{3+} $N_a = 1{,}4 \cdot 10^{20}$ см$^{-3}$ найдите коэффициенты поглощения и усиления материала в центре линии $\alpha(\nu_0) \equiv -\gamma(\nu_0)$. Считайте, что материал находится в состоянии теплового равновесия при $T = 300$ °K (т. е. накачка отсутствует).

в. Рассмотрим лазерный стержень из этого материала длиной 6 см и диаметром 2 мм. Торцы стержня полируются так, чтобы обеспечить коэффициенты отражения $\mathcal{R}_1 = 0{,}8$ и $\mathcal{R}_2 = 1{,}0$. Считая, что рассеяние отсутствует и других внешних потерь нет, определите коэффициент потерь резонатора α_r и время жизни фотона в резонаторе $\tau_ф$.

г. При включении накачки лазера коэффициент усиления $\gamma(\nu_0)$ растет от своего начального отрицательного значения при тепловом равновесии до положительного значения, обеспечивающего усиление. Определите пороговую разность заселенностей $N_п$ для лазерной генерации.

д. Почему выгодно, чтобы энергия уровня 3 была близка к энергии уровня 2?

е. Какие изменения в работе лазера произошли бы, если бы в качестве базового вещества был взят не иттрий-алюминиевый гранат (YAG), а ванадат иттрия (YVO_4)?

2. **Пороговая разность заселенностей в ионном аргоновом лазере.** Ионный аргоновый лазер имеет резонатор длиной 1 м с коэффициентами отражения зеркал 98 и 100 %. Другие механизмы потерь пренебрежимо малы. Атомный

переход имеет центральную длину волны $\lambda_0 = 515$ нм, спонтанное время жизни $t_{сп} = 10$ нс и ширину линии $\Delta\lambda = 0{,}003$ нм. Нижний энергетический уровень имеет очень короткое время жизни и, следовательно, нулевую заселенность. Диаметр моды генерации 1 мм. Определите:

а) время жизни фотона;

б) пороговую разность заселенностей для лазерной генерации.

3. **Спонтанное время жизни ВУФ-перехода.** Лазерный переход видимого диапазона с длиной волны $\lambda_0 = 500$ нм имеет спонтанное время жизни $t_{сп} = 10$ нс. Оцените спонтанное время жизни для лазерного перехода в ВУФ-диапазоне, если $\lambda_0 = 18{,}2$ нм, при той же силе перехода S. Сравните ваш результат с данными табл. 14.1.

К РАЗДЕЛУ 15.4

1. **Переходные явления в лазере с модуляцией усиления.**

а. Введем новые переменные $X = n/\tau_ф$, $Y = N/N_п$ и нормированное время $s = t/\tau_ф$. Тогда скоростные уравнения (15.49) и (15.52) принимают вид

$$\frac{dX}{ds} = -X + XY;$$

$$\frac{dY}{ds} = a(Y_0 - Y) - 2XY,$$

где $a = \tau_ф/t_{сп}$ и $Y_0 = N_0/N_п$.

б. Напишите компьютерную программу для решения этих двух уравнений при включении и выключении. Предположите, что Y_0 меняется от 0 до 2 при включении лазера и от 2 до 0 при его выключении. Считайте, что очень слабый поток фотонов, соответствующий $X = 10^{-5}$, запускает генерацию при $t = 0$. Подумайте, откуда может взяться этот поток. Определите постоянные времени включения и выключения при $a = 10^{-3}$, 1 и 10^3. Прокомментируйте значение полученных результатов.

2. **Мощность рубинового лазера с модулированной добротностью.** Рубиновый лазер с модулированной добротностью имеет стержень длиной 15 см с площадью поперечного сечения 1 см^2, помещенный в резонатор длиной 20 см. Зеркала имеют коэффициенты отражения $\mathcal{R}_1 = 0{,}95$ и $\mathcal{R}_2 = 0{,}7$. Плотность ионов Cr^{3+} составляет $1{,}58\ \ 10^{19}$ атомов/см^3, а сечение перехода $\sigma(\nu_0) = 2\ \ 10^{-20}$ см^2. Накачка лазера обеспечивает начальную заселенность верхнего лазерного уровня 10^{19} атомов/см^3 и пренебрежимо малую заселенность нижнего лазерного уровня. Энергетическая зона, через которую происходит накачка (уровень 3), имеет центр на ≈450 нм, а распад за счет переходов с уровня 3 на уровень 2 происходит быстро. Время жизни уровня 2 составляет ≈3 мс.

а. Какова должна быть мощность накачки для поддержания заселенности уровня 2 равной 10^{19}?

б. Какая мощность излучается спонтанно до срабатывания модулятора добротности?

в. Найдите пиковую мощность, энергию и длительность импульса при модуляции добротности.

3. **Генерация лазера в режиме открытия резонатора.** Изобразите графически изменения пороговой разности заселенностей $N_п$ (которая пропорциональна потерям), разности заселенностей $N(t)$, плотности числа фотонов в резонаторе $n(t)$ и плотности потока фотонов на выходе из лазера ϕ_0 на протяжении двух рабочих циклов открытия резонатора.

4. **Синхронизация мод при лоренцевом распределении амплитуд.** Предположим, что огибающие мод в лазере с синхронизацией мод определяются выражением

$$A_q = \sqrt{P}\,\frac{(\Delta\nu/2)^2}{(q\nu_F)^2 + (\Delta\nu/2)^2}, \quad q = -\infty, \ \ldots, \ \infty,$$

а их фазы равны. Найдите выражения для следующих параметров генерируемого лазерного импульса:

а) средней мощности;

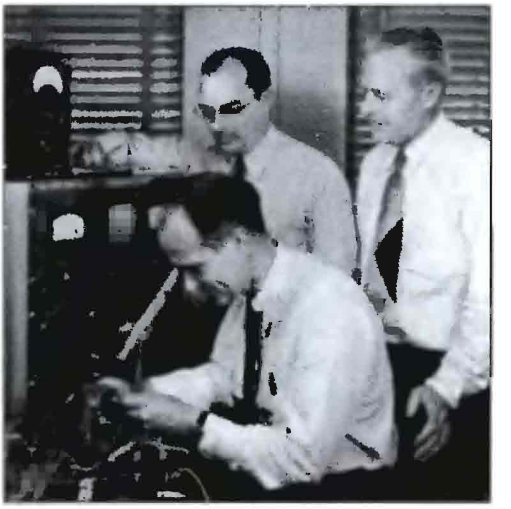
б) пиковой мощности;

в) длительности импульса на уровне полувысоты максимума.

5. **Генерация второй гармоники.** Кристаллы с нелинейными оптическими свойствами часто используются для генерации второй гармоники, которая рассматривается в гл. 21. В этом процессе два фотона с частотой ν преобразуются в один фотон с частотой 2ν. Предположим, что такой кристалл помещен внутрь резонатора лазера, активная среда которого обеспечивает усиление на частоте ν. Часто́ты ν и 2ν соответствуют двум модам резонатора. Пусть скорость преобразования во вторую гармонику составляет $\zeta\nu$ ($с^{-1}$ $м^{-3}$), а скорость рождения фотонов в процессе лазерной генерации (с учетом вынужденного излучения и поглощения) равна $\xi\nu$ ($с^{-1} \cdot м^{-3}$), где ζ и ξ — постоянные. Напишите скоростные уравнения для плотностей числа фотонов n и n_2 на частотах ν и 2ν. Считайте, что времена жизни фотонов на частотах ν и 2ν равны $\tau_ф$ и $\tau_{ф2}$ соответственно. Определите стационарные значения n и n_2.

ГЛАВА 16

ОПТИКА ПОЛУПРОВОДНИКОВ



Уильям Шокли (1910—1989) (*сидит*), **Джон Бардин (1908—1991)** (*слева*) и **Уолтер Браттейн (1902—1987)** (*справа*) были удостоены Нобелевской премии 1956 г. за демонстрацию того, что полупроводниковые устройства можно использовать для достижения усиления.

Фотоника управляет потоками фотонов в значительно мере так же, как электроника манипулирует потоками носителей заряда (электронов и дырок). Эти две технологии сливаются вместе в полупроводниковой оптоэлектронике: фотоны генерируют подвижные носители заряда, а носители заряда генерируют и контролируют поток фотонов. Устройства полупроводниковой оптоэлектроники действуют как источники фотонов (светоизлучающие диоды и диодные лазеры), усилители, приемники, волноводы, модуляторы, датчики и нелинейно-оптические элементы. Совместимость устройств полупроводниковой оптоэлектроники с электронными устройствами благоприятствовала развитию обеих областей.

Полупроводниковые материалы поглощают и излучают фотоны при переходах между разрешенными энергетическими уровнями. Хотя основные правила, управляющие такими взаимодействиями, такие же, как изложенные в разд. 13.3, полупроводники имеют ряд специфических свойств (см. подразд. 13.1.3):

- Полупроводниковый материал нельзя рассматривать как набор невзаимодействующих атомов, каждый из которых имеет свои собственные энергетические уровни. Из-за близости атомов в кристаллической решетке энергетические уровни принадлежат системе в целом.

• Наборы близко расположенных энергетических уровней образуют энергетические зоны. В отсутствие внешних возбуждений при $T = 0$ °K эти зоны либо полностью заполнены электронами, либо полностью свободны. Самая нижняя из незанятых энергетических зон называется зоной проводимости, а самая высшая из занятых зон — валентной зоной. Между этими зонами лежит запрещенная зона шириной E_g.

• Внешний источник энергии (тепловой, оптический или электронный) может сообщать энергию электронам валентной зоны, вызывая переходы в зону проводимости. Такой переход оставляет вакансию (дырку) в валентной зоне. При обратном процессе — электронно-дырочной рекомбинации — электрон покидает зону проводимости и заполняет свободное состояние в валентной зоне (при условии его доступности), испуская фотон и/или фононы. Таким образом, фотоны взаимодействуют с носителями заряда обоих типов, электронами и дырками.

Для работы почти всех оптоэлектронных полупроводниковых приборов основными являются два процесса:

1. *Поглощение фотона может создать электронно-дырочную пару.* Подвижные носители заряда, возникающие благодаря поглощению фотона, меняют электрические свойства полупроводника. Этот процесс лежит в основе действия фотопроводящих фотоприемников.

2. *Рекомбинация электрона и дырки может приводить к испусканию фотона.* Этот процесс отвечает за работу полупроводниковых источников фотонов. Спонтанное испускание при рекомбинации электрона и дырки порождает свет в светоизлучающих диодах, а вынужденное излучение — в диодных лазерах.

О данной главе

Предполагается, что читатель знаком с основными принципами физики полупроводников. В разд. 16.1 предлагается обзор полупроводников и их свойств, в особенности тех, которые важны для полупроводниковой оптоэлектроники. В разд. 16.2 дается введение в оптические свойства полупроводников. Дана упрощенная теория поглощения, спонтанного и вынужденного излучения по образцу теории радиационных переходов в атомах, представленной в разд. 13.3.

Эту и две последующие главы нужно рассматривать как единое целое. Глава 17 посвящена работе полупроводниковых источников света, таких как светоизлучающий диод и диодный лазер. Глава 18 посвящена полупроводниковым приемникам.

16.1. ПОЛУПРОВОДНИКИ

Как обсуждалось в подразд. 13.1.3, полупроводник — это кристаллическое или аморфное твердое вещество, электропроводность которого в типичном случае является промежуточной между металлом и изолятором. Его проводимость может существенно меняться за счет температуры или концентрации примесей либо при облучении светом. Зонная структура полупроводников и их способность к образованию переходов и гетероструктур открывают

уникальные возможности. Квантово-размерные полупроводниковые структуры еще более расширяют диапазон доступных свойств. Электронные полупроводниковые приборы главным образом изготавливают из кремния (Si), в то время как для оптоэлектронных полупроводниковых устройств часто используются трех- и четырехкомпонентные сложные полупроводники, такие как InGaAsP или AlInGaN (см. подразд. 16.1.2).

16.1.1. Энергетические зоны и носители заряда

Энергетические зоны в полупроводниках

Атомы, образующие твердые материалы, испытывают настолько сильные межатомные взаимодействия, что их нельзя рассматривать как отдельные сущности (см. подразд. 13.1.3). Их электроны проводимости не связаны с отдельным атомом, а принадлежат всему коллективу атомов как целому. Решение уравнения Шредингера для энергии электрона в периодическом потенциале, создаваемом атомами кристаллической решетки, приводит к расщеплению энергии атомных уровней и формированию энергетических зон. Каждая зона содержит большое число плотно расположенных дискретных энергетических уровней, совокупность которых хорошо аппроксимируется континуумом. Как видно на рис. 16.1, валентная зона отделена от зоны проводимости запрещенной зоной с шириной E_g, которая играет важную роль в определении электрических и оптических свойств материала.

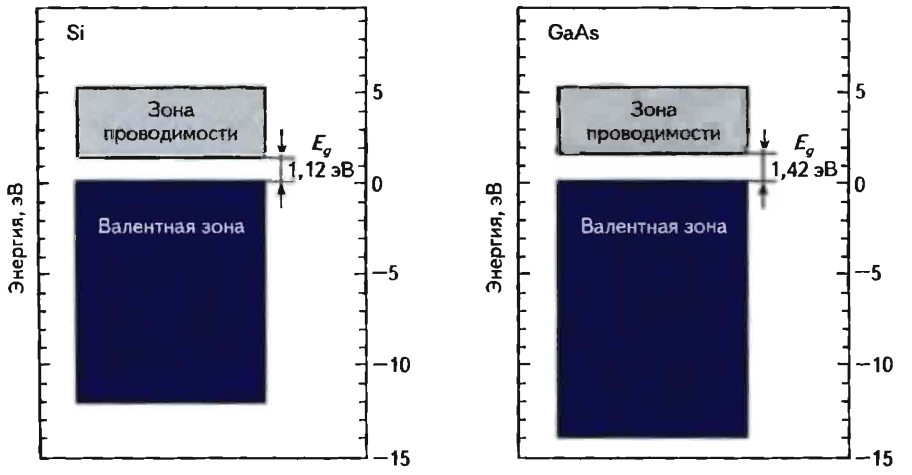


Рис. 16.1. Энергетические зоны в Si и GaAs. Ширина E_g запрещенной зоны, которая разделяет валентную зону и зону проводимости, равна 1,12 эВ для Si и 1,42 эВ для GaAs при комнатной температуре

Происхождение запрещенной зоны можно проиллюстрировать на **модели Кронига—Пенни**. В этой простой теории потенциал кристаллической решетки,

одномерный случай которой показан на рис. 16.2, *а*, аппроксимируется одномерной последовательностью прямоугольных потенциальных барьеров, показанной на рис. 16.2, *б*. Решение соответствующего уравнения Шредингера (13.3) для такого потенциала дает разрешенные энергетические зоны с волновыми функциями в виде бегущих волн и запрещенные зоны, в которых волновая функция экспоненциально затухает. Можно показать, что эти результаты являются общими и применимы к трем измерениям. Такой подход аналогичен тому, который использовался в оптике одномерно-периодических сред в разд. 7.2. Бегущие волны являются **модами Блоха** и имеют периодичность кристаллической решетки [см. (7.69)].

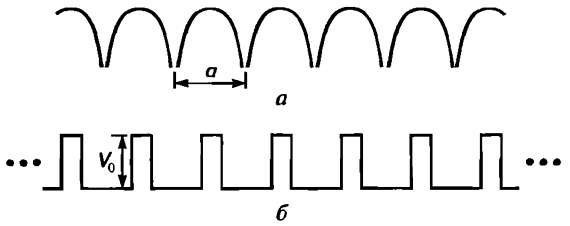


Рис. 16.2. Потенциал кристаллической решетки, связанный с бесконечной одномерной последовательностью атомов с постоянной решетки *a* (*а*). Идеализированный потенциал в виде последовательности прямоугольных барьеров (высотой V_0), используемый в модели Кронига и Пенни (*б*)

Электроны и дырки

Как обсуждалось в подразд. 13.1.3, волновые функции электронов в полупроводнике перекрываются, так что действует **принцип запрета Паули**. Этот принцип требует, чтобы в каждом квантовом состоянии находилось не более одного электрона. Как и в атомах, состояния с самой низкой энергией заполняются первыми. Одноэлементные полупроводники, такие как Si и Ge, имеют по 4 валентных электрона на каждый атом, которые образуют ковалентные связи. При $T = 0$ °K число квантовых состояний, помещающихся в валентной зоне, таково, что все они заполнены, в то время как валентная зона совершенно пуста. При этих условиях материал не проводит электрический ток.

Однако с ростом температуры некоторые электроны в результате теплового возбуждения попадают из валентной зоны в пустую до этого зону проводимости, которая изобилует незанятыми состояниями (рис. 16.3). Эти электроны становятся подвижными носителями заряда и могут дрейфовать по кристаллической решетке под действием внешнего поля и давать вклад в электрический ток. Кроме того, покидая валентную зону, электрон оставляет после себя незаполненное квантовое состояние, которое могут заполнять оставшиеся электроны валентной зоны, меняясь местами друг с другом под влиянием внешнего поля. Таким образом, оставшиеся в валентной зоне электроны приходят в движение. Это движение эквивалентно движению в противоположном направлении дырки, оставленной покинувшим зону электроном. Эта дырка ведет себя как положительный заряд $+e$.

В результате возбуждение каждого электрона создает свободный электрон в зоне проводимости и свободную дырку в валентной зоне. Оба носителя заряда

могут свободно дрейфовать под действием приложенного электрического поля и создавать электрический ток. Материал ведет себя как *полупроводник*, электропроводность которого резко возрастает с увеличением температуры по мере того, как все больше и больше подвижных носителей заряда появляется в результате нагревания.

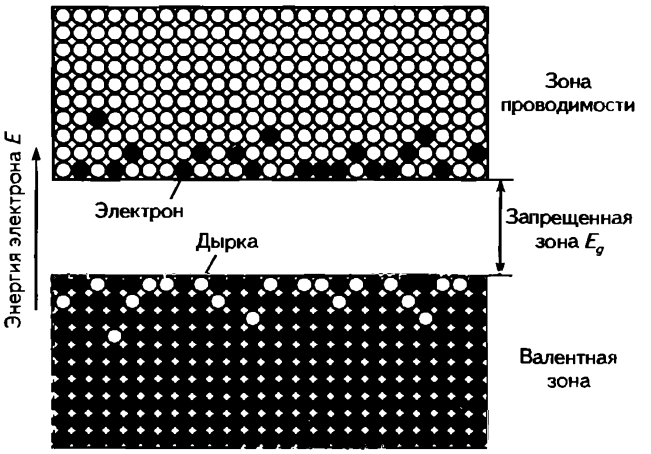


Рис. 16.3. Электроны в зоне проводимости и дырки в валентной зоне при $T > 0$ °K

Связь энергии с импульсом

В соответствии с волновой механикой энергия E и импульс $\boldsymbol{p}$ электрона в области постоянного потенциала, такой как свободное пространство, связаны соотношением

$$E = \frac{p^2}{2m_0} = \frac{\hbar^2 k^2}{2m_0},$$

где p — модуль импульса; k — модуль волнового вектора, $\boldsymbol{k} = \boldsymbol{p}/\hbar$; m_0 — масса электрона (9,1 10^{-31} кг). Таким образом, соотношение между E и k для свободного электрона — простая параболическая зависимость.

Упражнение 16.1

Связь энергии с импульсом для свободного электрона

a. Рассмотрим одномерный вариант временного уравнения Шредингера (13.3) для свободного электрона ($V = 0$) массы m_0. Используйте пробное решение вида $\psi(x) \propto \exp(-jkx)$, чтобы показать, что связь энергии с импульсом принимает квадратичный вид

$$E = \frac{\hbar^2 k^2}{2m_0}, \tag{16.1}$$

так что энергия в этом идеальном случае не квантуется.

б. Свободный фотон, напротив, имеет линейную связь между энергией и импульсом (12.10):

$$E = pc = c\hbar k, \tag{16.2}$$

где c — скорость света в среде. Каково происхождение и значение этого различия?

Движение электрона в полупроводниковом материале тоже описывается уравнением Шредингера, но с потенциалом, созданным зарядами в периодической кристаллической решетке материала. Как обсуждалось ранее, в результате получаются разрешенные зоны энергии, разделенные запрещенными зонами, как предсказывает модель Кронига—Пенни. Возникающая при этом связь между E и k для электронов в зоне проводимости и дырок в валентной зоне графически изображена на рис. 16.4 для Si и GaAs. E — периодическая функция компонент (k_1, k_2, k_3) волнового вектора $\boldsymbol{k}$ с периодами (p/a_1, p/a_2, p/a_3), где (a_1, a_2, a_3) — постоянные решетки кристалла. На рис. 16.4 показаны сечения поверхности, графически выражающей связь E и k, для двух частных направлений волнового вектора $\boldsymbol{k}$. Диапазон изменения k в интервале $[-p/a, p/a]$ определяет первую **зону Бриллюэна**. Таким образом, энергия электрона в зоне проводимости и дырки в валентной зоне зависят не только от величины импульса, но и от его направления относительно решетки кристалла. Зависимость E от k в полупроводнике имеет некоторое сходство с аналогичной зависимостью в фотонном кристалле (см. рис. 7.29).

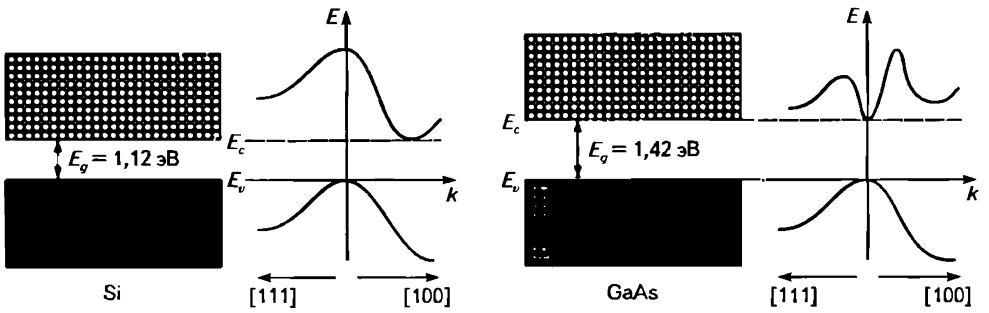


Рис. 16.4. Сечение дисперсионной E—k-диаграммы для Si и GaAs вдоль двух кристаллических осей: [111] слева и [100] справа

Эффективная масса

Из рис. 16.4 можно видеть, что вблизи дна зоны проводимости зависимость E от k можно аппроксимировать параболой

$$E = E_c + \frac{\hbar^2 k^2}{2m_c}, \tag{16.3}$$

где E_c — энергия дна зоны проводимости; k отсчитывается от волнового вектора, при котором энергия имеет минимум. Это соотношение показывает, что электрон зоны проводимости ведет себя так же, как свободный электрон, но с массой m_c, которая называется **эффективной массой** электрона (в зоне проводимости) и отличается от массы свободного электрона m_0. Таким образом, влияние ионов решетки на движение электрона в зоне проводимости содержится в эффективной массе m_c. Это поведение специально выделено на рис. 16.5.

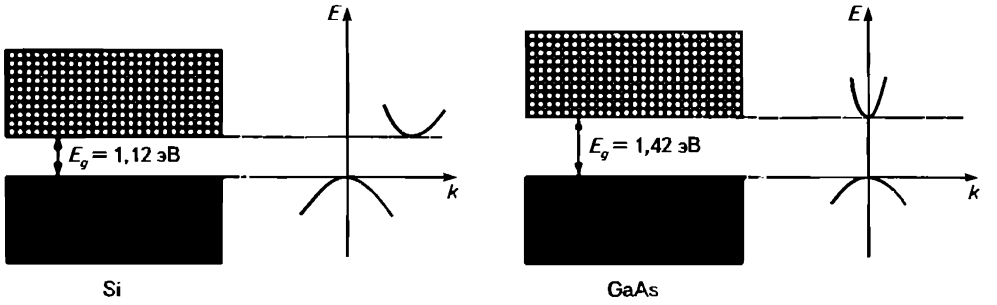


Рис. 16.5. *E—k*-диаграммы для Si и GaAs хорошо аппроксимируются параболами вблизи дна зоны проводимости и потолка валентной зоны

Аналогично, вблизи потолка валентной зоны имеем

$$E = E_v - \frac{\hbar^2 k^2}{2m_v}, \tag{16.4}$$

где $E_v = E_c - E_g$ — энергия потолка валентной зоны; m_v — эффективная масса дырки (в валентной зоне), как показано на рис. 16.5. Влияние ионов решетки на движение дырки в валентной зоне учитывается эффективной массой m_v. Эффективная масса зависит от кристаллической структуры материала и направления распространения по отношению к решетке, поскольку межатомные расстояния различаются в разных кристаллографических направлениях. Она также зависит от конкретной рассматриваемой зоны. Действительно, несколько парабол разной кривизны часто сосуществуют вблизи потолка валентной зоны; они соответствуют так называемым тяжелым дыркам, легким дыркам и дыркам, связанным с отщепленной зоной. Типичные отношения усредненных эффективных масс к массе свободного электрона m_0 приведены в табл. 16.1 для Si, GaAs и GaN.

Таблица 16.1. **Типичные значения эффективных масс электронов и дырок в некоторых полупроводниковых материалах**

Материал	m_c/m_0	m_v/m_0
Si	0,98	0,49
GaAs	0,07	0,50
GaN	0,20	0,80

Прямозонные и непрямозонные полупроводники

Полупроводники, у которых минимум энергии в зоне проводимости и максимум энергии в валентной зоне соответствуют одному и тому же значению волнового числа k (одинаковому импульсу), называются **прямозонными** материалами. Полупроводники, у которых это не так, называются **непрямозонными**. Как видно из рис. 16.5, GaAs является прямозонным полупроводником, в то время как Si — непрямозонный полупроводник. Это различие важно, потому что переход между дном зоны проводимости и потолком валентной зоны в непрямозонном полупроводнике происходит с существенным изменением импульса электрона. Впоследствии будет показано, что прямозонные полупроводники, такие как GaAs, являются эффективными излучателями фотонов, а непрямозонные полупроводники, такие как Si, не могут служить эффективными излучателями света при обычных условиях.

16.1.2. Полупроводниковые материалы

На рис. 16.6 воспроизведена часть периодической таблицы, содержащая бóльшую часть элементов, важных для полупроводниковой электроники и фотоники. Как одноэлементные, так и сложные полупроводники играют решающие роли в этих технологиях.

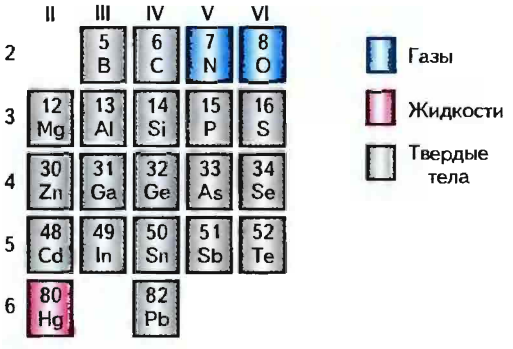


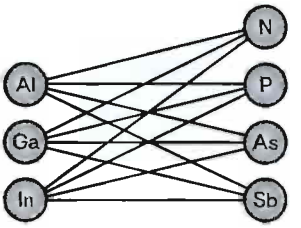
Рис. 16.6. Часть периодической таблицы, относящаяся к полупроводникам. Элементы, показанные голубым, розовым и серебристым цветом, при комнатной температуре находятся в газообразном, жидком и твердом состояниях, соответственно. Полная периодическая таблица показана на рис. 13.3

Одноэлементные полупроводники

Кремний (Si) и германий (Ge) — важные одноэлементные полупроводники группы IV периодической таблицы. По-видимому, все промышленные интегральные электронные схемы и приборы изготовлены с использованием Si. Как Si, так и Ge находят также широкое применение в фотонике, главным образом как фотоприемники. Эти материалы традиционно не использовались для изготовления источников света из-за их непрямозонной структуры. Однако некоторые формы Si способны излучать свет, так что кремниевая фотоника выходит на передовые позиции. Основные свойства Si и Ge приведены в табл. 16.2.

Двухкомпонентные полупроводники III—V групп

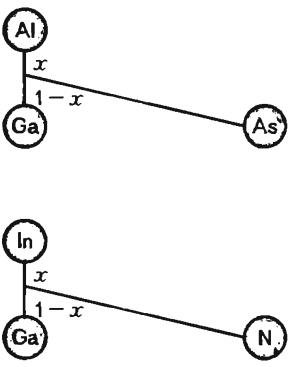
Соединения элемента группы III, таким как алюминий (**Al**), галлий (**Ga**) или индий (**In**), с элементом группы IV, таким как азот (**N**), фосфор (**P**), мышьяк (**As**) или сурьма (**Sb**), — важные для фотоники полупроводники. Эти 12 соединений $A^{III}B^{V}$ приведены в табл. 16.2 со своей кристаллической структурой (типа цинковой обманки или вюрцита), типом запрещенной зоны (прямая или непрямая), шириной E_g и длиной волны $l_g = hc_0/E_g$ запрещенной зоны (длина волны фотона, энергия которого равна ширине запрещенной зоны E_g). Ширины запрещенной зоны и постоянные решетки этих соединений показаны также на рис. 16.7 (см. цв. вклейку). Из многих этих соединений легко изготавливают источники фотонов (светоизлучающие диоды и лазеры), а также приемники света. Первым из двухкомпонентных полупроводников $A^{III}B^{V}$ нашел применение в фотонике арсенид галлия (**GaAs**), который также иногда используется как альтернатива кремнию в быстродействующих электронных приборах и схемах.



Нитрид галлия (**GaN**) играет центральную роль в фотонике благодаря тому, что длина волны запрещенной зоны у него попадает в ближний ультрафиолетовый диапазон. Благодаря своей способности выдерживать высокие температуры, он важен также в электронике. AlN, являющийся изолятором, имеет наибольшую ширину запрещенной зоны из всех соединений $A^{III}B^{V}$ и излучает свет с наименьшей длиной волны, в среднем ультрафиолетовом диапазоне.

Трехкомпонентные полупроводники III—V групп

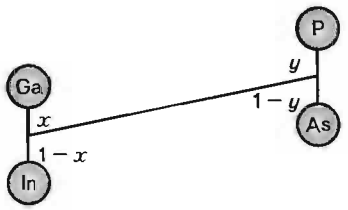
Соединения, образованные двумя элементами группы III и одним элементом группы V (или одним элементом группы III и двумя элементами группы V), являются важными трехкомпонентными полупроводниками. Например, $(Al_xGa_{1-x})As$ — соединение, свойства которого являются промежуточными между AlAs и GaAs в зависимости от соотношения компонентов x (доли атомов Ga в GaAs, замещенных атомами Al). Ширина запрещенной зоны этого материала меняется от 1,42 эВ для GaAs до 2,16 эВ для AlAs по мере того, как x меняется от 0 до 1 вдоль линии, соединяющей GaAs и AlAs на рис. 16.7, *а*. Поскольку эта линия существенно вертикальна, по строению решетки $(Al_xGa_{1-x})As$ соответствует GaAs; следовательно, слой с любым составом данного материала можно вырастить на слое с другим составом без напряжения решетки. Другие полезные трехкомпонентные соединения III—V групп, такие как $Ga(As_{1-x}P_x)$, также представлены на диаграмме «ширина запрещенной зоны — постоянная решетки», рис. 16.7, *а*. $(In_xGa_{1-x})As$ широко применяется в источниках и приемниках фотонов ближней инфракрасной области спектра. Анало-



гично, $(Al_xGa_{1-x})N$ и $(In_xGa_{1-x})N$ — важные трехкомпонентные полупроводники для устройств фотоники, работающих в ультрафиолетовой области, фиолетовой, синей и зеленой частях видимой области, как можно заключить из рис. 16.7, *б*. В области электроники биполярные транзисторы на гетеропереходе $(In_xGa_{1-x})As/InP$ имеют скорости переключения, приближающиеся к 1 ТГц. Различные соединения III—V групп можно использовать для изготовления сверхбыстрых транзисторов, излучающих свет.

Четырехкомпонентные полупроводники III—V групп

Эти соединения образованы из двух элементов III и двух элементов V групп (или трех элементов III группы и одного элемента V группы). Четырехкомпонентные полупроводники обладают большей гибкостью для изготовления материалов с желаемыми свойствами, чем трехкомпонентные полупроводники, благодаря наличию дополнительной степени свободы. Примером может служить $In_{1-x}Ga_xAs_{1-y}P_y$, у которого ширина запрещенной зоны меняется между 0,36 эВ (InAs) и 2, 26 эВ (GaP) по мере изменения соотношения компонентов x и y между 0 и 1. Постоянная решетки обычно линейно меняется с изменением соотношения компонентов (закон Вегарда). Светло-серая область на рис. 16.7, *а* указывает на диапазон ширин запрещенной зоны и постоянных решетки, охватываемый этим соединением. При соотношениях компонентов, удовлетворяющих условию $y = 2{,}16(1 - x)$, решетка $In_{1-x}Ga_xAs_{1-y}P_y$ согласуется с решеткой InP, который может служить удобным шаблоном (подложкой). Этот четырехкомпонентный полупроводник используется для изготовления светоизлучающих диодов, диодных лазеров и фотоприемников, особенно в окрестности длины волны 1550 нм, стандартно используемой в волоконно-оптической связи (см. гл. 17, 18 и 24). Другой пример — это $(Al_xIn_yGa_{1-x-y})P$, для которого GaAs служит эталонной структурой. Это соединение излучает яркий свет в красной, оранжевой и желтой спектральных областях [см. область, отмеченную серым цветом на рис. 16.7, *а*]. Еще один полезный четырехкомпонентный материал — это III-нитрид $(Al_xIn_yGa_{1-x-y})N$, который выполняет те же функции в зеленой, синей, фиолетовой и ультрафиолетовой областях спектра (см. рис. 16.7, *б*). Удобными шаблонами для III-нитридов являются сапфир и SiC.



Элементы IV группы также можно смешивать для получения сложных полупроводников. Двухкомпонентный сплав карбид кремния (SiC), известный также как карборунд, является непрямозонным и полезен для изготовления ультрафиолетовых фотоприемников, а также в качестве подложки для выращивания III-нитридов. Кремниевый германий ($Si_{1-x}Ge_x$) имеет разнообразные применения в электронике и фотонике, включая использование в качестве материала для фотоприемников. Трех- и четырехкомпонентные полупроводниковые соединения IV группы включают $Si_{1-x-y}Ge_xC_y$ и $Si_{1-y-z}Ge_xC_yZn_z$ соответственно.

Двухкомпонентные материалы II—VI групп, т. е. соединения элементов II группы (например, Zn, Cd, Hg) и VI группы (например, S, Se, Te) периодической таблицы, также используются как полупроводниковые материалы. Это семейство включает ZnS, ZnSe, ZnTe, CdS, CdSe, CdTe, HgS, HgSe и HgTe и показано на рис. 16.8 (см. цв. вклейку). Все эти материалы имеют решетку типа цинковой обманки и являются прямозонными полупроводниками; исключение составляют HgSe и HgTe, которые являются полуметаллами с малыми отрицательными значениями ширины запрещенной зоны. Особым достоинством ZnSe является возможность его нанесения на подложку из GaAs при сравнительно низкой плотности дефектов, поскольку постоянные решетки у этих материалов одинаковы. Более того, HgTe и CdTe также почти согласованы по структуре решетки, поэтому трехкомпонентный полупроводник $Hg_xCd_{1-x}Te$ можно вырастить без напряжений на подложке из CdTe. Эта система материалов часто используется для изготовления приемников фотонов, как и другие соединения II—VI групп (см. гл. 18). В отличие от сплавов III—V групп, соединения II—VI групп широко распространены в природе, однако изготовленные из этих материалов фотоприемники обычно имеют ограниченный срок службы. Тем не менее, двухкомпонентным полупроводниковым материалам II—VI групп легко придать форму квантовых точек с перестраиваемой длиной волны фотолюминесценции (см., например, рис. 13.12). Трехкомпонентные полупроводниковые соединения IV—VI групп, такие как $Pb_xSn_{1-x}Te$ и $Pb_xSn_{1-x}Se$, также использовались в фотоприемниках и лазерных диодах. Однако эти сплавы имеют более медленный отклик из-за больших диэлектрических проницаемостей. Они имеют также большие коэффициенты теплового расширения, поэтому переход от комнатных к криогенным температурам с ними может быть проблематичен.

Таблица 16.2. **Некоторые одноэлементные и двухкомпонентные полупроводники III—V групп. Указаны тип решетки, тип запрещенной зоны, ширина запрещенной зоны и соответствующая длина волны**

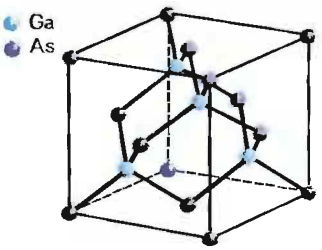
Материал	Тип решетки* (А/ЦО/В)	Тип запрещенной зоны** (П/НП)	Ширина запрещенной зоны*** E_g, эВ	Длина волны запрещенной зоны**** λ_g, мкм
Si	А	НП	1,12	1,11
Ge	А	НП	0,66	1,88
AlN	В	П	6,20	0,200
AlP	ЦО	НП	2,45	0,506
AlAs	ЦО	НП	2,16	0,574
AlSb	ЦО	НП	1,58	0,785

Окончание табл. 16.2

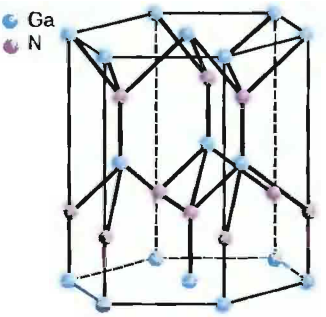
Материал	Тип решетки* (А/ЦО/В)	Тип запрещенной зоны** (П/НП)	Ширина запрещенной зоны*** E_g, эВ	Длина волны запрещенной зоны**** λ_g, мкм
GaN	В	П	3,39	0,366
GaP	ЦО	НП	2,26	0,549
GaAs	ЦО	П	1,42	0,873
GaSb	ЦО	П	0,73	1,70
InN	В	П	0,65	1,91
InP	ЦО	П	1,35	0,919
InAs	ЦО	П	0,36	3,44
InSb	ЦО	П	0,17	7,29

* Кристаллическая структура указана для наиболее употребительной формы материала: А — алмаз, ЦО — цинковая обманка, В — вюрцит (см. рис. ниже). Структура цинковой обманки включает две взаимопроникающие гранецентрированные кубические решетки, одна для каждого элемента, смещенные друг относительно друга на 1/4 диагонали куба. Решетка типа алмаза не отличается по форме, но все атомы в ней одинаковы. Зоны Бриллюэна для этих структур показаны на рис. 7.28. Решетка типа вюрцита состоит из двух гексагональных плотно упакованных решеток, одна для каждого элемента, смещенных друг относительно друга вдоль оси c третьего порядка на 3/8 ее длины. Все атомы тетраэдрически связаны со своими соседями.

Цинковая обманка и алмаз



Вюрцит



** П — прямозонный, НП — непрямозонный.

*** Данные приведены для $T = 300$ °K.

**** Длина волны связана с шириной запрещенной зоны E_g формулой $\lambda_g = hc_0/E_g$; если ширина запрещенной зоны выражена в электронвольтах, а длина волны в микрометрах, то $\lambda_g \approx 1{,}24/E_g$.

Примесные полупроводники

Электрические и оптические свойства полупроводников могут существенно меняться при контролируемом введении в материал небольшого количества специально подобранных примесей или **легирующих добавок**. Введение примесей может изменить концентрацию подвижных носителей заряда на много порядков величины. Легирующие добавки с избытком валентных электронов, называемые **донорами**, замещая небольшое число нормальных атомов кристаллической решетки, создают избыток подвижных электронов. В этом случае материал называется **полупроводником *n*-типа**. Так, атомы V группы (например, P или As), замещая атомы IV группы в одноэлементном полупроводнике (например, Si или Ge), или атомы VI группы (например, Se или Te), замещая атомы V группы в двухкомпонентном полупроводнике $A^{III}B^{V}$ (например, As или Sb), дают материал *n*-типа. Аналогично, **полупроводник *p*-типа** получается при добавлении примесей с недостатком валентных электронов, называемых **акцепторами**. В результате получается избыток свободных дырок. Замещение атомов одноэлементного полупроводника IV группы атомами III группы (например, B или In) или атомов III группы в двухкомпонентном III—V полупроводнике атомами II группы (например, Zn или Cd) дает материал *p*-типа. Атомы IV группы являются донорами для группы III, но акцепторами для группы V, поэтому их можно использовать для создания избытка как электронов, так и дырок в двухкомпонентных III—V материалах. Естественно, что при введении легирующих добавок материал остается электронейтральным.

Полупроводники без умышленно введенных примесей называются **собственными** материалами, а с введением легирующих добавок — **несобственными**. Концентрации подвижных электронов и дырок в собственных полупроводниках одинаковы, $n = p = n_i$, где собственная концентрация носителей n_i экспоненциально растет с увеличением температуры. С другой стороны, концентрация подвижных электронов в полупроводнике *n*-типа (**основных носителей**) намного больше, чем концентрация дырок (**неосновных носителей**), т. е. $n \gg p$. Противоположное верно для полупроводника *p*-типа, где дырки являются основными носителями и $p \gg n$. При комнатной температуре концентрация основных носителей в примесном полупроводнике, как правило, приблизительно равна концентрации примеси.

С помощью техники имплантации отдельных ионов можно изготовить полупроводниковый материал, в котором число атомов примеси и их положение точно контролируются. Получающиеся материалы демонстрируют свойства, намного более детерминированные, чем при случайном числе примесных атомов, что важно для некоторых приложений.

Пример 16.1

Энергия ионизации электрона донора

Рассмотрим кристалл германия с диэлектрической постоянной $\varepsilon/\varepsilon_0 = 16$ (см. табл. 16.1), легированный донорными атомами мышьяка. Эффективная масса электрона $m_c = 0{,}2m_0$, где m_0 — масса свободного электрона. Электрон

донора движется в поле однократно заряженного иона мышьяка (As^+) и имеет уровни энергии, аналогичные уровням энергии электрона в атоме водорода. Полагая $n = 1$ и $Z = 1$ в (13.4) и заменяя ε_0 на ε, а M_r на m_c, чтобы учесть плотность поляризации и кристаллическую решетку полупроводникового материала, соответственно, найдем выражение для энергии электрона донора

$$E_D = -\left(\frac{1}{4\pi\varepsilon}\right)^2 \frac{m_c e^4}{2\hbar^2}. \tag{16.5}$$

Поскольку энергия электрона в основном состоянии атома водорода равна −13,6 эВ (т. е. на 13,6 эВ ниже предела ионизации), энергия электрона донора для мышьяка:

$$E_D = -\frac{m_c}{m_0}\left(\frac{\varepsilon_0}{\varepsilon}\right)^2 \cdot 13{,}6 \text{ эВ} \approx -0{,}01 \text{ эВ}.$$

Таким образом, электрон донора находится в запрещенной зоне на уровне на 0,01 эВ ниже дна зоны проводимости. Однако, поскольку тепловая энергия при $T = 300\,^\circ\text{K}$ равна $k_B T \approx 0{,}026$ эВ, при комнатной температуре практически все доноры ионизируются и их электроны оказываются в зоне проводимости. Материал, таким образом, обладает электронами проводимости, концентрация которых равна концентрации примесных атомов.

Органические полупроводники

Органические полупроводники находят все большее применение в множестве областей. Эти области включают фотонику, где они используются для изготовления фотогальванических устройств, светоизлучающих диодов и дисплеев. Хотя они и не обеспечивают ни быстродействия, ни миниатюрности, присущих обычным полупроводниковым структурам, они допускают дешевое производство в виде тонких листов, благодаря чему механически гибкие и дешевые оптоэлектронные компоненты становятся реальностью. Эти материалы получаются в практически неограниченном множестве разновидностей, которые можно проектировать, приспосабливаясь к специфическим требованиям, а некоторые из них можно наносить на подходящую подложку с помощью технологии струйной печати.

Органически полупроводники бывают двух основных типов, схематически показанных на рис. 16.9.

1. Малые органические молекулы, такие как пентацен, который состоит из пяти бензольных колец, расположенных в виде цепочки (рис. 16.9, *а*).

2. Полимерные цепи сопряженных углеводородов, такие как полиацетилен, включающие сотни или тысячи атомов углерода (рис. 16.9, *б*).

Отличительным признаком этих аморфных материалов, который называется сопряжением, является наличие чередующихся одинарных и двойных связей между атомами углерода. Хотя на схемах, подобных рис. 16.9, двойные

связи изображаются как локализованные, в реальности соответствующие им электроны делокализованы и принадлежат одновременно множеству атомов или участку полимерной цепи примерно из десяти повторяющихся звеньев. Такая молекула или участок полимера ведет себя как одна система, в которой энергии возможных состояний электрона образуют зоны.

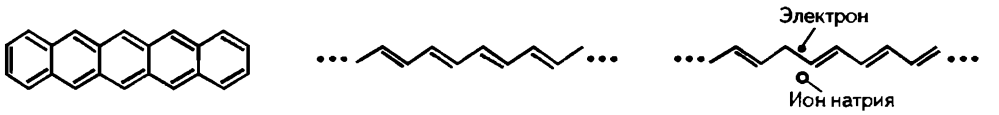


Рис. 16.9. Органические полупроводники бывают двух основных типов:

а — малые органические молекулы, такие как пентацен; *б* — полимерные цепи с сопряженными связями, такие как полиацетилен; *в* — легирование полиацетилена донорами натрия дает материал *n*-типа, а акцепторами йода — *p*-типа. Каждая линия изображает химическую связь между двумя атомами углерода; двойные линии изображают двойные связи. Связи с атомами водорода для простоты не показаны. Множество органических молекул и полимеров используется в электронике и фотонике

В отсутствие примесей валентная зона сопряженной полимерной цепи, как правило, заполнена, а зона проводимости пуста, поэтому материал ведет себя как изолятор. Однако, как изображено на рис. 16.9, *в*, такие примеси, как натрий и йод, являются донорами и акцепторами, соответственно, обеспечивая *n*- или *p*-тип проводимости. Малые органические молекулы нередко обладают собственной проводимостью.

16.1.3. Концентрации электронов и дырок

Определение концентрации носителей заряда (электронов и дырок) в зависимости от энергии требует знания двух характеристик, которые мы рассмотрим по очереди:

- плотность уровней энергии (плотность состояний);
- вероятность заселения каждого из этих уровней.

Плотность состояний

Квантовое состояние электрона в полупроводнике характеризуется его энергией E, вектором k [величина которого приближенно связана с E соотношением (16.3) или (16.4)] и спином. Состояние описывается волновой функцией, удовлетворяющей некоторым граничным условиям.

Вблизи края зоны проводимости электрон приближенно описывается как частица массы m_c, находящаяся в трехмерном кубическом ящике (размером d) с идеально отражающими стенками, т. е. в трехмерной бесконечно глубокой прямоугольной потенциальной яме. Решения в виде стоячих волн требуют, чтобы компоненты вектора $k = (k_x, k_y, k_z)$ имели дискретные значения $k = (q_1\pi/d, q_2\pi/d, q_3\pi/d)$, где соответствующие модовые числа (q_1, q_2, q_3) принимают целые положительные значения. Этот результат является обобщением одномерной

бесконечно глубокой прямоугольной потенциальной ямы, рассмотренной в упражнении 16.5. Конец вектора $\boldsymbol{k}$ должен находиться в точках решетки, кубическая ячейка которой имеет размер π/d. Таким образом, на единицу объема приходится $(d/\pi)^3$ точек в $\boldsymbol{k}$-пространстве. Число состояний, векторы $\boldsymbol{k}$ которых имеют модуль между 0 и k, определяется подсчетом числа точек, лежащих в положительном октанте сферы радиусом k [с объемом $\approx(1/8)4\pi k^3/3 = \pi k^3/6$]. Ввиду наличия двух возможных спиновых состояний каждая точка в $\boldsymbol{k}$-пространстве соответствует двум состояниям. Следовательно, имеется приблизительно

$$2\frac{\pi k^3/6}{(\pi/d)^3} = \frac{k^3}{3\pi^2}d^3$$

таких точек в объеме d^3 и $(k^3/3\pi^2)$ точек в единичном объеме. Отсюда следует, что число состояний с волновыми числами электрона между k и $k + \Delta k$, приходящееся на единичный объем,

$$\rho(k)\Delta k = \left(\frac{\mathrm{d}}{\mathrm{d}k}\frac{k^3}{3\pi^2}\right)\Delta k = \frac{k^2}{\pi^2}\Delta k,$$

так что плотность состояний выражается как

$$\boxed{\rho(k) = \frac{k^2}{\pi^2}.} \qquad (16.6)$$

Плотность состояний

Этот вывод идентичен использованному при подсчете числа мод в трехмерном электромагнитном резонаторе (см. разд. 10.3). В случае электромагнитных мод две степени свободы связаны с поляризацией поля (т. е. с двумя значениями проекции спина фотона), а в полупроводниках — с двумя значениями проекции спина электрона. В оптике резонаторов допустимые значения электромагнитного волнового вектора $\boldsymbol{k}$ преобразовывались в частоты с помощью линейного соотношения $\nu = ck/2\pi$. В физике полупроводников аналогичное преобразование возможных значений $\boldsymbol{k}$ в уровни энергии осуществляется квадратичными соотношениями (16.3) и (16.4).

Если $\rho_c(E)\Delta E$ — число энергетических уровней в зоне проводимости (на единицу объема), лежащих между E и $E + \Delta E$, то, в силу взаимно однозначной связи (16.3) между E и k, плотности $\rho_c(E)$ и $\rho(k)$ должны быть связаны соотношением $\rho_c(E)\mathrm{d}E = \rho(k)\mathrm{d}k$. Таким образом, плотность уровней энергии в зоне проводимости равна

$$\rho_c(E) = \frac{\rho(k)}{\mathrm{d}E/\mathrm{d}k}.$$

Аналогично, плотность состояний в валентной зоне

$$\rho_v(E) = \frac{\rho(k)}{\mathrm{d}E/\mathrm{d}k},$$

где E дается формулой (16.4). Приближенные квадратичные соотношения (16.3) и (16.4), применимые вблизи краев зоны проводимости и валентной зоны, соответственно, используются для вычисления производной $\mathrm{d}E/\mathrm{d}k$ в каждой зоне. Результат получается следующим:

$$\rho_c(E) = \frac{(2m_c)^{3/2}}{2\pi^2\hbar^3}\sqrt{E - E_c}, \quad E \geq E_c; \tag{16.7}$$

$$\rho_v(E) = \frac{(2m_c)^{3/2}}{2\pi^2\hbar^3}\sqrt{E_v - E}, \quad E \leq E_v. \tag{16.8}$$

Плотность состояний вблизи границ зон

Квадратные корни в этих формулах — следствие квадратичного закона связи энергии с волновым числом для электронов и дырок вблизи границ энергетических зон. Зависимость плотности состояний от энергии показана на рис. 16.10. Она равна нулю на границе зоны и возрастает с удалением от нее со скоростью, зависящей от эффективной массы электронов и дырок. Значения m_c и m_v, приведенные в табл. 16.1, в действительности представляют собой усредненные значения, применяемые при расчете плотности состояний.

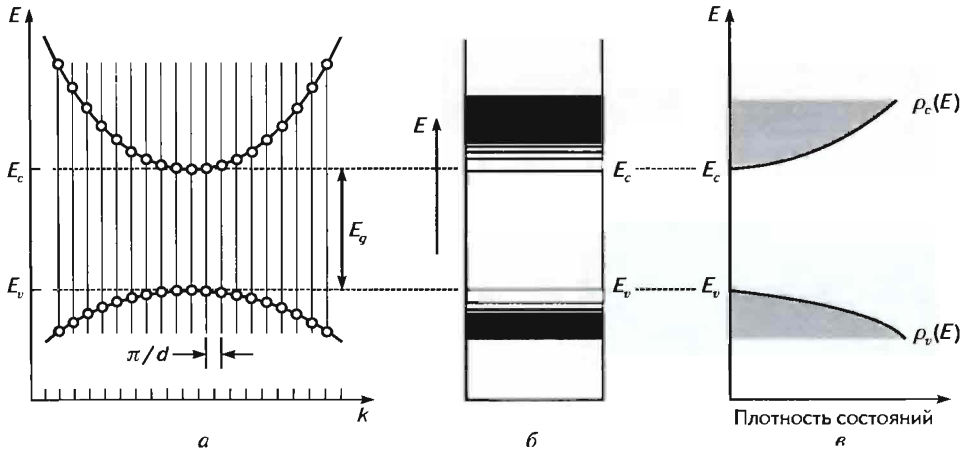


Рис. 16.10. Зависимость E от k_1 (при фиксированных k_2 и k_3) (*а*). Допустимые значения энергии (для всех k) (*б*). Плотность состояний вблизи краев зоны проводимости и валентной зоны (*в*). Величина $\rho_c(E)\mathrm{d}E$ — число квантовых состояний с энергией между E и $E + \mathrm{d}E$ на единицу объема в зоне проводимости. Величина $\rho_v(E)\mathrm{d}E$ — то же самое для валентной зоны

Вероятность заселения состояний

В отсутствие теплового возбуждения (при $T = 0$ °K) все электроны занимают наинизшие возможные уровни энергии в соответствии с принципом Паули. Валентная зона при этом полностью заполнена (нет дырок), а зона

проводимости полностью свободна (не содержит электронов). При повышении температуры тепловое возбуждение переводит некоторые электроны из валентной зоны в зону проводимости, оставляя в валентной зоне свободные состояния (дырки). Из законов статистической механики вытекает, что в состоянии теплового равновесия при температуре T вероятность того, что данное состояние с энергией E занято электроном, определяется функцией Ферми

$$f(E) = \frac{1}{\exp\left[\left(E - E_f\right)/k_B T\right] + 1}, \qquad (16.9)$$

Функция Ферми

где k_B — постоянная Больцмана (при $T = 300\,^\circ\text{K}$ $k_B T = 0{,}026$ эВ); E_f — постоянная, называемая **энергией** или **уровнем, Ферми**. Функция (16.9) называется также **распределением Ферми—Дирака**. Каждый энергетический уровень E либо занят [с вероятностью $f(E)$], либо пуст [с вероятностью $1 - f(E)$]. Вероятности $f(E)$ и $1 - f(E)$ зависят от энергии E, согласно формуле (16.9). Функция $f(E)$ не является плотностью вероятности, и ее интеграл по всем E не равен единице. Она дает последовательность вероятностей заселения следующих друг за другом уровней энергии.

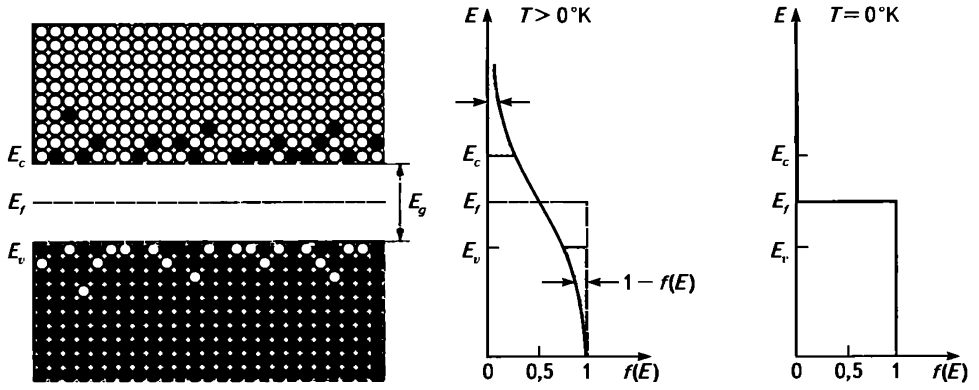


Рис. 16.11. Функция Ферми $f(E)$ — вероятность того, что энергетический уровень заполнен электроном; $1 - f(E)$ — вероятность того, что он пуст. В валентной зоне $1 - f(E)$ является вероятностью того, что уровень занят дыркой. При $T = 0\,^\circ\text{K}$ $f(E) = 0$, если $E > E_f$, и $f(E) = 1$, если $E \le E_f$; при этом в зоне проводимости вообще нет электронов, а в валентной зоне — дырок

Поскольку $f(E_f) = 1/2$, какой бы ни была температура T, уровень Ферми представляет собой ту энергию, при которой вероятность заселения энергетического уровня (если такой уровень существует) равна 1/2. Функция Ферми монотонно убывает с ростом энергии E (рис. 16.11). При $T = 0\,^\circ\text{K}$ $f(E) = 0$, если $E > E_f$, и $f(E) = 1$, если $E \le E_f$. Этим определяется физический смысл E_f: это энергия, отделяющая заселенные уровни от незаселенных при $T = 0\,^\circ\text{K}$. Поскольку $f(E)$ — вероятность того, что уровень заселен, есть вероятность

того, что он пуст, то $1 - f(E)$ есть вероятность того, что уровень занят дыркой, если он лежит в валентной зоне. Таким образом, для каждого уровня:

- $f(E)$ — вероятность того, что уровень занят электроном;
- $1 - f(E)$ — вероятность того, что уровень занят дыркой (в валентной зоне).

Эти функции симметричны относительно уровня Ферми.

Когда $E - E_f \gg k_B T$,

$$f(E) \approx \exp\left[-\frac{E - E_f}{k_B T}\right],$$

так что «хвост» функции Ферми в области высоких энергий в зоне проводимости экспоненциально затухает с ростом энергии. Тогда функция Ферми пропорциональна распределению Больцмана, которое описывает экспоненциальную энергетическую зависимость доли общего числа атомов, приходящейся на данный энергетический уровень (см. разд. 13.2). Из соображений симметрии вытекает, что когда $E < E_f$ и $E_f - E \gg k_B T$,

$$1 - f(E) \approx \exp\left[-\frac{E_f - E}{k_B T}\right];$$

вероятность заселения уровня дыркой в валентной зоне экспоненциально затухает по мере убывания энергии в области намного ниже уровня Ферми.

Концентрации носителей при тепловом равновесии

Пусть $n(E)\Delta E$ и $p(E)\Delta E$ — число электронов и дырок в единице объема, соответственно, с энергией между E и $E + \Delta E$. Плотности $n(E)$ и $p(E)$ можно получить, умножая плотности состояний на энергетическом уровне E на вероятность заселения уровня электроном или дыркой, так что

$$n(E) = \rho_c(E) f(E); \quad p(E) = \rho_v(E)[1 - f(E)]. \tag{16.10}$$

Концентрации (заселенности на единицу объема) электронов и дырок, n и p, получаются интегрированием

$$n = \int_{E_c}^{\infty} n(E)\mathrm{d}E; \quad p = \int_{-\infty}^{E_v} p(E)\mathrm{d}E. \tag{16.11}$$

В собственном (беспримесном) полупроводнике при любой температуре $n = p$, поскольку тепловое возбуждение всегда создает электроны проводимости и дырки парами. Следовательно, уровень Ферми должен совпадать с таким значением энергии, при котором $n = p$. У материалов, для которых $m_v = m_c$, функции $n(E)$ и $p(E)$ также симметричны, так что E_f должен лежать точно посередине запрещенной зоны (рис. 16.12). У большинства собственных полупроводников уровень Ферми действительно находится вблизи середины запрещенной зоны.

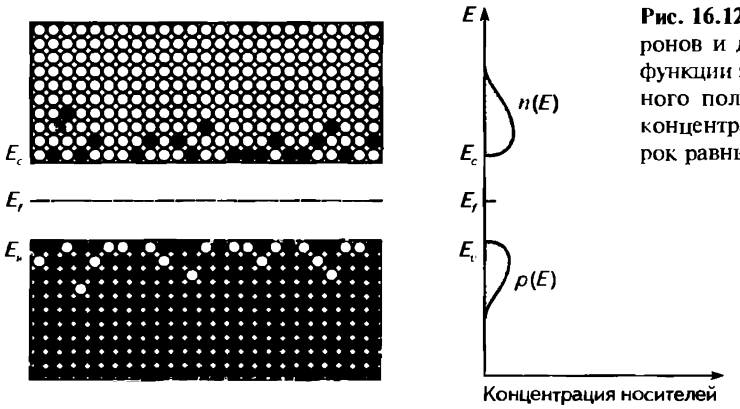


Рис. 16.12. Концентрации электронов и дырок $n(E)$ и $p(E)$ как функции энергии E для собственного полупроводника. Полные концентрации электронов и дырок равны n и p соответственно

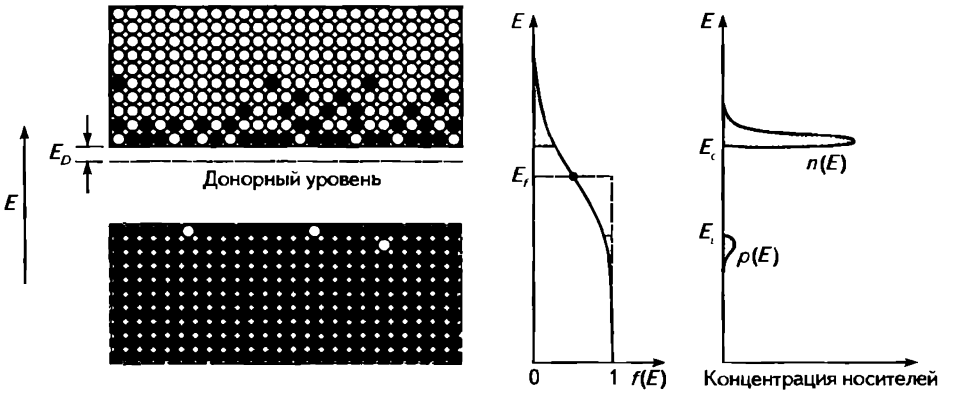


Рис. 16.13. Диаграмма энергетических зон, функция Ферми $f(E)$ и концентрации подвижных электронов и дырок $n(E)$ и $p(E)$, соответственно, для полупроводника n-типа

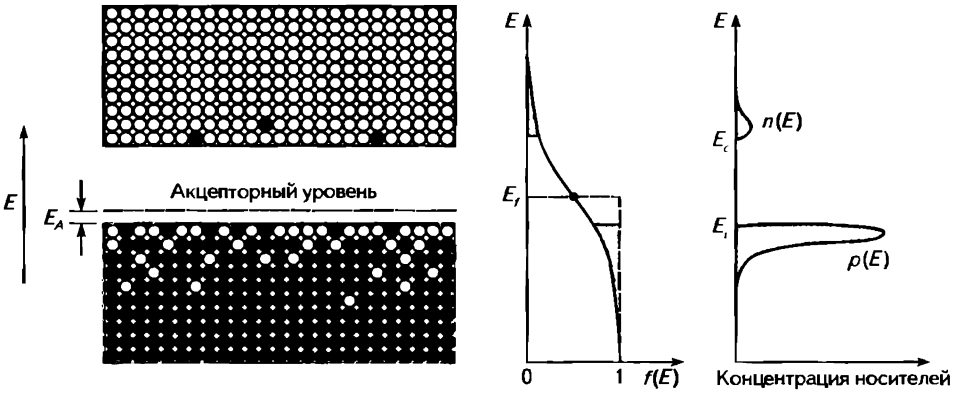


Рис. 16.14. Диаграмма энергетических зон, функция Ферми $f(E)$ и концентрации подвижных электронов и дырок $n(E)$ и $p(E)$, соответственно, для полупроводника p-типа

Диаграммы энергетических зон, функции Ферми и равновесные концентрации электронов и дырок для примесных полупроводников n- и p-типа представлены на рис. 16.13 и 16.14 соответственно. Электроны донорной примеси занимают уровни с энергией ниже дна зоны проводимости на небольшую величину E_D, поэтому они могут легко переходить в зону проводимости. Например, если $E_D = 0{,}01$ эВ, то при комнатной температуре ($k_B T = 0{,}026$ эВ) большинство таких электронов окажется в зоне проводимости из-за теплового возбуждения (см. пример 16.1). В результате уровень Ферми [энергия, при которой $f(E_f) = 1/2$] будет расположен выше середины запрещенной зоны. Для полупроводника p-типа уровни акцептора лежат выше потолка валентной зоны на небольшую величину E_A, так что уровень Ферми расположен ниже середины запрещенной зоны. Предметом нашего внимания являются подвижные носители заряда в примесных полупроводниках. Конечно, эти материалы электрически нейтральны, что обеспечивается неподвижными ионами донорной или акцепторной примеси, так что $n + N_A = p + N_D$, где N_A и N_D — число ионизированных атомов акцептора и донора в единице объема.

Упражнение 16.2

Экспоненциальное приближение функции Ферми

Когда $E - E_f \gg k_B T$, функцию Ферми $f(E)$ можно аппроксимировать экспоненциальной функцией. Аналогично, когда $E_f - E \gg k_B T$, экспоненциальную аппроксимацию допускает $1 - f(E)$. Условия такой аппроксимации выполняются, когда уровень Ферми лежит в запрещенной зоне, при энергиях, отстоящих от краев запрещенной зоны, по меньшей мере, на несколько $k_B T$ (при комнатной температуре $k_B T = 0{,}026$ эВ, тогда как $E_g = 1{,}12$ эВ для Si и 1,42 эВ для GaAs). С помощью этих приближений, применимых как к собственным, так и к примесным полупроводникам, покажите, что (16.11) дает

$$n = N_c \exp\left(-\frac{E_c - E_f}{k_B T}\right); \tag{16.12}$$

$$p = N_v \exp\left(-\frac{E_f - E_v}{k_B T}\right); \tag{16.13}$$

$$np = N_c N_v \exp\left(-\frac{E_g}{k_B T}\right), \tag{16.14}$$

где

$$N_c = 2\left(\frac{2\pi m_c k_B T}{h^2}\right)^{3/2} \qquad N_v = 2\left(\frac{2\pi m_v k_B T}{h^2}\right)^{3/2}$$

Проверьте, что если E_f находится ближе к зоне проводимости и $m_v = m_c$, то $n > p$, а если ближе к валентной зоне, то $n < p$.

Закон действующих масс

Из уравнения (16.14) следует, что при тепловом равновесии произведение

$$np = 4\left(\frac{2\pi k_B T}{h^2}\right)^3 (m_c m_v)^{3/2} \exp\left(-\frac{E_g}{k_B T}\right) \qquad (16.15)$$

не зависит от положения уровня Ферми E_f внутри запрещенной зоны и содержания примесей в полупроводнике при условии применимости экспоненциального приближения функции Ферми. Постоянство произведения концентраций называется **законом действующих масс**. Для собственного полупроводника $n = p \equiv n_i$. В сочетании с (16.14) последнее соотношение дает

$$\boxed{n_i \approx \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_B T}\right),} \qquad (16.16)$$

Концентрация собственных носителей

т. е. собственная концентрация электронов и дырок экспоненциально растет с температурой T. Тогда закон действующих масс можно записать в виде

$$\boxed{np = n_i^2.} \qquad (16.17)$$

Закон действующих масс

Значения n_i для различных материалов меняются из-за различия в ширине запрещенной зоны и эффективных массах. Концентрации собственных носителей при комнатной температуре для Si, GaAs и GaN приведены в табл. 16.3.

Таблица 16.3. **Концентрации собственных носителей при $T = 300$ °K***

Материал	n_i, см$^{-3}$
Si	$1{,}5 \cdot 10^{10}$
GaAs	$1{,}8 \cdot 10^{6}$
GaN	$1{,}9 \cdot 10^{-10}$

* Подстановка значений m_c и m_v из табл. 16.1 и значения E_g из табл. 16.2 не дает приведенных значений n_i из-за чувствительности формулы к точности значений параметров.

Закон действующих масс полезен для определения концентраций электронов и дырок в примесных полупроводниках. Умеренно легированный материал n-типа, например, имеет концентрацию электронов n, равную концентрации доноров N_D. Используя закон действующих масс, для концентрации дырок получаем $p = n_i^2/N_D$. Знание n и p позволяет определить уровень Ферми с помощью (16.11). Если уровень Ферми находится в запрещенной зоне на расстоянии больше чем несколько $k_B T$ от ее краев, то приближенные соотношения (16.12) и (16.13) можно использовать для его непосредственного определения.

Если уровень Ферми лежит внутри валентной зоны или зоны проводимости, то материал называется вырожденным полупроводником. В этом случае экспоненциальная аппроксимация функции Ферми непригодна, поэтому $np \neq n_i^2$. Тогда концентрации носителей нужно получать с помощью численного решения. При условии очень сильного легирования полоса донорной (акцепторной) примеси действительно накладывается на зону проводимости (валентную зону), образуя то, что называется **хвостом зоны**. Результатом является эффективное уменьшение ширины запрещенной зоны.

Квазиравновесные концентрации носителей

Вероятности заселения уровней и концентрации носителей заряда, рассмотренные выше, применимы только к полупроводникам, находящимся в состоянии теплового равновесия. Они не годятся в случае нарушения этого состояния. Однако существуют ситуации, при которых электроны в зоне проводимости находятся в тепловом равновесии между собой, как и дырки в валентной зоне, но между электронами и дырками отсутствует общее тепловое равновесие. Это может происходить, например, когда внешний электрический ток или поток фотонов индуцирует межзонные переходы со слишком большой скоростью для того, чтобы достичь межзонного равновесия. Такая ситуация, называемая **квазиравновесием**, возникает, когда времена релаксации (распада) для переходов внутри каждой из зон намного меньше, чем время межзонной релаксации. В типичном случае время внутризонной релаксации $<10^{-12}$ с, тогда как время радиационной электронно-дырочной рекомбинации $\approx 10^{-9}$ с.

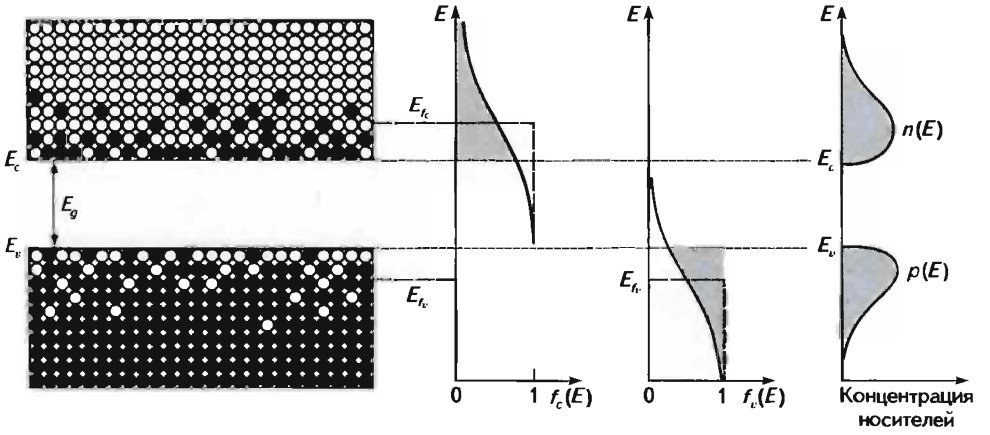


Рис. 16.15. Полупроводник в состоянии квазиравновесия. Вероятность того, что некоторый уровень E в зоне проводимости занят электроном, равна $f_c(E)$ — значению функции Ферми с уровнем Ферми E_{fc}. Вероятность того, что уровень E в валентной зоне занят дыркой, равна $1 - f_v(E)$, где $f_v(E)$ — функция Ферми с уровнем Ферми E_{fv}. Концентрации электронов равны $n(E)$ и $p(E)$ соответственно. Обе они могут быть велики

При таких обстоятельствах имеет смысл использовать отдельную функцию Ферми для каждой зоны; два соответствующих уровня Ферми E_{fc} и E_{fv} называются **квазиуровнями Ферми** (рис. 16.15). Когда E_{fc} и E_{fv} лежат достаточно глубоко внутри зоны проводимости и валентной зоны, соответственно, концентрация и электронов, и дырок может быть весьма большой.

Упражнение 16.3

Определение квазиуровней Ферми по данным концентрациям электронов и дырок

а. Для заданных концентраций электронов n и дырок p в полупроводнике при $T = 0$ °K с помощью (16.10) и (16.11) покажите, что квазиуровни Ферми выражаются как

$$E_{fc} = E_c + (3\pi^2)^{2/3} \frac{\hbar^2}{2m_c} n^{2/3}; \tag{16.18а}$$

$$E_{fv} = E_v - (3\pi^2)^{2/3} \frac{\hbar^2}{2m_v} p^{2/3} \tag{16.18б}$$

б. Покажите, что эти выражения приближенно справедливы при произвольной температуре T, если n и p достаточно велики, так что $E_{fc} - E_c \gg k_B T$ и $E_v - E_{fv} \gg k_B T$, т. е. квазиуровни Ферми лежат глубоко внутри зоны проводимости и валентной зоны.

16.1.4. Генерация, рекомбинация и инжекция

Генерация и рекомбинация при тепловом равновесии

Тепловое возбуждение электронов с переходом из валентной зоны в зону проводимости приводит к **электронно-дырочной генерации** (рис 16.16). Тепловое равновесие требует, чтобы генерация сопровождалась одновременным обратным процессом ликвидации возбуждения. Этот процесс, называемый **электронно-дырочной рекомбинацией**, характеризуется переходом электрона из зоны проводимости в валентную зону с заполнением дырки (см. рис. 16.16). Энергия электрона при этом может высвобождаться в виде испущенного фотона, тогда процесс называется **радиационной рекомбинацией**.

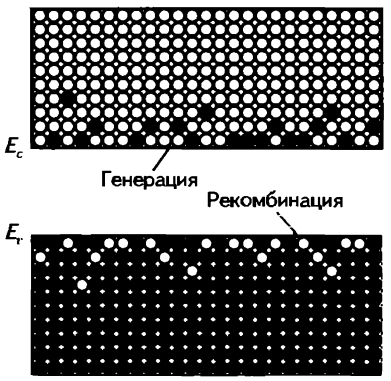


Рис. 16.16. Электронно-дырочная генерация и рекомбинация

Безызлучательная рекомбинация может происходить за счет целого ряда независимых конкурирующих процессов, включающих передачу энергии колебаниям решетки (с рождением одного или нескольких фононов) или другому свободному электрону (процесс Оже). Рекомбинация может происходить на поверхности, а также непрямым путем через ловушки — энергетические уровни, связанные с примесями, границами зерен, дислокациями или другими нарушениями структуры решетки и лежащие внутри запрещенной зоны. Примесь или дефект может вести себя как центр рекомбинации, если он способен захватывать электроны и дырки одновременно, увеличивая вероятность их рекомбинации (рис. 16.17). Примесная рекомбинация может быть радиационной или безызлучательной.

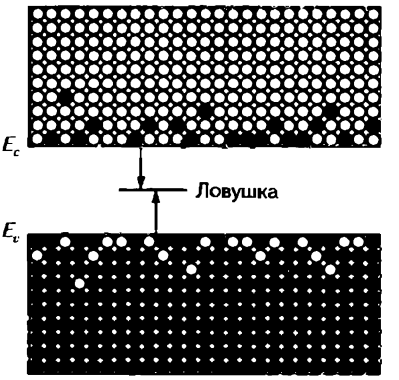


Рис. 16.17. Электронно-дырочная генерация и рекомбинация через ловушку

Поскольку для рекомбинации требуется одновременное присутствие как электрона, так и дырки, скорость рекомбинации пропорциональна произведению концентраций электронов и дырок, т. е.

$$\text{Скорость рекомбинации} = \gamma np, \tag{16.19}$$

где коэффициент рекомбинации γ зависит от характеристик материала, включая его состав и плотность числа дефектов, от температуры, а также сравнительно слабо — от уровня легирования.

Равновесные концентрации электронов и дырок n_0 и p_0 устанавливаются, когда скорость рекомбинации равна скорости генерации. Если G_0 — скорость тепловой электронно-дырочной генерации при данной температуре, то при тепловом равновесии

$$G_0 = \gamma n_0 p_0. \tag{16.20}$$

Произведение концентраций электронов и дырок n_0p_0 примерно одинаково для материалов n-типа, p-типа или собственных полупроводников. Таким образом $n_i^2 = G_0/\gamma$, что приводит прямо к закону действующих масс $n_0p_0 = n_i^2$. Видно, что этот закон — следствие баланса между генерацией и рекомбинацией при тепловом равновесии.

Электронно-дырочная инжекция

Полупроводник в состоянии теплового равновесия с концентрациями носителей n_0 и p_0 имеет одинаковые скорости генерации и рекомбинации $G_0 = \gamma n_0 p_0$. Пусть теперь дополнительно генерируются электронно-дырочные

пары с постоянной скоростью R (число пар в единице объема за единицу времени) за счет внешнего (нетеплового) механизма инжекции, такого как свет, падающий на материал. Тогда будет достигнуто новое состояние равновесия, при котором концентрации равны $n = n_0 + \Delta n$ и $p = p_0 + \Delta p$. Ясно, что $\Delta n = \Delta p$, поскольку электроны и дырки рождаются парами. Приравнивая новые скорости генерации и рекомбинации, получаем

$$G_0 + R = \gamma np. \tag{16.21}$$

Подставляя $G_0 = \gamma n_0 p_0$ в (16.21), получаем выражение

$$R = \gamma\left(np - n_0 p_0\right) = \gamma\left(n_0\Delta n + p_0\Delta n + \Delta n^2\right) = \gamma\Delta n\left(n_0 + p_0 + \Delta n\right), \tag{16.22}$$

которое мы перепишем в виде

$$R = \frac{\Delta n}{\tau}, \tag{16.23}$$

где

$$\tau = \frac{1}{\gamma\left[\left(n_0 + p_0\right) + \Delta n\right]}. \tag{16.24}$$

При скоростях инжекции, для которых $\Delta n \ll n_0 + p_0$,

$$\boxed{\tau \approx \frac{1}{\gamma\left(n_0 + p_0\right)}}. \tag{16.25}$$

Рекомбинационное время жизни избыточных носителей

В материале n-типа, где $n_0 \gg p_0$, рекомбинационное время жизни $\tau \approx 1/\gamma n_0$ обратно пропорционально концентрации электронов. Аналогично, для материала p-типа, где $p_0 \gg n_0$, получаем $\tau \approx 1/\gamma p_0$. Такая простая формулировка неприменима, если существенную роль в процессе рекомбинации играют ловушки.

Параметр τ можно рассматривать как **рекомбинационное время жизни** инжектируемых избыточных электронно-дырочных пар. Это легко понять, если заметить, что концентрация инжектируемых носителей подчиняется скоростному уравнению

$$\frac{\mathrm{d}\left(\Delta n\right)}{\mathrm{d}t} = R - \frac{\Delta n}{\tau}, \tag{16.26}$$

которое аналогично (14.17). В стационарном состоянии $\mathrm{d}(\Delta n)/\mathrm{d}t = 0$, в результате чего воспроизводится соотношение (16.23), аналогичное (14.17). Если источник инжекции внезапно убрать (R становится равным нулю) в момент времени t_0, то Δn экспоненциально затухает с постоянной времени τ, т. е.

$$\Delta n(t) = \Delta n(t_0)\exp\left[-\frac{t - t_0}{\tau}\right].$$

В присутствии сильной инжекции τ само становится функцией Δn, что очевидно из (16.24), так что скоростное уравнение нелинейно, а затухание не экспоненциально.

Если скорость инжекции R известна, то стационарные концентрации инжектируемых носителей можно определить из соотношения

$$\boxed{\Delta n = R\tau,} \tag{16.27}$$

позволяющего определить полные концентрации $n = n_0 + \Delta n$ и $p = p_0 + \Delta p$. Более того, в предположении квазиравновесия соотношения (16.11) можно использовать для нахождения квазиуровней Ферми. Квазиравновесие не противоречит балансу генерации и рекомбинации, который предполагается в проведенных выше рассуждениях; оно просто требует, чтобы время внутризонной релаксации было мало по сравнению с рекомбинационным временем τ.

Анализ такого типа будет полезен при развитии теории полупроводникового светоизлучающего диода и полупроводникового диодного лазера, основанных на усилении испускания света за счет инжекции носителей, как станет ясно в гл. 17.

Упражнение 16.4

Инжекция электронно-дырочных пар в GaAs

Предположим, что электронно-дырочные пары инжектируются в GaAs n-типа ($E_g = 1{,}42$ эВ, $m_c \approx 0{,}07\, m_0$, $m_v \approx 0{,}50 m_0$) со скоростью $R = 10^{23}$ см$^{-3} \cdot$ с$^{-1}$. Концентрация электронов в состоянии теплового равновесия $n_0 = 10^{16}$ см$^{-3}$. При условии, что коэффициент рекомбинации $r = 10^{-11}$ см^3/с и $T = 300$ °K, определите:

а) равновесную концентрацию дырок p_0;

б) рекомбинационное время жизни τ;

в) стационарную избыточную концентрацию Δn;

г) расстояние между квазиуровнями Ферми $E_{fc} - E_{fv}$, предполагая, что $T = 300$ °K.

Внутренний квантовый выход

Внутренний квантовый выход η_i полупроводникового материала определяется как отношение коэффициента радиационной электронно-дырочной рекомбинации к полному коэффициенту рекомбинации (радиационной и безызлучательной). Этот параметр важен, поскольку он определяет эффективность генерации света в полупроводниковом материале. Полный коэффициент рекомбинации дается (16.19). Если разбить коэффициент рекомбинации r на сумму радиационной и безызлучательной частей, $r = r_{\text{рад}} + r_{\text{без}}$, то внутренний квантовый выход выражается как

$$\eta_i = \frac{r_{\text{рад}}}{r} = \frac{r_{\text{рад}}}{r_{\text{рад}} + r_{\text{без}}}. \tag{16.28}$$

Внутренний квантовый выход можно выразить через рекомбинационные времена жизни, поскольку τ обратно пропорционально r [см. (16.25)]. Если ввести радиационное и безызлучательное времена жизни $\tau_{рад}$ и $\tau_{без}$ соответственно, то

$$\frac{1}{\tau} = \frac{1}{\tau_{рад}} + \frac{1}{\tau_{без}}. \tag{16.29}$$

Тогда для внутреннего квантового выхода получаем

$$\frac{r_{рад}}{r} = \frac{1}{\tau_{рад}}\frac{1}{\tau}$$

или

$$\boxed{\eta_i = \frac{\tau}{\tau_{рад}} = \frac{\tau_{без}}{\tau_{рад} + \tau_{без}}.} \tag{16.30}$$

Внутренний квантовый выход

Радиационно-рекомбинационное время жизни $\tau_{рад}$ определяет скорость поглощения и испускания фотона, как объясняется в подразд. 16.2.3. Его величина зависит от концентраций носителей и материального параметра $r_{рад}$. Для малых и средних скоростей инжекции

$$\tau_{рад} \approx \frac{1}{r_{рад}(n_0 + p_0)} \tag{16.31}$$

в соответствии с (16.25). Безызлучательно-рекомбинационное время жизни определяется аналогичным равенством. Однако если безызлучательная рекомбинация происходит через уровни дефектов в запрещенной зоне, то $\tau_{без}$ более чувствительно к концентрации этих центров, чем к концентрации электронов и дырок.

Типичные порядки величины коэффициентов рекомбинации $r_{рад}$, времен жизни $\tau_{рад}$, $\tau_{без}$ и τ, а также внутреннего квантового выхода η_i приведены в табл. 16.4.

Таблица 16.4. **Типичные значения коэффициента радиационной рекомбинации $r_{рад}$, рекомбинационных времен жизни и внутреннего квантового выхода η_i для некоторых полупроводников***

Материал	$r_{рад}$, см^3/с	$\tau_{рад}$	$\tau_{без}$	τ	η_i
Si	10^{-15}	10 мс	100 нс	100 нс	10^{-5}
GaAs	10^{-10}	100 нс	100 нс	50 нс	0,5
GaN**	10^{-8}	20 нс	0,1 нс	0,1 нс	0,005

* Предполагается материал n-типа с концентрацией носителей $n_0 = 10^{17}$ см$^{-3}$ и концентрацией дефектов 10^{15} см$^{-3}$ при $T = 300$ °K.

** На практике чаще используется InGaN, что повышает внутренний квантовый выход до $\eta_i \approx 0{,}3$.

Радиационное время жизни для объемного Si на порядки больше, чем его полное время жизни, главным образом из-за его непрямозонной структуры. Это приводит к низкому внутреннему квантовому выходу. Для GaAs и GaN, напротив, распад происходит в значительной степени за счет радиационных переходов (эти материалы имеют прямозонную структуру) и, следовательно, внутренний квантовый выход велик. Следовательно, прямозонные полупроводники хороши для изготовления светоизлучающих структур, а непрямозонные — нет.

16.1.5. Переходы

Сосединения областей с различным легированием в одном полупроводниковом материале называются **гомопереходами**. Важным примером является *p*—*n*-переход, обсуждаемый в этом разделе. Переходы между различными полупроводниковыми материалами называются **гетеропереходами**. Они будут рассмотрены далее.

p—n-Переход

p—*n*-Переход — это гомопереход между полупроводниками *p*- и *n*-типа. Он действует как диод, который в электронике может служить выпрямителем, логической ячейкой, регулятором напряжения (стабилитроном) или тюнером (варактором). В оптоэлектронике он может служить основой светоизлучающего диода (СИД), диодного лазера, фотоприемника или элемента солнечной батареи.

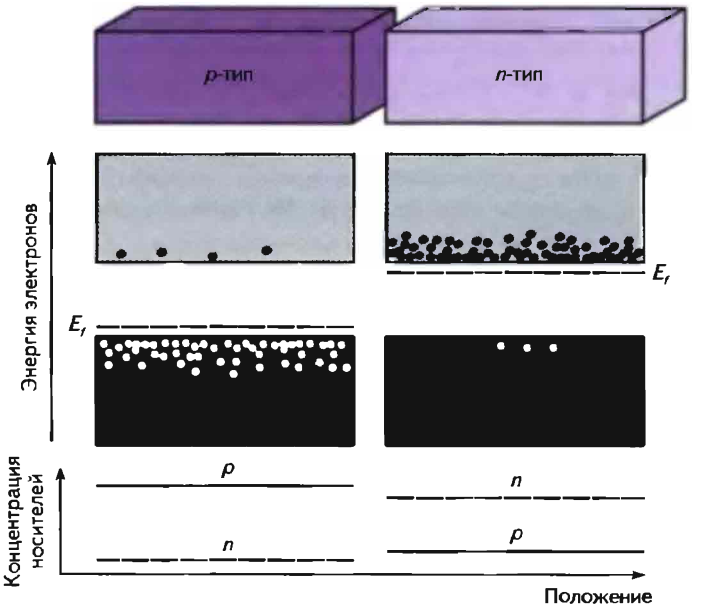


Рис. 16.18. Уровни энергии и концентрации носителей в полупроводниках *p*- и *n*-типа до контакта

p—n-Переход состоит из контактирующих областей p- и n-типа в одном и том же полупроводнике. Область p-типа имеет большое количество дырок (основных носителей) и мало подвижных электронов (неосновных носителей). В области n-типа много подвижных электронов и мало дырок (рис. 16.18). Оба типа носителей заряда находятся в непрерывном тепловом движении во всех направлениях.

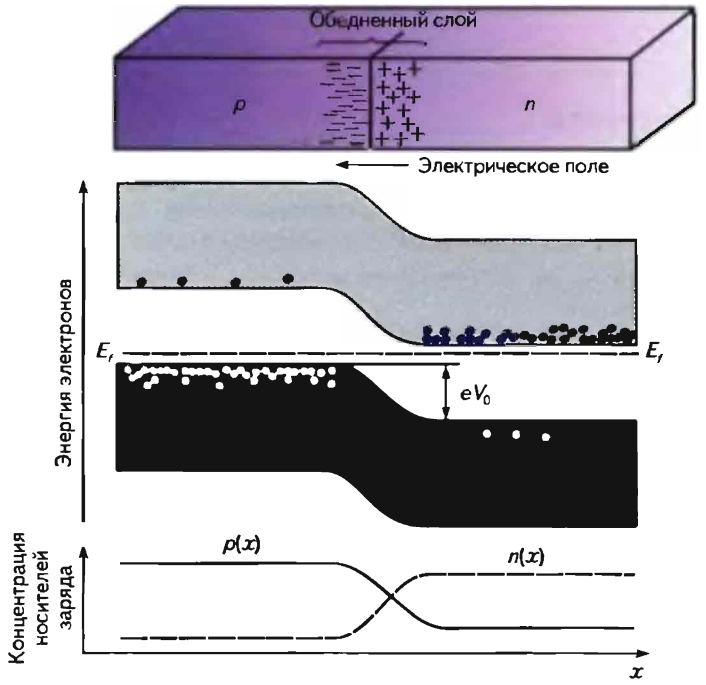


Рис. 16.19. p—n-Переход в состоянии теплового равновесия при $T > 0$ °K. Обедненный слой, схема энергетических зон и концентрации (в логарифмической шкале) подвижных электронов $n(x)$ и дырок $p(x)$ показаны в зависимости от координаты x. Контактная разность потенциалов V_0 соответствует энергии eV_0, где e — величина заряда электрона

Когда две области приводятся в контакт друг с другом (рис. 16.19), происходит следующее:

• Электроны и дырки диффундируют из области высокой концентрации в область низкой. Электроны диффундируют из n-области в p-область, оставляя после себя положительно заряженные ионизированные атомы донора. В p-области электроны рекомбинируют с дырками, которых там много. Аналогично, дырки диффундируют из p-области в n-область, оставляя за собой отрицательно заряженные ионизированные атомы акцептора. В n-области дырки рекомбинируют с подвижными электронами, которыми она богата. Однако этот процесс диффузии не продолжается бесконечно, поскольку он вызывает нарушение баланса зарядов в обеих областях.

• В результате узкая область по обе стороны перехода становится почти лишенной *подвижных* носителей заряда. Эта область называется **обедненным слоем**. Он содержит только неподвижные заряды (положительные ионы со стороны *n*-области и отрицательные — со стороны *p*-области). Толщина обедненного слоя в каждой области обратно пропорциональна концентрации соответствующей примеси в ней.

• Неподвижные заряды создают электрическое поле в обедненном слое, направленное из *n*-области в *p*-область. Это поле препятствует дальнейшей диффузии подвижных носителей заряда через переход.

• Устанавливается равновесие, результатом которого является возникновение разности потенциалов V_0 между двумя сторонами обедненного слоя, причем потенциал со стороны *n*-области выше, чем со стороны *p*-области.

• Из-за указанной разности потенциалов потенциальная энергия электрона со стороны *n*-области ниже, чем со стороны *p*-области. В результате этого энергетические зоны изгибаются, как показано на рис. 16.19. В состоянии теплового равновесия существует единственная функция Ферми для всей структуры, поэтому уровни Ферми в *n*- и *p*-областях должны совпадать.

• Полный ток через переход равен нулю. Токи, связанные с диффузией, и токи, вызванные разностью потенциалов (дрейфовые токи), компенсируют друг друга как для электронов, так и для дырок.

Смещенный p—n-переход

Приложенный извне потенциал изменяет разность потенциалов между *p*- и *n*-областями. Это, в свою очередь, меняет поток основных носителей, так что переход может действовать как «затвор». Если переход подвергается **прямому смещению** путем приложения положительного напряжения V к *p*-области (рис. 16.20), то ее потенциал увеличивается относительно *n*-области, так что электрическое поле создается в направлении, противоположном собственному полю *p*—*n*-перехода. Присутствие внешнего смещающего напряжения вызывает выход из состояния равновесия и рассогласование уровней Ферми в *p*- и *n*-областях, а также в обедненном слое. Наличие двух уровней Ферми E_{fc} и E_{fv} в обедненном слое представляет состояние квазиравновесия.

Полный эффект от прямого смещения состоит в понижении потенциальной ступени на величину eV. Ток основных носителей возрастает в $\exp(eV/k_BT)$ раз, так что полный ток становится равным

$$i = i_s \exp\left(\frac{eV}{k_BT}\right) - i_s,$$

где i_s — постоянная величина. Избыточные основные носители — дырки и электроны, которые попадают в *n*- и *p*-области, соответственно, становятся неосновными носителями и рекомбинируют с местными основными носителями. Поэтому их концентрация убывает с увеличением расстояния до перехода, как показано на рис. 16.20. Этот процесс называется **инжекцией неосновных носителей**.

При **обратном смещении** перехода путем приложения отрицательного напряжения V к p-области высота потенциальной ступени увеличивается на eV. Это препятствует потоку основных носителей. Соответствующий ток умножается на $\exp(eV/k_BT)$, где V отрицательно, т. е. уменьшается. Полный ток становится равным $i = i_s \exp(eV/k_BT) - i_s$, так что небольшой ток $\approx i_s$ течет в обратном направлении при $|V| \gg k_BT/e$.

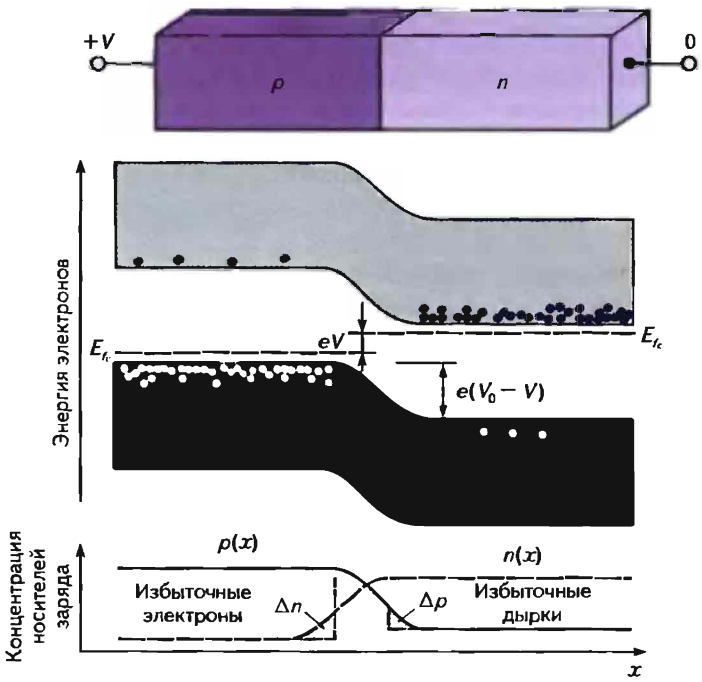


Рис. 16.20. Схема энергетических зон и концентрация носителей для p—n-перехода, смещенного в прямом направлении

Таким образом, p—n-переход действует как диод с вольт-амперной характеристикой

$$i = i_s\left[\exp\left(\frac{eV}{k_BT}\right) - 1\right], \qquad (16.32)$$

Характеристика идеального диода

как показано на рис. 16.21. Характеристика идеального диода (16.32) известна как **уравнение Шокли**.

Динамический отклик p—n-перехода на переменное напряжение определяется путем решения дифференциальных уравнений, которым подчиняется процесс диффузии электронов и дырок, их дрейф под действием собственного электрического поля p—n-перехода и внешнего приложенного поля, а также

рекомбинация. Эти эффекты важны для определения скорости, с которой диод может работать. Удобно моделировать их с помощью двух емкостей: емкости перехода и диффузионной емкости, включенных параллельно идеальному диоду. Емкость перехода учитывает время, необходимое для изменения величины неподвижных положительных и отрицательных зарядов в обедненном слое при изменении приложенного напряжения. Толщина l обедненного слоя оказывается пропорциональной $\sqrt{V_0 - V}$; она растет при обратном (отрицательное V) и уменьшается при прямом (положительное V) смещении. Следовательно, емкость перехода $C = \varepsilon A/l$, где A — площадь перехода, обратно пропорциональна $\sqrt{V_0 - V}$. Емкость перехода при обратном смещении меньше (и, следовательно, время RC-отклика короче), чем при прямом смещении. Зависимость C от V используется при изготовлении конденсаторов, управляемых напряжением (варакторов).

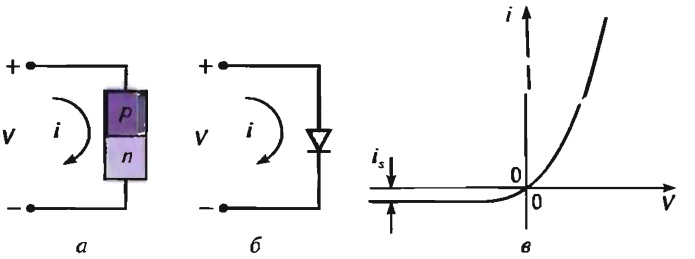


Рис. 16.21. Напряжение и ток в p—n-переходе (*а*). Электрическая схема диода на p—n-переходе (*б*). Вольт-амперная характеристика идеального диода на p—n-переходе (*в*)

Инжекция неосновных носителей в диоде с прямым смещением описывается **диффузионной емкостью**, которая зависит от времени жизни неосновных носителей и рабочего тока.

Диод на p—i—n-переходе

Диод на p—i—n-переходе получается добавлением слоя собственного (или слаболегированного) полупроводника между областями p- и n-типа (рис. 16.22). Поскольку толщина обедненного слоя по каждую сторону от перехода обратно пропорциональна концентрации соответствующей примеси, этот слой далеко проникает в i-область как со стороны p—i-, так и со стороны i—n-перехода. В результате p—i—n-диод может вести себя как p—n-переход, у которого обедненный слой охватывает всю область собственного полупроводника. Энергия электрона, плотность неподвижных зарядов и электрическое поле в p—i—n-диоде при тепловом равновесии показаны на рис. 16.22. Одним из достоинств диода с большой шириной обедненного слоя является малая емкость перехода и, соответственно, быстрый отклик. По этой причине p—i—n-диоды часто предпочтительны для использования в качестве полупроводниковых фотоприемников по сравнению с p—n-диодами. Большая ширина обедненного слоя позволяет захватить большую долю падающего света, увеличивая таким образом эффективность приемника (см. подразд. 18.3.2).

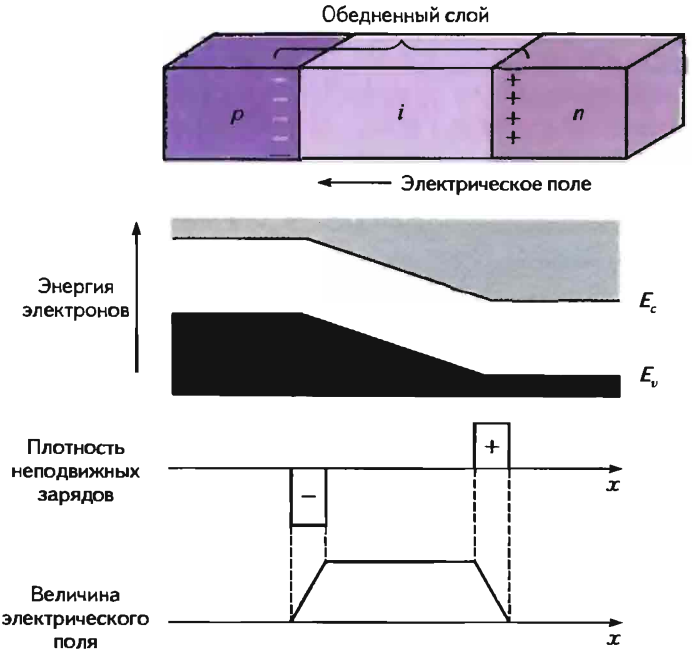


Рис. 16.22. Энергия электрона, плотность неподвижных зарядов и величина электрического поля в p—i—n-диоде при тепловом равновесии

16.1.6. Гетеропереходы

Переходы между различными полупроводниковыми материалами называются гетеропереходами. Гетеропереходы широко используются в источниках и приемниках света, причем не только как активные элементы, но и как контактные слои и волноводные области. Сродство к электрону у материалов определяет расположение краев зоны проводимости и валентной зоны. Зачастую выгодно согласовать полупроводниковые материалы по структуре решетки и использовать плавные, а не резкие переходы. Контакт различных полупроводников может иметь множество преимуществ для фотоники:

- Переход между материалами с различной шириной запрещенной зоны создает локальный скачок зонной структуры. Разрыв потенциальной энергии создает барьер, который может быть полезным для предотвращения проникновения определенного типа носителей в области, где их присутствие нежелательно. Это свойство может быть использовано, например, в p—n-переходе для уменьшения части тока, переносимого неосновными носителями, и, следовательно, для повышения эффективности инжекции (рис. 16.23).
- Разрывы энергетических зон, создаваемые двумя гетеропереходами, могут быть полезны для удержания носителей заряда в заданной области пространства. Например, слой материала с узкой запрещенной зоной можно поместить

между двумя слоями материалов с большей шириной запрещенной зоны, как в *p*—*p*—*n*-структуре, показанной на рис. 16.23 и состоящей из гетеропереходов *p*—*p* и *p*—*n*. Такая **двойная гетероструктура** эффективно используется при изготовлении светодиодов, полупроводниковых усилителей света и диодных лазеров (см. гл. 17).

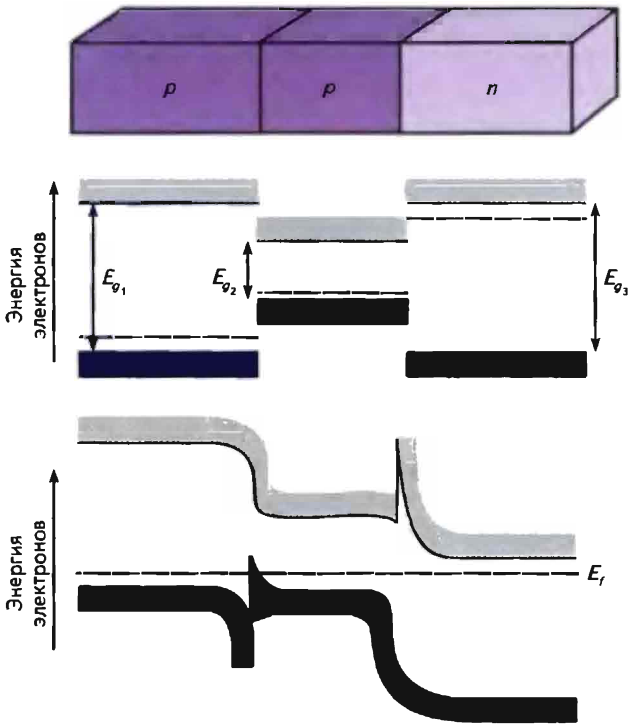


Рис. 16.23. *p*—*p*—*n*-Структура с двойной гетероинжекцией. Средний слой имеет более узкую запрещенную зону, чем крайние. При равновесии уровни Ферми совпадают, так что край зоны проводимости резко падает на *p*—*p*-переходе, а край валентной зоны — на *p*—*n*-переходе. Разрывы зоны проводимости и валентной зоны известны как **скачки потенциала зоны** (band offsets). При прямом смещении эти скачки действуют как барьеры, запирающие инжектируемые неосновные носители в области с меньшей шириной запрещенной зоны. Например, для электронов, инжектируемых из *n*-области, предотвращается диффузия через энергетический барьер на *p*—*p*-переходе. Аналогично, дыркам, инжектируемым из *p*-области, не позволяется диффундировать за энергетический барьер на *p*—*n*-переходе. Таким образом, данная двойная гетероструктура заставляет электроны и дырки занимать общую узкую область. Это существенно увеличивает эффективность светоизлучающих диодов, полупроводниковых оптических усилителей и диодных лазеров (см. гл. 17)

• Гетеропереходы полезны для создания разрывов энергетических зон, которые ускоряют носители заряда в нужных местах. Дополнительная кинетичес-

кая энергия, внезапно сообщаемая носителю заряда, может быть использована для селективного повышения вероятности ударной ионизации в многослойном лавинном фотодиоде (см. подразд. 18.4.1).

• Полупроводники с различным типом запрещенной зоны (прямозонные и непрямозонные) можно использовать в таком же устройстве для выделения областей структуры, в которых поглощается свет. Полупроводниковые материалы, у которых ширина запрещенной зоны больше энергии фотона падающего света, будут прозрачными и служат **оптическими окнами**.

• Гетеропереходы между материалами с различными показателями преломления можно использовать для создания фотонно-кристаллических структур и оптических волноводов, захватывающих и направляющих фотоны, как обсуждалось в гл. 7 и 8.

16.1.7. Квантово-размерные структуры

Гетероструктуры из тонких слоев полупроводниковых материалов можно выращивать посредством эпитаксии, т. е. как слои одного полупроводникового материала поверх другого с использованием таких методов, как молекулярно-пучковая эпитаксия (МПЭ), эпитаксия из жидкой фазы (ЭЖФ), а также различные виды эпитаксии из паровой фазы (ЭПФ), из которых общепринятыми являются химическое осаждение из паровой фазы металлоорганических соединений и гибридная эпитаксия из паровой фазы. **Гомоэпитаксия** — это наращивание на подложку материалов того же состава, в то время как **гетероэпитаксия** — наращивание на подложку материалов с другим составом, как согласованных по структуре решетки, так и не согласованных. В МПЭ используются молекулярные пучки составляющих элементов, направляемые на подготовленную подложку в условиях высокого вакуума. При ЭЖФ используется охлаждение насыщенного раствора, содержащего составляющие элементы и находящегося в контакте с подложкой. При ЭПФ используются газы в реакторе. Состав и легирование отдельных слоев, которые могут быть настолько тонкими, что являются монослоями, определяются путем манипулирования интенсивностями потока молекул и температурой поверхности подложки.

Если толщина слоя сравнима с длиной волны де Бройля движущегося с тепловой скоростью электрона или меньше ее, необходимо учитывать квантование энергии электрона в слое и соотношение между энергией и импульсом, справедливое в объемном полупроводниковом материале, становится непригодным. Длина волны де Бройля выражается как $\lambda = h/p$, где h — постоянная Планка, а p — импульс электрона ($\lambda \approx 50$ нм для GaAs). Три типа структур обладают ценными достоинствами для применений в фотонике: квантовые ямы, квантовые проволоки и квантовые точки (см. подразд. 13.1.3). Ниже выведены соотношения между энергией и импульсом для этих структур. Применения этих структур отложим до последующих гл. 17 и 18.

Квантовые ямы

Структура типа **квантовой ямы**, показанная на рис. 16.24, — двойная гетероструктура, состоящая из ультратонкого ($\lesssim$50 нм) слоя полупроводникового материала, ширина запрещенной зоны которого меньше, чем у окружающего материала. В качестве примера может служить тонкий слой GaAs, окруженный AlGaAs (см. рис. 13.11). В зоне проводимости и валентной зоне такой сэндвич-структуры формируются одномерные прямоугольные потенциальные ямы, в которых электроны и дырки находятся в связанных состояниях: электроны — в зоне проводимости, а дырки — в валентной зоне.

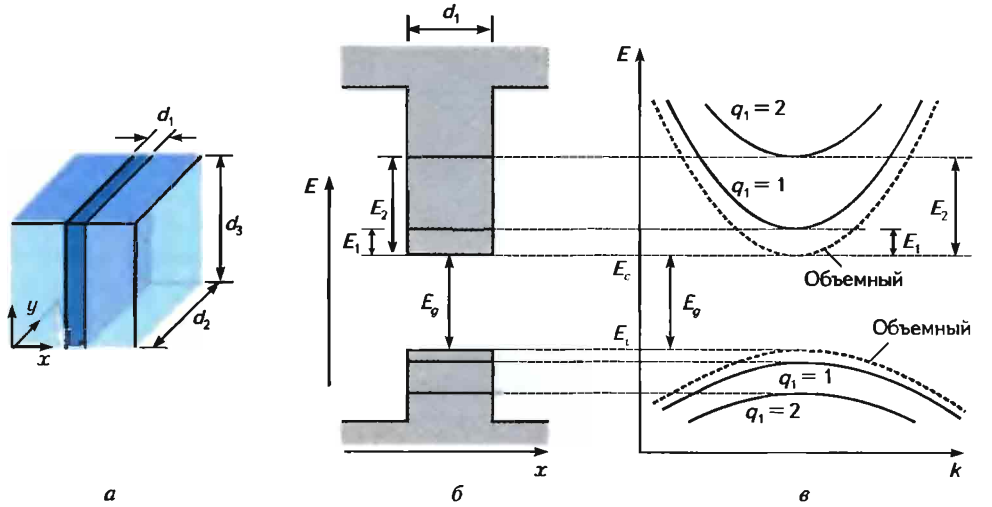


Рис. 16.24. Геометрия структуры с квантовой ямой (*а*). Схема энергетических уровней электронов и дырок в квантовой яме (*б*). Сечение E—k-диаграммы в направлении k_2 или k_3 (*в*). Энергетические подзоны нумеруются своим квантовым числом $q_1 = 1, 2, 3,$ Сечение E—k-диаграммы для объемного полупроводника показано штриховыми линиями

Достаточно глубокую потенциальную яму можно аппроксимировать бесконечно глубокой прямоугольной потенциальной ямой (рис. 16.25). Энергетические уровни частицы массы m (m_c для электронов и m_v для дырок), находящейся в одномерной прямоугольной бесконечно глубокой потенциальной яме ширины d, определяются решением стационарного уравнения Шредингера (13.3). Как показано в упражнении 16.5, собственные значения энергии оказываются равными

$$E_q = \frac{\hbar^2 (q\pi/d)^2}{2m}, \quad q = 1, 2, 3, \tag{16.33}$$

Например, для первых трех уровней энергия электрона в бесконечно глубокой яме в GaAs ($m_c = 0{,}07m_0$) шириной $d = 10$ нм составляет $E_q = 54$, 216 и 486 мэВ

соответственно (напомним, что $k_B T = 26$ мэВ при $T = 300$ °K). Чем меньше ширина ямы, тем больше расстояние между уровнями энергии.

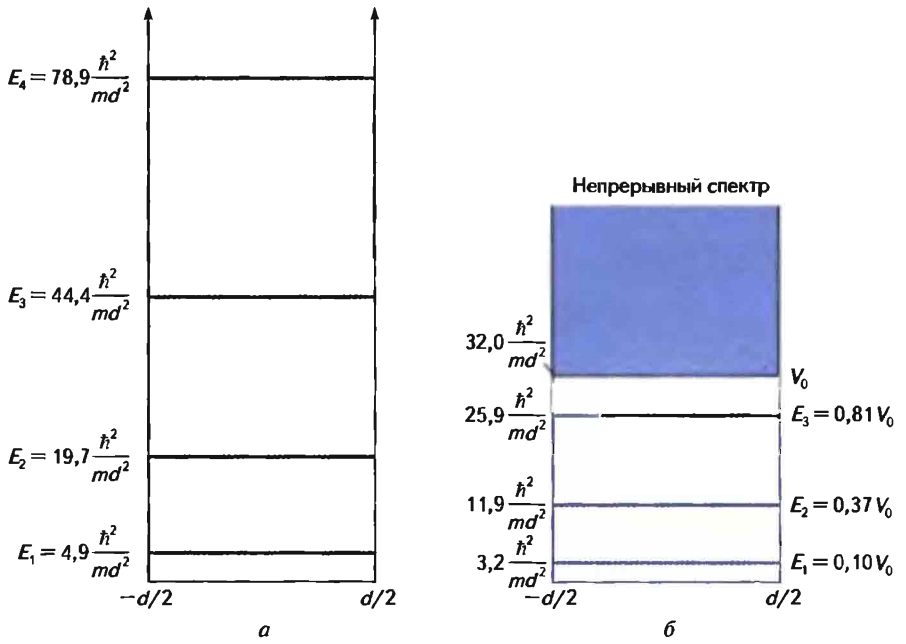


Рис. 16.25. Энергетические уровни для одномерной бесконечной прямоугольной потенциальной ямы (*а*) и конечной квадратной ямы глубиной $V_0 = 32\hbar^2/md^2$ (*б*)

Упражнение 16.5

Энергетические уровни квантовой ямы

Решите уравнение Шредингера (13.3) и определите возможные значения энергии электрона массы m в бесконечно глубокой прямоугольной потенциальной яме [$V(x) = 0$ при $0 < x < d$ и $V(x) = \infty$ в остальных случаях]. Проверьте, что

$$E_q = \frac{\hbar^2 (q\pi/d)^2}{2m}, \quad q = 1,\ 2,\ 3,$$

(см. рис. 16.25, *а*). Сравните эти значения со значениями для квадратной ямы конечной глубины, показанной на рис. 16.25, *б*.

Однако квантовые ямы в полупроводниках на самом деле — трехмерные объекты. В структуре, показанной на рис. 16.24, движение электронов и дырок ограничено в направлении x расстоянием d_1 (шириной ямы) и намного большими расстояниями d_2, $d_3 \gg d_1$ в плоскости слоя. Таким образом, в плоскости y—z они ведут себя так же, как в объемном полупроводнике.

Соотношение между энергией и импульсом электрона имеет вид

$$E = E_c + \frac{\hbar^2 k_1^2}{2m_c} + \frac{\hbar^2 k_2^2}{2m_c} + \frac{\hbar^2 k_3^2}{2m_c}, \tag{16.34}$$

где

$$k_1 = \frac{q_1 \pi}{d_1}; \quad k_2 = \frac{q_2 \pi}{d_2}; \quad k_3 = \frac{q_3 \pi}{d_3}, \quad q_1, q_2, q_3 = 1, 2, 3,$$

Поскольку d_2, $d_3 \ll d_1$, параметр k_1 принимает далеко отстоящие друг от друга значения, а k_2 и k_3 — очень близкие друг к другу значения, которые можно аппроксимировать сплошным спектром. Отсюда следует, что соотношение между энергией и импульсом для электронов в зоне проводимости квантовой ямы имеет вид

$$E = E_c + E_{q_1} + \frac{\hbar^2 k^2}{2m_c}, \quad q_1 = 1, 2, 3, \tag{16.35}$$

где k — модуль двумерного вектора $\boldsymbol{k} = (k_1, k_2)$ в плоскости y—z. Каждое значение квантового числа q_1 соответствует **подзоне** с наименьшим значением энергии $E_c + E_{q_1}$. Аналогичные соотношения имеют место для валентной зоны.

Соотношение между энергией и импульсом в объемном полупроводнике дается формулой (16.3), где k — модуль трехмерного вектора $\boldsymbol{k} = (k_1, k_2, k_3)$. Главное отличие в том, что для квантовой ямы k_1 принимает далеко отстоящие друг от друга дискретные значения. В результате плотность состояний для структуры с квантовой ямой отличается от таковой для объемного полупроводника, где она определяется модулем трехмерного вектора с компонентами

$$k_1 = \frac{q_1 \pi}{d}; \quad k_2 = \frac{q_2 \pi}{d}; \quad k_3 = \frac{q_3 \pi}{d}, \quad d_1 = d_2 = d_3 = d.$$

В результате получается $\rho(k) = k^2/\pi^2$ на единицу объема [см. (16.6)], откуда плотность состояний в зоне проводимости [см. (16.7) и рис. 16.10]

$$\rho_c(E) = \frac{\sqrt{2} m_c^{3/2}}{\pi^2 \hbar^3} \sqrt{E - E_c}, \quad E > 0. \tag{16.36}$$

В структуре с квантовой ямой плотность состояний получается из модуля двумерного вектора (k_1, k_2). Следовательно, для каждого значения квантового числа q_1 плотность состояний равна $\rho(k) = k/\pi$ на единицу площади в плоскости y—z и, следовательно, $k/\pi d_1$ на единицу объема. Плотности $\rho_c(E)$ и $\rho(k)$ связаны соотношением

$$\rho_c(E)\mathrm{d}E = \rho(k)\mathrm{d}k = \frac{k}{\pi d_1}\mathrm{d}k.$$

Окончательно, с использованием соотношения (16.35) получаем

$$\frac{dE}{dk} = \frac{\hbar^2 k}{m_c},$$

откуда

$$\rho_c(E) = \begin{cases} \dfrac{m_c}{\pi\hbar^2 d_1}, & E > E_c + E_{q_1}; \\ 0, & E < E_c + E_{q_1}, \end{cases} \quad q_1 = 1,\ 2,\ 3, \qquad (16.37)$$

Таким образом, для каждого значения квантового числа q_1 плотность состояний на единицу объема постоянна при $E > E_c + E_{q_1}$. Полная плотность состояний есть сумма плотностей состояний для всех q_1, поэтому она имеет ступенчатый вид, показанный на рис. 16.26. Каждая ступень соответствует значению квантового числа q_1 и может рассматриваться как подзона внутри зоны проводимости (см. рис. 16.24). Дно этих подзон сдвигается все выше и выше с ростом квантового числа. При подстановке $E = E_c + E_{q_1}$ в (16.36) с использованием (16.33) находим, что плотность состояний для квантовой ямы такая же, как для объемного материала. Плотность состояний в валентной зоне имеет такое же ступенчатое распределение.

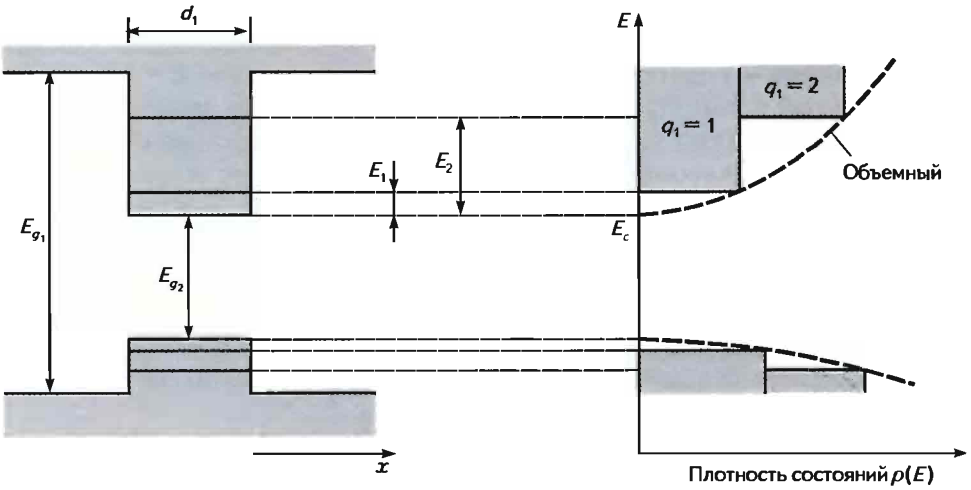


Рис. 16.26. Плотность состояний для структуры с квантовой ямой (сплошная линия) и для объемного полупроводника (штриховая линия)

В отличие от объемного полупроводника структура с квантовой ямой имеет значительную плотность состояний при наименьшем возможном значении энергии в зоне проводимости и при наибольшем возможном — в валентной зоне. Это свойство оказывает существенное влияние на оптические характеристики материала, что обсуждается в подразд. 17.2.4.

Множественные квантовые ямы и сверхрешетки

Многослойные структуры, состоящие из чередующихся полупроводниковых материалов, известны как **множественные квантовые ямы (МКЯ)** (рис. 16.27). Они могут быть изготовлены так, что ширина запрещенной зоны меняется в зависимости от положения любым желательным образом (см., например, рис. 13.11).

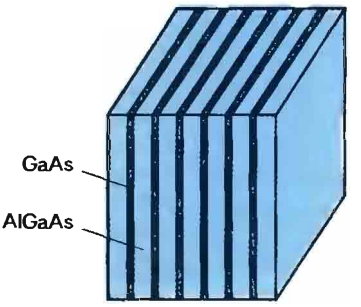


Рис. 16.27. Структура с множественными квантовыми ямами, изготовленная из чередующихся слоев материалов с различными ширинами запрещенной зоны, таких как AlGaAs и GaAs. Эти конкретные материалы часто используются как пример структуры с множественными квантовыми ямами, поскольку они согласованы по структуре решетки в широком диапазоне составов (см. рис. 16.7, *а*), что сводит к минимуму механические напряжения между двумя решетками, а также поскольку у них большая разница в ширине запрещенной зоны (см. табл. 16.2), что обеспечивает существенную связь носителей. Другие комбинации материалов для структур с МКЯ включают AlInAs/InGaAs, AlInGaP/InGaP, GaN/InGaN и $Al_xGa_{1-x}N/Al_yGa_{1-y}N$

Структура с МКЯ может иметь любое число слоев, от нескольких до сотен. Например, МКЯ-структура из 100 слоев, каждый из которых имеет толщину ≈10 нм и содержит примерно 40 атомных плоскостей, имеет полную толщину ≈1 мкм. Как обсуждалось в подразд. 13.1.3, если энергетические барьеры между соседними ямами достаточно тонкие, чтобы электроны могли легко туннелировать из ямы в яму, то дискретные энергетические уровни расширяются с образованием мини-зон. В этом случае структуру с множественными квантовыми ямами принято называть **сверхрешеткой**. Переход от подзон в МКЯ к минизонам в сверхрешетках происходит аналогично тому, как дискретные уровни атома превращаются в зоны твердого тела по мере того, как атомы приходят в непосредственную близость и начинают взаимодействовать (см. рис. 13.6 и 13.7). Квантовые ямы и сверхрешетки можно изготовить также путем пространственного изменения степени легирования материала, создавая таким образом поля пространственных зарядов, образующие потенциальные барьеры.

Структуры со смещенными множественными квантовыми ямами

Типичные схемы энергетических зон несмещенных и смещенных множественных квантовых ям и сверхрешеток показаны на рис. 16.28. Электрическое поле делает ямы скошенными и меняет уровни энергии. В сверхрешетках дискретные энергетические уровни размываются в мини-зоны. Структуры с множественными квантовыми ямами находят множество применений в фотонике, например, в качестве активных областей светоизлучающих диодов, в полупроводниковых оптических усилителях и диодных лазерах (см.

подразд. 17.1.3, 17.2.4 и разд. 17.4 соответственно). Они служат также фотоприемниками (см. подразд. 18.2.3) и модуляторами (см. разд. 20.5).

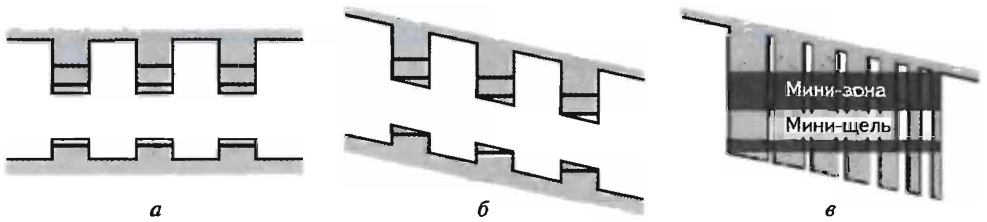


Рис. 16.28. Схемы энергетических зон МКЯ и сверхрешетки из чередующихся слоев материалов с различной шириной запрещенной зоны, таких как AlGaAs и GaAs. МКЯ-структура без смещения (*а*). МКЯ-структура со смещением (*б*). Сверхрешетка со смещением, показаны две разрешенные и одна запрещенная минизоны (*в*)

Квантовые провода

Полупроводниковый материал, которому придана форма тонкой проволоки, окруженной материалом с большей шириной запрещенной зоны, называется **квантовым проводом** (рис. 16.29). Провод обладает потенциалом, ограничивающим движение электронов и дырок в двух поперечных измерениях x и y. Предполагая, что провод имеет прямоугольное сечение с площадью $d_1 d_2$, можно записать связь энергии с импульсом в зоне проводимости как

$$E = E_c + E_{q_1} + E_{q_2} + \frac{\hbar^2 k^2}{2m_c}, \tag{16.38}$$

где

$$E_{q_1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_c}; \quad E_{q_2} = \frac{\hbar^2 (q_2 \pi / d_2)^2}{2m_c}, \quad q_1, q_2 = 1, 2, 3, \tag{16.39}$$

k — компонента волнового вектора в направлении z (вдоль оси провода).

Каждая пара квантовых чисел (q_1, q_2) связана с энергетической подзоной, имеющей плотность состояний $\rho(k) = 1/\pi$ на единицу длины провода и, следовательно, $1/\pi d_1 d_2$ на единицу объема. Соответствующая плотность состояний на единицу объема как функция энергии:

$$\rho_c(E) = \begin{cases} \dfrac{(1/d_1 d_2)\left(\sqrt{m_c}/\sqrt{2}\pi\hbar\right)}{\sqrt{E - E_c - E_{q_1} - E_{q_2}}}, & E > E_c + E_{q_1} + E_{q_2}; \quad q_1, q_2 = 1, 2, 3, \\ 0, & E < E_c + E_{q_1} + E_{q_2}, \end{cases} \tag{16.40}$$

Это убывающие функции энергии, как показано на рис. 16.29.

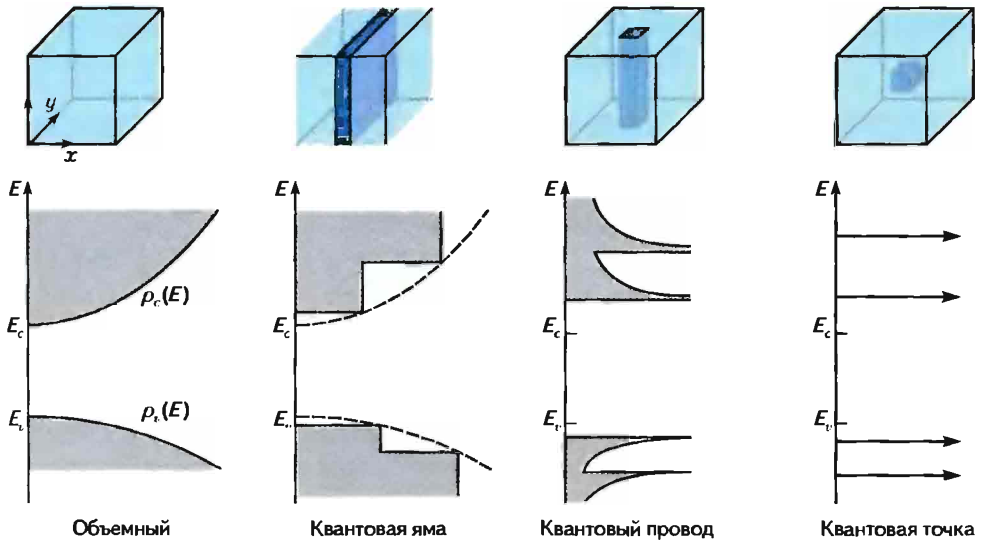


Рис. 16.29. Плотность состояний для конфигураций, ограниченных в различном числе измерений. Зона проводимости и валентная зона расщепляются на перекрывающиеся подзоны и становятся все уже по мере того, как движение электрона ограничивается в большем числе измерений

Квантовые точки

В структурах типа **квантовой точки** движение электронов сильно ограничено во всех трех измерениях. Положим, что это прямоугольный объем $d_1 d_2 d_3$. Тогда энергия квантуется как

$$E = E_c + E_{q_1} + E_{q_2} + E_{q_3}, \tag{16.41}$$

где

$$E_{q_1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_c}; \quad E_{q_2} = \frac{\hbar^2 (q_2 \pi / d_2)^2}{2m_c}; \quad E_{q_3} = \frac{\hbar^2 (q_3 \pi / d_3)^2}{2m_c}, \tag{16.42}$$

$$q_1,\ q_2,\ q_3 = 1,\ 2,\ 3,$$

Возможные значения энергии дискретны и достаточно далеки друг от друга, так что плотность состояний представляет собой последовательность дельта-функций, как показано на рис. 16.29. Квантовые точки часто называют искусственными атомами (см. подразд. 13.1.4). Хотя они могут содержать огромное число сильно взаимодействующих атомов, их дискретный энергетический спектр можно создавать по своему желанию путем соответствующего проектирования.

16.2. ВЗАИМОДЕЙСТВИЕ ФОТОНОВ С НОСИТЕЛЯМИ ЗАРЯДА

Перейдем к рассмотрению некоторых основных оптических свойств полупроводников с акцентом на процессы поглощения и испускания, важные для работы приборов фотоники. Эта область исследований известна как **оптика полупроводников**.

16.2.1. Взаимодействие фотонов с объемными полупроводниками

В объемных полупроводниках целый ряд механизмов может приводить к поглощению и испусканию фотонов. Вот важнейшие из них.

• *Межзонные переходы.* Поглощение фотона может вызвать переход электрона из валентной зоны в зону проводимости и, следовательно, рождение электронно-дырочной пары (рис. 16.30, *а*). Электронно-дырочная рекомбинация может приводить к испусканию фотона. Межзонные переходы могут происходить с участием одного или нескольких фононов. Фонон — это квант колебаний решетки, связанный с молекулярными или акустическими колебаниями атомов материала.

• *Переходы «примесь—зона».* Поглощение фотона может вызвать переход между донорным (акцепторным) уровнем и зоной в примесном полупроводнике. В материале *p*-типа, например, фотон с малой энергией может перевести электрон из валентной зоны на акцепторный уровень, где он захватывается атомом акцептора (рис. 16.30, *б*). В валентной зоне остается дырка, а атом акцептора ионизуется. Либо дырка может быть захвачена ионизованным атомом акцептора, в результате чего электрон освобождает акцепторный уровень, рекомбинируя с дыркой. Энергия может высвобождаться радиационным путем (в виде испущенного фотона) или без излучения (в виде фононов). Такой переход может происходить с помощью ловушек в примесных состояниях, как показано на рис. 16.17.

• *Переходы между состояниями свободных носителей* (*внутризонные переходы*). Поглощенный фотон может отдать свою энергию электрону, движущемуся внутри зоны, так что электрон перейдет на более высокий энергетический уровень внутри зоны. Например, находясь в зоне проводимости, электрон может поглотить энергию фотона и перейти на более высокий уровень в этой же зоне (рис. 16.30, *в*). Затем следует термализация — процесс, в ходе которого электрон постепенно релаксирует к равновесному состоянию вблизи дна зоны проводимости, отдавая энергию фононам. Величина поглощения свободными носителями пропорциональна их концентрации; она убывает с ростом энергии фотона по степенно́му закону.

• *Фононные переходы.* Длинноволновые фотоны могут отдать свою энергию непосредственно колебаниям решетки, т. е. путем рождения фононов.

• *Экситонные переходы.* Поглощение фотона может привести к образованию **экситона**. Это объект напоминает атом водорода, в котором роль протона играет дырка. Электрон и дырка образуют связанное состояние за счет куло-

новского притяжения. Фотон может быть испущен при рекомбинации дырки и электрона, т. е. уничтожения экситона.

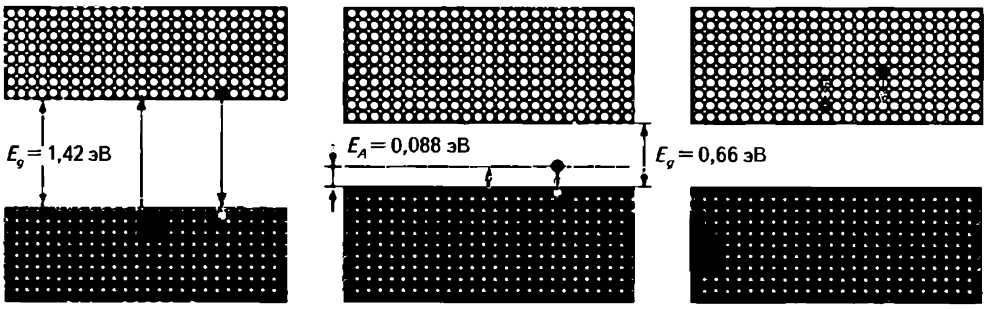


Рис. 16.30. Примеры поглощения и испускания фотонов в объемных полупроводниках:

а — межзонные переходы в GaAs могут приводить к поглощению и испусканию фотонов с длиной волны $\lambda_0 < \lambda_g = hc_0/E_g = 0.87$ мкм; *б* — поглощение фотона с длиной волны $\lambda_A = hc_0/E_A = 14$ мкм приводит к переходу из валентной зоны на акцепторный уровень в Ge, легированном Hg (Ge : Hg); *в* — переходы между состояниями свободных носителей заряда в зоне проводимости Ge

Все эти виды переходов вносят вклад в полный коэффициент поглощения, изображенный на рис. 16.31 (см. цв. вклейку) для Si и GaAs, и, с большим увеличением, на рис. 16.32 (см. цв. вклейку) для ряда полупроводниковых материалов. Для энергий фотона больше, чем ширина запрещенной зоны E_g, в поглощении доминируют межзонные переходы, что составляет основу действия множества приборов в фотонике. Спектральная область, в которой свойства материала меняются от относительной прозрачности ($h\nu < E_g$) к сильному поглощению ($h\nu > E_g$), называется **краем поглощения**. Прямозонные полупроводники имеют более резкие края поглощения, чем непрямозонные, как видно из рис. 16.31 и 16.32.

16.2.2. Межзонные переходы в объемных полупроводниках

Перейдем к развитию простой теории прямого межзонного поглощения и испускания фотонов в объемных полупроводниках, игнорируя другие типы переходов.

Длина волны, соответствующая ширине запрещенной зоны

Прямое межзонное поглощение и испускание могут происходить только на частотах, при которых энергия фотона $h\nu > E_g$. Минимальная частота n, необходимая для этого, равна $\nu_g = E_g/h$, так что соответствующая максимальная длина волны:

$$\lambda_g = \frac{c_0}{\nu_g} = \frac{hc_0}{E_g}.$$

Если ширина запрещенной зоны выражается в электронвольтах, а не в джоулях, то соответствующая длина волны

$$\lambda_g = \frac{hc_0}{eE_g}$$

в микрометрах выражается как

$$\boxed{\lambda_g \approx \frac{1,24}{E_g}.} \qquad (16.43)$$

Длина волны λ_g (мкм), соответствующая ширине запрещенной зоны E_g (эВ)

Величина λ_g называется **длиной волны запрещенной зоны**, или **длиной волны отсечки**.

Значения длины волны отсечки λ_g и ширины запрещенной зоны E_g для ряда полупроводниковых материалов, важных для фотоники, приведены в табл. 16.2 и на рис. 16.7 и 16.8. Трех- и четырехкомпонентные полупроводники III—V групп с различным составом имеют длины волны отсечки от среднего инфракрасного до среднего ультрафиолетового диапазона, как хорошо видно из рис. 16.33 (см. цв. вклейку).

Условия поглощения и испускания

Возбуждение электрона из валентной зоны в зону проводимости возможно в результате поглощения фотона с соответствующей энергией ($h\nu > E_g$ или $\lambda < \lambda_g$). Рождается электронно-дырочная пара (рис. 16.34, *а*). Это вносит добавку в концентрацию подвижных носителей заряда и повышает электропроводность материала. Материал ведет себя как фотопроводник с проводимостью, пропорциональной потоку фотонов. Этот эффект используется для детектирования света, которое обсуждается в гл. 18.

Переход электрона из зоны проводимости в валентную зону (электронно-дырочная рекомбинация) может иметь своим результатом испускание фотона с энергией $h\nu > E_g$ (рис. 16.34, *б*), причем либо спонтанное, либо вынужденное, если фотон с такой энергией уже имеется (см. разд. 13.3). Спонтанное излучение лежит в основе действия светоизлучающего диода, как будет видно из разд. 17.1. Вынужденное излучение отвечает за работу полупроводниковых оптических усилителей и диодных лазеров, которые будут рассмотрены в разд. 17.2, 17.3 и 17.4. Условия, при которых происходит поглощение и испускание, можно сформулировать следующим образом:

• *Сохранение энергии.* Для поглощения или испускания фотона с энергией $h\nu$ требуется, чтобы энергии двух состояний, участвующих во взаимодействии (скажем, E_1 в валентной зоне и E_2 в зоне проводимости, как показано на рис. 16.34), были разделены интервалом $h\nu$. Например, для того, чтобы произошло испускание фотона за счет электронно-дырочной рекомбинации, электрон, занима-

ющий энергетический уровень E_2, должен взаимодействовать с дыркой, занимающей энергетический уровень E_1, так чтобы энергия сохранялась, т. е.

$$E_2 - E_1 = h\nu. \tag{16.44}$$

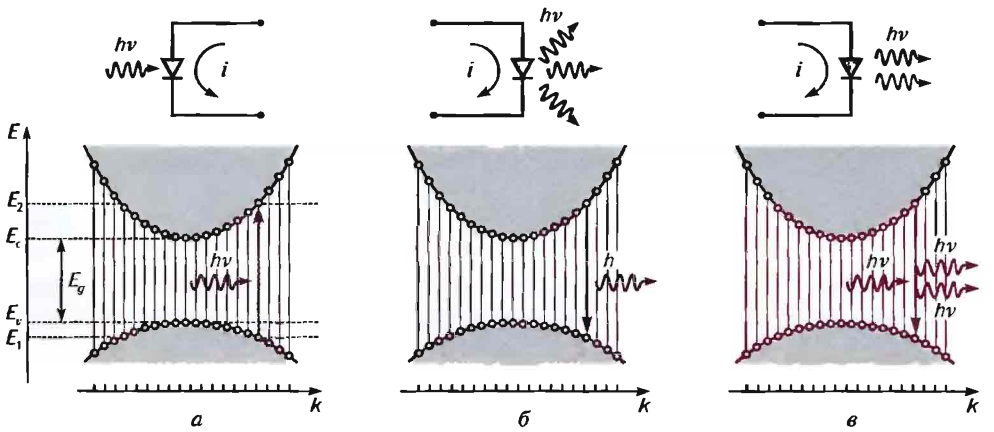


Рис. 16.34. Поглощение фотона приводит к рождению электронно-дырочной пары. Процесс используется в фотоприемниках (*а*). Рекомбинация электронно-дырочной пары сопровождается спонтанным испусканием фотона (*б*). На этой основе работают светоизлучающие диоды. Рекомбинация электрона и дырки может быть индуцирована фотоном (*в*). В результате происходит вынужденное испускание идентичного фотона. Этот процесс лежит в основе действия полупроводниковых диодных лазеров

• *Сохранение импульса.* Импульс также должен сохраняться в процессе испускания или поглощения фотона, так что

$$p_2 - p_1 = \frac{h\nu}{c} = \frac{h}{\lambda} \quad \text{или} \quad k_2 - k_1 = \frac{2\pi}{\lambda}.$$

Однако величина импульса фотона h/λ очень мала по сравнению с возможными значениями импульса электронов и дырок. E—k-диаграмма полупроводника простирается до значений k порядка $2\pi/a$, где постоянная решетки a мала по сравнению с длиной волны λ, так что $2\pi/\lambda = 2\pi/a$. Следовательно, импульсы электрона и дырки, участвующие во взаимодействии, должны быть приблизительно равны. Условие $k_2 \approx k_1$ называется **правилом отбора по** k. Переходы, подчиняющиеся этому правилу, изображаются на E—k-диаграмме (см. рис. 16.34) вертикальными линиями, отмечая тот факт, что изменение импульса в масштабе диаграммы пренебрежимо мало.

• *Энергии и импульсы электрона и дырки, с которыми взаимодействует фотон.* Как видно из рис. 16.34, законы сохранения энергии и импульса требуют, чтобы фотон с частотой ν взаимодействовал с электронами и дырками со специфическими значениями энергии и импульса, определяемыми соотношением между E и k для данного полупроводника. С использованием (16.3) и (16.4) для

аппроксимации этого соотношения в прямозонном полупроводнике двумя параболами, записывая $E_c - E_v = E_g$, соотношение (16.44) можно записать в виде

$$E_2 - E_1 = \frac{\hbar^2 k^2}{2m_v} + E_g + \frac{\hbar^2 k^2}{2m_c} = h\nu, \tag{16.45}$$

откуда

$$k^2 = \frac{2m_r}{\hbar^2}(h\nu - E_g), \tag{16.46}$$

где

$$\frac{1}{m_r} = \frac{1}{m_v} + \frac{1}{m_c}. \tag{16.47}$$

Подставляя (16.46) в (16.45), находим, что энергии уровней E_1 и E_2, с которыми взаимодействует фотон, выражаются как

$$E_2 = E_c + \frac{m_r}{m_c}(h\nu - E_g); \tag{16.48}$$

$$E_1 = E_v - \frac{m_r}{m_v}(h\nu - E_g) = E_2 - h\nu. \tag{16.49}$$

В частном случае, когда $m_c = m_v$, получаем

$$E_2 = E_c + \frac{1}{2}(h\nu - E_g),$$

в соответствии с требованиями симметрии.

• *Оптическая совместная плотность состояний.* Определим теперь плотность состояний $\rho(\nu)$, с которыми фотон с энергией $h\nu$ взаимодействует при выполнении условий сохранения энергии и импульса в прямозонном полупроводнике. Эта величина включает плотности состояний как в валентной зоне, так и в зоне проводимости и называется **оптической совместной плотностью состояний**. Взаимно однозначное соответствие между E_2 и ν, выраженное формулой (16.48), позволяет легко связать $\rho(\nu)$ с плотностью состояний $\rho_c(E_2)$ в зоне проводимости с помощью дифференциального соотношения

$$\rho_c(E_2)\mathrm{d}E_2 = \rho(\nu)\mathrm{d}\nu,$$

откуда

$$\rho(\nu) = \frac{\mathrm{d}E_2}{\mathrm{d}\nu}\rho_c(E_2),$$

так что

$$\rho(\nu) = \frac{hm_r}{m_c}\rho_c(E_2). \tag{16.50}$$

С помощью (16.7) и (16.48) окончательно получаем число состояний на единицу объема и на единицу частоты:

$$\rho(\nu) = \frac{(2m_r)^{3/2}}{\pi\hbar^2}\sqrt{h\nu - E_g}, \quad h\nu \geq E_g, \tag{16.51}$$

Оптическая совместная плотность состояний

зависимость которого от энергии фотона графически показана на рис. 16.35. Взаимно однозначное соответствие (16.49) между $\rho_\nu(E_1)$ и ν, совместно с $\rho_\nu(E_1)$ из (16.8), приводит к выражению для $\rho(\nu)$, идентичному (16.51).

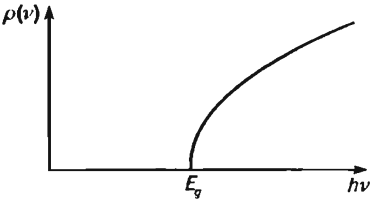


Рис. 16.35. Плотность состояний, с которыми взаимодействует фотон с энергией $h\nu$, растет с увеличением $h\nu - E_g$ как квадратный корень из $h\nu - E_g$

• *Испускание фотона происходит иначе в непрямозонном полупроводнике.* Радиационная рекомбинация электрона и дырки происходит по-другому в непрямозонном полупроводнике. Разница в том, что переходы между состояниями вблизи дна зоны проводимости и потолка валентной зоны, где преимущественно находятся электроны и дырки, соответственно, требуют обмена импульсом, который не может обеспечить испускаемый фотон. Сохранение импульса в этом случае может быть обеспечено за счет участия фононов во взаимодействии. Фононы могут иметь сравнительно большой импульс, при этом их типичные энергии малы (≈0,01—0,1 эВ, см. рис. 16.31), так что переходы с их испусканием или поглощением выглядят как горизонтальные на E—k-диаграмме (рис. 16.36). В итоге импульс сохраняется, но с нарушением правила отбора по k. Поскольку требуется участия трех частиц (электрона, фотона и фонона), вероятность такого процесса весьма мала. Так, например, Si, являющийся непрямозонным полупроводником, имеет существенно меньший коэффициент радиационной рекомбинации, чем прямозонный GaAs (см. табл. 16.4). Таким образом, кремний не является эффективным излучателем света, а GaAs является.

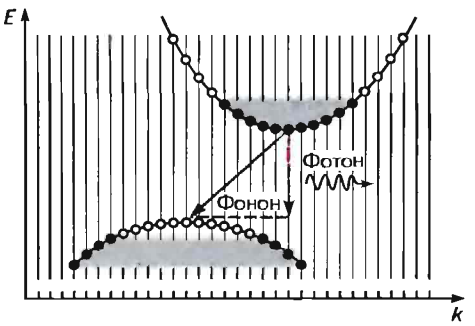


Рис. 16.36. Испускание фотона в непрямозонном полупроводнике. Рекомбинация электрона вблизи дна зоны проводимости с дыркой вблизи потолка валентной зоны требует обмена энергией и импульсом. Энергию может забрать фотон, однако для сохранения импульса требуется один или несколько фононов. Поэтому взаимодействие является многочастичным и переход маловероятен

• *Поглощение фотона в непрямозонном полупроводнике не отличается от прямозонного.* Хотя при поглощении фотона в непрямозонном полупроводнике также требуется сохранение энергии и импульса, оно легко достигается посредством двухступенчатого процесса (рис. 16.37). Сначала электрон возбуждается на верхний энергетический уровень в зоне проводимости в результате вертикального перехода с сохранением k. Затем он быстро релаксирует к дну зоны проводимости в результате процесса, называемого термализацией, при котором его импульс передается фононам. Образующаяся дырка ведет себя аналогично. Поскольку процесс происходит последовательно, он не требует одновременного присутствия трех тел и не является маловероятным. Поэтому кремний — хороший приемник излучения, как и GaAs.

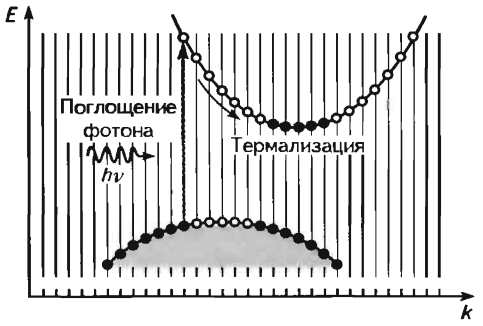


Рис. 16.37. Поглощение фотона в непрямозонном полупроводнике с вертикальным (сохраняющим k) переходом. Фотон рождает возбужденный электрон в зоне проводимости и дырку в валентной зоне. Затем электрон и дырка претерпевают быстрые переходы в состояния с наинизшей и наивысшей энергиями, соответственно, высвобождая избыток энергии в виде фононов. Так как процесс последовательный, его вероятность не мала

16.2.3. Поглощение, испускание и усиление в объемных полупроводниках

Перейдем к определению плотностей вероятности испускания или поглощения фотона с энергией $h\nu$ объемным прямозонным полупроводником. Законы сохранения энергии и импульса в виде (16.33), (16.35) и (16.36) определяют энергии E_1 и E_2 и импульс $\hbar k$ электронов и дырок, с которыми фотон может взаимодействовать. Как будет видно ниже, искомые плотности вероятности определяются тремя факторами:

1) вероятности заселения состояний;
2) вероятности переходов;
3) оптической совместной плотностью состояний.

Вероятности заселения состояний

Условия заселения уровней для испускания и поглощения фотона при переходе между дискретными энергетическим уровнями E_2 и E_1 формулируются так следующим образом:

Условие испускания: Состояние с энергией E_2 в зоне проводимости заполнено (электроном) и состояние с энергией E_1 в валентной зоне свободно (т. е. занято дыркой).

Условие поглощения: Состояние с энергией E_2 в зоне проводимости свободно и состояние с энергией E_1 в валентной зоне занято.

Вероятности реализации этих условий для различных E_2 и E_1 определяются из соответствующих функций Ферми $f_c(E)$ и $f_v(E)$ для зоны проводимости и валентной зоны полупроводника в состоянии квазиравновесия. Так, например, вероятность $f_e(\nu)$ того, что условие испускания выполняется для фотона с энергией $h\nu$, есть произведение вероятностей того, что верхнее состояние заполнено, а нижнее свободно (это независимые события), т. е.

$$\boxed{f_e(\nu) = f_c(E_2)\left[1 - f_v(E_1)\right].} \tag{16.52}$$

Энергии E_2 и E_1 связаны с ν посредством (16.35) и (16.36). Аналогично, вероятность $f_a(\nu)$ выполнения условия поглощения равна

$$\boxed{f_a(\nu) = \left[1 - f_c(E_2)\right] f_v(E_1).} \tag{16.53}$$

Упражнение 16.6

Условие, при котором скорость испускания превышает скорость поглощения

а. Для объемного полупроводника в состоянии теплового равновесия покажите, что $f_e(\nu)$ всегда меньше $f_a(\nu)$, так что скорость испускания фотонов не может превышать скорость их поглощения.

б. Для полупроводника в состоянии квазиравновесия ($E_{f_c} \neq E_{f_v}$) с радиационными переходами между состоянием зоны проводимости с энергией E_2 и состоянием валентной зоны с энергией E_1, имеющими одинаковое значение k, покажите, что испускание более вероятно, чем поглощение, если расстояние между квазиуровнями Ферми больше, чем энергия фотона, т. е.

$$\boxed{E_{fc} - E_{fv} > h\nu.} \tag{16.54}$$

Условие преобладания излучения над поглощением

Что означает это условие для расположения E_{f_c} по отношению к E_c и E_{f_v} по отношению к E_v?

Вероятности переходов

Выполнение условий поглощения/излучения не гарантирует того, что эти процессы действительно произойдут. Они подчиняются вероятностным законам взаимодействия между фотонами и атомными системами, которые подробно рассматривались в подразд. 13.3.1—13.3.3 (см. также упражнение 13.1). В применении к полупроводникам эти законы обычно формулируются для излучения или поглощения в узкой полосе частот между ν и $\nu + d\nu$.

Резюме

Радиационный переход между двумя дискретными энергетическими уровнями E_1 и E_2 характеризуется сечением перехода $\sigma(\nu) = (\lambda^2/8\pi t_{сп})g(\nu)$, где ν — частота; $t_{сп}$ — спонтанное время жизни; $g(\nu)$ — функция формы линии, которая имеет ширину $\Delta\nu$, центральную частоту $\nu_0 = (E_2 - E_1)/h$ и единичную площадь. В полупроводниках время радиационной электронно-дырочной рекомбинации $\tau_{рад}$, которое обсуждалось в подразд. 16.1.4, играет роль $t_{сп}$, так что

$$\sigma(\nu) = \frac{\lambda^2}{8\pi\tau_{рад}} g(\nu). \tag{16.55}$$

• Если заселенности удовлетворяют условиям, необходимым для излучения, то плотность вероятности (на единицу времени) спонтанного испускания фотона в любую из доступных мод в узком частотном интервале между ν и $\nu + d\nu$ выражается как

$$P_{сп}(\nu)d\nu = \frac{1}{\tau_{рад}} g(\nu)d\nu. \tag{16.56}$$

• Если заселенности удовлетворяют условиям, необходимым для излучения, и присутствует поток фотонов со средней спектральной плотностью ϕ_ν (число фотонов в единицу времени на единицу площади и на единицу частоты), то плотность ν (на единицу времени) вероятности вынужденного излучения в узкую полосу частот между ν и $\nu + d\nu$ выражается формулой

$$W_i(\nu)d\nu = \phi_\nu\sigma(\nu)d\nu = \phi_\nu \frac{\lambda^2}{8\pi\tau_{рад}} g(\nu)d\nu. \tag{16.57}$$

• Если заселенности удовлетворяют условиям, необходимым для поглощения, и присутствует поток фотонов со средней спектральной плотностью ϕ_ν, то плотность вероятности поглощения из узкой полосы частот между ν и $\nu + d\nu$ выражается той же формулой (16.57).

Поскольку каждый переход имеет свою центральную частоту ν_0, и поскольку мы рассматриваем набор таких переходов, то явно отметим значение центральной частоты перехода, записывая $g(\nu)$ как $g_{\nu_0}(\nu)$. В полупроводниках формирование однородно уширенной линии $g_{\nu_0}(\nu)$, связанной с парой энергетических уровней, обычно связано с электрон-фононными столкновениями. Поэтому, как правило, форма линии лоренцева [см. (13.45) и (13.48)] с шириной $\Delta\nu \approx 1/\pi T_2$, где время T_2 между электрон-фононными столкновениями порядка пикосекунды. Если, например, $T_2 = 1$ пс, то $\Delta\nu = 318$ ГГц, что соответствует энергии $h\Delta\nu \approx 1{,}3$ мэВ. Радиационное уширение уровней пренебрежимо мало по сравнению со столкновительным.

Полные скорости излучения и поглощения

Для пары энергетических уровней, отстоящих друга от друга на $E_1 - E_2 = h\nu_0$, скорости спонтанного излучения, вынужденного излучения и поглощения фотонов с энергией $h\nu$ [число фотонов/(с · Гц · см^3)] на частоте ν получаются следующим образом. Соответствующая плотность вероятности перехода $P_{сп}(\nu)$ или $W_i(\nu)$ [формулы (16.56) или (16.57)] умножается на соответствующую вероятность заселения $f_e(\nu_0)$ или $f_a(\nu_0)$ [формулы (16.52) или (16.53)] и на плотность состояний, которые могут взаимодействовать с фотоном, $\rho(\nu_0)$ [формула (16.51)]. Полная скорость перехода для всех допустимых частот затем получается интегрированием по ν_0.

Например, скорость спонтанного излучения на частоте ν дается выражением

$$r_{сп}(\nu) = \int \left[\frac{1}{\tau_{рад}} g_{\nu_0}(\nu)\right] f_e(\nu_0)\rho(\nu_0)\,d\nu_0. \tag{16.58}$$

Когда столкновительная ширина $\Delta\nu$ существенно меньше ширины произведения $f_e(\nu_0)\rho(\nu_0)$, что является обычным, $g_{\nu_0}(\nu)$ можно аппроксимировать функцией $\delta(\nu - \nu_0)$, после чего скорость перехода упрощается до

$$r_{сп}(\nu) = \frac{1}{\tau_{рад}} \rho(\nu) f_e(\nu).$$

Скорости вынужденного излучения и поглощения получаются аналогичным образом, что приводит к следующим формулам:

$$r_{сп}(\nu) = \frac{1}{\tau_{рад}} \rho(\nu) f_e(\nu); \tag{16.59}$$

$$r_{изл}(\nu) = \phi_\nu \frac{\lambda^2}{8\pi\tau_{рад}} \rho(\nu) f_e(\nu); \tag{16.60}$$

$$r_{погл}(\nu) = \phi_\nu \frac{\lambda^2}{8\pi\tau_{рад}} \rho(\nu) f_a(\nu). \tag{16.61}$$

Скорости излучения и поглощения

Эти уравнения вместе с (16.51)—(16.53) позволяют рассчитать скорости спонтанного и вынужденного испускания, а также поглощения [число фотонов/(с · Гц · см^3)], обусловленного прямыми межзонными переходами, при наличии средней плотности потока фотонов ϕ_ν [число фотонов/(с · Гц · см^2)]. Произведения $\rho(\nu)f_e(\nu)$ и $\rho(\nu)f_a(\nu)$ аналогичны произведениям функции формы линии и плотности числа атомов на верхнем и нижнем энергетических уровнях $g(\nu)N_2$ и $g(\nu)N_1$, соответственно, использованным в гл. 13—15 при исследовании испускания и поглощения в атомных системах.

Определение вероятностей заселения $f_e(\nu)$ и $f_a(\nu)$ требует знания квазиуровней Ферми E_{f_c} и E_{f_v}. Управляя этими двумя параметрами (например, с помощью внешнего смещающего напряжения, подаваемого на переход), можно менять скорости поглощения и испускания при создании полупроводниковых устройств для выполнения различных функций в фотонике. Уравнение (16.59) является основой описания действия светоизлучающего диода (СИД) — полупроводникового источника света, основанного на спонтанном излучении (см. разд. 17.1). Уравнение (16.60) применимо к полупроводниковым оптическим усилителям и диодным лазерам, работающим на основе вынужденного излучения (см. разд. 17.2—17.4). Уравнение (16.61) описывает действие полупроводниковых приемников, основанное на поглощении фотонов.

Спектральная интенсивность спонтанного излучения при тепловом равновесии

Полупроводник в состоянии теплового равновесия имеет единственную функцию Ферми, так что (16.52) превращается в $f_e(\nu) = f(E_2)[1 - f(E_1)]$. Если уровень Ферми лежит внутри запрещенной зоны на расстоянии по крайней мере в несколько $k_B T$ от ее краев, то можно использовать экспоненциальную аппроксимацию функций Ферми

$$f(E_2) \approx \exp\left[-\frac{E_2 - E_f}{k_B T}\right]; \quad 1 - f(E_1) \approx \exp\left[-\frac{E_f - E_1}{k_B T}\right],$$

в результате чего

$$f_e(\nu) \approx \exp\left[-\frac{E_2 - E_1}{k_B T}\right],$$

т. е.

$$f_e(\nu) \approx \exp\left(-\frac{h\nu}{k_B T}\right). \tag{16.62}$$

Подставляя (16.51) для $\rho(\nu)$ и (16.62) для $f_e(\nu)$ в (16.59), получаем

$$\boxed{r_{\text{сп}}(\nu) \approx D_0 \sqrt{h\nu - E_g} \exp\left(-\frac{h\nu - E_g}{k_B T}\right), \quad h\nu \geq E_g,} \tag{16.63}$$

где D — параметр, который экспоненциально растет с увеличением температуры,

$$D_0 = \frac{(2m_r)^{3/2}}{\pi\hbar^2 \tau_{\text{рад}}} \exp\left(-\frac{E_g}{k_B T}\right). \tag{16.64}$$

Скорость спонтанного излучения (16.63), график которой как функции $h\nu$ приведен на рис. 16.38, содержит два сомножителя: функцию, связанную с плот-

ностью состояний, которая растет как квадратный корень из $h\nu - E_g$, и экспоненциально убывающую функцию $h\nu - E_g$, получающуюся из функции Ферми. Скорость спонтанного излучения можно повысить за счет увеличения $f_e(\nu)$. В соответствии с (16.52), этого можно добиться, целенаправленно выводя материал из состояния теплового равновесия таким образом, чтобы сделать $f_c(E_2)$ большим, а $f_v(E_1)$ малым. Это обеспечивает изобилие как электронов, так и дырок, что является желательным условием для работы СИД, как будет видно в разд. 17.1.

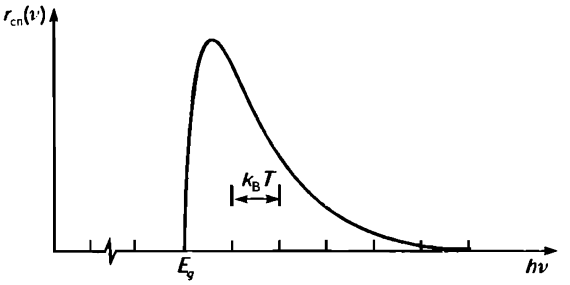


Рис. 16.38. Спектральная плотность скорости прямого межзонного спонтанного испускания $r_{сп}(\nu)$ [число фотонов/(с Гц · см^3)] для полупроводника в состоянии теплового равновесия в зависимости от $h\nu$. Спектр имеет низкочастотную отсечку при $\nu = E_g/h$ и простирается в область высоких частот на ширину порядка $2k_BT/h$

Коэффициент усиления при квазиравновесии

Для определения полного коэффициента усиления $\gamma_0(\nu)$, соответствующего скоростям вынужденного излучения и поглощения (16.60) и (16.61), рассмотрим цилиндр с единичной площадью основания и бесконечно малой длиной dz, ось которого направлена вдоль средней спектральной плотности потока фотонов (см. рис. 14.1). Если $\phi_\nu(z)$ и $\phi_\nu(z + dz)$ — средние спектральные плотности потока фотонов, входящих в цилиндр и выходящих из него, соответственно, то $d\phi_\nu(z) = \phi_\nu(z + dz) - \phi_\nu(z)$ — средняя спектральная плотность потока фотонов, испускаемых внутри цилиндра. Приращение числа фотонов в единицу времени, на единицу частоты и на единицу площади поверхности есть просто плотность числа фотонов, добавляемых единицей объема среды за единицу времени и приходящихся на единицу частоты $[r_{изл}(\nu) - r_{погл}(\nu)]$, умноженная на длину цилиндра dz. Таким образом, $d\phi_\nu(z) = [r_{изл}(\nu) - r_{погл}(\nu)]dz$. Подставляя сюда скорости (16.60) и (16.61), получаем

$$\frac{d\phi_\nu(z)}{dz} = \frac{\lambda^2}{8\pi\tau_{рад}}\rho(\nu)[f_e(\nu) - f_a(\nu)]\phi_\nu(z) = \gamma_0(\nu)\,\phi_\nu(z). \qquad (16.65)$$

Тогда полный коэффициент усиления равен

$$\boxed{\gamma_0(\nu) = \frac{\lambda^2}{8\pi\tau_{рад}}\rho(\nu)f_g(\nu),} \qquad (16.66)$$

Коэффициент усиления

где фактор инверсии Ферми $f_g(\nu)$ дается выражением

$$f_g(\nu) \equiv f_e(\nu) - f_a(\nu) = f_c(E_2) - f_v(E_1), \qquad (16.67)$$

как легко видеть из (16.52) и (16.53), учитывая, что E_1 и E_2 связаны с ν соотношениями (16.48) и (16.49). Сравнивая (16.66) с (14.4), обнаруживаем, что $\rho(\nu)f_g(\nu)$ в полупроводнике играет ту же роль, что и $Ng(\nu)$ в атомной системе. Используя (16.51), можно представить коэффициент усиления в виде

$$\gamma_0(\nu) = D_1\sqrt{h\nu - E_g}\,f_g(\nu), \quad h\nu > E_g, \tag{16.68а}$$

где

$$D_1 = \frac{\sqrt{2}m_r^{3/2}\lambda^2}{h^2\tau_{\text{рад}}}. \tag{16.68б}$$

Знак и спектральный профиль фактора инверсии Ферми $f_g(\nu)$ зависят от квазиуровней Ферми E_{f_c} и E_{f_v}, которые, в свою очередь, зависят от состояния возбуждения носителей заряда в полупроводнике. Как показано в упражнении 16.6, этот фактор положителен (что соответствует инверсии заселенностей и усилению), только когда $E_{f_c} - E_{f_v} > h\nu$. При накачке полупроводника внешним источником энергии до достаточно высокого уровня это условие может быть выполнено и достигается усиление, как мы увидим в разд. 17.2. Таковы физические основы действия полупроводниковых оптических усилителей и диодных лазеров.

Коэффициент поглощения при тепловом равновесии

Полупроводник в состоянии теплового равновесия имеет единственный уровень Ферми $E_f = E_{f_c} = E_{f_v}$, так что

$$f_c(E) = f_v(E) = f(E) = \frac{1}{\exp\left[(E - E_f)/k_BT\right] + 1}. \tag{16.69}$$

Множитель $f_g(\nu) = f_c(E_2) - f_v(E_1) = f(E_2) - f(E_1) < 0$, следовательно, коэффициент усиления $\gamma_0(\nu)$ всегда отрицателен [поскольку $E_2 > E_1$ и $f(E)$ монотонно убывает с ростом E]. Это справедливо при любом положении уровня Ферми E_f. Таким образом, в состоянии теплового равновесия полупроводник, будь он собственным или примесным, всегда ослабляет свет. Коэффициент ослабления (поглощения) $\alpha(\nu) = -\gamma_0(\nu)$ равен

$$\boxed{\alpha(\nu) = D_1\sqrt{h\nu - E_g}\left[f(E_1) - f(E_2)\right],} \tag{16.70}$$

Коэффициент поглощения

где E_1 и E_2 определяются выражениями (16.48) и (16.49) соответственно, а D_1 — выражением (16.58).

Если E_f лежит внутри запрещенной зоны, причем расстояние до ее краев составляет по крайней мере несколько k_BT, то $f(E_1) \approx 1$, а $f(E_2) \approx 0$, так что $f(E_1) - f(E_2) \approx 1$. В этом случае вклад прямых межзонных переходов в коэффициент поглощения равен

$$\alpha(\nu) \approx \frac{\sqrt{2}c^2m_r^{3/2}}{\tau_{\text{рад}}}\frac{1}{(h\nu)^2}\sqrt{h\nu - E_g}. \tag{16.71}$$

График функции (16.71) показан на рис 16.39 для GaAs при следующих значениях параметров: $n = 3{,}6$, $m_c = 0{,}07m_0$, $m_v = 0{,}50m_0$, $m_0 = 9{,}1\ \ 10^{-31}$ кг, уровень легирования таков, что $\tau_{рид} = 0{,}4$ нс (это отличается от приведенного в табл. 16.4 из-за другого уровня легирования), $E_g = 1{,}42$ эВ, а температура такова, что $f(E_1) - f(E_2) \approx 1$. По мере роста температуры $f(E_1) - f(E_2)$ уменьшается и становится меньше единицы, тогда коэффициент поглощения, даваемый выражением (16.70), убывает.

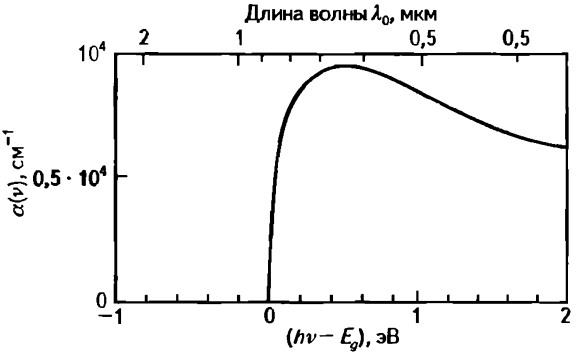


Рис. 16.39. Рассчитанный коэффициент поглощения $\alpha(\nu)$ (см$^{-1}$) за счет прямых межзонных переходов как функция энергии фотона $h\nu$ (эВ) и его длины волны λ_0 (мкм) для GaAs. Эту кривую нужно сравнить с экспериментальным результатом, показанным на рис. 16.32, который включает все механизмы поглощения

В соответствии с (16.71) поглощение вблизи края зоны в прямозонном полупроводнике следует функциональной зависимости $\sqrt{h\nu - E_g}$. Однако резкое включение поглощения при $h\nu = E_g$ является идеализацией. Как видно из рис. 16.32, у прямозонных полупроводников обычно наблюдается экспоненциальный хвост поглощения, известный как **хвост Урбаха**, с характерной шириной $\approx k_B T$, слегка заходящий в запрещенную зону. Это связано с наличием теплового и статического беспорядка в кристалле, возникающего из-за нескольких факторов, включающих поглощение с участием фононов, случайность распределения примеси и вариации состава материала. Поглощение вблизи края зоны в непрямозонных полупроводниках (например, Ge, Si и GaP на рис. 16.32) обычно следует закону $(h\nu - E_g)^2$, а не закону квадратного корня, как у прямозонных полупроводников.

Упражнение 16.7

Длина волны максимального межзонного поглощения

С помощью (16.71) определите длину волны λ_p (в свободном пространстве), при которой коэффициент поглощения полупроводника в состоянии теплового равновесия максимален. Вычислите значение λ_p для GaAs. Не забывайте, что этот результат применим только к поглощению за счет прямых межзонных переходов.

16.2.4. Взаимодействие фотонов с квантово-размерными структурами

Множественные квантовые ямы (МКЯ) и сверхрешетки рассматривались в подразд. 16.1.7. Взаимодействие фотонов с этими структурами имеет определенное сходство с взаимодействием с объемными полупроводниками

(см. подразд. 16.2.1). Важную роль в поглощении и испускании фотонов квантово-размерными структурами играют несколько механизмов:

- межзонные переходы;
- экситонные переходы;
- переходы между подзонами;
- переходы между мини-зонами.

Они проиллюстрированы на рис. 16.40 и обсуждаются далее.

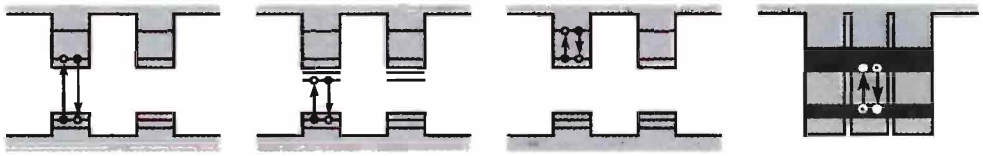


Рис. 16.40. Поглощение и испускание фотонов в структурах с множественными квантовыми ямами:

а — межзонные переходы; *б* — экситонные переходы; *в* — переходы между подзонами; *г* — переходы между мини-зонами в сверхрешетке

• *Межзонные переходы.* Межзонное испускание и поглощение происходит между состояниями валентной зоны и зоны проводимости (рис. 16.40, *а*) в значительной степени так же, как в объемных подупроводниках. Однако из-за квантово-размерных эффектов совместную оптическую плотность состояний (16.51) нужно заменить на (17.40). Межзонные переходы отвечают за работу светоизлучающих диодов с МКЯ, суперлюминесцентных диодов и диодных лазеров (см. рис. 17.21, 17.34 и 17.52 соответственно), а также электропоглощающих модуляторов (см. рис. 20.36).

• *Экситонные переходы.* Одномерное ограничение движения носителей в структурах с МКЯ приводит к увеличению энергии связи экситонов. Это приводит к сильным экситонным переходам даже при $T = 300\,^{\circ}\mathrm{K}$, как схематически показано на рис. 16.40, *б*. Экситонные переходы играют важную роль во многих устройствах на основе квантово-размерных структур, включая электропоглощающие модуляторы с МКЯ.

• *Переходы между подзонами.* Переходы между энергетическими уровнями в пределах одной зоны структуры с МКЯ (рис. 16.40, *в*) называются **переходами между подзонами**. К устройствам, работающим на таких внутризонных переходах, относятся каскадный лазер на квантовых ямах (см. рис. 14.26, *а*) и инфракрасный фотоприемник на квантовых ямах (см. рис. 18.10). В последнем случае поглощение фотона вызывает переход с уровня связанного состояния в сплошной спектр. Пикосекундная динамика носителей в системах с переходами между подзонами обеспечивает очень большую ширину полосы.

• *Переходы между мини-зонами.* В сверхрешетках дискретные уровни МКЯ расширяются в разрешенные мини-зоны, разделенные запрещенными минизонами. Переходы между мини-зонами (рис. 16.40, *г*) играют решающую роль в работе квантовых каскадных лазеров на сверхрешетках (см. рис. 17.50, *б*). Такие переходы, как и межзонные переходы, обладают быстрой релаксацией и

большими нелинейностями и поэтому перспективны для таких приложений, как чисто-оптическая коммутация и демультиплексирование.

16.2.5. Показатель преломления

Возможность управления показателем преломления полупроводника важна при разработке многих устройств фотоники, особенно тех, в которых используются оптические волноводы, интегральная оптика и диодные лазеры. Полупроводниковые материалы обладают дисперсией, так что показатель преломления зависит от длины волны. В самом деле, показатель преломления связан с коэффициентом поглощения $\alpha(\nu)$, поскольку действительная и мнимая части восприимчивости должны удовлетворять соотношениям Крамерса—Кронига (см. подразд. 5.5.2 и разд. Б.1 приложения Б). Групповой и фазовый показатели преломления GaAs, рассчитанные с помощью уравнения Зельмейера и обсуждавшиеся в подразд. 5.5.3, показаны на рис. 16.41. Показатели преломления некоторых одноэлементных и многокомпонентных объемных полупроводников при определенных условиях вблизи длины волны, соответствующей ширине запрещенной зоны, приведены в табл. 16.5.

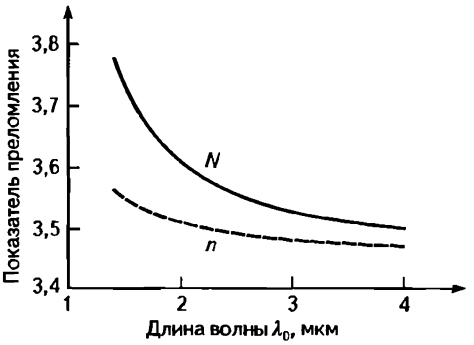


Рис. 16.41. Фазовый n и групповой N показатели преломления для GaAs как функции длины волны λ_0. Результаты расчета по формуле Зельмейера из табл. 5.1

Таблица 16.5. **Показатели преломления некоторых полупроводниковых материалов***

Материал	Показатель преломления	Материал	Показатель преломления
Одноэлементные полупроводники			
Ge	4,0	Si	3,5
Двухкомпонентные полупроводники III—V групп			
AlN	2,2	GaAs	3,6
AlP	3,0	GaSb	4,0
AlAs	3,2	InN	3,0
AlSb	3,8	InP	3,5
GaN	2,5	InAs	3,8
GaP	3,3	InSb	4,2

* Результаты представлены для энергии фотона, близкой к ширине запрещенной зоны материала ($h\nu \approx E_g$) при $T = 300$ °K. Показатели преломления трех- и четырехкомпонентных полупроводников можно получить с помощью линейной интерполяции между показателями преломления их компонент.

Рекомендуемая литература

КНИГИ ПО ФИЗИКЕ, ПРИБОРАМ И ОПТИКЕ ПОЛУПРОВОДНИКОВ

См. также литературу к гл. 17 и 18.

Sze S.M., Ng K.K. Physics of Semiconductor Devices. Wiley, 3rd ed. 2006.

Islam S.S. Semiconductor Physics and Devices. Oxford University Press, 2006.

Neamen D.A. An Introduction to Semiconductor Devices. McGraw-Hill, 2006.

Kasap S., Capper P., eds. Springer Handbook of Electronic and Photonic Materials. Springer-Verlag, 2006.

Olafsen L.J., Biefeld R.M., Wanke M.C., Saxler A.W., eds. Progress in Semiconductor Materials V-Novel Materials and Electronic and Optoelectronic Applications. Materials Research Society Symposium Proceedings Volume 891, Materials Research Society, 2006.

Kuball M., Myers T.H., Redwing J.M., Mukai T., eds. GaN, AlN, InN and Related Materials. Materials Research Society Symposium Proceedings Volume 892, Materials Research Society, 2006.

Moliton A. Optoelectronics of Molecules and Polymers. Springer-Verlag, 2006.

Streetman B.G., Banerjee S. Solid State Electronic Devices. Prentice Hall, 6th ed. 2005.

Klingshirn C.F. Semiconductor Optics. Springer-Verlag, 1997; 2nd ed. 2005.

Kasap S.O. Principles of Electronic Materials and Devices. McGraw-Hill, 3rd ed. 2005.

Barford W. Electronic and Optical Properties of Conjugated Polymers. Oxford University Press, 2005.

Würfel P. Physics of Solar Cells: From Principles to New Concepts. Wiley-VCH, 2005.

Kittel C. Introduction to Solid State Physics. Wiley, 8th ed. 2004.

Seeger K. Semiconductor Physics: An Introduction. Springer-Verlag, 9th ed. 2004.

Haug H., Koch S.W. Quantum Theory of the Optical and Electronic Properties of Semiconductors. World Scientific, 4th ed. 2004.

Singh J. Electronic and Optoelectronic Properties of Semiconductor Structures. Cambridge University Press, 2003.

Landsberg P.T. Recombination in Semiconductors. Cambridge University Press, paperback ed. 2003.

Toyozawa Y. Optical Processes in Solids. Cambridge University Press, paperback ed. 2003.

Basu P.K. Theory of Optical Processes in Semiconductors: Bulk and Microstructures. Oxford University Press, paperback ed. 2003.

Ng K.K. Complete Guide to Semiconductor Devices. Wiley-IEEE, 2nd ed. 2002.

Schäfer W., Wegener M. Semiconductor Optics and Transport Phenomena. Springer-Verlag, 2002.

Sze S.M. Semiconductor Devices: Physics and Technology. Wiley, 2nd ed. 2001.

Fox M. Optical Properties of Solids. Oxford University Press, 2001.

Jackson K.A., Schröter W., eds. Handbook of Semiconductor Technology. Wiley-VCH, 2000.

Brennan K.F. The Physics of Semiconductors with Applications to Optoelectronic Devices. Cambridge University Press, 1999.

Hess K. Advanced Theory of Semiconductor Devices. Wiley-IEEE, 1999.

Fukuda M. Optical Semiconductor Devices. Wiley, 1999.

Adachi S. Optical Properties of Crystalline and Amorphous Semiconductors: Materials and Fundamental Principles. Kluwer, 1999.

Chow W.W., Koch S.W. Semiconductor-Laser Fundamentals: Physics of the Gain Materials. Springer-Verlag, 1999.

Morkoç H. Nitride Semiconductors and Devices. Springer-Verlag, 1999.

Pope M., Swenberg C.E. Electronic Processes in Organic Crystals and Polymers. Oxford University Press, 2nd ed. 1999.

Willardson R.K., Weber E.R., eds. Semiconductors and Semimetals. Vol. 57. Gallium Nitride (GaN) II. J.I. Pankove and T.D. Moustakas, eds. Academic Press, 1999.

Willardson R.K., Weber E.R., eds. Semiconductors and Semimetals. Vol. 50, Gallium Nitride (CaN) I. J.I. Pankove and T.D. Moustakas, eds. Academic Press, 1998.

Rotman S.R., ed. Wide-Gap Luminescent Materials: Theory and Applications. Kluwer, 5th ed. 1997.

Schubert E.F., ed. Delta-Doping of Semiconductors. Cambridge University Press, 1996.

Schubert E.F. Doping in III—V Semiconductors. Cambridge University Press, 1993.

Peyghambarian N., Koch S.W., Mysyrowicz A. Introduction to Semiconductor Optics. Prentice Hall, 1993.

Eisberg R., Resnick R. Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles. Wiley, 2nd ed. 1985.

Ziman J.M. Principles of the Theory of Solids. Cambridge University Press, paperback ed. 1979.

Casey, Jr. H.C., Panish M.B. Heterostructure Lasers. Part A: Fundamental Principles. Academic Press, 1978.

Ashcroft N.W., Mermin N.D. Solid State Physics. Holt, Rinehart and Winston, 1976.

van der Ziel A. Solid State Physical Electronics. Prentice Hall, 3rd ed. 1976.

Pankove J.I. Optical Processes in Semiconductors. Prentice Hall, 1971; Dover, reissued 1975.

КНИГИ ПО КВАНТОВО-РАЗМЕРНЫМ МАТЕРИАЛАМ И НАНОСТРУКТУРАМ

Paiella R., ed. Intersubband Transitions in Quantum Structures. McGraw-Hill, 2006.

Ren S.Y. Electronic States in Crystals of Finite Size: Quantum Confinement of Bloch Waves. Springer-Verlag, 2006.

Grundmann M. The Physics of Semiconductors: An Introduction Including Devices and Nanophysics. Springer-Verlag, 2006.

O'Connell M.J., ed. Carbon Nanotubes. CRC Press, 2006.

Novotny L., Hecht B. Principles of Nano-Optics. Cambridge University Press, 2006.

Rao C.N.R., Govindaraj A. Nanotubes and Nanowires. Royal Society of Chemistry, 2005.

Ivchenko E.L. Optical Spectroscopy of Semiconductor Nanostructures. Alpha Science, 2005.

Prasad P.N. Nanophotonics. Wiley, 2004.

Kalt H., Hetterich M., eds. Optics of Semiconductors and Their Nanostructures. Springer-Verlag, 2004.

Steiner T., ed. Semiconductor Nanostructures for Optoelectronic Applications. Artech, 2004.

Masumoto Y., Takagahara T., eds. Semiconductor Quantum Dots: Physics, Spectroscopy and Applications. Springer-Verlag, 2002.

Harrison P. Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures. Wiley, 2000.

Vasko F.T., Kuznetsov A.V. Electronic States and Optical Transitions in Semiconductor Heterostructures. Springer-Verlag, 1999.

Mitin V., Kochelap V., Stroscio M.A. Quantum Heterostructures: Microelectronics and Optoelectronics. Cambridge University Press, 1999.

Willardson R., Weber E.R., eds. Semiconductors and Semimetals. Vol. 62. Intersubband Transitions in Quantum Wells: Physics and Device Applications I. H.C. Liu, F. Capasso, eds. Academic Press, 1999.

Willardson R.K., Weber E.R., eds. Semiconductors and Semimetals. Vol. 66. Intersubband Transitions in Quantum Wells: Physics and Device Applications II. H.C. Liu, F. Capasso, eds. Academic Press, 1999.

Londergan J.T., Carini J.P., Murdock D.P. Binding and Scattering in Two-Dimensional Systems: Application to Quantum Wires, Waveguides and Photonic Crystals. Springer-Verlag, 1999.

Gaponenko S.V. Optical Properties of Semiconductor Nanocrystals. Cambridge University Press, 1998.

Ruf T. Phonon Raman-Scattering in Semiconductors, Quantum Wells and Superlattices: Basic Results and Applications. Springer-Verlag, 1998.

Yokoyama H., Ujihara K., eds. Spontaneous Emission and Laser Oscillation in Microcavities. CRC Press, 1995.

Bányai L., Koch S.W. Semiconductor Quantum Dots. World Scientific, 1993.

СТАТЬИ

Issue on optoelectronic materials and processing and nanostructures. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 11, № 6, 2005.

Shinada T., Okamoto S., Kobayashi T., Ohdomari I. Enhancing Semiconductor Device Performance Using Ordered Dopant Arrays. Nature. Vol. 437, 2005. P. 1128—1131.

Malliaras G., Friend R. An Organic Electronics Primer. Physics Today. Vol. 58, № 5, 2005. P. 53—58.

Collins G.P. Next Stretch for Plastic Electronics. Scientific American. Vol. 291, № 2, 2004. P. 74—81.

Wu J., Walukiewicz W., Yu K.M., Ager III J.W., Li S.X., Haller E.E., Lu H., Schaff W.J. Universal Bandgap Bowing in Group-III Nitride Alloys. Solid State Communications. Vol. 127, 2003. P. 411—414.

Issue on optoelectronic materials and processing. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, № 4, 2002.

Issue on nanostructures and quantum dots. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, № 5, 2002.

Gammon D., Steel D.G. Optical Studies of Single Quantum Dots. Physics Today. Vol. 55, № 10, 2002. P. 36—41.

Issue on organics for photonics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, № 5, 2001.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Issue on nanostructures and quantum dots. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 3, 2000.

Issue on organic electroluminescence. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 4, № 4, 1998.

Riordan M., Hoddeson L. The Origins of the *p—n*-Junction. IEEE Spectrum. Vol. 34, № 6, 1997. P. 46—51.

Miller D.A.B. Optoelectronic Applications of Quantum Wells. Optics & Photonics News. Vol. 1, № 2, 1990. P. 7—15.

Schmitt-Rink S., Chemla D.S., Miller D.A.B. Linear and Nonlinear Optical Properties of Semiconductor Quantum Wells. Advances in Physics. Vol. 38, 1989. P. 89—188.

Esaki L. A Bird's-Eye View on the Evolution of Semiconductor Superlattices and Quantum Wells. IEEE Journal of Quantum Electronics. Vol. QE-22, 1986. P. 1611—1624.

Задачи

К РАЗДЕЛУ 16.1

1. **Энергии ионизации и радиусы допоров.** Оцените энергии ионизации E_D и боровские радиусы r_1 для доноров в полупроводниковых материалах, перечисленных ниже (см. подразд. 13.1.1 и пример 16.1). В каждом случае прокомментируйте роль тепловых возбуждений и использование объемных диэлектрических постоянных в вычислениях.
   а. Кристаллический кремний с эффективной массой электрона $m_c = 0{,}98m_0$ (см. табл. 16.1) и диэлектрической постоянной $\varepsilon/\varepsilon_0 = 12{,}3$ (см. табл. 16.2).
   б. Кристаллический арсенид галлия с эффективной массой электрона $m_c = 0{,}07m_0$ (см. табл. 16.1) и диэлектрической постоянной $\varepsilon/\varepsilon_0 = 13$ (см. табл. 16.2).
   в. Кристаллический нитрид галлия с эффективной массой электрона $m_c = 0{,}20m_0$ (см. табл. 16.1) и диэлектрической постоянной $\varepsilon/\varepsilon_0 = 6, 25$ (см. табл. 16.2).
   г. Образец легированного Na^+-полиацетилена — подупроводника n-типа на основе сопряженного полимера с эффективной массой электрона $m_c = m_0$ и диэлектрической постоянной $\varepsilon/\varepsilon_0 = 3$. Органические светоизлучающие диоды работают на основе рекомбинационного излучения связанных экситонов.

2. **Уровень Ферми собственного полупроводника.** На основе выражений (16.12) и (16.13) для концентрации носителей заряда в зоне проводимости и в валентной зоне при тепловом равновесии.
   а. Найдите выражение для уровня Ферми E_f собственного полупроводника и покажите, что он попадает точно в середину запрещенной зоны только в случае точного равенства эффективных масс электрона и дырки.
   б. Определите выражение для уровня Ферми примесного полупроводника в зависимости от уровня легирования и уровня Ферми, найденного в пункте а.

3. **Электронно-дырочная рекомбинация в условиях сильной инжекции.** Рассмотрим электронно-дырочную рекомбинацию в условиях сильной инжекции электронно-дырочных пар, настолько что время рекомбинации можно аппроксимировать как $\tau = 1/r\Delta n$, где r — коэффициент рекомбинации; Δn — созданная инжекцией избыточная концентрация носителей. Считая, что источник инжекции R становится равным нулю при $t = t_0$, найдите аналитическое выражение для $\Delta n(t)$, показывающее, что эта зависимость степенная, а не экспоненциальная.

4. **Параметры изгиба для трехкомпонентных полупроводников.** Постоянная решетки трехкомпонентного полупроводникового сплава $A_xB_{1-x}C$ в типичном случае меняется линейно в зависимости от параметра состава x в соответствии с законом Вегарда. С другой стороны, ширина запрещенной зоны E_g обычно нелинейно зависит от x, так что график зависимости ширины запрещенной зоны от постоянной решетки имеет изогнутую форму. Обычно это соотношение моделируется квадратным уравнением

$$E_g^{ABC}(x) = E_g^{AC}x + E_g^{BC}(1-x) - bx(1-x),$$

где b называется **параметром изгиба**. Пользуясь кривыми, представленными на рис. 16.7 и 16.8, определите параметры изгиба для $Al_xGa_{1-x}As$, $Ga_xAs_{1-x}P$, $Al_xGa_{1-x}N$, $In_xGa_{1-x}N$, $Al_xIn_{1-x}N$ и $Hg_xCd_{1-x}Te$. Какое значение имеет параметр изгиба для согласования решетки трехкомпонентного соединения с подложкой?

5. **Уровни энергии в квантовой яме GaAs/AlGaAs.**

а. Нарисуйте диаграмму энергетических зон для квантовой ямы в монокристаллической структуре GaAs/AlGaAs с соблюдением масштаба по оси энергии, когда AlGaAs имеет состав $Al_{0,3}Ga_{0,7}As$. Ширина запрещенной зоны GaAs E_g(GaAs) равна 1,42 эВ; ширина запрещенной зоны AlGaAs, начиная с этой величины, растет линейно, увеличиваясь на ≈12,47 мэВ при увеличении содержания Al на 1 %. Благодаря собственным свойствам обоих материалов, глубина квантовой ямы в зоне проводимости GaAs составляет около 60 % от суммы глубин квантовых ям в зоне проводимости и валентной зоне.

б. Пусть яма в зоне проводимости GaAs имеет глубину, полученную в пункте а, и в точности такие же уровни энергии, как прямоугольная яма конечной глубины, показанная на рис. 16.25, *б*, для которой

$$\left(\frac{mV_0d^2}{2\hbar^2}\right)^{1/2} = 4,$$

где V_0 — глубина ямы. Найдите полную ширину d ямы в зоне проводимости GaAs. Эффективная масса электрона в зоне проводимости GaAs равна $m_c \approx 0{,}07m_0 = 0{,}64 \cdot 10^{-31}$ кг.

К РАЗДЕЛУ 16.2

1. **Справедливость приближения для скоростей излучения/поглощения.** При выводе скорости спонтанного испускания использовалось приближение $g_{\nu_0}(\nu) \approx \delta(\nu - \nu_0)$ в ходе вычисления интеграла

$$r_{\text{сп}}(\nu) = \int \left[\frac{1}{\tau_{\text{рад}}} g_{\nu_0}(\nu)\right] f_e(\nu_0)\rho(\nu_0)\,\mathrm{d}\nu_0.$$

а. Покажите, что это приближение является удовлетворительным для GaAs, путем построения графиков функций $g_{\nu_0}(\nu)$, $f_e(\nu_0)$ и $\rho(\nu_0)$ при $T = 300$ °K и сравнения их ширин. Уширение в GaAs столкновительное, $T_2 \approx 1$ пс.

б. Повторите задание пункта а для скорости поглощения при тепловом равновесии.

2. **Максимальная скорость спонтанного излучения при тепловом равновесии.**

а. Определите энергию фотона $h\nu$, при которой скорость спонтанного излучения за счет прямых межзонных переходов в полупроводниковом материале при тепловом равновесии достигает максимума, если уровень Ферми лежит внутри запрещенной зоны и удален от ее краев на несколько k_BT.

б. Покажите, что эта максимальная скорость [число фотонов/(с · Гц · см^3)] дается выражением

$$r_{\text{сп}}(\nu_p) = \frac{D_0}{\sqrt{2e}}\sqrt{k_B T} = \frac{2(m_r)^{3/2}}{\sqrt{e}\pi\hbar^2\tau_{\text{рад}}}\sqrt{k_B T}\exp\left(-\frac{E_g}{k_B T}\right).$$

в. Как влияет легирование на этот результат?

г. Считая, что $\tau_{\text{рад}} = 0{,}4$ нс, $m_c = 0{,}07m_0$, $m_v = 0{,}50m_0$ и $E_g = 1{,}42$ эВ, найдите максимальную скорость в GaAs при $T = 300$ °K.

3. **Скорость радиационной рекомбинации при тепловом равновесии.**

а. Покажите, что скорость прямого межзонного спонтанного испускания, проинтегрированная по всем испускаемым частотам, определяется выражением

$$\int_0^\infty r_{\text{сп}}(\nu)\,d\nu = D_0\frac{\sqrt{\pi}}{2h}(k_B T)^{3/2} = \frac{(m_r)^{3/2}}{\sqrt{2}\pi^{3/2}\hbar^3\tau_{\text{рад}}}(k_B T)^{3/2}\exp\left(-\frac{E_g}{k_B T}\right),$$

при условии, что уровень Ферми находится внутри запрещенной зоны вдали от краев зон.

Примечание:

$$\int_0^\infty x^{1/2}e^{-\mu x}\,dx = \frac{\sqrt{\pi}}{2}\mu^{-3/2}.$$

б. Сравните этот результат с приближенной интегральной скоростью, полученной умножением максимальной скорости, найденной в задаче 2 к разд. 16.2, на приближенную частотную ширину $2k_B T/h$, показанную на рис. 16.38.

в. Используя (16.15) и полагая феноменологическую равновесную скорость радиационной рекомбинации $r_{\text{рад}}np = r_{\text{рад}}n_i^2$ (число фотонов на 1 см^3/с), введенную в подразд. 16.1.4, равной результату, полученному в пункте а, получите выражение для коэффициента радиационной рекомбинации

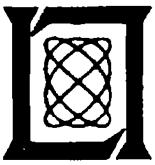
$$r_{\text{рад}} = \frac{\sqrt{2}\pi^{3/2}\hbar^3}{(m_c + m_v)^{3/2}}\frac{1}{(k_B T)^{3/2}\tau_{\text{рад}}}.$$

г. Используйте результат, полученный в пункте в, для нахождения $r_{\text{рад}}$ для GaAs при $T = 300$ °K, $m_c = 0{,}07m_0$, $m_v = 0{,}5m_0$ и $\tau_{\text{рад}} = 0{,}4$ нс. Сравните результат со значением, приведенным в табл. 16.4 ($r_{\text{рад}} = 10^{-10}$ см^3/с).

ГЛАВА 17 ПОЛУПРОВОДНИКОВЫЕ ИСТОЧНИКИ ФОТОНОВ







О создании действующих диодных лазеров почти одновременно сообщили в 1962 г. независимые исследовательские группы **Дженерал Электрик**, **ИБМ** и **Лаборатории Линкольна** Массачусетского Технологического Института.

Свет может излучаться полупроводником в результате электронно-дырочной рекомбинации. Однако материалы, способные излучать такой свет, не светятся при комнатной температуре, так как концентрации термически возбужденных электронов и дырок слишком малы, чтобы породить заметное излучение. С помощью внешнего источника энергии можно вызвать рождение электронно-дырочных пар в достаточном количестве, чтобы за счет процессов их рекомбинации возникло свечение (люминесценция) материала. Удобным способом является прямое смещение p—n-перехода, вызывающее инжекцию электронов и дырок в одну и ту же область пространства в окрестности перехода. Возникающее при этом рекомбинационное излучение называется **инжекционной электролюминесценцией** (см. разд. 13.5).

Светоизлучающий диод (СИД) представляет собой прямосмещенный p—n-переход, изготовленный из прямозонного полупроводникового материала, который излучает свет за счет инжекционной электролюминесценции (рис. 17.1, *а*). Если прямое напряжение превышает некоторое значение, число электронов и дырок в зоне инжекции может стать достаточно большим, так что достигается инверсия заселенностей, в результате чего вынужденное излучение начинает преобладать над поглощением. В этом случае область перехода можно использовать как полупроводниковый оптический усилитель (ПОУ) (рис. 17.1, *б*) или, при обеспечении должной обратной связи, как лазерный диод (ЛД) (рис. 17.1, *в*).

Полупроводниковые источники фотонов в виде СИД или ЛД служат высокоэффективными электронно-фотонными преобразователями. Они стали незаменимыми во многих приложениях, благодаря своим малым размерам, большой яркости, высокой эффективности, надежности, прочности и долговечнос-

ти. СИД видимого диапазона широко используются как индикаторные лампы, в сотовых телефонах, компьютерах, телевизорах, игровых устройствах, информационных дисплеях, лампах-вспышках, вывесках, автомобильных огнях, дорожных знаках, архитектурном освещении и подсветке жидкокристаллических дисплеев. СИД инфракрасного диапазона часто используются для дистанционного управления в таких потребительских товарах, как оптические мыши, наушники, микрофоны и клавиатуры. Ультрафиолетовые СИД применяются в таких областях, как очистка воды, стерилизация хирургических инструментов, дезинфекция оборудования и персонала и невидимая связь. Они полезны также для обнаружения химических и биологических агентов, многие их которых флуоресцируют на специфических длинах волн под действием ультрафиолетового света. Лазерные диоды находят широкое применение в оптических системах хранения данных с высокой плотностью, таких как проигрыватели цифровых видеодисков (DVD), в дальних линиях оптоволоконной связи, а также в системах сканирования, считывания и цветной печати высокого разрешения. Они служат также высокоэффективными источниками оптической накачки для оптоволоконных усилителей и твердотельных лазеров. Особенно удобно, что они легко модулируются за счет управления током инжекции.

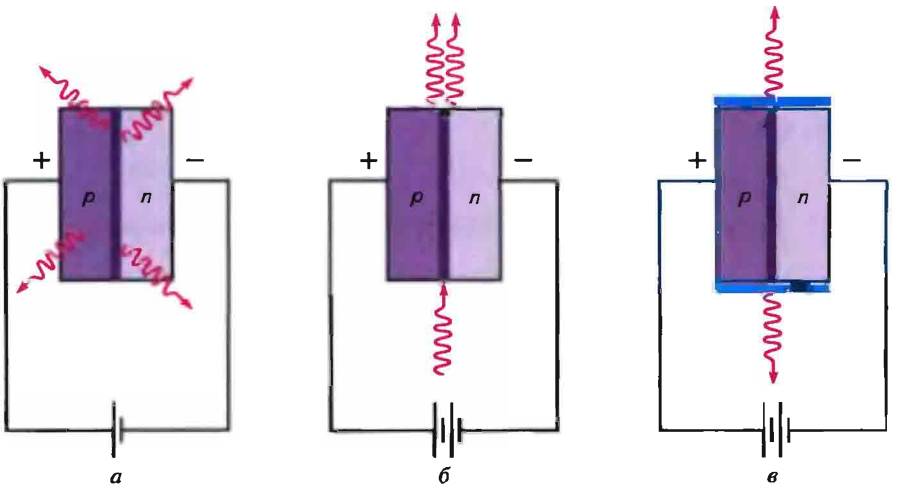


Рис. 17.1. Прямосмещенный полупроводниковый диод на *p*—*n*-переходе, действующий как: *а* — светоизлучающий диод (СИД); *б* — полупроводниковый оптический усилитель (ПОУ); *в* — лазерный диод (ЛД)

О данной главе

Эта глава посвящена изучению светоизлучающих диодов (см. разд. 17.1), полупроводниковых оптических усилителей (см. разд. 17.2), лазерных диодов (см. разд. 17.3), а также кванто-воразмерных лазеров и лазеров с микрорезонаторами (см. разд. 17.4). В качестве основы широко привлекается ма-

териал гл. 16. Теоретическое рассмотрение полупроводниковых оптических усилителей и лазерных диодов — генераторов проводится параллельно анализу лазерных усилителей и генераторов, проведенному в гл. 14 и 15 соответственно.

17.1. СВЕТОИЗЛУЧАЮЩИЕ ДИОДЫ

Электролюминесценция, впервые наблюдавшаяся в 1907 г., — это явление, при котором свет излучается материалом под действием электрического поля (см. разд. 13.5). Инжекционная электролюминесценция лежит в основе действия светоизлучающих диодов — высокоэффективных устройств, способных излучать свет любого цвета. СИД приобрели огромную важность в множестве областей фотоники. Мы рассмотрим теорию инжекционной электролюминесценции в подразд. 17.1.1, характеристики светоизлучающих диодов в подразд. 17.1.2 и типичные материалы и структуры устройств в подразд. 17.1.3.

17.1.1. Инжекционная электролюминесценция

Электролюминесценция при тепловом равновесии

Электронно-дырочная радиационная рекомбинация приводит к излучению света полупроводниковым материалом. Однако при комнатной температуре тепловая концентрация электронов и дырок так низка, что возникающий поток фотонов очень мал.

Пример 17.1

Испускание фотонов GaAs при тепловом равновесии

При комнатной температуре собственная концентрация электронов и дырок в GaAs составляет $n_i \approx 1{,}8 \cdot 10^6$ см$^{-3}$ (см. табл. 16.3). Поскольку коэффициент радиационной электронно-дырочной рекомбинации $r_{рад} \approx 10^{-10}$ см^3/с (как указано в табл. 16.4 для некоторых условий), скорость электролюминесценции $r_{рад}np = r_{рад}n_i^2 \approx 324$ фотона/(см$^3 \cdot$ с), как обсуждалось в подразд. 16.1.4. С учетом того, что ширина запрещенной зоны GaAs $E_g = 1{,}42$ эВ $= 1{,}42 \cdot (1{,}6 \cdot 10^{-19})$ Дж, эта скорость излучения соответствует плотности оптической мощности $324 \cdot 1{,}42 \ (1{,}6 \cdot 10^{19}) \approx 7{,}4 \ 10^{-17}$ Вт/см^3. Слой GaAs толщиной 2 мкм создает интенсивность $I \approx 1{,}5 \cdot 10^{-20}$ Вт/см^2, что пренебрежимо мало. Свет, испускаемый более толстыми слоями GaAs, чем 2 мкм, испытывает повторное поглощение.

Если тепловое равновесие не нарушается, то эту интенсивность невозможно существенно увеличить (или уменьшить) за счет легирования материала. В соответствии с законом действующих масс (16.17) произведение np равно n_i^2, если только материал не легирован слишком сильно, так что скорость рекомбинации $r_{рад}np = r_{рад}n_i^2$ зависит от уровня легирования только через $r_{рад}$. Для

большой скорости рекомбинации требуется *одновременное* наличие большого числа и электронов, и дырок; в полупроводниках n-типа n велико, но p мало, а в полупроводниках p-типа — наоборот.

Электролюминесценция в присутствии инжекции носителей

Скорость испускания фотонов можно существенно увеличить с помощью внешних воздействий, увеличивающих избыточную концентрацию электронно-дырочных пар в материале. Это можно осуществить, например, освещая материал светом, но, как правило, это достигается с помощью прямого смещения диода на p—n-переходе, которое способствует инжекции носителей в область перехода. Процесс иллюстрируется рис. 16.20 и будет объяснен далее в подразд. 17.1.2. Скорость испускания фотона можно вычислить, зная скорость инжекции электронно-дырочных пар R [число пар/(см^3 с)], где R играет роль скорости накачки лазера (см. разд. 14.2). Поток фотонов Ф (число фотонов в секунду), создаваемый в объеме V полупроводникового материала, прямо пропорционален скорости инжекции пар носителей (рис. 17.2).

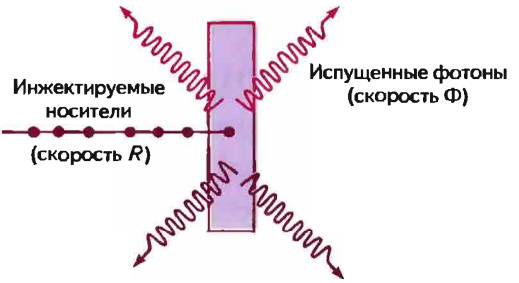


Рис. 17.2. Спонтанное испускание фотонов в результате электронно-дырочной радиационной рекомбинации, которая может происходить в прямосмещенном p—n-переходе

Обозначая равновесные концентрации электронов и дырок в отсутствие накачки через n_0 и p_0, соответственно, представим стационарные концентрации носителей в присутствии накачки в виде $n = n_0 + \Delta n$ и $p = p_0 + \Delta p$ (см. подразд. 16.1.4). Избыточная концентрация электронов Δn в точности совпадает с избыточной концентрацией дырок Δp, поскольку электроны и дырки рождаются парами. Предполагается, что избыточные электронно-дырочные пары рекомбинируют со скоростью $1/\tau$, где τ — полное время рекомбинации за счет радиационных и безызлучательных процессов. В стационарных условиях скорость рождения пар (накачки) находится в точном балансе со скоростью их уничтожения (рекомбинации), так что $R = \Delta n/\tau$. Таким образом, стационарная концентрация избыточных носителей заряда пропорциональна скорости накачки, т. е.

$$\Delta n = R\tau. \tag{17.1}$$

Для достаточно низких скоростей инжекции носителей, как объясняется подразд. 16.1.1, имеем $\tau \approx 1/r(n_0 + p_0)$, где r — коэффициент рекомбинации (радиационной и безызлучательной), так что $R \approx r\Delta n\,(n_0 + p_0)$.

Только радиационная рекомбинация рождает фотоны, и внутренний квантовый выход $\eta_i = r_{рад}/r = \tau/\tau_{рад}$, определяемый формулами (16.28) и (16.30), учитывает тот факт, что только часть рекомбинационных переходов по своей природе являются радиационными. Следовательно, инжекция RV пар носителей в секунду приводит к созданию потока $\Phi = \eta_i RV$ фотонов в секунду, т. е.

$$\boxed{\Phi = \eta_i RV = \eta_i \frac{V\Delta n}{\tau} = \frac{V\Delta n}{\tau_{рад}}.} \tag{17.2}$$

Внутренний поток фотонов Φ пропорционален скорости инжекции пар носителей R и, следовательно, стационарной концентрации избыточных электронно-дырочных пар Δn.

Внутренний квантовый выход η_i играет решающую роль в работе нашего электронно-фотонного преобразователя. Для изготовления СИД (и лазерных диодов) обычно используются прямозонные полупроводники, поскольку у них η_i значительно выше, чем у непрямозонных полупроводников (например, при комнатной температуре для GaAs $\eta_i \approx 0{,}5$, тогда как для Si $\eta_i \approx 10^{-5}$, как показано в табл. 16.4). Внутренний квантовый выход зависит от легирования, температуры и концентрации дефектов в полупроводнике.

Пример 17.2

Излучение при инжекционной электролюминесценции в GaAs

При определенных условиях для GaAs $\tau = 50$ нс и $\eta_i = 0{,}5$ (см. табл. 16.4), так что стационарная избыточная концентрация инжектируемых электронно-дырочных пар $\Delta n = 10^{17}$ см$^{-3}$ обеспечит концентрацию потока фотонов $\eta_i \Delta n/\tau \approx 10^{24}$ фотонов/(см$^3 \cdot$ с). Это соответствует плотности оптической мощности $\approx 2{,}3 \cdot 10^5$ Вт/см^3 для фотонов с энергией, равной ширине запрещенной зоны $E_g = 1{,}42$ эВ. Слой GaAs толщиной 2 мкм создает оптическую интенсивность ≈ 46 Вт/см^2, что в 10^{21} раз больше, чем значение при тепловом равновесии, рассчитанное ранее в примере 17.1. При данных условиях мощность, излучаемая с поверхности площадью 200 × 10 мкм, составляет $\approx 0{,}9$ мВт.

Спектральная интенсивность фотонов электролюминесценции

Спектральную интенсивность инжекционной электролюминесценции можно определить с помощью теории прямого межзонного испускания, развитой в разд. 16.2. Скорость спонтанного излучения $r_{сп}(\nu)$ (число фотонов в единицу времени на единицу объема и на единицу частоты), согласно (16.59), равна

$$r_{сп}(\nu) = \frac{1}{\tau_{рад}} \rho(\nu) f_e(\nu), \tag{17.3}$$

где $\tau_{\text{рад}}$ — время радиационной электронно-дырочной рекомбинации. Совместная оптическая плотность состояний для взаимодействия с фотонами частоты ν, согласно (16.51),

$$\rho(\nu) = \frac{(2m_r)^{3/2}}{\pi\hbar^2}\sqrt{h\nu - E_g}, \tag{17.4}$$

где m_r связано с эффективными массами электронов и дырок соотношением

$$\frac{1}{m_r} = \frac{1}{m_v} + \frac{1}{m_c}$$

[см. (16.47)]; E_g — ширина запрещенной зоны. Условие излучения (16.52) дает

$$f_e(\nu) = f_c(E_2)\left[1 - f_v(E_1)\right], \tag{17.5}$$

что представляет собой вероятность того, что состояние в зоне проводимости с энергией

$$E_2 = E_c + \frac{m_r}{m_c}(h\nu - E)_g \tag{17.6}$$

заполнено, а состояние в валентной зоне с энергией

$$E_1 = E_2 - h\nu \tag{17.7}$$

свободно [см. (16.48) и (16.49) и рис. 17.3].

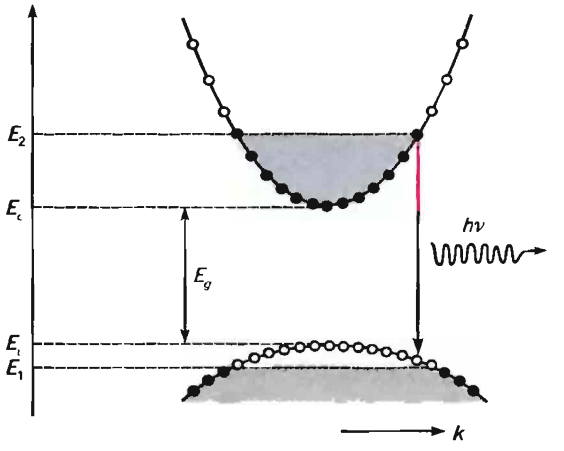


Рис. 17.3. Спонтанное излучение фотона в результате рекомбинации электрона с энергией E_2 и дырки с энергией $E_1 = h\nu$. Переход изображен вертикальной линией, поскольку импульс $h\nu/c$, уносимый фотоном, пренебрежимо мал в масштабе рисунка

Соотношения (17.6) и (17.7) гарантируют сохранение энергии и импульса. Функции Ферми

$$f_c(E) = \frac{1}{\exp\left[\left(E - E_{f_c}\right)/k_{\text{B}}T\right] + 1}; \quad f_v(E) = \frac{1}{\exp\left[\left(E - E_{f_v}\right)/k_{\text{B}}T\right] + 1},$$

входящие в (17.5), с квазиуровнями Ферми E_{f_c} и E_{f_v} относятся к зоне проводимости и валентной зоне, соответственно, в условиях квазиравновесия.

Параметры полупроводника E_g, $\tau_{\text{рад}}$, m_v и m_c, а также температура T определяют спектральное распределение $r_{\text{сп}}(\nu)$ при заданных квазиуровнях Ферми E_{f_c} и E_{f_v}. Последние, в свою очередь, определяются концентрациями электронов и дырок, которые выражаются формулами (16.10) и (16.11),

$$\begin{aligned} \int_{E_c}^{\infty} \rho_c(E) f_c(E)\,\mathrm{d}E &= n = n_0 + \Delta n; \\ \int_{\infty}^{E_v} \rho_v(E)\left[1 - f_v(E)\right]\mathrm{d}E &= p = p_0 + \Delta n. \end{aligned} \tag{17.8}$$

Плотности состояний вблизи краев зоны проводимости и валентной зоны, согласно (16.7) и (16.8), выражаются как

$$\rho_c(E) = \frac{(2m_c)^{3/2}}{2\pi^2\hbar^3}\sqrt{E - E_c}; \quad \rho_v(E) = \frac{(2m_c)^{3/2}}{2\pi^2\hbar^3}\sqrt{E - E_v}, \tag{17.9}$$

где n_0 и p_0 — концентрации электронов и дырок при тепловом равновесии (в отсутствие инжекции); $\Delta n = R\tau$ — стационарная концентрация инжектируемых носителей. Для достаточно слабой инжекции, такой чтобы уровни Ферми лежали внутри запрещенной зоны на расстоянии в несколько $k_{\text{B}}T$ от ее краев, функции Ферми можно аппроксимировать их экспоненциальными хвостами. Тогда поток фотонов спонтанного излучения (проинтегрированный по частотам) получается из скорости спонтанного излучения $r_{\text{сп}}(\nu)$ как

$$\Phi = V\int_0^{\infty} r_{\text{сп}}(\nu)\,\mathrm{d}\nu = \frac{V(m_r)^{3/2}}{\sqrt{2}\pi^{3/2}\hbar^3\tau_{\text{рад}}}(k_{\text{B}}T)^{3/2}\exp\left(\frac{E_{f_c} - E_{f_v} - E_g}{k_{\text{B}}T}\right), \tag{17.10}$$

что легко получить экстраполяцией задачи 3 к разд. 16.2.

Повышение уровня накачки R вызывает рост Δn, что, в свою очередь, смещает E_{f_c} в сторону зоны проводимости (или внутрь нее), а E_{f_v} — в сторону валентной зоны (или внутрь нее). Это приводит к росту вероятности $f_c(E_2)$ того, что состояние в валентной зоне с энергией E_2 заполнено электроном, и вероятности $1 - f_v(E_1)$ того, что состояние в зоне проводимости с энергией E_1 свободно (заполнено дыркой). Полный результат состоит в том, что вероятность

$$f_e(n) = f_c(E_2)[1 - f_v(E_1)],$$

определяющая условие излучения, растет с увеличением R, таким образом увеличивая скорость спонтанного излучения (17.3) и соответствующий поток фотонов Ф (см. ранее).

Упражнение 17.1

Квазиуровни Ферми полупроводника под действием накачки

а. В идеальных условиях при $T = 0\,^{\circ}\mathrm{K}$, когда тепловых носителей заряда нет (рис. 17.4, *а*), покажите, что квазиуровни Ферми связаны с концентрациями инжектируемых электронно-дырочных пар Δn формулами

$$E_{f_c} = E_c + \left(3\pi^2\right)^{2/3} \frac{\hbar^2}{2m_c} (\Delta n)^{2/3}; \tag{17.11а}$$

$$E_{f_v} = E_v - \left(3\pi^2\right)^{2/3} \frac{\hbar^2}{2m_t} (\Delta n)^{2/3} \tag{17.11б}$$

так что

$$E_{f_c} - E_{f_v} = E_g + \left(3\pi^2\right)^{2/3} \frac{\hbar^2}{2m_r} (\Delta n)^{2/3}, \tag{17.11в}$$

где $\Delta n \gg n_0, p_0$. При этих условиях все Δn электронов занимают наинизшие допустимые уровни энергии в зоне проводимости, а все Δp дырок — наивысшие допустимые уровни энергии в валентной зоне. Сравните результат с полученным в упражнении 16.3.

б. Постройте примерные графики функций $f_e(n)$ и $r_{\text{сп}}(n)$ для двух значений Δn. Учитывая действие температуры на функции Ферми, показанное на рис. 17.4, *б*, опишите действие роста температуры на $r_{\text{сп}}(n)$.

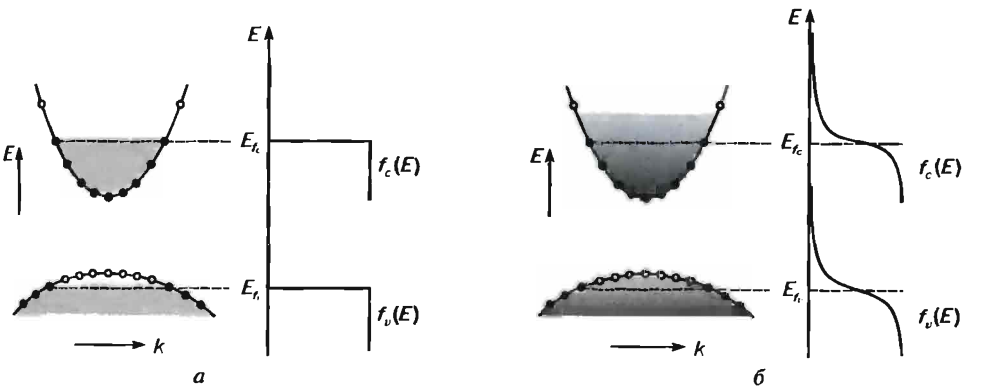


Рис. 17.4. Энергетические зоны и функции Ферми для полупроводника в состоянии квазиравновесия:

а — при $T = 0\,^{\circ}\mathrm{K}$; *б* — при $T > 0\,^{\circ}\mathrm{K}$

Упражнение 17.2

Спектральная интенсивность инжекционной электролюминесценции при слабой инжекции

При достаточно слабой инжекции, такой, что $E_c - E_{f_c} \gg k_B T$ и $E_{f_v} - E_v \gg k_B T$, функции Ферми можно аппроксимировать их экспоненциальными хвостами. Покажите, что скорость люминесценции в этом случае можно выразить как

$$r_{сп}(\nu) = D\sqrt{h\nu - E_g}\,\exp\left(-\frac{h\nu - E_g}{k_B T}\right), \quad h\nu \geq E_g, \tag{17.12a}$$

где D — экспоненциально возрастающая функция расстояния между квазиуровнями Ферми $E_{f_c} - E_{f_v}$,

$$D = \frac{(2m_r)^{3/2}}{\pi\hbar^2 \tau_{рад}} \exp\left(\frac{E_{f_c} - E_{f_v} - E_g}{k_B T}\right). \tag{17.12б}$$

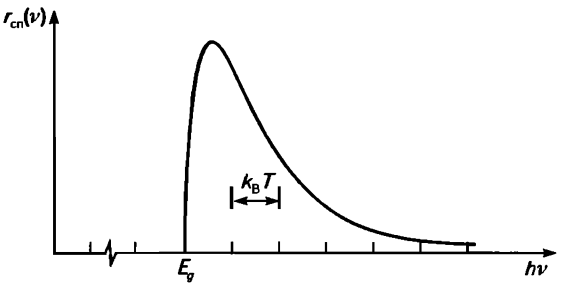


Рис. 17.5. Спектральная плотность скорости прямой межзонной инжекционной электролюминесценции $r_{сп}(\nu)$ [число фотонов/(с · Гц · см^3)] в зависимости от $h\nu$, рассчитанная с помощью (17.12) в условиях слабой инжекции

Спектральная плотность скорости спонтанного излучения показана на рис. 17.5; она имеет в точности такую же форму, как при тепловом равновесии (см. рис. 16.38), однако ее величина увеличена на множитель

$$\frac{D}{D_0} = \exp\left[\frac{E_{f_c} - E_{f_v}}{k_B T}\right],$$

который может быть очень большим при наличии инжекции. При тепловом равновесии $E_{f_c} = E_{f_v}$, поэтому мы возвращаемся к соотношениям (16.63) и (16.64).

Упражнение 17.3

Спектральная ширина электролюминесценции

а. Покажите, что спектральная интенсивность испущенного света, описываемая формулами (17.12), достигает максимального значения на частоте ν_p, определяемой выражением

$$h\nu_p = E_g + \frac{1}{2}k_B T. \tag{17.13}$$

Частота пика

б. Покажите, что полная ширина на уровне полумаксимума спектральной интенсивности

$$\boxed{\Delta\nu \approx 1{,}8\frac{k_B T}{h}.} \qquad (17.14)$$

Спектральная ширина (Гц)

Значение $\Delta\nu$ в активных материалах из многокомпонентных полупроводников может быть больше, чем даваемое (17.14) из-за случайности химического состава; это явление известно как **примесное уширение**.

в. Покажите, что полученная ширина соответствует разбросу длин волн

$$\Delta\lambda \approx 1{,}8\lambda_p^2 \frac{k_B T}{hc},$$

где $\lambda_p = c/\nu_p$. Выражая $k_B T$ в электровольтах, а длину волны в микрометрах, покажите, что

$$\Delta\lambda \approx 1{,}45\lambda_p^2 k_B T. \qquad (17.15)$$

г. Вычислите $\Delta\nu$ и $\Delta\lambda$ при $T = 300$ °K для $\lambda_p = 0{,}8$ и 1,6 мкм.

17.1.2. Характеристики СИД

Как ясно из предыдущего обсуждения, одновременная доступность электронов и дырок существенно увеличивает поток фотонов, спонтанно испущенных полупроводником. Электронов достаточно много в материале *n*-типа, а дырок — в материале *p*-типа, однако для испускания интенсивного света требуется одновременное присутствие и электронов, и дырок в одной области пространства. Это легко достигается в области перехода прямосмещенного *p*—*n*-диода (см. подразд. 16.1.5). Как показано на рис. 17.6, прямое смещение за счет процесса инжекции неосновных носителей заставляет электроны с *n*-стороны и дырки с *p*-стороны перехода собираться в общем переходном слое, где они рекомбинируют и испускают фотоны.

Светоизлучающий диод (СИД) представляет собой *прямосмещенный p—n-переход* с большой скоростью радиационной рекомбинации за счет инжекции неосновных носителей. Полупроводниковый материал обычно *прямозонный*, что гарантирует высокий квантовый выход. В этом разделе мы определим выходную мощность, а также спектральное и пространственное распределение света, испущенного СИД, и выведем выражения для его эффективности, токовой чувствительности и времени отклика.

Внутренний поток фотонов и внутренний квантовый выход

Схема простого диода на *p*—*n*-гомопереходе приведена на рис. 17.7. Постоянный ток инжекции i приводит к увеличению стационарной концент-

рации носителей Δv, что, в свою очередь, вызывает радиационную рекомбинацию в объеме активной области V.

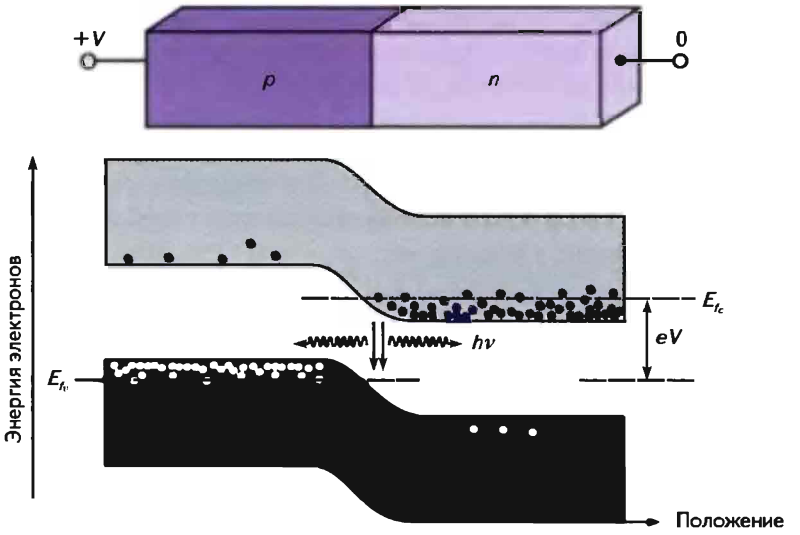


Рис. 17.6. Зонная диаграмма сильно легированного p—n-перехода, к которому приложено большое прямое смещающее напряжение V (сравните с аналогичной зонной диаграммой при меньшем прямом смещении на рис. 16.20). Штриховыми линиями показаны квазиуровни Ферми, разделенные в результате смещения. Одновременное изобилие электронов и дырок внутри области перехода приводит к сильной электронно-дырочной радиационной рекомбинации (инжекционной электролюминесценции)

Поскольку полное число носителей, пересекающих область перехода за 1 с, равно i/e, где e — величина заряда электрона, скорость инжекции (накачки) носителей (число носителей в секунду на см^3) составляет

$$R = \frac{i/e}{V}. \tag{17.16}$$

Уравнение (17.1) дает $\Delta n = R\tau$, что приводит к стационарной концентрации носителей

$$\Delta n = \frac{i/e}{V\tau}. \tag{17.17}$$

В соответствии с (17.2) **внутренний поток фотонов** Φ тогда равен $\eta_i RV$, что с использованием (17.3) дает

$$\boxed{\Phi = \eta_i \frac{i}{e}}. \tag{17.18}$$

Внутренний поток фотонов

Эта простая и интуитивно привлекательная формула определяет рождение фотонов электронами в СИД: часть η_i потока инжектируемых электронов i/e (число электронов в секунду) преобразуется в поток фотонов. Следовательно, **внутренний квантовый выход** η_i есть просто отношение потока создаваемых фотонов к потоку инжектируемых электронов.

Внутренний поток фотонов можно усилить за счет использования СИД с двойной гетероструктурой (см. подразд. 16.1.6) и с квантовыми ямами (КЯ) (см. подразд. 16.1.7) в активной зоне. Выигрыш обусловлен тем, что двойные гетероструктуры обеспечивают большие концентрации носителей, что усиливает радиационную рекомбинацию (сокращает радиационное время жизни $\tau_{рад}$) и, следовательно, повышает внутренний квантовый выход η_i [см. (16.30), (16.31) и (17.18)]. Для достижения максимального η_i ограничивающие слои гетероструктуры должны иметь согласованные решетки с активной областью.

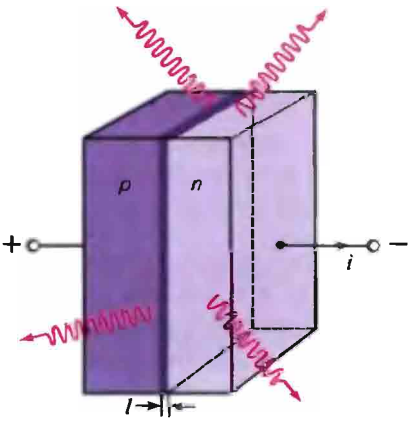


Рис. 17.7. Простой прямосмещенный СИД. Фотоны спонтанно испускаются из области перехода

Узкие квантовые ямы еще сильнее ограничивают движение носителей, что приводит к дальнейшему повышению η_i. Число квантовых ям, используемых в приборах, часто ограничено трудностью заселения их всех. Чтобы достичь хороших рабочих характеристик, важно использовать материалы наивысшего кристаллического качества с минимальным количеством дефектов, что сводит к минимуму количество поверхностей, доступных носителям обоих типов, и, следовательно, минимизирует безызлучательную рекомбинацию.

Коэффициент вывода

Поток фотонов, создаваемых в переходе, излучается равномерно по всем направлениям; однако поток, выходящий из прибора, зависит от направления излучения. Это легко проиллюстрировать, рассматривая поток фотонов, проходящий через плоский слой материала в трех направлениях, обозначенных *A*, *B* и *C* на рис. 17.8.

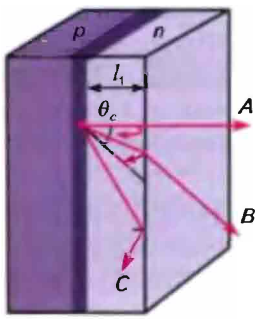


Рис. 17.8. Не весь свет, генерируемый в СИД с плоской поверхностью, может выйти наружу. Луч *A* частично отражается, луч *B* испытывает более сильное отражение. Луч *C* падает под углом больше критического, претерпевает полное внутреннее отражение и остается внутри структуры

• Поток фотонов, распространяющийся вдоль луча A, ослабляется на множитель

$$\eta_1 = \exp(-\alpha l_1), \tag{17.19}$$

где α — коэффициент поглощения материала n-типа; l_1 — расстояние от перехода до поверхности диода. Далее, при нормальном падении через границу полупроводник—воздух проходит только часть излучения

$$\eta_2 = 1 - \frac{(n-1)^2}{(n+1)^2} = \frac{4n}{(n+1)^2}, \tag{17.20}$$

где n — показатель преломления полупроводника [см. формулы Френеля (6.41)]. Для GaAs $n = 3{,}6$, так что $\eta_2 = 0{,}68$. Полный коэффициент пропускания для потока фотонов в направлении A равен $\eta_A = \eta_1\eta_2$.

• Поток фотонов в направлении B проходит больший путь в толще среды и поэтому испытывает большее поглощение. Потери за счет отражения для него также больше. Поэтому $\eta_B < \eta_A$.

• Поток фотонов вдоль направлений, лежащих вне конуса с критическим углом $\theta_c = \arcsin(1/n)$ при вершине (луч C на рис. 17.8), испытывает полное внутреннее отражение в идеальном материале и не выходит из него [см. (1.8)]. Площадь части сферы, ограниченная этим конусом, равна

$$A = \int_0^{\theta_c} 2\pi r \sin\theta \, r\,\mathrm{d}\theta = 2\pi r^2 (1 - \cos\theta_c),$$

а площадь всей сферы $4\pi r^2$. Таким образом, доля излучаемого света, лежащего внутри телесного угла, ограниченного конусом, равна $A/4\pi r^2$, так что

$$\eta_3 = \frac{1}{2}(1 - \cos\theta_c) = \frac{1}{2}\left(1 - \sqrt{1 - \frac{1}{n^2}}\right) \approx \frac{1}{4n^2}. \tag{17.21}$$

Например, из материала с показателем преломления $n = 3{,}6$ только 1,9 % полного генерируемого потока фотонов может выйти наружу. Для параллелепипеда с показателем преломления $n > \sqrt{2}$ отношение энергии изотропного излучения, которое может выйти наружу, к полной энергии генерируемого света равно $3[1 - (1 - 1/n^2)^{1/2}]$, как показано в упражнении 1.7. Однако некоторая доля фотонов, испускаемая вне критического угла, может поглощаться и переизлучаться внутрь этого угла, так что на практике коэффициент η_3 может быть больше, чем рассчитанный по формуле (17.21). Потери и френелевское отражение также должны учитываться для таких лучей.

Эффективность вывода внутренних фотонов из структуры СИД характеризуется **коэффициентом вывода** η_e. Для уменьшения френелевского отражения и повышения η_e могут быть использованы антиотражающие покрытия (см. упражнение 7.3).

Упражнение 17.4

Вывод излучения из СИД с плоской поверхностью

а. Выведите (17.21).

б. Определите критические углы для света, выходящего в воздух из GaAs (n = 3,6), GaN (n = 2,5) и прозрачного полимера (n = 1,5). Рассчитайте долю света, которую можно вывести из материала в этих трех случаях, если пренебречь поглощением и френелевским отражением.

в. Насколько увеличится доля выводимого света, если планарный СИД покрыть прозрачным полимером с коэффициентом преломления n = 1,5? Поглощение и френелевское отражение на границе полупроводник—полимер не учитывать.

г. Определите показатель преломления полимера, обеспечивающий максимальную долю света, выходящего из СИД в воздух, если не учитывать поглощение, но учесть френелевское отражение на границе полупроводник—полимер и полимер—воздух.

Коэффициент вывода можно повысить множеством способов. Один из подходов заключается в подборе геометрии, позволяющей выходить наружу большей части света. Например, сферический купол с точечным источником в центре позволяет выходить всем лучам, хотя они продолжают испытывать френелевское отражение. Несколько других геометрических форм обеспечивают повышение коэффициента вывода по сравнению с параллелепипедом, как показано на рис. 17.9: полусферический купол, цилиндрические структуры (которые имеют кольцо вывода по периметру плюс к конусу вывода через верхнее основание), перевернутые конусы, усеченные перевернутые пирамиды. Однако геометрических форм, требующих сложной многоступенчатой обработки, на практике стараются избегать, поскольку они удорожают производство. Простые СИД с плоскими излучающими поверхностями можно использовать, когда предполагаемый угол зрения мало отличается от нормали или когда свет подается в оптическое волокно, как в телекоммуникационных приложениях.

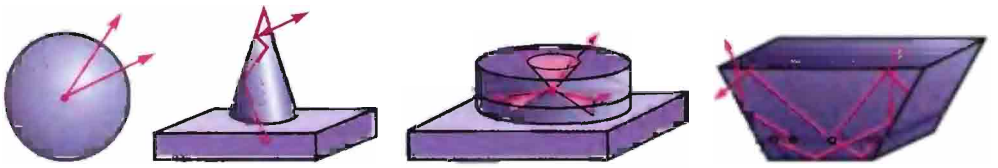


Рис. 17.9. Формы СИД, обеспечивающие повышение коэффициента вывода по сравнению с параллелепипедом

Альтернативный подход состоит в том, чтобы сделать выходную поверхность шероховатой или придать ей некоторую текстуру. Это увеличивает коэффициент вывода за счет того, что лучам, падающим под углами больше критического, разрешается выходить наружу за счет рассеяния, как показано

на рис. 17.10. В действительности текстура поверхности естественно появляется при некоторых условиях роста кристалла.

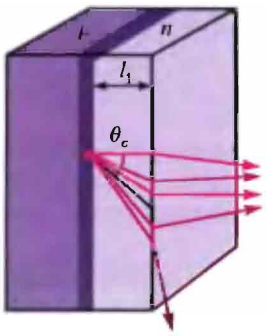


Рис. 17.10. СИД с шероховатой плоской поверхностью дает возможность лучам, падающим вне критического конуса, покидать кристалл, что повышает коэффициент вывода η_c

В СИД с верхней излучающей поверхностью часто используются расширяющие ток слои (называемые также оконными слоями), которые представляют собой прозрачные проводящие полупроводниковые слои, которые расширяют область испускания света за границы области, непосредственно окружающей электрический контакт. Слои, блокирующие ток и предотвращающие его проникновение в активную область ниже верхнего контакта, также могут использоваться для управления излучением света. Геометрия контакта может быть спроектирована так, чтобы добиться максимального пропускания света.

Существует еще множество методов увеличения коэффициента вывода. К ним относится использование отражающих и прозрачных контактов, прозрачных подложек и распределенных брэгговских отражателей (см. гл. 7) между активным слоем и поглощающей подложкой, отражающих свет назад в нужном для вывода направлении. Когда подложка прозрачна, другим удачным решением является компоновка методом перевернутого кристалла, которая позволяет выводить свет через подложку, а не верхнюю поверхность прибора.

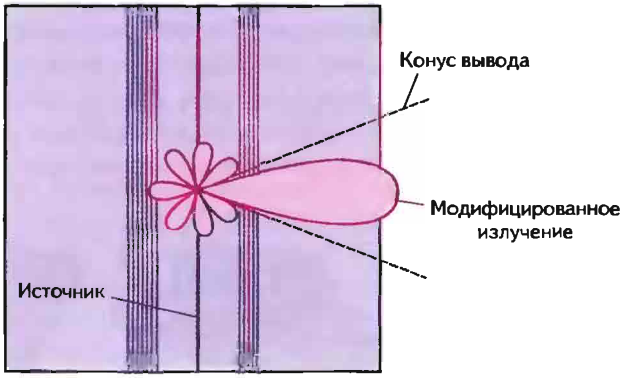


Рис. 17.11. СИД с плоскопараллельным зеркальным микрорезонатором. Два близко расположенных отражателя (слева с коэффициентом отражения 100 %, справа — 50 %) образуют резонатор длиной порядка длины волны, который удерживает свет, выпуская бо́льшую его часть внутри конуса вывода

Коэффициент вывода также можно повысить, направляя свет на поверхность прибора через двумерный фотонный кристалл (см. подразд. 7.3.1), обра-

зованный регулярно расположенными отверстиями диаметром 100—250 нм в токорасширяющем слое.

Еще один способ повышения коэффициента вывода излучения из СИД — использование микрорезонатора (см. подразд. 10.1.1 и разд. 10.4). Пара зеркал (например, распределенных брэгговских отражателей) удерживает свет в области размером порядка длины волны в одном измерении. Как показано на рис. 17.11, это существенно сужает угловое распределение света, так что бóльшая его часть излучается в резонансную моду и в основном находится внутри критического конуса вывода (см. разд. 10.1). В микрорезонаторный СИД можно включить фотонно-кристаллическую структуру, направляющую бóльшую часть остаточного света на поверхность прибора и дополнительно увеличивающую η_e. Дальнейшее обсуждение применения микрорезонаторов для улучшения свойств источников фотонов содержится в подразд. 17.4.2.

Пространственное распределение испущенного света

Пространственное распределение излучения СИД с плоской излучающей поверхностью в дальней зоне аналогично распределению для излучателя Ламберта (источника диффузного света). Интенсивность меняется как $\cos\theta$, где угол θ отсчитывается от нормали к плоскости излучателя; интенсивность падает до половины максимума при $\theta = 60°$. Такое распределение — следствие закона Снелла: лучи отклоняются в сторону от нормали при прохождении границы полупроводник—воздух.

Для многих приложений СИД герметизируют, помещая в прозрачную линзу из эпоксидного компаунда (рис. 17.12). Линзы различной формы по-разному влияют на формирование углового распределения испускаемого света, что схематически показано на рис. 17.13 на примере полусферических и параболических линз. Эпоксидные линзы увеличивают также коэффициент вывода η_e. Линза, показатель преломления которой близок к показателю преломления полупроводника, оптимизирует вывод света через его поверхность в вещество линзы. Форма поверхности линзы подбирается таким образом, чтобы обеспечить максимальный последующий вывод света из линзы в воздух. Значения показателя преломления эпоксидных материалов обычно лежат между таковыми для полупроводников и воздуха, что на практике позволяет увеличить коэффициент вывода света в 2—3 раза.

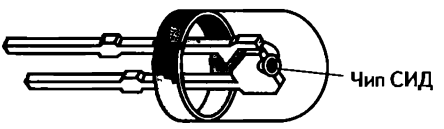


Рис. 17.12. СИД, герметизированный эпоксидным компаундом. Герметизирующая среда предохраняет чип СИД, увеличивает долю выводимого излучения за счет уменьшения перепада показателей преломления, а также служит линзой для формирования пучка

Излучение СИД с торцевым излучателем и лазерных диодов обычно остро направлено и его угловое распределение часто описывается эмпирической функ-

цией $\cos^s \theta$, где $s > 1$. Например, если $s = 10$, то интенсивность спадает до половины максимума при $\theta \approx 21°$

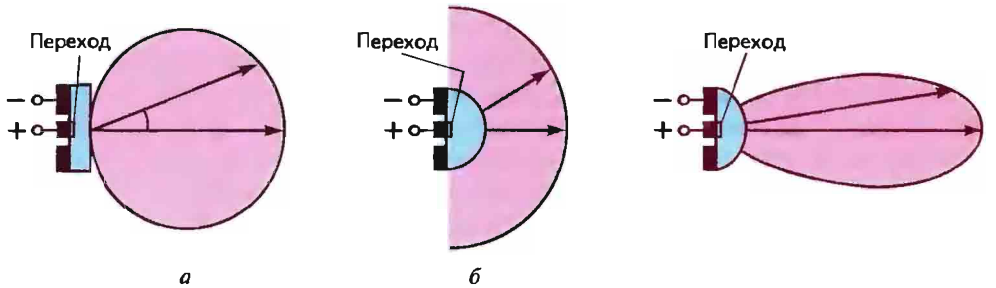


Рис. 17.13. Угловое распределение излучения для СИД с поверхностным излучателем:

а — лабертовское распределение в отсутствие линзы; *б* — распределение, полученное с помощью полусферической линзы; *в* — распределение, полученное с помощью параболической линзы

Выходной поток фотонов и внешний квантовый выход

Выходной поток фотонов Φ_0 (называемый также **внешним потоком фотонов**) связан с внутренним потоком фотонов соотношением

$$\Phi_0 = \eta_e \Phi = \eta_e \eta_i \frac{i}{e}, \tag{17.22}$$

где внутренний квантовый выход связывает внутренний поток фотонов с потоком инжектируемых электронов, а коэффициент вывода η_e указывает, какая часть внутреннего потока фотонов выводится из структуры. Полная эффективность, учитывающая оба эти процесса, характеризуется внешним квантовым выходом $\eta_{вн}$:

$$\boxed{\eta_{вн} \equiv \eta_e \eta_i.} \tag{17.23}$$

Внешний квантовый выход

Тогда выходной поток фотонов (17.22) можно записать в виде

$$\boxed{\Phi_0 = \eta_{вн} \frac{i}{e},} \tag{17.24}$$

Внешний поток фотонов

так что внешний квантовый выход $\eta_{вн}$ есть просто отношение внешнего потока фотонов Φ_0 к потоку инжектируемых электронов i/e. Поскольку скорость накачки, вообще говоря, меняется от точки к точке в области перехода, то же самое происходит и с создаваемым потоком фотонов. Выходная оптическая мощность СИД P_0 непосредственно связана с выходным потоком фотонов, поскольку каждый фотон несет энергию $h\nu$:

$$\boxed{P_0 = h\nu \Phi_0 = \eta_{вн} h\nu \frac{i}{e}.} \tag{17.25}$$

Выходная мощность

Внутренний квантовый выход η_i СИД меняется от 50 до почти 100 %, коэффициент вывода η_e для хорошо спроектированных приборов может достигать 50 %. Таким образом, типичный внешний квантовый выход СИД ниже 50 %.

Как говорилось в подразд. 15.2.1, другой характеристикой эффективности прибора является **коэффициент преобразования мощности (коэффициент полезного действия)**, определяемый как отношение мощности излучаемого света P_0 к потребляемой электрической мощности

$$\eta_c \equiv \frac{P_0}{iV} = \eta_{\text{вн}} \frac{h\nu}{eV}, \tag{17.26}$$

где V — напряжение, приложенное к прибору. Для $h\nu \approx eV$, как в случае некоторых часто встречающихся СИД, получаем $\eta_c \approx \eta_{\text{вн}}$.

Чувствительность

Чувствительность СИД $\mathcal{R}$ определяется как отношение испускаемой оптической мощности P_0 к току инжекции i, т. е. $\mathcal{R} = P_0/i$. С помощью (17.25) получаем

$$\mathcal{R} = \frac{P_0}{i} = \frac{h\nu\Phi_0}{i} = \eta_{\text{вн}} \frac{h\nu}{e}. \tag{17.27}$$

Чувствительность $\mathcal{R}$ (Вт/А) выражается через длину волны λ_0 (мкм) формулой

$$\boxed{\mathcal{R} = \eta_{\text{вн}} \frac{1{,}24}{\lambda_0}}. \tag{17.28}$$

Чувствительность СИД (Вт/А, λ_0 в мкм)

Например, если $\lambda_0 = 1{,}24$ мкм, то $\mathcal{R} = \eta_{\text{вн}}$, Вт/А; если бы $\eta_{\text{вн}}$ равнялось единице, то максимальная оптическая мощность, создаваемая током в 1 мА, была бы равна 1 мВт. Для $\eta_{\text{вн}} = 1/2$ и $\lambda_0 = 1{,}24$ мкм имеем в этом случае $\mathcal{R} = 1/2$ мВт/мА.

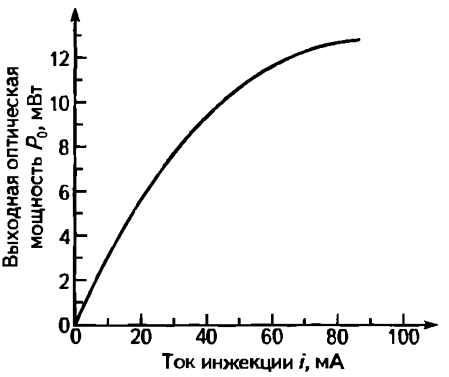


Рис. 17.14. Оптическая мощность на выходе СИД в зависимости от тока инжекции. Этот СИД на InGaN/GaN с КЯ излучает в фиолетовой области спектра на $\lambda_0 = 420$ нм, структура прибора показана на рис. 17.21

В соответствии с (17.25) выходная мощность СИД P_0 пропорциональна току инжекции i. На практике, однако, это соотношение верно только для не слишком больших i. Например, для конкретного прибора, вольт-амперная характеристика которого показана на рис. 17.14, мощность испускаемого света

пропорциональна току инжекции только при условии, что этот ток не превышает 20 мА. В этом диапазоне чувствительность имеет постоянное значение, составляющее около 0,3 мВт/мА, что легко определить по наклону кривой. Для больших значений инжекционного тока линейная зависимость нарушается из-за насыщения, и чувствительность начинает зависеть от тока. Поскольку для данного СИД $\lambda_0 = 0{,}420$ мкм, по формуле (17.28) получаем внешний квантовый выход $\eta_{вн} = 0{,}10$.

Спектральное распределение

Спектральная интенсивность $r_{сп}(\nu)$ света, спонтанно испущенного полупроводником в состоянии квазиравновесия, была определена как функция концентрации инжектируемых носителей Δn в упражнениях 17.2 и 17.3. Эта теория применима к электролюминесцентному излучению СИД, в котором условия квазиравновесия устанавливаются за счет инжекции тока в p—n-переход.

В условиях слабой накачки, при которых квазиуровни Ферми остаются внутри запрещенной зоны на расстоянии по крайней мере в несколько $k_B T$ от ее краев, спектральная интенсивность достигает своего пикового значения на частоте

$$\nu_p = \frac{E_g + k_B T/2}{h}$$

(см. упражнение 17.3). В соответствии с (17.14) и (17.15) ширина кривой спектральной интенсивности на уровне половины максимума равна $\Delta n \approx 1{,}8 k_B T/h$ ($\Delta n = 10$ ТГц при $T = 300$ °K) и не зависит от ν. Однако в шкале длин волн эта ширина зависит от λ,

$$\boxed{\Delta\lambda \approx 1{,}45 \lambda_p^2 k_B T,} \qquad (17.29)$$

Спектральная ширина (мкм)

где $k_B T$ выражено в электронвольтах, а длина волны в микрометрах, причем $\lambda_p = c/\nu_p$.

Зависимость $\Delta\lambda$ от λ_p^2 хорошо видна на рис. 17.15 (см. цв. вклейку), где показаны наблюдаемые спектральные интенсивности излучения ряда СИД, работающих в УФ (показаны пурпурным цветом) и видимой областях спектра. AlN имеет наибольшую ширину запрещенной зоны среди III-нитридов и излучает на 210 нм; AlGaN, как правило, используется в среднем и ближнем ультрафиолетом диапазоне; InGaN — материал, который выбирают для получения фиолетового, синего и зеленого излучения, а AlInGaP — желтого, оранжевого и красного. Для СИД, работающих в ИК-области спектра, типичные спектральные интенсивности приведены на рис. 17.6; эти приборы обычно делаются на основе InGaAsP. Спектральная ширина растет примерно как λ_p^2, что соответствует формуле (17.29). Однако дальнейшее увеличение спектральной ширины может быть вызвано примесным уширением, как это хорошо видно из спектральной кривой для зеленого СИД. Например, если $\lambda_p = 1$ мкм, то при $T = 300$ °K (17.29) дает $\Delta\lambda \approx 36$ нм.

Время отклика

Время отклика СИД, используемых для освещения, обычно ограничено постоянной времени эквивалентного *RC*-контура, поскольку площадь перехода и, соответственно, его емкость велика. Время отклика СИД, используемых в системах связи, напротив, ограничено временем жизни инжектируемых неосновных носителей, ответственных за радиационную рекомбинацию. При достаточно малых скоростях накачки R процесс инжекции/рекомбинации описывается линейным дифференциальным уравнением первого порядка (см. подразд. 16.1.4) и, следовательно, откликом на синусоидальные сигналы. Экспериментальное определение наибольшей частоты, при которой СИД можно эффективно модулировать, легко достигается путем измерения мощности света на выходе в ответ на синусоидальные электрические токи различной частоты. При токе инжекции вида $i = i_0 + i_1 \cos(\Omega t)$, где i_1 достаточно мало, чтобы выходная оптическая мощность P линейно зависела от тока инжекции, величина P меняется как $P = P_0 + P_1 \cos(\Omega t + \varphi)$.

Соответствующая передаточная функция, определяемая как

$$H(\Omega) = \frac{P_1}{i_1} \exp(j\varphi),$$

принимает вид

$$H(\Omega) = \frac{\mathcal{R}}{1 + j\Omega\tau}, \tag{17.30}$$

характерной для *RC*-цепи. Время установления равно τ (с), а ширина полосы на уровне 3 дБ $B = 1/2\pi\tau$ (Гц). Большая ширина полосы достигается уменьшением времени установления τ, вклады в которое дают радиационное $\tau_{рад}$ и безызлучательное $\tau_{без}$ времена жизни через соотношение

$$\frac{1}{\tau} = \frac{1}{\tau_{рад}} + \frac{1}{\tau_{без}}.$$

Однако сокращение $\tau_{без}$ приводит к нежелательному уменьшению внутреннего квантового выхода $\eta_i = \tau/\tau_{рад}$. Поэтому может быть желательно максимизировать произведение внутреннего квантового выхода на ширину полосы $\eta_i B = 1/2\pi\tau_{рад}$, а не только ширину полосы саму по себе. Для этого требуется уменьшение только радиационного времени жизни $\tau_{рад}$ без изменения $\tau_{без}$, что может быть достигнуто тщательным подбором полупроводникового материала и уровня легирования. Типичные времена установления сигнала СИД лежат в интервале от 1 до 50 нс, что соответствует ширине полосы до нескольких сотен мегагерц.

Электрические схемы

СИД обычно питается от источника тока, как показано на рис. 17.16, *а*. Часто для этого используется источник постоянного напряжения с последовательно включенным сопротивлением, как показано на рис. 17.16, *б*. Излучае-

мый свет легко модулировать, просто модулируя ток инжекции. Аналоговая и цифровая модуляция проиллюстрированы на рис. 17.16, *в* и *г* соответственно. Работа источников питания СИД может быть улучшена путем добавления схем, регулирующих ток смещения, согласующих импедансы и обеспечивающих нелинейную компенсацию для ограничения максимального тока. Флуктуации интенсивности испускаемого света можно стабилизировать, отслеживая их фотоприемником, выходной сигнал которого используется в качестве обратной связи для управления током инжекции.

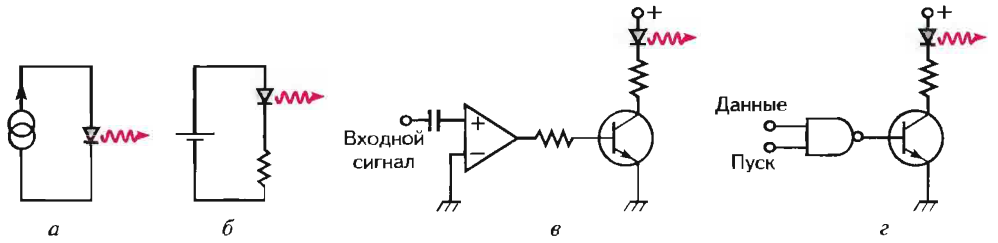


Рис. 17.16. Различные схемы питания СИД:

а — идеальный источник постоянного тока; *б* — источник постоянного напряжения с последовательным сопротивлением; *в* — транзисторная схема управления током инжекции для аналоговой модуляции испускаемого света; *г* — транзисторное переключение тока инжекции СИД для цифровой модуляции испускаемого света

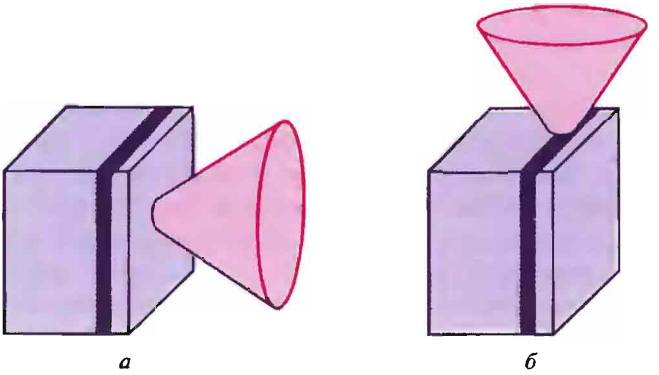
Для применения в освещении архитектурных сооружений множество СИД какого-либо цвета обычно соединяют последовательно и возбуждают током с широтно-импульсной модуляцией (ШИМ) от задающего транзистора. Уровень света определяется средним током, протекающим через СИД, который, в свою очередь, зависит от рабочего цикла тока при ШИМ. Наборы СИД базовых цветов (например, красного, зеленого и синего) используются для синтеза света произвольного цвета, включая белый. Адресуемый микропроцессор можно использовать для управления относительным уровнем света, генерируемого СИД разных цветов, обеспечивая общие цвет и яркость, меняющиеся заданным образом в зависимости от времени и положения. Набор таких осветительных элементов можно объединить в осветительную сеть.

17.1.3. Материалы и структуры устройств

Фотоника пережила революцию в 1950-е гг., когда научились выращивать монокристаллические двухкомпонентные полупроводники III—V групп — соединения, не встречающиеся в природе. Многие из этих соединений имеют прямозонную структуру и, следовательно, дают высокие значения внутреннего квантового выхода. Источники фотонов, сделанные из этих материалов, имеют также длительный срок службы, в отличие от изготовленных из сплавов элементов II—IV групп. В 1962 г. GaAs был впервые использован для изготовления СИД и ЛД (см. начало данной главы).

Современное производство СИД связано с трех- и четырехкомпонентными соединениями III—V групп, особенно InGaAsP, AlInGaP и AlInGaN. С помощью СИД легко генерируется свет высокой яркости всех цветов радуги, от инфракрасного до ультрафиолетового (см. рис. 17.15 и 17.17, см. цв. вклейку).

СИД может иметь конфигурацию поверхностного или торцевого излучателя (рис. 17.18). СИД с поверхностным излучателем испускает свет с грани прибора, параллельной плоскости активной области. СИД с торцевым излучателем испускает свет с торца активного слоя.



а *б*

Рис. 17.18. СИД с поверхностным излучателем (*а*). СИД с торцевым излучателем (*б*)

Перейдем к краткому описанию основных классов III—V материалов и схем некоторых структур СИД.

GaAs

Первый материал классов III—V, который стал играть важную роль в фотонике — это GaAs. Этот двухкомпонентный прямозонный полупроводник был использован в 1962 году для изготовления первого лазерного диода, который излучал на длине волны $\lambda_0 = 0{,}873$ нм вблизи длины волны, соответствующей ширине запрещенной зоны λ_g. СИД из GaAs был побочным продуктом разработки этого лазерного диода. Вскоре после этого электролюминесцентные свойства и лазерный эффект на длинах волн вблизи λ_g были обнаружены у нескольких других двухкомпонентных прямозонных III—V-полупроводников, выращенных эпитаксией из паровой или жидкой фазы: GaSb ($\lambda_g = 1{,}70$ мкм), InP ($\lambda_g = 0{,}919$ мкм), InAs ($\lambda_g = 3{,}44$ мкм), InSb ($\lambda_g = 7{,}29$ мкм).

GaAsP

Длина волны, соответствующая ширине запрещенной зоны трехкомпонентного полупроводника $GaAs_{1-x}P_x$, сдвигается в видимый диапазон с увеличением молярного содержания фосфора, обеспечивая излучение в красной части спектра (см. рис. 16.7, *а*). При использовании модификаций этих материалов, легированных азотом, GaAsP : N и GaP : N, структура становится непрямозонной в красной области, тем не менее достигается испускание в оранжевой,

желтой и зеленой областях. Азотные примеси (вместо азота иногда используют цинк и кислород) включаются в основной материал в жестко локализованных положениях и способны принимать на себя значительные изменения импульса, связанные с непрямыми переходами. Однако внешний квантовый выход СИД на GaAsP обычно мал (0,02—0,5 %), частично из-за несовпадения постоянных решетки с подложкой из GaAs. Тем не менее СИД из GaAs, GaAsP, GaAsP : N и GaP : N недороги в производстве и поэтому продолжают использоваться в приложениях, не требующих высокой яркости, таких как пульты дистанционного управления для бытовых приборов и индикаторные лампы.

InGaAsP

Добавление индия уменьшает ширину запрещенной зоны GaAsP. Четырехкомпонентный полупроводник $In_{1-x}Ga_xAs_{1-y}P_y$ — функционально гибкий сплав, широко используемый в ближней инфракрасной области спектра. Ширина его запрещенной зоны за счет изменения состава может существенно перестраиваться [0,549 мкм (GaP) $\leq \lambda_g \leq$ 3,44 мкм (InAs)], при этом согласование решетки с подложкой из InP сохраняется при разумном подборе композиционных отношений x и y (участок, покрытый точками на рис. 16.7, *а*). Однако лишь часть этого диапазона обладает достоинством прямозонной структуры. InGaAsP применяется при изготовлении СИД для систем связи ближнего действия с умеренной битовой скоростью передачи, работающих на длинах волн вблизи $\lambda_0 = 1330$ нм (рис. 17.19). Системы дальней связи с высокой битовой скоростью передачи обычно работают в окрестности $\lambda_0 = 1550$ нм и используют лазерные диоды, а не СИД, поскольку хорошо коллимированный свет от лазерного диода намного легче вводится в одномодовое волокно (см. гл. 24). Дешевые СИД на $In_xGa_{1-x}As$ используются при изготовлении разнообразных потребительских изделий.

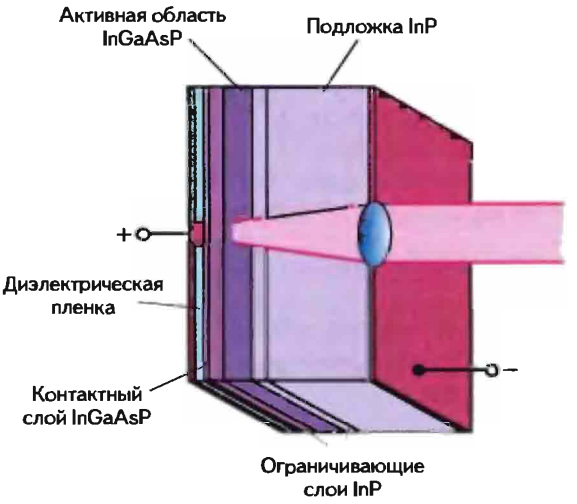


Рис. 17.19. СИД на InGaAsP с поверхностным излучателем, разработанный для применения в системе волоконно-оптической связи, работающей на длине волны 1,3 мкм. Активная область согласована по решетке с подложкой из InP; прибор монтируется методом перевернутого кристалла, так что свет выходит через подложку. Интегрированная в прибор линза коллимирует свет для ввода в волокно. В СИД для систем связи свет излучается с одной поверхности

AlGaAs

Добавление фосфора в GaAs увеличивает ширину запрещенной зоны. Точно такой же эффект производится добавлением алюминия. Как и $GaAs_{1-x}P_x$, трехкомпонентный сплав $Al_xGa_{1-x}As$ является прямозонным полупроводником в красной и ближней инфракрасной областях, допускающим перестройку за счет изменения состава. Однако, в отличие от GaAsP, он обладает тем достоинством, что его решетка согласуется с GaAs при любой молярной доле алюминия (см. рис 16.7, *а*), поэтому этот материал может служить источником красного света высокой яркости. Поскольку $Al_xGa_{1-x}As/GaAs$ — структуры с множественными квантовыми ямами страдают склонностью к образованию неоднородного распределения носителей в активной области, СИД часто изготавливают с использованием двойной гетероструктурной конфигурации вида $Al_xGa_{1-x}As/Al_yGa_{1-y}As$, у которой состав барьеров и ямы различен. Отрицательным качеством этого материала является ограниченный срок службы, связанный с окислением и коррозией материала с течением времени при достаточно высоком содержании алюминия.

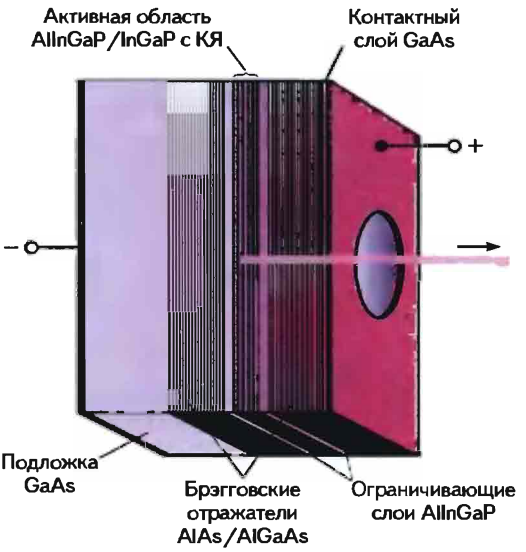


Рис. 17.20. СИД на AlInGaP/InGaP с КЯ и микрорезонатором, излучающий с поверхности на 650 мкм для использования в системе ближней связи на пластиковом оптоволокне. Излучение происходит с верхней поверхности, поскольку подложка из GaAs непрозрачна. Распределенные брэгговские отражатели сделаны из слоев AlAs/AlGaAs, причем содержание алюминия достаточно велико для того, чтобы пропускать свет на 650 нм. Линза улучшает условия ввода излучения в волокно

AlInGaP

Четырехкомпонентный полупроводник $(Al_xGa_{1-x})_yIn_{1-y}P$ — прямозонный материал, у которого интервал изменения ширины запрещенной зоны покрывает значительную часть ближнего инфракрасного и видимого спектра (затененная область на рис. 16.7, *а*). Согласование решеток с GaAs достигается для составов в пределах $(Al_xGa_{1-x})_{0.5}In_{0,5}P$. Согласованное по решетке тройное соединение $In_{0,5}Ga_{0,5}P$ имеет ширину запрещенной зоны, соответствующую длине волны 650 нм, что используется в таких устройствах, как лазерные

указки и DVD-плейеры. AlInGaP — наиболее подходящий материал для приложений, требующих высокой яркости, таких как светофоры и вывески. Он излучает свет красного, оранжевого, желто-оранжевого (янтарного) и желтого цветов. Квантовый выход повышается с использованием сращенной с кремниевой пластиной подложки из прозрачного GaP вместо GaAs, активной области с множественными квантовыми ямами и микрорезонаторов, обеспечивающих сужение полосы и повышение направленности излучения. СИД на AlInGaP/InGaP часто используются в системах связи на пластиковом волокне, работающих в области 600—650 нм (рис. 17.20).

GaN

GaN — двухкомпонентный прямозонный полупроводник с длиной волны λ_g = 0,366 мкм, попадающей в ближнюю ультрафиолетовую область спектра. Будучи сравнительным новичком в фотонике, GaN, возможно, является одним из самых важных материалов. Его можно выращивать методами молекулярно-пучковой эпитаксии, химического осаждения из паровой фазы методом разложения металлоорганических соединений или эпитаксии из газовой фазы гидрида. Хотя электролюминесценция этого материала наблюдалась еще в 1971 г., только в 1992 г. был изготовлен первый СИД на *p—n*-гомопереходе в GaN. Материал, как правило, выращивают на сапфировой подложке, которая имеет существенное рассогласование решетки с GaN. Однако, в отличие от арсенидных и фосфидных III—V-соединений, GaN хорошо переносит высокую концентрацию дислокаций, так что указанное рассогласование оказывается несущественным. GaN является предшественником важнейших соединений InGaN, AlGaN и AlInGaN, так же как GaAs — предшественником InGaAsP и AlInGaP.

InGaN

Трехкомпонентный полупроводник $In_xGa_{1-x}N$ — прямозонный материал с шириной запрещенной зоны, соответствующей 366 нм (GaN) $\leq \lambda_g \leq$ 1,61 мкм (InN). Однако вырастить InGaN с высоким содержанием InN тяжело, так как в ходе этого процесса образуются кластеры InN. Из-за этого происходит значительное примесное уширение спектра зеленого СИД, показанного на рис. 17.15. InGaN — подходящий материал для СИД высокой яркости в диапазоне длин волн 366 $\leq \lambda_g \leq$ 580 нм, включающем ближнюю ультрафиолетовую, фиолетовую, синюю и зеленую области спектра (см. рис 16.7, *б*). Таким образом, этот III-нитридный сплав дополняет AlInGaP, покрывающий красную, оранжевую и желтую области. Как и в случае AlInGaP, повышение квантового выхода достигается с использованием структур GaN/InGaN с КЯ, показанных на рис. 17.21. Число квантовых ям часто ограничено 3—5 из-за ограничений возможности их заселения, накладываемых длиной диффузии дырок. Подложкой обычно служит GaN на сапфире. Для улучшения рабочих характеристик

используются резонаторные структуры. Еще одна интересная конфигурация включает наборы квантовых точек, упорядоченных в результате самосборки в процессе выращивания.

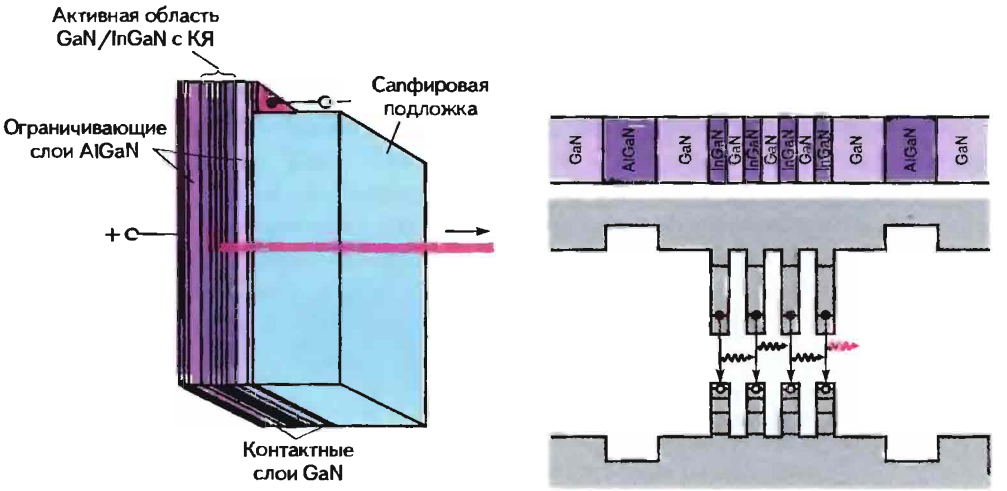


Рис. 17.21. СИД на GaN/InGaN с КЯ, излучающий с поверхности на $\lambda_0 = 420$ нм фиолетовой области спектра. Активная область включает барьеры из GaN шириной 5 нм и четыре ямы из $In_xGa_{1-x}N$ шириной по 2,5 нм (для простоты влияние характерных внутренних электрических полей III-нитрид-полупроводников не показано). Свет выводится через подложку GaN на сапфире, которая прозрачна на 420 нм

AlGaN

$Al_xGa_{1-x}N$ также является трехкомпонентным III-нитридным прямозонным полупроводником, однако длина волны, соответствующая ширине его запрещенной зоны, лежит в интервале 200 нм (AlN) $\leq \lambda_g \leq$ 366 нм (GaN) (см. рис. 16.7, *б*), так что его излучение покрывает среднюю и ближнюю ультрафиолетовую области ($200 \leq \lambda_0 \leq 390$ нм). Одна из трудностей, мешающих изготовлению приборов для работы на меньших длинах волн, — уменьшение проводимости *p*-типа с ростом концентрации AlN, приводящее к недостатку дырок, доступных для рекомбинации в активной зоне. Тем не менее СИД на AlN успешно работают на 210 нм.

Как и у InGaN, квантовый выход СИД повышается посредством использования в активной области двойных гетероструктур, квантовых ям или КЯ со слоями $Al_xGa_{1-x}N/Al_yGa_{1-y}N$. Шаблоны AlGaN/AlN/сапфир служат прозрачными согласованными по решетке подложками для ультрафиолетовых излучателей на основе AlGaN. Использования GaN в качестве подложки избегают ввиду его поглощения на длинах волн ниже 366 нм.

AlInGaN

Как ясно из предыдущего обсуждения, такие тройные III-нитридные соединения, как InGaN и AlGaN, весьма подходят для изготовления источников, покрывающих видимую и ультрафиолетовую области спектра. Однако, четырехкомпонентный полупроводник $(Al_xIn_yGa_{1-x-y})N$ имеет перед ними преимущество в том, что его решетка хорошо согласуется с подложкой GaN при подходящих значениях x и y (см. рис. 16.7, *б*), что способствует повышению квантового выхода. Аналогичная ситуация уже имела место для пар AlInGaP—GaAs и InGaAsP—InP. СИД на AlInGaN на подложке GaN с согласованными постоянными решетки часто используются в области длин волн от 366 нм (длина волны GaN) примерно до 250 нм (длина волны AlInN, постоянная решетки которого согласована с GaN). Квантовые ямы AlInGaN/InGaN/AlInGaN служат активными зонами. Включение индия в AlGaN дает увеличение внутреннего квантового выхода. AlInGaN используется также в качестве прозрачного контактного слоя.

СИД белого света

Подходящая смесь красного, зеленого и синего света воспринимается как белый свет. Для получения белого света от СИД применяются два основных подхода. СИД ближнего УФ и синего диапазона, изготовленные из III-нитридов, можно использовать для освещения фосфоров, которые затем генерируют различные другие цвета посредством фотолюминесценции (см. подразд. 13.5.2 и рис. 17.22). Альтернативный способ состоит в смешении света от СИД с разными составами с целью синтеза заданного цвета. Хотя внутренний квантовый выход в зеленой области обычно ниже, чем в других областях спектра, последний подход позволяет получать белый свет с отличной цветопередачей. СИД вытесняют лампы накаливания в качестве средств освещения домов и рабочих мест и все более широко используются в архитектуре. Преимущество СИД перед лампами накаливания состоит в высоком коэффициенте полезного действия и светоотдаче, долговечности, низкой стоимости и компактности.

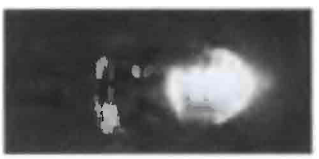


Рис. 17.22. Белый свет от содержащего фосфор полупроводникового СИД на III-нитриде, излучающем в синей области

Органические СИД

Органические светоизлучающие диоды могут быть изготовлены на основе малых органических молекул или сопряженных полимерных цепей (см. подразд. 16.1.2). Матрицы из органических СИД-пикселей могут выпус-

каться в виде тонких светоизлучающих пластиковых листов, генерирующих диффузный свет с большой поверхности. Они служат в качестве недорогих, гибких, сворачиваемых, высокоэффективных самосветящихся дисплеев. Эти приборы можно использовать в цифровых фотокамерах, сотовых телефонах, мониторах компьютеров, телевизорах, а также в архитектурном освещении. Они менее сложны и более тонки, чем жидкокристаллические дисплеи (ЖКД), которые требуют подсветки. Органические СИД могут служить источниками такой подсветки для ЖКД.

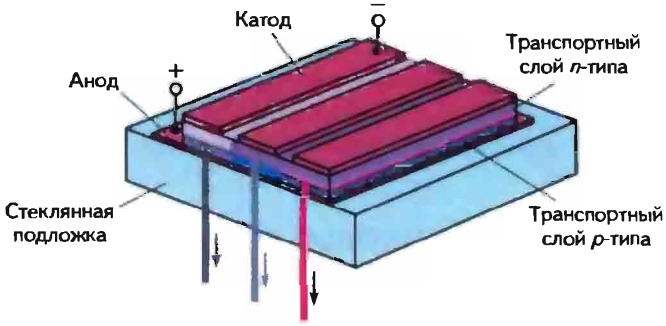


Рис. 17.23. Типичная структура ОСИД. В качестве материалов для катода и прозрачного анода обычно используются оксиды олова с добавлением кальция и индия соответственно. Излучение за счет экситонной рекомбинации из области органического гетероперехода выходит через прозрачный анод и стеклянную подложку. К органическими полупроводникам, используемым для изготовления ОСИД, относятся производная трифенилдиамина с дырочной проводимостью и Alq_3 (алюминий-трис[8-гидроксихинолин]) с электронной проводимостью. Люминесцирующие добавки могут внедряться в активную область для повышения внутреннего квантового выхода и получения белого света

Органические светоизлучающие диоды (ОСИД) на малых молекулах — эффективные источники электролюминесценции в красной, зеленой и синей областях. Две тонкие (≈10 нм) пленки органического полупроводника накладываются друг на друга, образуя органическую гетероструктуру. Как показано на рис. 17.23, эта структура вставляется между двумя электродами: анодом, инжектирующим дырки, и одним или несколькими катодами, инжектирующими электроны. Инжектируемые носители переносятся в активную область гетероперехода, образуя связанные экситоны, которые генерируют спонтанное излучение при рекомбинации. Различные материалы гетероструктур дают различные длины волн рекомбинационного излучения, так что несколько гетероструктур можно разместить на подложке для получения многоцветного ОСИД. Вместо этого в активную область можно внедрить флуоресцентную добавку для получения синего света и фосфоресцентные добавки для зеленого и красного света, повышая таким образом внутренний квантовый выход за счет использования как синглетных, так и триплетных экситонов (см. подразд. 13.5.2). ОСИД белого света, изготовленные таким

методом, имеют внутренний квантовый выход, близкий к единице, и хорошо сбалансированную цветопередачу.

Полимерные светоизлучающие диоды (ПСИД) по конструкции аналогичны ОСИД за исключением того, что обычно они имеют активную область *n*-типа, куда дырки инжектируются органическим слоем *p*-типа. Для изготовления ПСИД часто используется PPV (полифениленвинилен). Эти приборы обычно проще в изготовлении и более эффективны, чем ОСИД, но дают более ограниченный набор цветов. Достоинства органических материалов из малых молекул и полимеров можно совместить в молекулах, известных как фосфоресцентные дендримеры. Они имеют вид больших молекулярных клубков, содержащих внутри ион тяжелого металла, например Ir(2-фенилпиридин)$_3$. Тяжелый ион способствует радиационной рекомбинации экситонов. С ним связаны слои ветвящихся кольцевых структур. Такие гибридные приборы обеспечивают получение ярких многоцветных изображений с высоким разрешением.

17.2. ПОЛУПРОВОДНИКОВЫЕ ОПТИЧЕСКИЕ УСИЛИТЕЛИ

Принципы, на которых основано действие **полупроводниковых оптических усилителей (ПОУ)**, называемых также **полупроводниковыми лазерными усилителями**, те же самые, что и у других лазерных усилителей: создание инверсии заселенностей приводит к преобладанию вынужденного испускания над поглощением. Инверсия заселенностей обычно создается током инжекции через *p*—*n*-переход в диоде; прямое напряжение смещения вызывает инжекцию пар носителей в область перехода, где они рекомбинируют посредством вынужденного испускания фотонов.

Однако теория ПОУ несколько сложнее теории, представленной в гл. 14 для других лазерных усилителей, поскольку переходы происходят между зонами плотно расположенных энергетических уровней, а не между далеко отстоящими друг от друга дискретными уровнями. Тем не менее для целей сравнения ПОУ можно рассматривать как четырехуровневую лазерную систему (см. рис. 14.20), у которой верхние два уровня лежат в зоне проводимости, а нижние два — в валентной зоне.

Распространение теории лазерного усилителя, приведенной в гл. 14, на полупроводниковые структуры было проведено в гл. 16. В данном разделе мы используем результаты, полученные в разд. 16.2, для получения выражений для усиления и ширины полосы полупроводникового оптического усилителя. Затем мы рассмотрим схемы накачки, используемые для получения инверсии заселенностей, и подчеркнем преимущества использования гетероструктур и квантовых ям при разработке усилителей. В конце мы дадим краткий обзор рабочих характеристик полупроводниковых оптических усилителей и сравним их с оптоволоконными усилителями. Теоретические основы описания работы ПОУ не отличаются от основ теории лазерных диодов, которым посвящены разд. 17.3 и 17.4.

17.2.1. Усиление и ширина полосы

Свет частоты ν может взаимодействовать с носителями заряда в полупроводниковом материале с шириной запрещенной зоны E_g за счет межзонных переходов, при условии что $\nu > E_g/h$. Падающие фотоны могут быть поглощены с рождением электронно-дырочной пары либо вызвать рождение новых фотонов путем вынужденной рекомбинации электрона и дырки (рис. 17.24). Когда испускание становится более вероятным, чем поглошение, материал может служить когерентным оптическим усилителем.

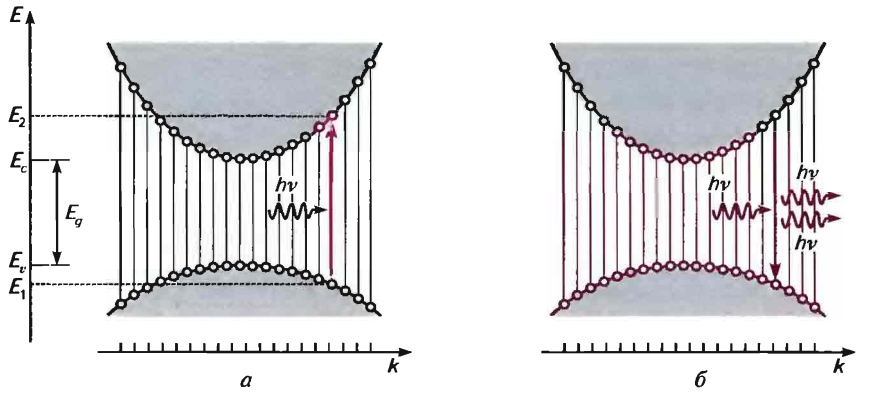


Рис. 17.24. Поглощение фотона приводит к рождению электронно-дырочной пары (*а*). Рекомбинация электрона и дырки может быть индуцирована фотоном; в результате происходит вынужденное испускание точно такого же фотона (*б*)

Выражения для скорости вынужденного излучения $r_{изл}$ и скорости поглощения фотонов $r_{погл}$ даются формулами (16.60) и (16.61) соответственно. Эти величины зависят от спектральной интенсивности потока фотонов ϕ_ν; квантово-механической силы перехода рассматриваемого материала (которая неявно содержится во времени электронно-дырочной радиационной рекомбинации $\tau_{рад}$); оптической совместной плотности состояний $\rho(\nu)$ и вероятностей заселения для испускания $f_e(\nu)$ и поглощения $f_a(\nu)$.

Оптическая совместная плотность состояний $\rho(\nu)$ определяется соотношением между волновым числом k и энергией E для электронов и дырок, а также законами сохранения энергии и импульса. С помощью параболического приближения зависимости E от k вблизи краев зоны проводимости и валентной зоны было показано [см. 16.48) и (16.49)], что энергии электрона и дырки, которые взаимодействуют с фотоном, имеющим энергию $h\nu$, равны

$$E_2 = E_c + \frac{m_r}{m_c}\left(h\nu - E_g\right), \quad E_1 = E_2 - h\nu, \tag{17.31}$$

соответственно, где m_c и m_v — их эффективные массы и $1/m_r = 1/m_c + 1/m_v$. Полученная в результате оптическая совместная плотность состояний, взаимодействующих с фотоном энергии $h\nu$, выражается так [см. (16.51)]:

$$\rho(\nu) = \frac{(2m_r)^{3/2}}{\pi\hbar^2}\sqrt{h\nu - E_g}, \quad h\nu \geq E_g. \tag{17.32}$$

Очевидно, что $\rho(\nu)$ растет как корень квадратный из превышения энергии фотона над шириной запрещенной зоны.

Вероятности заселения $f_e(\nu)$ и поглощения $f_a(\nu)$ определяются скоростью накачки через квазиуровни Ферми E_{f_c} и E_{f_v}. Величина $f_e(\nu)$ — вероятность того, что состояние с энергией E_2 в зоне проводимости занято электроном, а состояние с энергией E_1 в валентной зоне — дыркой. Величина $f_a(\nu)$ — вероятность того, что состояние с энергией E_2 в зоне проводимости свободно, а состояние с энергией E_1 в валентной зоне занято электроном. Фактор инверсии Ферми [см. (16.67)]

$$f_g(\nu) = f_e(\nu) - f_a(\nu) = f_c(E_2) - f_v(E_1) \tag{17.33}$$

характеризует степень инверсии заселенностей. Величина $f_g(\nu)$ зависит как от функции Ферми для зоны проводимости

$$f_c(E) = \frac{1}{\exp\left[\left(E - E_{f_c}\right)/k_B T\right] + 1},$$

так и от функции Ферми для валентной зоны

$$f_v(E) = \frac{1}{\exp\left[\left(E - E_{f_v}\right)/k_B T\right] + 1}.$$

Она является функцией температуры и квазиуровней Ферми E_{f_c} и E_{f_v}, которые, в свою очередь, определяются скоростью накачки. Поскольку в полупроводниковом оптическом усилителе в принципе можно достичь полной инверсии заселенностей [$f_g(\nu) = 1$], он ведет себя как четырехуровневая система.

Результаты, приведенные выше, позволяют получить выражение (16.66) для полного коэффициента усиления $\gamma_0(\nu) = [r_{изл}(\nu) - r_{погл}(\nu)]/\phi_\nu$

$$\boxed{\gamma_0(\nu) = \frac{\lambda^2}{8\pi\tau_{рад}}\rho(\nu) f_g(\nu),} \tag{17.34}$$

Коэффициент усиления

Из сравнения (17.34) и (14.4) ясно, что величина $\rho(\nu)f_g(\nu)$ в полупроводниковом оптическом усилителе играет ту же роль, что и величина $Ng(\nu)$ в других лазерных усилителях, и что $\sigma(\nu) \approx \gamma_0(\nu)/\Delta n$.

Ширина полосы усилителя

В соответствии с (17.33) и (17.34) полупроводниковая среда обеспечивает полное оптическое усиление на частоте ν, когда $f_c(E_2) > f_v(E_1)$. И наоборот, при $f_c(E_2) < f_v(E_1)$ происходит ослабление света. Таким образом, полупроводниковый материал в состоянии теплового равновесия не может давать усиления ни при какой температуре. Это происходит потому, что квазиуровни Ферми валентной зоны и зоны проводимости совпадают ($E_{f_c} = E_{f_v} = E_f$). Для разделения квазиуровней Ферми двух зон и достижения усиления требуется внешняя накачка.

Условие $f_c(E_2) > f_v(E_1)$ эквивалентно требованию, чтобы энергия фотона была меньше расстояния между квазиуровнями Ферми, т. е. $h\nu < E_{f_c} - E_{f_v}$, как показано в упражнении 16.6. Конечно, энергия фотона должна быть больше, чем ширина запрещенной зоны ($h\nu > E_g$), чтобы было возможно усиление за счет межзонных переходов. Таким образом, если скорость накачки остаточно велика, чтобы расстояние между квазиуровнями Ферми превысило ширину запрещенной зоны E_g, то среда может действовать как усилитель в полосе оптических частот

$$\boxed{\frac{E_g}{h} < \nu < \frac{E_{f_c} - E_{f_v}}{h}}. \qquad (17.35)$$

Полоса усиления

Для $h\nu < E_g$ среда является прозрачной, а для $h\nu > E_{f_c} - E_{f_v}$ она является поглощающей, а не усиливающей. Неравенство (17.35) показывает, что ширина полосы усиления растет с увеличением $E_{f_c} - E_{f_v}$ и, следовательно, с повышением уровня накачки. В этом полупроводниковый усилитель отличается от атомного лазерного усилителя, который имеет ненасыщенную ширину полосы $\Delta\nu$, не зависящую от уровня накачки (см. рис. 14.4).

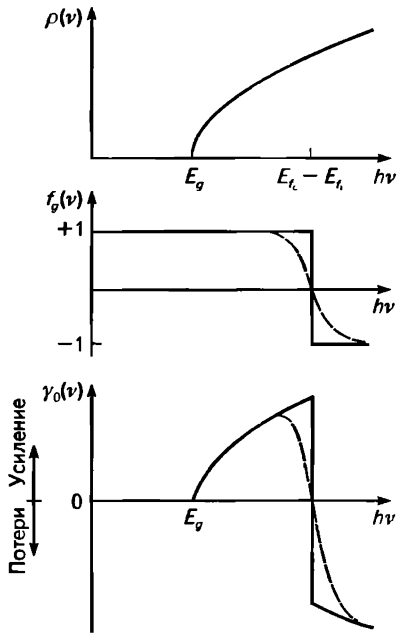


Рис. 17.25. Зависимость оптической совместной плотности состояний $\rho(\nu)$, фактора инверсии Фермн $f_g(\nu)$ и коэффициента усиления $\gamma_0(\nu)$ от энергии фотона $h\nu$ при $T = 0$ °K (сплошные линии) и при комнатной температуре (штриховые линии). Фотоны с энергией между E_g и $E_{f_c} - E_{f_v}$ испытывают лазерное усиление

Расчет усиления заметно упрощается, если можно пренебречь тепловыми возбуждениями (т. е. если $T = 0$ °K). Тогда функции Ферми упрощаются: $f_c(E_2) = 1$ для $E_2 < E_{f_c}$ и 0 в остальных случаях; $f_v(E_1) = 1$ при $E_1 < E_{f_v}$ и 0 в остальных случаях. При этих условиях фактор инверсии Ферми выражается как

$$f_g(\nu) = \begin{cases} +1, & h\nu < E_{f_c} - E_{f_v}; \\ -1, & \text{в остальных случаях.} \end{cases} \tag{17.36}$$

Схематические графики функций $\rho(\nu)$, $f_g(\nu)$ и коэффициента усиления $\gamma_0(\nu)$ представлены на рис. 17.25. Видно, как $\gamma_0(\nu)$ меняет знак и превращается в коэффициент потерь, когда $h\nu > E_{f_c} - E_{f_v}$. Зависимость от ν^{-2} в $\gamma_0(\nu)$, происходящая от множителя λ^2 в знаменателе (17.34), является достаточно медленной и ее можно не учитывать. При конечной температуре графики функций $f_g(\nu)$ и $\gamma_0(\nu)$ сглаживаются, как показано штриховыми линиями на рис. 17.25.

Зависимость коэффициента усиления от уровня накачки

Коэффициент усиления $\gamma_0(\nu)$ растет как по величине, так и по ширине полосы с повышением скорости накачки R. Как следует из (17.1), постоянная скорость накачки R (число инжектируемых избыточных электронно-дырочных пар в см^3 за 1 с) устанавливает стационарную концентрацию электронно-дырочных пар в соответствии с формулой $\Delta n = \Delta p = R\tau$, где τ — рекомбинационное время жизни электронно-дырочной пары (с учетом как радиационной, так и безызлучательной рекомбинации). Знание полных стационарных концентраций электронов $n = n_0 + \Delta n$ и дырок $p = p_0 + \Delta p$ позволяет определить положение квазиуровней Ферми E_{f_c} и E_{f_v} с помощью (17.8). Зная квазиуровни Ферми, можно перейти к расчету коэффициента усиления $\gamma_0(\nu)$ по формуле (17.34). Зависимость $\gamma_0(\nu)$ от Δn и, следовательно, от R иллюстрирует упражнение 17.3.

Пример 17.3

Коэффициент усиления для ПОУ на InGaAsP

Образец четырехкомпонентного материала $In_{0,72}Ga_{0.28}As_{0.6}P_{0.4}$ с шириной запрещенной зоны $E_g = 0{,}95$ эВ используется в качестве полупроводникового оптического усилителя на длине волны $\lambda_0 = 1300$ нм при $T = 300$ °K. Образец не легирован, но имеет остаточную концентрацию доноров и акцепторов $\approx 2 \cdot 10^{17}$ см$^{-3}$. Время жизни радиационной электронно-дырочной рекомбинации $\tau_{рад} \approx 2{,}5$ нс. Эффективные массы электрона и дырки равны $m_c \approx 0{,}06m_0$ и $m_v \approx 0{,}4m_0$ соответственно, а показатель преломления $n \approx 3{,}5$. При известной стационарной концентрации носителей Δn (которая определяется скоростью накачки R и полным временем рекомбинации τ) коэффициент усиления $\gamma_0(\nu)$ можно рассчитать по формуле (17.34) в сочетании с (17.8). Как показано на рис. 17.26, как ширина полосы, так и максимальное значение коэффициента усиления γ_p растет с повышением Δn. Энергия, при которой коэффициент усиле-

ния максимален, также растет с увеличением Δn, что можно ожидать на основании зависимостей, показанных на рис. 17.25. Наименьшая энергия, при которой наблюдается усиление, слабо убывает с ростом Δn вследствие наличия у зоны хвоста, уменьшающего ширину запрещенной зоны. При наибольшем из показанных значений Δn ($\Delta n = 1{,}8\ \ 10^{18}$ см$^{-3}$) усиление происходит с фотонами, энергии которых попадают в интервал от 0,91 до 0,97 эВ. Это соответствует полной ширине полосы усиления по частоте 14,5 ТГц, а по длине волны 80 нм. Более удобной мерой является ширина на высоте половины максимума профиля усиления, которая еще называется шириной на уровне усиления 3 дБ и в данном случае составляет 10 ТГц, что соответствует примерно 50 нм при $\lambda_0 = 1300$ нм (см. табл. 14.1 для сравнения с другими лазерными переходами). Рассчитанный при данном значении Δn пиковый коэффициент усиления $\gamma_p = 270$ см$^{-1}$ велик по сравнению с коэффициентами усиления большинства атомных лазерных усилителей.

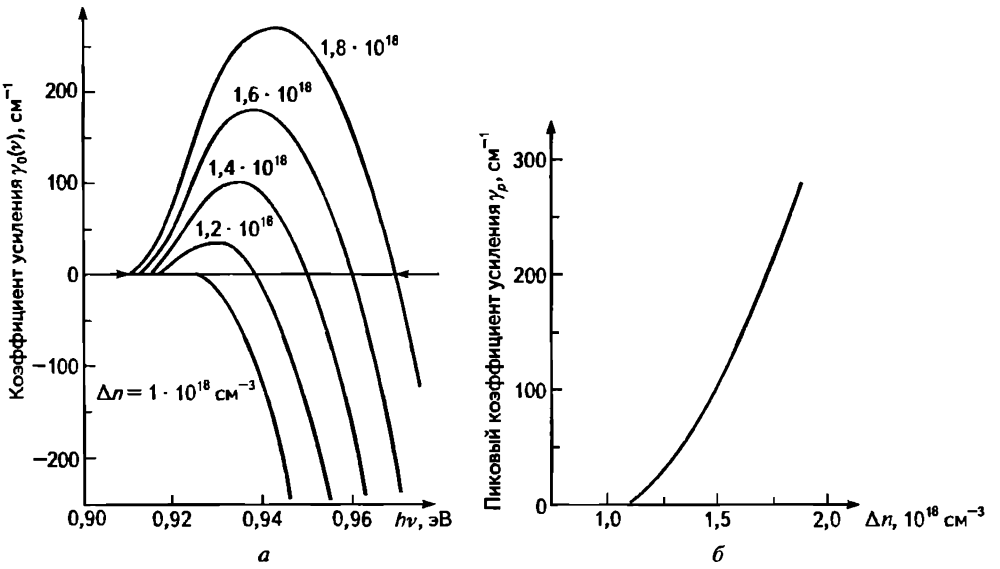


Рис. 17.26. Рассчитанный коэффициент усиления $\gamma_0(\nu)$ для ПОУ на InGaAsP в зависимости от энергии фотона $h\nu$ с концентрацией инжектируемых носителей Δn в качестве параметра ($T = 300$ °K) (*а*). Полоса частот, на которых происходит усиление (с центром около 1300 нм), расширяется с ростом Δn. При наибольшем из показанных Δn, полная ширина на половине высоты максимума полосы составляет 10 ТГц, что соответствует 0,04 эВ по энергии и 50 нм по длине волны. (Заимствовано из работы *Dutta N.K.* Calculated Absorption, Emission and Gain in $In_{0,72}Ga_{0,28}As_{0,6}P_{0,4}$. Journal of Applied Physics. Vol. 51, 1980. P. 6095—6100, рис. 8.) Рассчитанные значения пикового коэффициента усиления γ_p как функции Δn (*б*). При максимальном значении Δn пиковый коэффициент усиления составляет ≈270 см$^{-1}$ (заимствовано из *Dutta N.K., Nelson R.J.* The Case for Auger Recombination in $In_xGa_{1-x}As_yP_{1-}$ Journal of Applied Physics. Vol. 53, 1982. P. 74—92, рис. 17)

Наступление насыщения в полупроводниковых оптических усилителях мало отличается от такового в других лазерных усилителях с однородным уширением, рассмотренного в разд. 14.4. Сравнительно большое сечение перехода в полупроводниках (см. табл. 14.1) приводит к низким значениям насыщающей плотности потока фотонов [$\phi_s \approx 1/\tau_{\text{рад}}\sigma(\nu)$] и, следовательно, пониженному коэффициенту усиления [см. (14.47) и (14.48)]. Это, в свою очередь, ограничивает полное усиление, которое могут обеспечивать ПОУ.

Как и другие оптические усилители, ПОУ страдают от шума усиленного спонтанного излучения (см. разд. 14.5), однако на их работе сказываются также шумы, связанные с флуктуациями температуры и концентрации носителей.

Приближенный расчет пикового коэффициента усиления

Сложная зависимость коэффициента усиления от концентрации инжектируемых носителей затрудняет анализ работы полупроводникового усилителя (и лазера). По этой причине принято использовать эмпирический подход, при котором предполагается линейная зависимость пикового коэффициента усиления γ_p от Δn для значений Δn вблизи рабочей точки. Как показывает пример, представленный на рис. 17.26, *б*, это приближение разумно при больших γ_p. Зависимость пикового коэффициента усиления γ_p от Δn в этом приближении моделируется линейным соотношением

$$\gamma_p \approx \alpha\left(\frac{\Delta n}{\Delta n_{\text{T}}} - 1\right), \qquad (17.37)$$

Пиковый коэффициент усиления (линейное приближение)

которое графически изображено на рис. 17.27. Параметры α и Δn_{T} выбираются так, чтобы выполнялись следующие предельные условия:

- когда $\Delta n = 0$, $\gamma_p = -\alpha$, где α — коэффициент поглощения полупроводника в отсутствие тока инжекции;
- когда $\Delta n = \Delta n_{\text{T}}$, $\gamma_p = 0$. Таким образом, Δn_{T} — значение концентрации инжектируемых носителей, при котором поглощение и вынужденное излучение уравновешивают друг друга, так что среда прозрачна.

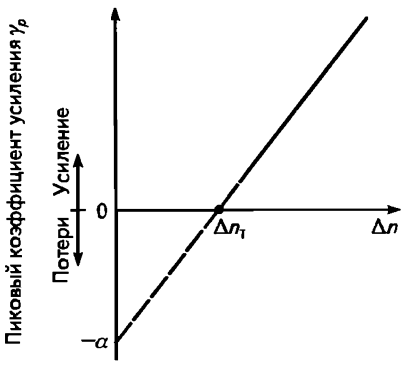


Рис. 17.27. Пиковый коэффициент усиления γ_p как функция концентрации инжектируемых носителей Δn в приближенной линейной модели. Величина α — коэффициент поглощения в отсутствие инжекции, а Δn_{T} — концентрация инжектируемых носителей, при которой усиление и поглощение в точности компенсируют друг друга. Сплошная часть прямой линии соответствует более реалистическому расчету, рассмотренному в предыдущем подразделе

Пример 17.4

Приближенный пиковый коэффициент усиления для ПОУ на InGaAsP

Пиковый коэффициент усиления γ_p для InGaAsP, показанный как функция Δn на рис. 17.26, *б*, можно аппроксимировать линейным соотношением вида (17.37) с параметрами $\Delta n_T \approx 1,25 \cdot 10^{18}$ см$^{-3}$ и $\alpha = 600$ см$^{-1}$. Для $\Delta n = 1{,}4\Delta n_T = 1{,}75 \cdot 10^{18}$ см$^{-3}$ линейная модель дает пиковый коэффициент усиления $\gamma_p = 240$ см. Для кристалла InGaAsP длиной $d = 350$ мкм это соответствует полному усилению $\exp(\gamma_p d) \approx 4447$, или 36,5 дБ. На практике это значение уменьшается за счет насыщения усиления, как обсуждалось выше, а также за счет потерь согласования, которые обычно составляют от 3 до 5 дБ на каждую грань.

Увеличение концентрации инжектируемых носителей с переходом через значение Δn_T, соответствующее прозрачности $[f_g(\nu) < 0]$, приводит к тому, что полупроводник из сильного поглотителя света превращается в сильный усилитель $[f_g(\nu) > 0]$. Та же самая высокая вероятность перехода, которая делает полупроводник хорошим поглотителем, делает его же хорошим усилителем, что легко понять из сравнения (16.60) и (16.61).

17.2.2. Накачка

Оптическая накачка

Накачку можно осуществить с использованием внешнего источника света, как показано на рис. 17.28, при условии, что энергия его фотонов достаточно велика ($>E_g$). Фотоны накачки поглощаются полупроводником, в результате чего рождаются пары носителей заряда. Образовавшиеся электроны и дырки релаксируют к дну зоны проводимости и потолку валентной зоны соответственно. Если время внутризонной релаксации много меньше времени межзонной релаксации, как обычно и бывает, то между зонами устанавливается стационарная инверсия заселенностей, как обсуждалось в разд. 14.2.

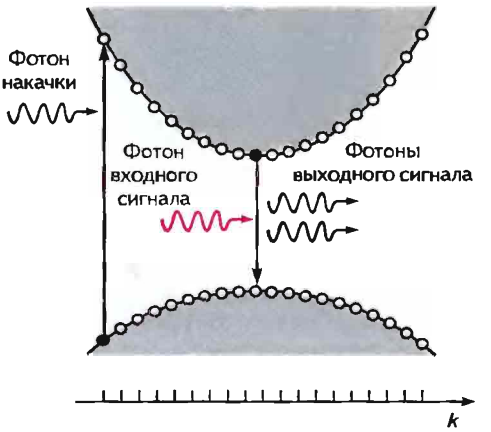


Рис. 17.28. Оптическая накачка полупроводникового оптического усилителя

Накачка током

Более практичная схема накачки полупроводникового оптического усилителя основана на инжекции электроном и дырок в сильно легированном p—n-переходе — диоде. Как и в СИД (см. разд. 17.1), переход смещается в прямом направлении, так что неосновные носители инжектируются в область перехода (электроны в p-область, а дырки в n-область). На рис. 17.6 показана схема энергетических зон для прямосмещенного сильно легированного p—n-перехода. Квазиуровни Ферми для зоны проводимости и валентной зоны, E_{f_c} и E_{f_v}, соответственно, лежат внутри своих зон, и область перехода находится в состоянии квазиравновесия. Квазиуровни Ферми разнесены достаточно далеко друг от друга, так что достигается инверсия заселенностей и полное усиление может быть получено в полосе $E_g \le h\nu \le E_{f_c} - E_{f_v}$ в активной области. Толщина этой области l является важным параметром диода, который определяется, в основном, длиной диффузии неосновных носителей с обеих сторон от перехода. Типичные значения l для InGaAsP составляют 1—3 мкм.

Если электрический ток i протекает через площадь $A = wd$, где w и d — ширина и высота прибора соответственно, в объем lA (как показано на рис. 17.29), то стационарная скорость инжекции составляет

$$R = \frac{i}{elA} = \frac{J}{el}$$

за секунду в единичном объеме, где J — плотность тока инжекции. Получающаяся в результате концентрация инжектируемых носителей равна

$$\Delta n = \tau R = \frac{\tau}{elA} i = \frac{\tau}{el} J. \qquad (17.38)$$

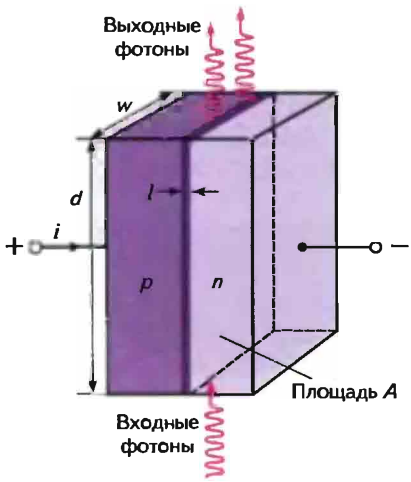


Рис. 17.29. Геометрия простого полупроводникового оптического усилителя. Носители заряда движутся перпендикулярно p—n-переходу, а фотоны — в плоскости перехода

Таким образом, концентрация инжектируемых носителей прямо пропорциональна плотности тока инжекции, так что результаты, показанные на рис. 17.26 и 17.27 с Δn в качестве параметра, можно применять, взяв в качестве параметра J. В частности, из (17.30) и (17.31) следует, что в пределах применимости линей-

ного приближения пиковый коэффициент усиления линейно связан с плотностью тока инжекции J, т. е.

$$\gamma_p \approx \alpha\left(\frac{J}{J_T} - 1\right). \qquad (17.39)$$

Пиковый коэффициент усиления

Плотность тока J_T, соответствующая прозрачности среды, равна

$$J_T = \frac{el}{\eta_i \tau_{рад}} \Delta n_T, \qquad (17.40)$$

Плотность тока, обеспечивающая прозрачность

где $\eta_i = \tau/\tau_{рад}$, как и раньше, обозначает внутренний квантовый выход.

При $J = 0$ пиковый коэффициент усиления $\gamma_p = -\alpha$ становится коэффициентом поглощения, как очевидно из рис. 17.30. Когда $J = J_T$, $\gamma_p = 0$ и материал прозрачен, он не поглощает и не усиливает свет. Полное усиление может быть достигнуто, только если плотность тока инжекции J превышает значение J_T, соответствующее прозрачности среды. Заметим, что J_T прямо пропорционально толщине области перехода l, так что меньшее значение J_T достигается при использовании меньших толщин области перехода. Это важное соображение используется при конструировании полупроводниковых оптических усилителей и лазеров.

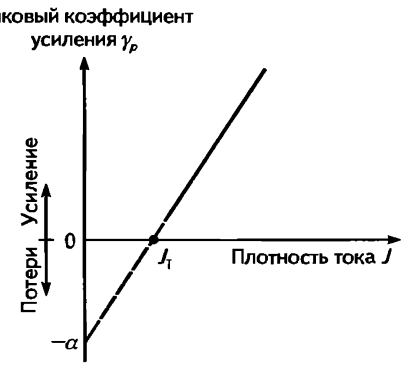


Рис. 17.30. Пиковый коэффициент усиления γ_p как функция плотности тока J для приближенной линейной модели. Когда $J = J_T$, материал прозрачен и не вносит ни усиления, ни потерь

Пример 17.5

Усиление в ПОУ на InGaAsP

Полупроводниковый оптический усилитель на InGaAsP работает при $T = 300$ °K и имеет следующие параметры: $\tau_{рад} = 2{,}5$ нс, $\eta_i = 0{,}5$, $\Delta n_T \approx 1$, $25 \cdot 10^{18}$ см$^{-3}$ и $\alpha = 600$ см$^{-1}$. Толщина области перехода $l = 2$ мкм, длина $d = 200$ мкм и ширина $w = 10$ мкм. С помощью (17.39) находим, что плотность тока инжекции, обеспечивающая прозрачность, $J_T = 3{,}2 \cdot 10^4$ А/см^2. Немного большая плотность тока $J = 3{,}5 \cdot 10^4$ А/см^2 обеспечивает пиковый коэффициент усиления $\gamma_p \approx 56$ см$^{-1}$, как следует из (17.39). Это дает полное усиление $G = \exp(\gamma_p d) = \exp(1{,}12) \approx 3$. Однако, поскольку площадь перехода $A = wd = 2 \cdot 10^{-5}$ см^2, для создания такой плотности тока требуется весьма высокий ток инжекции $i = JA = 700$ мА.

Соображения в пользу гетероструктур

Если в примере 17.5 толщину активной области l уменьшить с 2 мкм до, скажем, 0,1 мкм, то плотность тока J_T уменьшится в 20 раз до более разумного значения 1600 А/см^2. Поскольку пропорционально меньший объем будет подлежать накачке, усилитель мог бы обеспечивать такое же усиление при меньшей плотности тока инжекции. Такое уменьшение толщины активной области, однако, представляет собой дополнительную проблему, так как длина диффузии электронов и дырок в InGaAsP составляет несколько микрометров, так что носители будут диффундировать наружу из области меньших размеров. Однако существует возможность запереть носители в активной области с использованием гетероструктуры, как описано в подразд. 17.2.3. Одновременно такая структура может удерживать и свет, что дает дополнительное преимущество.

17.2.3. Гетероструктуры

Как ясно из (17.39) и (17.40), пиковый коэффициент усиления диодного лазера γ_p меняется обратно толщине l активной области. Поэтому выгодно уменьшать эту толщину, насколько возможно. Размер активной области определяется диффузионной длиной неосновных носителей по обе стороны перехода. Концепция двойной гетероструктуры состоит в формировании потенциальных барьеров по обе стороны p—n-перехода; получается потенциальная яма, ограничивающая расстояние, на которое могут диффундировать неосновные носители заряда. Барьеры определяют область пространства, в которой оказываются запертыми неосновные носители, что позволяет получать активные области малой толщины $l \approx 0{,}1$ мкм. Еще более тонкие активные слои, до ≈0,01 мкм, получаются с помощью квантовых ям, что будет обсуждаться в подразд. 17.2.4.

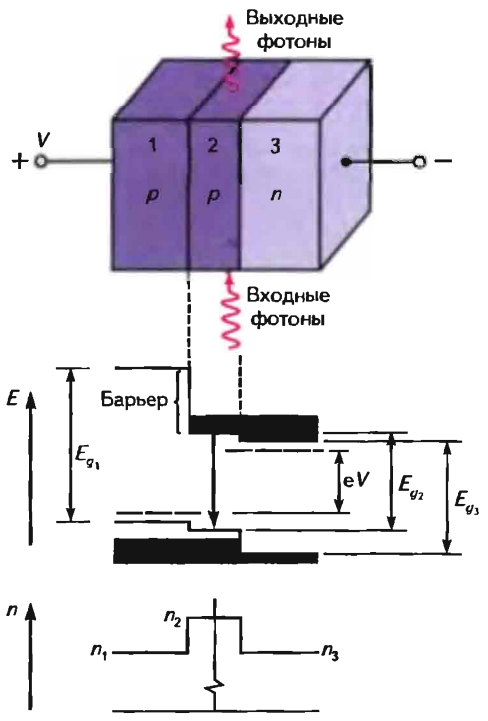


Рис. 17.31. Схема энергетических зон и показатель преломления в зависимости от положения для полупроводникового оптического усилителя с двойной гетероструктурой

Одновременно может быть достигнуто и электромагнитное удержание усиливаемого оптического пучка, если материал активного слоя подобрать так, чтобы его показатель преломления был немного больше окружающих слоев. Тогда структура ведет себя как оптический волновод (см. разд. 8.2).

Таким образом, для изготовления двойной гетероструктуры требуется три слоя из различных, но согласованных по решетке материалов, как показано на рис. 17.31:

Слой 1: p-тип, ширина запрещенной зоны E_{g_1}, показатель преломления n_1;

Слой 2: p-тип, ширина запрещенной зоны E_{g_2}, показатель преломления n_2;

Слой 3: n-тип, ширина запрещенной зоны E_{g_3}, показатель преломления n_3.

Материалы подбираются так, чтобы E_{g_1} и E_{g_3} были больше, чем E_{g_2}, что обеспечивает удержание носителей заряда, и одновременно чтобы n_1 и n_3 были немного меньше n_2, что обеспечивает удержание фотонов. Активный слой (слой 2) делается очень тонким (0,1—0,2 мкм) для минимизации плотности инжекционного тока, обеспечивающего прозрачность, и, следовательно, достижения наибольшего пикового коэффициента усиления. Вынужденное испускание наблюдается в p—n-переходе между слоями 2 и 3.

Итак, перечислим преимущества, получаемые при использовании двойной гетероструктуры.

• Повышение усиления при заданной плотности тока инжекции в результате уменьшения толщины активного слоя в соответствии с (17.32) и (17.33). Инжектируемые неосновные носители удерживаются в тонком активном слое между двумя барьерами, создаваемыми гетеропереходами и предотвращающими диффузию в окружающие слои.

• Повышение усиления за счет удержания фотонов внутри активного слоя, имеющего больший показатель преломления. Активная среда ведет себя как волновод.

• Понижение потерь в результате неспособности слоев 1 и 3 поглощать удерживаемые волноводом фотоны, так как ширины запрещенных зон этих слоев E_{g_1} и E_{g_3} больше энергии фотонов ($h\nu = E_{g_2} < E_{g_1}, E_{g_3}$).

Приведем два примера полупроводниковых оптических усилителей с двойной гетероструктурой.

• *Усилительный лазерный диод с двойной гетероструктурой InGaAsP/InP.* Активный слой $In_{1-x}Ga_xAs_{1-y}P_y$ окружен слоями InP. Параметры состава x и y выбраны так, чтобы постоянные решетки материалов совпадали. Таким образом, рабочий интервал параметров x и y определяется требованием соответствия E_{g_2} длине волны, лежащей в полосе 1,1—1,7 мкм.

• *Усилительный лазерный диод с двойной гетероструктурой GaAs/AlGaAs.* Активный слой (слой 2) изготовляется из GaAs (E_{g_2} = 1,42 эВ, n_2 = 3,6). Окружающие слои (1 и 3) изготовлены из $Al_xGa_{1-x}As$ с $E_g > 1{,}43$ эВ и $n < 3{,}6$ (на 5—10 %). Этот усилитель обычно работает в полосе длин волн 0,82—0,88 мкм при значениях параметра состава x от 0,35 до 0,5.

17.2.4. Структуры с квантовыми ямами

Как обсуждалось в подразд. 17.2.3, гетероструктуры обеспечивают уменьшение толщины активного слоя, в котором удерживаются носители заряда и фотоны. Это, в свою очередь, обеспечивает повышение усиления и снижение потерь усилителя. При дальнейшем уменьшении толщины активного слоя, скажем, до 5—10 нм (что меньше, чем длина волны де Бройля тепловых электронов) ключевую роль начинают играть квантовые эффекты. Поскольку активный слой в двойной гетероструктуре имеет меньшую ширину запрещенной зоны, чем окружающие слои, структура действует как **квантовая яма** (см. подразд. 16.1.7).

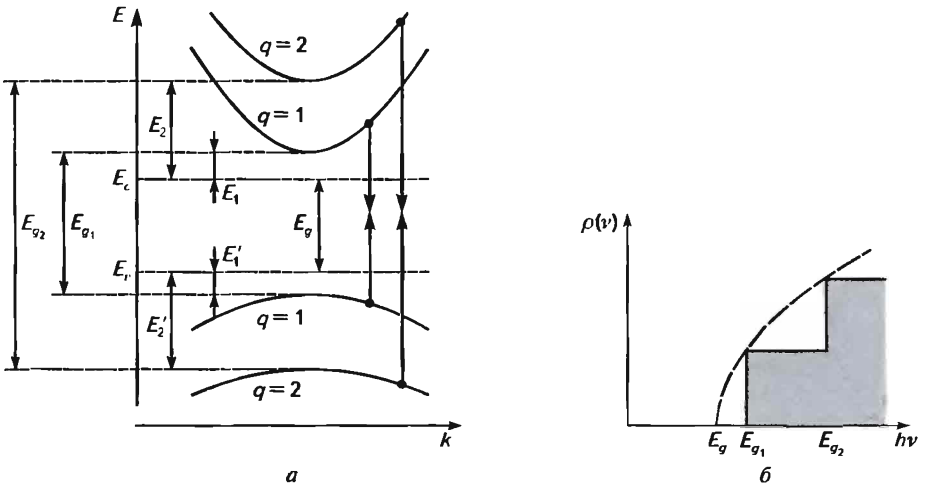
Зонная структура и соотношение между энергией и импульсом (E—k) для квантовой ямы отличаются от таковых для объемного материала. Зона проводимости расщепляется на ряд подзон, нумеруемых квантовым числом $q = 1, 2, 3,$ каждая из которых имеет собственное E—k-соотношение и свою плотность состояний. Дно каждой из этих подзон имеет энергию $E_c + E_q$, где E_q — энергия электрона с эффективной массой m_c в одномерной квантовой яме ширины l,

$$E_q = \hbar^2 \frac{(q\pi/l)^2}{2m_c}, \quad q = 1,\ 2,\ 3,$$

(см. рис. 16.24 и 16.26; q_1 и d_1 в гл. 16 соответствуют q и l здесь). Каждая подзона имеет параболическое E—k-соотношение и постоянную плотность состояний, не зависящую от энергии. Полная плотность состояний в зоне проводимости $\rho_c(E)$ вследствие этого принимает вид ступенчатого распределения [см. (16.37)], у которого ступени соответствуют энергиям $E_c + E_q$, $q = 1, 2, 3,$ Валентная зона имеет аналогичные подзоны при энергиях $E_v + E'_q$, где E'_q — энергия дырки с эффективной массой m_v в квантовой яме ширины l,

$$E'_q = \hbar^2 \frac{(q\pi/l)^2}{2m_v}.$$

Взаимодействие фотонов с электронами и дырками в квантовой яме происходит в виде переходов между валентной зоной и зоной проводимости с сохранением энергии и импульса. При этих переходах должно также сохраняться квантовое число q, как видно на рис. 17.32; они подчиняются таким же правилам, как и переходы между валентной зоной и зоной проводимости в объемном полупроводнике. Выражения для вероятностей переходов и коэффициента усиления в объемном материале (см. разд. 16.2) применимы и в этом случае, если заменить ширину запрещенной зоны E_g энергетическим расстоянием между подзонами $E_{gq} = E_g + E_q + E'_q$, и использовать постоянную плотность состояний, а не плотность состояний, меняющуюся как квадратный корень из энергии. Полный коэффициент усиления равен сумме коэффициентов усиления для каждой подзоны ($q = 1, 2, 3, \ldots$).



а *б*

Рис. 17.32. E—k-соотношения для различных подзон (*а*). Оптическая совместная плотность состояний для структуры с квантовой ямой (ступенчатая линия) и для объемного полупроводника (штриховая линия) (*б*). Первый скачок происходит при энергии $E_{g_1} = E_g + E_1 + E_1'$ (где E_1 и E_1', соответственно, наименьшие энергии электрона и дырки в квантовой яме)

Плотность состояний

Рассмотрим переходы между двумя подзонами с квантовым числом q. Согласно законам сохранения энергии и импульса, фотон $h\nu$ может взаимодействовать с состояниями, имеющими энергию

$$E = E_c + E_q + \frac{m_r}{m_c}\left(h\nu - E_{gq}\right)$$

в верхней подзоне и $E - h\nu$ в нижней подзоне. Оптическая совместная плотность состояний $\rho(\nu)$ связана с $\rho_c(E)$ формулой

$$\rho(\nu) = \frac{\mathrm{d}E}{\mathrm{d}\nu}\rho_c(E) = \frac{hm_r}{m_c}\rho_c(E).$$

Из (16.37) следует, что

$$\rho(\nu) = \begin{cases} \dfrac{hm_r}{m_c}\dfrac{m_c}{\pi\hbar^2 l} = \dfrac{2m_r}{\hbar l}, & h\nu > E_g + E_q + E_q'; \\ 0, & \text{в остальных случаях.} \end{cases} \qquad (17.41)$$

Включая переходы между всеми подзонами $q = 1, 2, 3, \ldots$, мы приходим к плотности $\rho(\nu)$, имеющей вид ступенчатого распределения, причем ступени соответствуют энергетическим расстояниям между подзонами с одним и тем же квантовым числом (см. рис. 17.32).

Коэффициент усиления

Коэффициент усиления данного устройства дается обычным выражением [см. (16.66)]:

$$\gamma_0(\nu) = \frac{\lambda^2}{8\pi\tau_{\text{рад}}}\rho(\nu) f_g(\nu), \tag{17.42}$$

где фактор инверсии Ферми $f_g(\nu)$ зависит от положения квазиуровней Ферми и температуры и одинаков для объемных лазеров и лазеров с квантовыми ямами. Однако плотность состояний $\rho(\nu)$ в этих случаях различна, как было только что показано. Частотные зависимости $\rho(\nu)$, $f_g(\nu)$ и их произведения показаны на рис. 17.33 для квантовой ямы и объемной двойной гетероструктуры. Структура с квантовой ямой имеет меньшее пиковое значение и более узкий профиль коэффициента усиления.

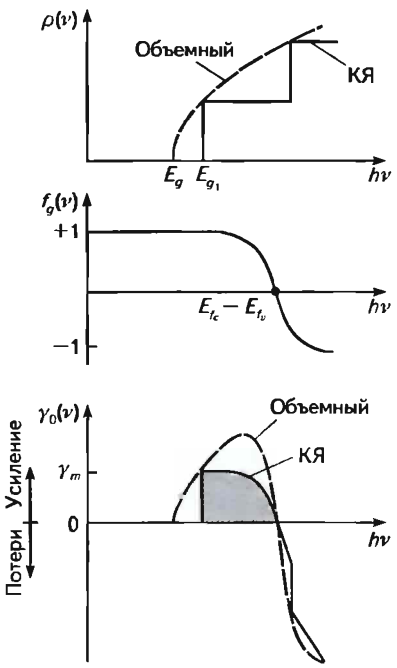


Рис. 17.33. Плотность состояний $\rho(\nu)$, фактор инверсии Ферми $f_g(\nu)$ и коэффициент усиления $\gamma_0(\nu)$ для квантовой ямы (сплошные линии) и объемной структуры (штриховые линии)

При построении рис. 17.33 предполагалось, что при энергии меньше $E_{f_c} - E_{f_v}$ происходит всего одно ступенчатое изменение $\rho(\nu)$. Именно так бывает при обычных условиях инжекции. Максимальное усиление γ_m тогда можно определить, подставляя $f_g(\nu) = 1$ и $\rho(\nu) = 2m_r/\hbar l$ в (17.42), что дает

$$\gamma_m = \frac{\lambda^2 m_r}{2\tau_{\text{рад}} hl}. \tag{17.43}$$

Связь коэффициента усиления с плотностью тока

При увеличении тока инжекции J увеличивается концентрация избыточных электронов и дырок Δn и, следовательно, расстояние между квазиуровнями Ферми $E_{f_c} - E_{f_v}$. Влияние этого увеличения на коэффициент усиления можно понять из анализа графиков на рис. 17.33. При достаточно малых J

усиления нет. Когда J увеличивается настолько, чтобы разность $E_{f_c} - E_{f_v}$ слегка превысила расстояние E_{g1} между подзонами с $q = 1$, среда становится усиливающей. Максимальный коэффициент усиления резко увеличивается и насыщается на уровне γ_m. Дальнейшее увеличение J вызывает рост ширины спектрального контура усиления, но не его максимального значения. Однако когда J достигает значения, для которого $E_{f_c} - E_{f_v}$ превосходит расстояние E_{g2} между подзонами с $q = 2$, максимальный коэффициент усиления испытывает новый скачок и т. д. Таким образом, профиль усиления может быть весьма широким, что открывает возможность перестройки в широкой полосе частот для таких устройств.

Материалы и структуры приборов

Структуры полупроводниковых оптических усилителей напоминают структуру лазерного диода, работающего выше порога прозрачности, но ниже порога генерации (см. разд. 17.3). Путем тщательного подбора материала полупроводниковые оптические усилители разрабатываются для любой части оптического спектра. Центральная длина волны, ширина полосы и величина усиления зависят как от материала, так и от структуры прибора.

ПОУ для целей передачи в ближней ИК-области обычно изготавливают из InGaAsP, InGaAs или InP. В телекоммуникационной полосе 1300—1600 нм достигается ширина полосы усиления $\Delta\lambda \approx 50$ нм, что в частотной шкале соответствует $\Delta\nu \approx 10$ ТГц при $\lambda_0 = 1300$ нм (см. пример 17.3). Это больше, чем у усилителей из волокна, легированного эрбием, и столько же, сколько у волоконных ВКР-усилителей (см. подразд. 14.3.3 и 14.3.4). ПОУ с квантовыми ямами достигают прозрачности при значительно меньших токах, но в остальном ведут себя аналогично объемным устройствам. Ширина полосы усилителей на квантовых точках может достигать $\Delta\lambda \approx 200$ нм, что соответствует $\Delta\nu \approx 25$ ТГц при $\lambda_0 = 1550$ нм.

Усиление, обеспечиваемое ПОУ, обычно ограничивается 15 дБ из-за насыщения коэффициента усиления и потерь излучения при вводе 3—5 дБ на каждой грани (см. пример 17.4). Насыщение приводит к межканальной и межсимвольной интерференции, что мешает использовать ПОУ в коммуникационных системах с мультиплексированием по длине (см. подразд. 24.3.3). Более того, малое время рекомбинации полупроводников (см. табл. 14.1) делает ПОУ уязвимыми для высокочастотных шумов, которые могут присутствовать как в токе накачки, так и в оптическом сигнале, что приводит к уровню шума ≈8—10 дБ в отличие от 3 дБ для усилителей из волокна, легированного эрбием. Важно отметить, что если ПОУ должен работать как широкополосное однопроходное устройство (т. е. как усилитель бегущей волны), то отражения на гранях должны быть сведены к минимуму. Если этого не сделать, то происходит многократное отражение и профиль усиления модулируется модами резонатора; это может также привести к генерации, что, конечно, исключает возможность контролируемого усиления. Методы уменьшения отражения включают использование противоотражательных покрытий и наклонных волноводов.

В результате всех обстоятельств, обсужденных выше, применения ПОУ для передачи оптических сигналов, по большей части, ограничены внутригородскими сетями оптической связи, где небольшого усиления достаточно для компенсации потерь, связанных с наличием множества узлов ввода—вывода. ПОУ обладают гораздо большей привлекательностью для таких приложений, как нелинейно-оптические устройства и оптические переключатели (см. подразд. 23.3.3), а также для преобразования длины волны.

Пример 17.6

Волноводные усилители

Полупроводниковый оптический усилитель с множественными квантовыми ямами может быть сконструирован в виде оптического волновода, обеспечивающего работу на основной оптической моде с повышением выходной насыщающей мощности и использованием прямого соединения встык с одномодовым волокном. Такие приборы имеют сравнительно низкие потери и небольшой коэффициент оптического удержания. Например, усилитель с квантовыми ямами из InGaAsP/InP длиной 1 см, работающий на 1550 нм, обеспечивает усиление «волокно—волокно» в 13 дБ[1].

Сравнение ПОУ и ВОУ

Полупроводниковые оптические усилители имеют достоинства и недостатки по сравнению с волоконно-оптическими усилителями, такими как волокно, легированное эрбием, и ВКР-усилитель.

Достоинства:

- центральную длину волны можно менять путем подбора материала;
- совместимость с интегрально-оптическими схемами;
- электрическая накачка;
- малые размеры;
- низкая цена.

Недостатки:

- низкое усиление;
- низкая мощность насыщения;
- высокий уровень шума;
- существенная межканальная и межсимвольная интерференция;
- чувствительность к тепловым эффектам;
- чувствительность к отражениям на гранях;
- чувствительность к поляризации сигнала;
- управление характеристиками поперечных мод;
- высокие потери ввода;
- несовместимость с волоконной геометрией.

[1] См. *Juodawlkis P.W., Plant J.J., Huang R.K., Missaggia L.J., Donnelly J.P.* High-Power 1.5-mm InGaAsP-InP Slab-Coupled Optical Waveguide Amplifier. IEEE Photonics Technology Letters. Vol. 17, 2005. P. 279—281.

Суммируя, видим, что рабочие характеристики ПОУ уступают таковым у эрбиевого и ВКР-волоконных лазеров, поэтому использование ПОУ ограничено специальными приложениями (см. подразд. 24.1.3). Сравнение эрбиевого и ВКР-волоконных лазеров было проведено в подразд. 14.3.4.

17.2.5. Сверхлюминесцентные диоды

Сверхлюминесцентные диоды (СЛД) представляют собой полупроводниковые лазерные диоды со сравнительно высоким током инжекции, при котором вынужденное излучение преобладает над спонтанным. СЛД отличаются от ПОУ тем, что на прибор не подается входного сигнала. Выходное излучение представляет собой усиленное спонтанное излучение самого прибора (см. разд. 14.5).

Примером СЛД является InGaAsP/InP-структура с множественными квантовыми ямами, изображенная на рис. 17.34. Выходная оптическая мощность СЛД в общем случае выше, чем у СИД, но меньше, чем у лазерных диодов (ЛД) (см. рис. 17.39), а оптический спектр обычно уже, чем у СИД, но шире, чем у ЛД (см. рис. 17.41). Как и в случае полупроводникового оптического усилителя, важно минимизировать оптическую обратную связь для предотвращения лазерной генерации. Этого можно достичь разными способами, такими как использование полоскового контакта, инжектирующего ток только в часть прибора, использование клиновидно-полосковой геометрии, нанесение на грани прибора антиотражающих покрытий или придание им наклона.

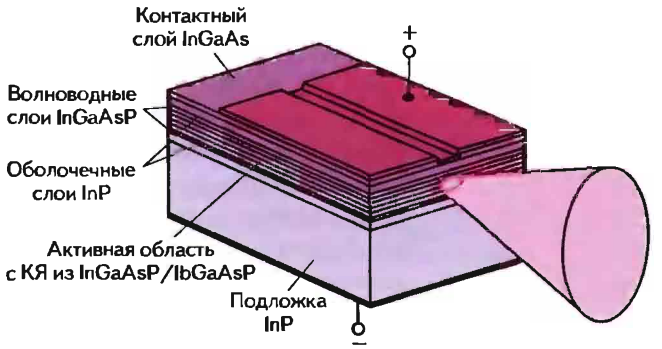


Рис. 17.34. Сверхлюминесцентный диод с КЯ на InGaAsP/InP. СЛД генерируют свет значительной мощности с шириной спектра, промежуточной между СИД и ЛД. Важно минимизировать оптическую обратную связь, чтобы не возникала лазерная генерация. Один из способов достичь этого — использование полоскового контакта, инжектирующего ток только в часть прибора

Сверхлюминесцентные диоды находят применение там, где большое время когерентности лазерного света нежелательно из-за вредного влияния случайно возникающей интерференции. Примерами являются такие интерференционные методы, как оптическая когерентная томография (см. подразд. 11.2.2),

волоконно-оптическая гироскопия и некоторые волоконно-оптические датчики. Волоконно-оптические усилители иногда тоже используются в качестве источников сверхлюминесцентного света.

17.3. ЛАЗЕРНЫЕ ДИОДЫ

В этом разделе мы рассмотрим общие характеристики лазерных диодов, которые бывают многих видов. Полупроводниковые лазеры с квантово-размерными структурами и микрорезонаторами будут рассмотрены позже в разд. 17.4.

17.3.1. Усиление, обратная связь и генерация

Лазерный диод представляет собой полупроводниковый оптический усилитель с оптической обратной связью. Как обсуждалось в предыдущем разделе, полупроводниковый оптический усилитель основан на прямосмещенном сильно легированном p—n-переходе из прямозонного полупроводникового материала. Ток инжекции достаточно велик для обеспечения оптического усиления. Оптическая обратная связь обеспечивается зеркалами, которые обычно создаются путем раскалывания полунроводникового материала вдоль его кристаллических плоскостей. Резкий скачок показателя преломления на границе раздела полупроводника и окружающего воздуха обусловливает отражающее действие поверхностей скола кристалла. Таким образом, полупроводниковый кристалл одновременно служит и усиливающей средой, и оптическим резонатором Фабри—Перо, как показано на рис. 17.35. При условии достаточно высокого коэффициента усиления обратная связь превращает усилитель в генератор, т. е. лазер. Устройство называется **лазерным диодом**, или **диодным лазером** (иногда употребляется также термин **полупроводниковый инжекционный лазер**).

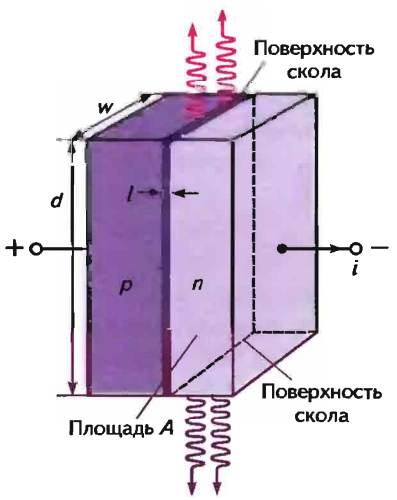


Рис. 17.35. Лазерный диод — прямосмещенный p—n-переход с двумя параллельными поверхностями, действующими как резонатор

Лазерный диод (ЛД) имеет существенное сходство со светоизлучающим диодом (СИД), рассмотренным в разд. 17.1. В обоих приборах источником энергии служит ток инжекции, протекающий через p—n-переход. Однако свет, излучаемый СИД, возникает благодаря спонтанному испусканию, а излучение ЛД происходит за счет вынужденного испускания.

Лазерные диоды имеют ряд преимуществ по сравнению с другими типами лазеров: высокую мощность, высокий коэффициент полезного действия, малые размеры, совместимость с элементной базой электроники, легкость накачки и модуляции через ток инжекции. Однако большая спектральная ширина и более низкая когерентность излучения могут быть препятствием для использования ЛД в некоторых приложениях. Лазерные диоды имеют множество применений, как будет видно из дальнейшего изложения.

Начнем рассмотрение условий лазерной генерации и свойств испускаемого света с краткой сводки основных результатов описания полупроводникового оптического усилителя и оптического резонатора.

Лазерное усиление

Коэффициент усиления $\gamma_0(\nu)$ полупроводникового оптического усилителя имеет максимальное значение γ_p, которое приблизительно прямо пропорционально концентрации инжектируемых носителей, в свою очередь, пропорциональной плотности тока инжекции. Таким образом, в соответствии с формулами (17.39) и (17.40), которые иллюстрируется рис. 17.30, имеем

$$\gamma_p \approx \alpha\left(\frac{J}{J_T}-1\right); \quad J_T = \frac{el}{\eta_i \tau_{рад}}\Delta n_T, \tag{17.44}$$

где $\tau_{рад}$ — время электронно-дырочной радиационной рекомбинации; $\eta_i = \tau/\tau_{рад}$ — внутренний квантовый выход; l — толщина активной области; α — коэффициент поглощения в состоянии теплового равновесия; Δn_T и J_T — концентрация инжектируемых носителей и плотность тока инжекции, необходимые, чтобы сделать полупроводник прозрачным.

Обратная связь

Обратная связь получается посредством скалывания кристалла вдоль его плоскостей, перпендикулярных плоскости p—n-перехода, или путем полировки двух параллельных поверхностей кристалла. Активная область p—n-перехода, показанная на рис. 17.35, в этом случае служит также плоско-параллельным зеркальным резонатором с длиной d и площадью поперечного сечения lw. Полупроводниковые материалы обычно имеют большие показатели преломления, так что коэффициент отражения по мощности на границе раздела полупроводник—воздух

$$\mathcal{R} = \left(\frac{n-1}{n+1}\right)^2 \tag{17.45}$$

имеет значительную величину [см. (6.41) и табл. 16.5]. Поэтому, если усиление среды достаточно велико, перепад показателя преломления сам по себе служит адекватной отражающей поверхностью и внешние зеркала не требуются. Например, для GaAs n = 3,6, так что (17.45) дает $\mathcal{R}$ = 0,32.

Потери резонатора

Основным источником потерь резонатора является частичное отражение на поверхности кристалла. Эти потери представляют собой полезный испущенный лазерный свет. Для резонатора длины d коэффициент потерь за счет частичного отражения равен [см. (10.22)]

$$\alpha_m = \alpha_{m1} + \alpha_{m2} = \frac{1}{2d} \ln\left(\frac{1}{\mathcal{R}_1 \mathcal{R}_2}\right); \tag{17.46}$$

если обе поверхности имеют одинаковые коэффициенты отражения $\mathcal{R}_1 = \mathcal{R}_2 = \mathcal{R}$, то

$$\alpha_m = \frac{1}{d} \ln\left(\frac{1}{\mathcal{R}}\right).$$

Полный коэффициент потерь выражается как

$$\alpha_r = \alpha_s + \alpha_m, \tag{17.47}$$

где α_s учитывает другие источники потерь, включая поглощение свободными носителями в полупроводниковом материале (см. рис. 16.31) и рассеяние на оптических неоднородностях. Величина α_s растет с увеличением концентрации примесей и дефектов на границах раздела в гетероструктурах. Она может достигать значений от 10 до 100 $см^{-1}$.

Конечно, член $-\alpha$ в выражении для коэффициента усиления (17.43), соответствующий поглощению в материале, тоже вносит существенный вклад в потери. Этот вклад, однако, уже учитывается в выражении (17.44) для полного максимального коэффициента поглощения γ_p. Это очевидно из выражения (16.66) для коэффициента $\gamma_0(\nu)$, который пропорционален $f_g(\nu) = f_e(\nu) - f_a(\nu)$ (т. е. вынужденному излучению минус поглощение).

Другим важным источником потерь является распределение оптической энергии, выходящее за пределы активного слоя усилителя (в направлении, перпендикулярном плоскости перехода). Это особенно вредно сказывается, если толщина активного слоя l мала. Тогда свет распространяется в тонком усиливающем слое (активной области), окруженном поглощающей средой, и большие потери вероятны. Эта проблема может быть сделана менее острой, если использовать двойную гетероструктуру (см. подразд. 17.2.3 и рис. 17.31), в которой средний слой сделан из материала с повышенным показателем преломления и действует как волновод, удерживающий оптическую энергию.

Потери за счет распределения излучения за пределами активной зоны можно учесть феноменологически путем введения **фактора удержания** Г, показывающего, какая часть оптической энергии находится в активной области (рис. 17.36). В предположении, что оказавшаяся за пределами активного слоя энергия полностью пропадает, Г представляет собой коэффициент, показывающий, во сколько раз уменьшается коэффициент усиления и, соответственно,

во сколько раз увеличивается коэффициент потерь. С учетом сказанного уравнение (17.47) нужно заменить на

$$\alpha_r = \frac{1}{\Gamma}(\alpha_s + \alpha_m). \qquad (17.48)$$

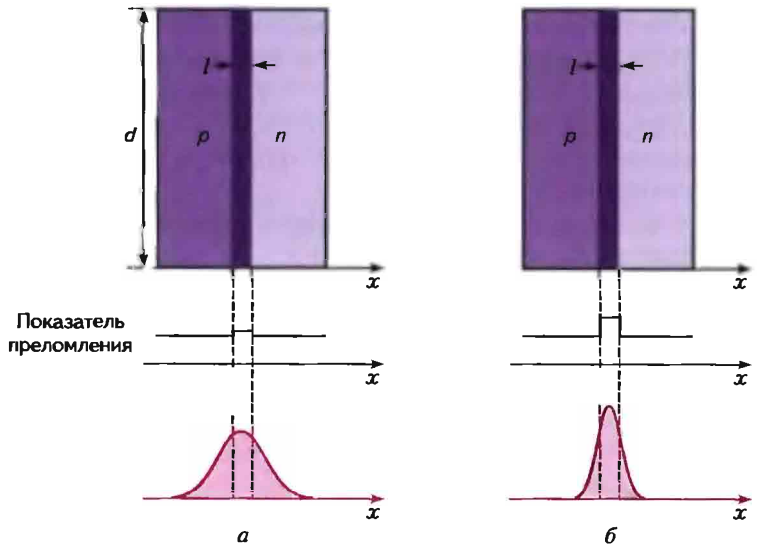


Рис. 17.36. Пространственное расплывание лазерного света в направлении, перпендикулярном плоскости перехода для гомоструктурных (*а*) и гетероструктурных (*б*) лазеров

Существует три типа структур лазерного диода, препятствующих растеканию носителей или света в поперечном направлении (т. е. способствующих их удержанию в плоскости перехода). В этих структурах используется: увеличение площади поперечного сечения активной зоны (при этом удержание становится ненужным), направление за счет усиления (для удержания используется поперечная зависимость коэффициента усиления) и направление за счет рефракции (для удержания используются поперечная зависимость показателя преломления).

Условие усиления: порог генерации

Условие лазерной генерации состоит в превышении усиления над потерями, $\gamma_p > \alpha_r$, в соответствии с (15.12). Таким образом, пороговый коэффициент усиления равен α_r. Подстановка $\gamma_p = \alpha_r$ и $J = J_{\text{п.пл}}$ в (17.44) соответствует пороговому значению плотности тока инжекции $J_{\text{п.пл}}$, выражаемому как

$$\boxed{J_{\text{п.пл}} = \frac{\alpha_r + \alpha}{\alpha} J_{\text{т}},} \qquad (17.49)$$

Пороговая плотность тока

где $J_т$ — плотность тока, при которой среда становится прозрачной

$$J_т = \frac{el}{\eta_i \tau_{рад}} \Delta n_т. \qquad (17.50)$$

Плотность тока, соответствующая прозрачности

Пороговая плотность тока больше той, которая необходима для придания среде прозрачности на множитель $(\alpha_r + \alpha)/\alpha$, который почти равен единице, когда $\alpha \gg \alpha_r$. Поскольку ток $i = JA$, где $A = wd$ — площадь поперечного сечения активного слоя, мы можем ввести величины $i_т = J_т A$ и $i_{п.пл} = J_{п.пл} A$, характеризующие токи, необходимые для достижения прозрачности и порога лазерной генерации соответственно

Пороговая плотность тока $J_{п.пл}$ — ключевой параметр, характеризующий действие лазерного диода; меньшие значения $J_{п.пл}$ соответствуют лучшим эксплуатационным качествам. В соответствии с (17.49) и (17.50), минимальное значение $J_{п.пл}$ достигается за счет максимального увеличения внутреннего квантового выхода η_i и минимизации коэффициента потерь резонатора α_r, концентрации инжектируемых носителей $\Delta n_т$, отвечающей состоянию прозрачности среды, и толщины l активного слоя. Однако при уменьшении l до некоторого значения начинает увеличиваться коэффициент потерь α_r, поскольку снижается коэффициент удержания Г [см. (17.48)]. Вследствие этого $J_{п.пл}$ понижается с уменьшением l лишь до достижения некоторого минимального значения, и попытки дальнейшего уменьшения l вызывают рост $J_{п.пл}$ (рис. 17.37). Однако в лазерах с двойной гетероструктурой коэффициент удержания остается близким к единице и для меньших значений l, поскольку активный слой ведет себя как оптический волновод (см. рис. 17.36). В результате достигается снижение минимального значения $J_{п.пл}$, как показано на рис. 17.37, и, следовательно, более высокие технические характеристики. Далее мы приведем ряд примеров снижения $J_{п.пл}$.

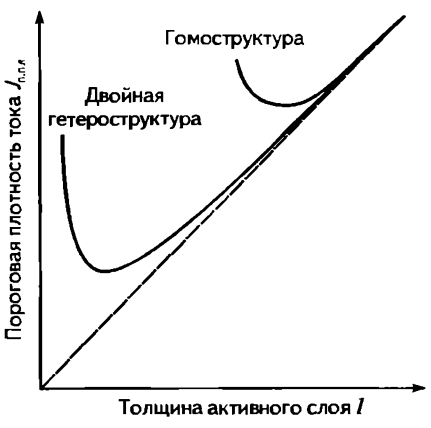


Рис. 17.37. Зависимость пороговой плотности тока $J_{п.пл}$ от толщины активного слоя l. Лазер с двойной гетероструктурой достигает меньших значений $J_{п.пл}$, чем гомоструктурный лазер, и, следовательно, лучших эксплуатационных характеристик. Увеличение $J_{п.пл}$ при малых значениях l происходит в результате худшего удержания света в активном слое

Поскольку параметры $\Delta n_т$ и α в (17.44) зависят от температуры, от нее зависит и пороговая плотность тока $J_{п.пл}$, а также частота, на которой усиление максимально. Для стабилизации выходных характеристик лазера и управления его частотой требуется температурный контроль.

Пример 17.7

Пороговый ток в гомоструктурном InGaAsP лазерном диоде

Рассмотрим гомоструктурный лазерный диод на InGaAsP с такими же материальными параметрами, как в примерах 17.3 и 17.4: $\Delta n_T \approx 1{,}25 \cdot 10^{18}$ см$^{-3}$, $\alpha = 600$ см$^{-1}$, $\tau_{рад} = 2{,}5$ нс, $n = 3{,}5$ и $\eta_i = 0{,}5$ при $T = 300$ °K. Пусть размеры перехода $d = 200$ мкм, $w = 10$ мкм и $l = 2$ мкм. Тогда можно рассчитать плотность тока, необходимую для прозрачности $J_T = 3{,}2 \cdot 10^4$ А/см^2. Определим теперь пороговую плотность тока для лазерной генерации. С помощью (17.45) находим коэффициент отражения на поверхности $\mathcal{R} = 0{,}31$. Соответствующий коэффициент потерь за счет зеркал резонатора

$$\alpha_m = \frac{1}{d}\ln\left(\frac{1}{\mathcal{R}}\right) = 59 \text{ см}^{-1}.$$

Предполагая, что коэффициент всех других потерь тоже $\alpha_s = 59$ см$^{-1}$ и что коэффициент удержания $\Gamma \approx 1$, находим полный коэффициент потерь $\alpha_r = 118$ см$^{-1}$. Тогда пороговая плотность тока

$$J_{п.пл} = \frac{\alpha_r + \alpha}{\alpha} J_T = \frac{118 + 600}{60}\left(3{,}2 \cdot 10^4\right) = 3{,}8 \cdot 10^4 \text{ А/см}^2$$

Соответствующий ток $i_{п.пл} = J_{п.пл} wd \approx 760$ мА, что достаточно много. Гомоструктурные лазеры используются редко из-за трудности достижения стационарного режима генерации без специальных мер охлаждения.

Пример 17.8

Пороговый ток в гетероструктурном InGaAsP лазерном диоде

Обратимся теперь к рассмотрению гетероструктурного лазерного диода на InGaAsP с такими же материальными параметрами и размерами, как в примере 17.7, кроме толщины активного слоя, которая теперь равна $l = 0{,}1$ мкм вместо 2 мкм. Если считать удержание света в активном слое полным ($\Gamma = 1$), можно использовать то же значение коэффициента потерь резонатора α_r. Плотность тока прозрачности тогда уменьшится в 20 раз и станет равной $J_T = 1600$ А/см, а пороговая плотность тока приобретет более разумное значение $J_{п.пл} = 1915$ А/см. Соответствующий пороговый ток равен $i_{п.пл} = 38$ мА. Мы видим существенное снижение порогового тока, которое делает возможной работу лазерного диода с двойной гетероструктурой в стационарном режиме при комнатной температуре.

17.3.2. Мощность и коэффициент преобразования

Внутренний поток фотонов

Когда плотность тока превышает пороговое значение (т. е. $J > J_{п.пл}$), максимальное значение коэффициента усиления γ_p становится больше коэффициента потерь α_r. Вынужденное излучение перевешивает поглощение и

другие потери в резонаторе, начинается лазерная генерация и поток фотонов Ф в резонаторе может увеличиваться. Как и в других лазерах с однородным уширением, рост потока фотонов приводит к насыщению коэффициента усиления из-за снижения разности заселенностей [см. (15.2)]. Как показано на рис. 15.8, коэффициент усиления при этом понижается до тех пор, пока не станет равным коэффициенту потерь, после чего достигается стационарный режим генерации.

Как и внутренняя плотность потока фотонов, а также внутренняя плотность числа фотонов, рассмотренные ранее для лазеров других типов [см. (15.25) и (15.26)], стационарный внутренний поток Ф пропорционален разности между скоростью накачки R и ее пороговым значением $R_{\text{п.пл}}$. Поскольку $R \propto i$ и $R_{\text{п.пл}} \propto i_{\text{п.пл}}$, в соответствии с (17.31) Ф можно записать как

$$\Phi = \begin{cases} \eta_i \dfrac{i - i_{\text{п.пл}}}{e}, & i > i_{\text{п.пл}}; \\ 0, & i \le i_{\text{п.пл}}. \end{cases} \qquad (17.51)$$

Стационарный внутренний поток фотонов

Таким образом, стационарный внутренний поток фотонов (число фотонов, генерируемых в активной среде в секунду) равен превышению потока электронов (число инжектируемых электронов в секунду) над его пороговым значением, умноженному на внутренний квантовый выход η_i.

Внутренняя лазерная мощность над порогом связана с внутренним потоком фотонов Ф простым соотношением $P = h\nu\Phi$, так что получается

$$P = \eta_i \left(i - i_{\text{п.пл}}\right) \frac{1{,}24}{\lambda_0}, \qquad (17.52)$$

Внутренняя лазерная мощность λ_0 (мкм), P (Вт), i (А)

где λ_0 выражено в микрометрах; i — в амперах; P — в ваттах.

Выходной поток фотонов и коэффициент преобразования

Выходной поток фотонов Φ_0 равен произведению внутреннего потока фотонов Ф и коэффициента вывода η_e [см. (15.39)], который равен отношению коэффициента потерь, связанных с выводом полезного излучения через зеркала резонатора, к полному коэффициенту потерь резонатора α_r. Если используется только свет, выходящий через зеркало 1, то $\eta_e = \alpha_{m_1}/\alpha_r$; если же используется свет, выходящий через оба зеркала, то $\eta_e = \alpha_m/\alpha_r$. В последнем случае, если оба зеркала имеют одинаковый коэффициент отражения $\mathcal{R}$, получаем

$$\eta_e = \frac{(1/d)\ln(1/\mathcal{R})}{\alpha_r}.$$

Выходящий из лазера поток фотонов тогда дается выражением

$$\boxed{\Phi_0 = \eta_e \eta_i \frac{i - i_{\text{п.пл}}}{e}.} \qquad (17.53)$$

Выходной поток фотонов лазера

Пропорциональность между выходящим из лазера потоком фотонов и потоком инжектируемых электронов выше порога, выражаемая формулой (17.53), определяется величиной, которая называется **внешним дифференциальным квантовым выходом**

$$\boxed{\eta_d = \eta_e \eta_i.} \qquad (17.54)$$

Внешний дифференциальный квантовый выход

Величина η_d характеризует скорость изменения выходного потока в ответ на изменение потока инжектируемых электронов над порогом:

$$\eta_d = \frac{\mathrm{d}\Phi_0}{\mathrm{d}(i/e)}. \qquad (17.55)$$

Выходная мощность лазера над порогом

$$P_0 = h\nu\Phi_0 = \eta_d (i - i_{\text{п.пл}}) \frac{h\nu}{e},$$

что в более простом виде записывается как

$$\boxed{P_0 = \eta_d (i - i_{\text{п.пл}}) \frac{1{,}24}{\lambda_0},} \qquad (17.56)$$

Выходная мощность лазера λ_0 (мкм), P (Вт), i (А)

при условии, что длина волны λ_0 выражена в микрометрах. Зависимость P_0 от i называется **ватт-амперной характеристикой** лазера. Наклон кривой $P_0(i)$ выше порога называется **дифференциальной чувствительностью** лазера и обычно выражается в Вт/А:

$$\mathcal{R}_d = \frac{\mathrm{d}P_0}{\mathrm{d}i} = \eta_d \frac{1{,}24}{\lambda_0} \quad [\lambda_0 \text{ (мкм)}, \ P \text{ (Вт)}, \ i \text{ (А)}]. \qquad (17.57)$$

Графики зависимости выходной мощности от тока возбуждения показаны сплошными линиями на рис. 17.38 для двух лазерных диодов: диода с КЯ и направленным усилением из InGaAsP/InGaAsP, работающего на 1550 нм (*а*), и диода с КЯ из GaN/InGaN, работающего на 405 нм (*б*). Результаты теоретической подгонки с помощью (17.56) показаны штриховыми линиями.

Параметры этих лазерных диодов можно легко извлечь из данных измерений с помощью (17.56) и (17.57), их значения приведены в табл. 17.1. Хотя внешний дифференциальный квантовый выход η_d почти одинаков для обоих

приборов, дифференциальная чувствительность $\mathcal{R}_d$ для GaN/InGaN-диода примерно в 4 раза выше благодаря более короткой рабочей длине волны, что легко понять на основании (17.57).

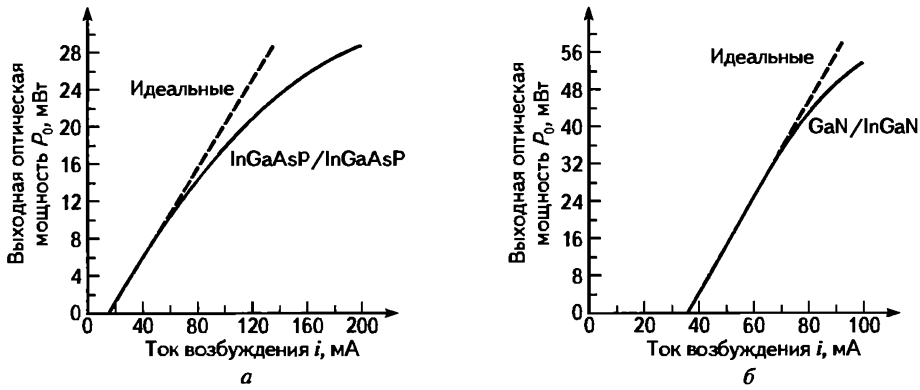


Рис. 17.38. Измеренные (сплошные линии) и идеальные (штриховые линии) ватт-амперные характеристики диода с КЯ и направленным усилением из InGaAsP/InGaAsP, работающего на 1550 нм в ближней инфракрасной области спектра (структура прибора изображена на рис. 17.52) (*а*), и диода с КЯ из GaN/InGaN, работающего на 405 нм в фиолетовой области (*б*). Нелинейности, не учитываемые простой теорией, вызывают насыщение выходной оптической мощности

Таблица 17.1. **Рабочие параметры лазерных диодов, извлеченные из ватт-амперных характеристик в инфракрасной и фиолетовой областях спектра, приведенных на рис. 17.38, *а* и *б* соответственно**

Материал	λ_0, нм	$i_{\text{п.пл}}$, мА	η_d	$\mathcal{R}_d$, Вт/А
InGaAsP/InGaAsP	1550	15	0,33	0,26
GaN/InGaN	405	35	0,33	1,0

Эффективность преобразования мощности (или коэффициент полезного действия) η_c определяется как отношение мощности излучаемого лазером света к потребляемой электрической мощности iV, где V — напряжение прямого смещения, приложенное к диоду. Поскольку $P_0 = \eta_d(i - i_{\text{п.пл}})(h\nu/e)$, имеем

$$\eta_c = \eta_d\left(1 - \frac{i_{\text{п.пл}}}{i}\right)\frac{h\nu}{eV}. \qquad (17.58)$$

Эффективность преобразования мощности

При работе намного выше порога, когда $i \gg i_{\text{п.пл}}$, и при $eV \approx h\nu$ имеем $\eta_c \approx \eta_d$. Лазерные диоды могут иметь коэффициент полезного действия более 50 %, что намного выше, чем у лазеров других типов (см. табл. 15.1). Электрическая

энергия, не преобразуемая в свет, идет на нагревание прибора. Поскольку на практике лазерные диоды выделяют значительное количество теплоты, их обычно устанавливают на радиаторы для отвода тепла и стабилизации температуры.

Пример 17.9

Сравнение эффективности лазерных InGaAsP-диодов на квантовых ямах и двойной гетероструктуре

Рассмотрим еще раз пример 17.8 лазерного InGaAsP/InP-диода с двойной гетероструктурой, у которого $\eta_i = 0{,}5$, $\alpha_m = 59$ см$^{-1}$, $\alpha_r = 118$ см$^{-1}$ и $i_{\text{п.пл}} = 38$ мА. Если используется свет, выходящий с обоих торцов, то коэффициент вывода равен $\eta_e = \alpha_m/\alpha_r = 0{,}5$, а внешний дифференциальный квантовый выход $\eta_d = \eta_e\eta_i = 0{,}25$. При $\lambda_0 = 1300$ нм дифференциальная чувствительность этого лазера $\mathcal{R}_d = dP_0/di = 0{,}24$ Вт/А. Если, например, $i = 50$ мА, имеем $i - i_{\text{п.пл}} = 12$ мА и $P_0 = 12 \cdot 0{,}24 = 2{,}9$ мВт. Сравнение этих результатов с приведенными на рис. 17.38, *а* и в табл. 17.1 для лазерного InGaAsP/InGaAsP-диода на квантовых ямах с рабочей длиной волны 1550 нм показывает, что, как и ожидалось, лазер на КЯ имеет более низкий пороговый ток и более высокий внешний дифференциальный квантовый выход, чем лазер с двойной гетероструктурой.

Резюме

Эффективность лазерного диода характеризуется четырьмя параметрами:

- Внутренний квантовый выход $\eta_i = \tau_{\text{рад}}/\tau = \tau/\tau_{\text{рад}}$, учитывающий то, что не все акты электронно-дырочной рекомбинации являются радиационными.
- Коэффициент вывода η_e, учитывающий то, что используется только часть света, теряемого в резонаторе.
- Внешний дифференциальный квантовый выход $\eta_d = \eta_e\eta_i$, учитывающий оба вышеупомянутых эффекта.
- Коэффициент преобразования мощности (коэффициент полезного действия) η_e, представляющий собой отношение мощности излучаемого света к электрической мощности, потребляемой прибором.

Полезной рабочей характеристикой является также дифференциальная чувствительность $\mathcal{R}_d$ (Вт/А).

Сравнение СИД, СЛД и ЛД по мощности и эффективности

Интересно сравнить СИД, СЛД и ЛД по мощности и эффективности. При работе ниже порога лазерные диоды генерируют спонтанное излучение и ведут себя как светоизлучающие диоды (см. разд. 17.1). Наличие спонтанного излучения видно на ватт-амперных характеристиках при низких значениях тока.

Четыре параметра, характеризующие эффективность действия ЛД, выделены в резюме выше. Эти параметры характеризуют также СИД, как отмечалось в

разд. 17.1. Это внутренний квантовый выход η_i, характеризующий тот факт, что лишь часть актов электронно-дырочной рекомбинации сопровождается излучением фотонов, коэффициент вывода η_e, учитывающий тот факт, что лишь небольшая часть света, генерируемого в области перехода, может покинуть среду с высоким коэффициентом преломления, внешний квантовый выход $\eta_{вн} = \eta_e \eta_i$, учитывающий оба эффекта, и коэффициент преобразования мощности η_e. Чувствительность $\mathcal{R}$ также используется для характеристики рабочих свойств СИД.

Имеется взаимно однозначное соответствие между величинами η_i, η_e и η_c для СИД и ЛД. Более того, имеется соответствие между $\eta_{вн}$ и η_d, $\mathcal{R}$ и $\mathcal{R}_d$, i и $(i - i_{п.пл})$. Превосходство рабочих характеристик лазера связано с большим значением η_e по сравнению с СИД. Причина этого состоит в том, что работа основана на вынужденном излучении, энергия которого сконцентрирована в определенных модах, что облегчает вывод света из прибора. Окончательный результат заключается в том, что лазерный диод над порогом имеет значения η_d большие, чем $\eta_{вн}$ у СИД.

Сверхлюминесцентные лазерные диоды (СЛД), которые работают при токах инжекции, достаточно больших, чтобы вынужденное излучение преобладало над спонтанным, демонстрируют промежуточное поведение между ЛД и СИД. Как обсуждалось в подразд. 17.2.5, обратная связь в этих приборах расстраивается, чтобы предотвратить лазерную генерацию.

Для сравнения рабочих характеристик на рис. 17.39 приведены ватт-амперные характеристики светоизлучающего диода, сверхлюминесцентного диода и лазерного диода. Все они изготовлены на основе структур InGaAsP/InP с квантовыми ямами и работают на длине волны 1600 нм. Чувствительность и эффективность ЛД намного выше, чем у двух других приборов. На врезке видно, что ватт-амперная характеристика для СЛД характерно загибается вверх, тогда как для СИД эта кривая загибается вниз из-за насыщения.

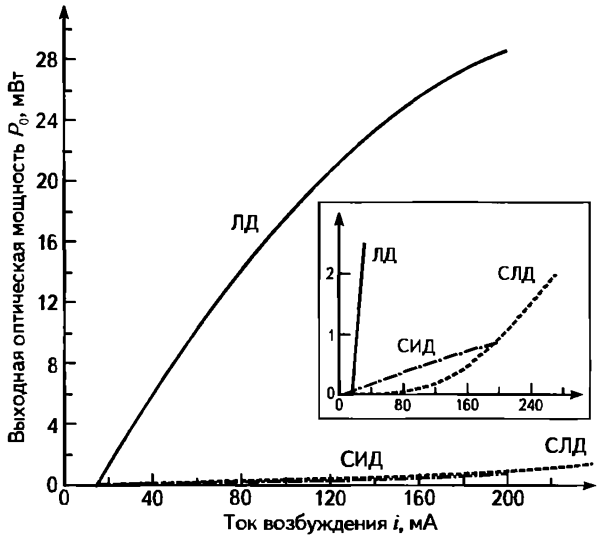


Рис. 17.39. Ватт-амперные характеристики светоизлучающего диода (СИД), сверхлюминесцентного диода (СЛД) и лазерного диода (ЛД). Все три прибора изготовлены на основе структур InGaAsP/InP с квантовыми ямами и работают на длине волны 1600 нм. На врезке в увеличенном масштабе показано поведение кривых для СЛД и СИД

17.3.3. Спектральные и пространственные характеристики

Спектральные характеристики

Спектральная интенсивность лазерного света определяется тремя факторами, как описано в подразд. 15.2.2.

1. Шириной полосы B, в которой коэффициент ненасыщенного усиления активной среды $\gamma_0(\nu)$ превышает коэффициент потерь α_r.

2. Однородной или неоднородной природой уширения спектральной линии (см. подразд. 13.3.4).

3. Модами резонатора, в частности, приблизительным частотным расстоянием между соседними продольными модами $\nu_F = c/2d$, где d — длина резонатора.

В частности, полупроводниковые лазеры характеризуются следующими тремя свойствами.

1. Спектральная ширина коэффициента усиления сравнительно велика, так как переходы происходят между энергетическими зонами, а не уровнями.

2. Внутризонные процессы происходят очень быстро, так что тип уширения в полупроводнике ближе к однородному. Тем не менее выжигание пространственных провалов делает возможной одновременную генерацию множества продольных мод (см. подразд. 15.2.2). Выжигание пространственных провалов особенно существенно в коротких резонаторах с небольшим числом пространственных периодов стоячей волны. При этом поля различных продольных мод, распределенные вдоль оси резонатора, меньше перекрываются, создавая условия для выжигания пространственных провалов.

3. Длина полупроводникового резонатора d значительно меньше, чем у большинства лазеров других типов. Поэтому межмодовое расстояние $\nu_F = c/2d$ значительно больше, чем у других лазеров. Тем не менее на большой ширине полосы B, в пределах которой ненасыщенный коэффициент усиления превышает коэффициент потерь, укладывается большое число продольных мод $M = B/\nu_F$ в соответствии с (15.44).

Пример 17.10

Число продольных мод лазерного диода на InGaAsP

Кристалл InGaAsP (n = 3,5) длиной d = 400 мкм имеет межмодовое расстояние

$$\nu_F = \frac{c}{2d} = \frac{c_0}{nd} \approx 107 \text{ ГГц}.$$

Вблизи центральной длины волны $\lambda_0 = 1300$ нм это межмодовое расстояние соответствует интервалу длин волн λ_F в свободном пространстве, где $\lambda_F/\lambda_0 = \nu_F/\nu$, так что

$$\lambda_F = \frac{\lambda_0 \nu_F}{\nu} = \frac{\lambda_0^2}{2nd} \approx 0{,}6 \text{ нм}.$$

Если спектральная ширина $B = 1{,}2$ ТГц (что соответствует интервалу длин волн $\Delta\lambda = 7$ нм), то одновременно возможна генерация нримерно 11 нродольных мод. Тиничный образец снектрального распределения интенсивности с одной понеречной и 11 продольными модами показан на рис. 17.40. Ширина линии отдельных продольных мод в типичном случае составляет десятки МГц для лазеров со световодом, сформированным нутем раснределения нказателя нреломления, и несколько гигагерц для лазеров со световодом, сформированным нутем раснределения усиления. Полная снектральная ширина света, излучаемого лазерными диодами, больше, чем у большинства других лазеров (см. табл. 14.1). Для уменьшения числа мод до единицы длину резонатора нужно уменьшать, и чтобы достичь $B = c/2d$, требуется резонатор длиной $d \approx 36$ мкм.

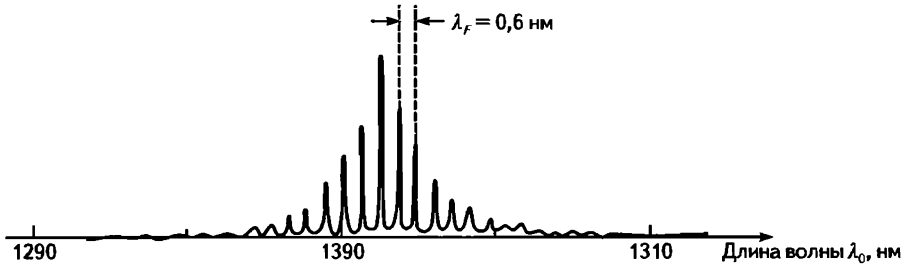


Рис. 17.40. Спектральная интенсивность излучения InGaAsP-лазера со световодом, сформированным путем распределения показателя преломления и скрытой гетероструктурой, работающего на 1300 нм. Это распределение значительно уже и отличается по форме от такового для InGaAsP-СИД с той же рабочей длиной волны $\lambda_0 \approx 1300$ нм (см. рис. 17.60). Число мод уменьшается с ростом тока инжекции: мощность моды, ближайшей к пику усиления, растет, в то время как боковые пики насыщаются. (Заимствовано из *Nelson R.J., Wilson R.B., Wright P.D., Barnes P.A., Dutta N.K.* CW Electrooptical Properties of InGaAsP (λ = 1,3 mm) Buried-Heterostructure Lasers. IEEE Journal of Quantum Electronics. Vol. QE-17, 1981. P. 202—207, Fig. 6 IEEE.)

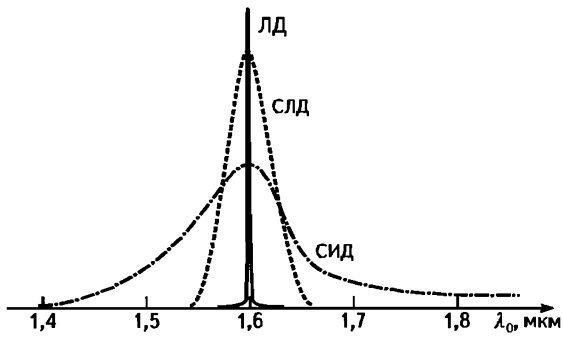


Рис. 17.41. Нормированная спектральная интенсивность для светоизлучающего диода (СИД), сверхлюминесцентного диода (СЛД) и лазерного диода (ЛД). Все три прибора изготовлены на основе структур InGaAsP/InP и работают на 1600 нм. СИД имеет широкий спектр, ЛД — узкий, а СЛД — промежуточный

Сравнение спектральных интенсивностей СИД, СЛД и ЛД

Спектральные интенсивности светоизлучающего диода, сверхлюминесцентного диода и лазерного диода на InGaAsP/InP сравниваются на рис. 17.41. Сужение спектра за счет вынужденного излучения видно для СЛД и в еще большей степени для ЛД.

Пространственные характеристики

Как и у других лазеров, генерация в лазерных диодах происходит в виде продольных и поперечных мод. В подразд. 15.2.3 были использованы индексы (l, m), характеризующие пространственное распределение поля в поперечной плоскости, и индекс q, характеризующий изменения поля в направлении распространения волны или во времени. В большинстве других лазеров световой пучок полностью находится в активной среде, так что поперечное распределение различных мод определяется формой зеркал и расстоянием между ними. Для систем с осевой симметрией поперечные моды могут быть представлены в виде пучков Эрмита—Гаусса или Лагерра—Гаусса (см. подразд. 10.2.4). У полупроводниковых лазеров ситуация иная, поскольку лазерный пучок простирается за пределы активного слоя. Поэтому поперечные моды представляют собой моды диэлектрического волновода, образованного различными слоями лазерного диода.

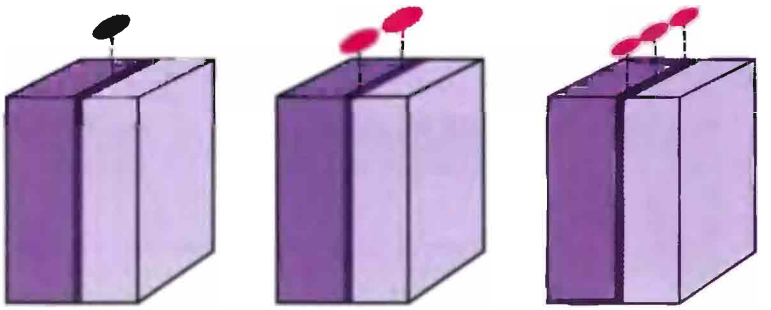


Рис. 17.42. Схема пространственного распределения оптической интенсивности для лазерных волноводных мод (l, m) = (1, 1), (1, 2), (1, 3)

Определить поперечные моды можно с помощью теории, изложенной в разд. 8.3, для оптического волновода с поперечным сечением в виде прямоугольника с размерами l и w. Если l/λ_0 достаточно мало, волновод имеет всего одну моду в поперечном направлении, перпендикулярном плоскости перехода. Однако w обычно намного больше λ_0, так что волновод поддерживает несколько мод в направлении, параллельном плоскости перехода, как показано на рис. 17.42. Моды в направлении, параллельном плоскости перехода, называются **латеральными модами**. Чем больше отношение w/λ_0, тем больше число возможных латеральных мод.

Распределение излучения в дальней зоне

Лазерный диод с размерами активного слоя l и w излучает свет с углом расходимости в дальней зоне $\approx\lambda_0/l$ (радиан) в плоскости, перпендикулярной переходу, и $\approx\lambda_0/w$ в плоскости, параллельной переходу, как показано на рис. 17.43. Это аналогично формуле (3.21) для гауссова пучка диаметром $2W_0$, у которого угол расходимости равен

$$\theta \approx \frac{2}{\pi}\frac{\lambda_0}{2W_0} = \frac{\lambda_0}{\pi W_0}, \quad \theta \ll 1.$$

Угловая расходимость определяет распределение поля в дальней зоне, как обсуждалось в разд. 4.3. Из-за малых размеров активного слоя лазерный диод характеризуется большей угловой расходимостью, чем большинство других тинов лазеров. Например, при $l = 2$ мкм, $w = 10$ мкм и $\lambda_0 = 800$ нм рассчитанные углы расходимости составляют ≈23° и 5° Свет от лазерного диода с одной поперечной модой, у которого w меньше, имеет еще большую угловую расходимость. Пространственное распределение света в дальней зоне в пределах конуса излучения зависит от числа поперечных мод и их относительно оптической мощности. Сильно асимметричное эллиптическое распределение излучения лазерного диода превращает его коллимацию в непростую задачу.

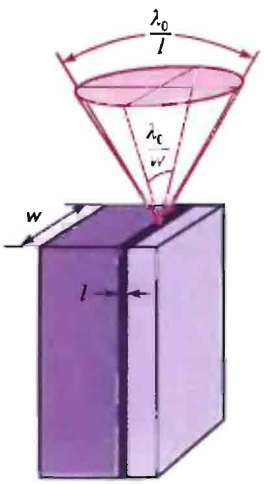


Рис. 17.43. Угловое распределение оптического пучка, испущенного лазерным диодом

Одномодовый режим

Поскольку латеральные моды высоких порядков имеют более широкие распределения поля, они менее локализованы и, следовательно, их коэффициент потерь α_r больше, чем у мод низших порядков. Следовательно, для некоторых поперечных мод наивысших порядков условия генерации оказываются невыполненными, а генерация других будет происходить с меньшей мощностью, чем у основной моды самого низкого порядка. Для достижения одномодовой генерации с высокой мощностью нужно уменьшать число мод волновода путем сокращения размеров сечения активного слоя (l и w), чтобы сделать

волновод одномодовым. Сопутствующее сокращение площади перехода понижает также пороговый ток. Латеральные моды высших порядков можно исключить, используя конфигурации лазерного диода со световодом, сформированным путем распределения усиления или показателя преломления.

Режим одной продольной моды, обеспечивающий одночастотное выходное излучение, может быть достигнут путем сокращения длины резонатора d до значений, при которых частотное расстояние между соседними продольными модами превышает спектральную ширину усиления активной среды. Одномодовый режим можно получить также с использованием многозеркальных резонаторов, как обсуждалось в подразд. 15.2.4 и показано на рис. 15.22.

Другой путь к достижению одномодовой генерации включает использование распределенных отражателей вместо сколотых граней кристалла, служащих сосредоточенными зеркалами резонатора Фабри—Перо. При обеспечении такой обратной связи на поверхности кристалла наносится антиотражающее покрытие, минимизирующее отражение. Например, в плоскости перехода можно поместить частотно-селективные отражатели, такие как решетки Брэгга (рис. 17.44, *а*). Как обсуждалось в подразд. 2.4.2 и 7.1.3, решетка Брэгга отражает свет, когда период решетки Λ удовлетворяет условию $\Lambda = q\lambda/2$, где q — целое число. Прибор, изображенный на рис. 17.44, *а*, называется **лазером с распределенными брэгговскими отражателями (РБО-лазером)**. В качестве альтернативы РБО может располагаться над или под активной областью, как показано на рис. 17.44, *б*.

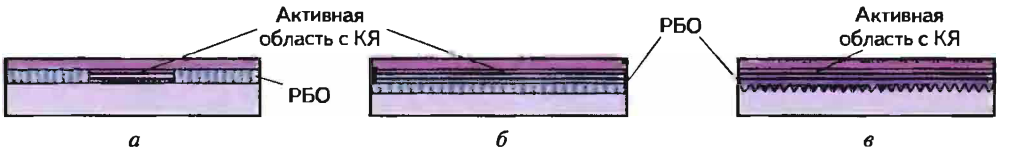


Рис. 17.44. Схема лазерного диода на квантовых ямах с РБО-зеркалами вне активной зоны (*а*). Схема РОС-лазера на квантовых ямах с РБО-структурой, расположенной под активным слоем и обеспечивающей распределенное отражение (*б*). Структура для создания распределенной обратной связи (РОС) в лазерном диоде на квантовых ямах с гофрированной границей между активным и волноводным слоями, служащей распределенным отражателем (*в*)

Еще один метод обеспечения одномодового режима основан на использовании гофрированной границы раздела активного и волноводного слоев, как показано на рис. 17.44, *в*, в результате чего возникают периодическая модуляция показателя преломления и, следовательно, дифракционная решетка. Структуры такого типа называются **лазерами с распределенной обратной связью (РОС-лазерами)**. Этот класс лазеров обеспечивает малую спектральную ширину и широкополосную модуляцию. Они широко используются как источники света в оптических системах связи в диапазоне длин волн 1300—1600 нм (см. подразд. 17.4.3).

17.4. КВАНТОВО-РАЗМЕРНЫЕ ЛАЗЕРЫ И ЛАЗЕРЫ С МИКРОРЕЗОНАТОРАМИ

17.4.1. Квантово-размерные лазеры

Квантово-размерные лазеры, в которых носители заключены в областях с размерами меньше длины волны де Бройля тенлового электрона (≈50 нм в GaAs), имеют нревосходные рабочие характеристики и являются самыми распространенными из всех полупроводниковых лазерных диодов. Удержание носителей в одном, двух и трех измерениях соответствует конфигурациям тина квантовой ямы, квантового проволоки и квантовой точки (рис. 17.45). Несколько нримеров ЛД, СЛД, СИД и ПОУ на квантовых ямах уже ноявлялись ранее.

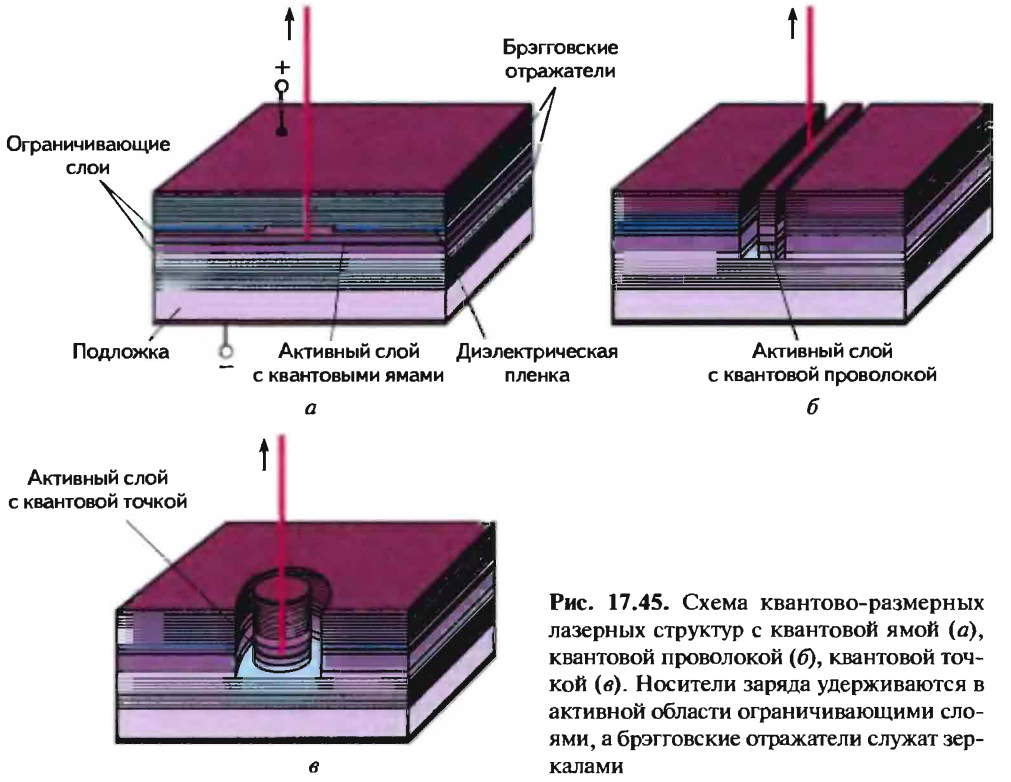


Рис. 17.45. Схема квантово-размерных лазерных структур с квантовой ямой (*а*), квантовой проволокой (*б*), квантовой точкой (*в*). Носители заряда удерживаются в активной области ограничивающими слоями, а брэгговские отражатели служат зеркалами

С уменьшением размерности нолупроводниковой структуры снектральный контур коэффициента усиления, как нравило, становится выше и у́же, обеспечивая меньший пороговый ток, больший внешний дифференциальный квантовый выход и более узкую линию излучения. Однако в то же время уменьшается объем области взаимодействия, что приводит к снижению выходной мощности лазеров на квантовых проволоках и квантовых точках.

В данном разделе мы последовательно обсудим полупроводниковые лазеры на квантовых ямах, квантовых проволоках и квантовых точках. Затем мы перейдем к квантово-каскадным лазерам, представляющим собой специально спроектированные униполярные приборы с множественными квантовыми ямами, которые позволяют получать значительную оптическую мощность в инфракрасной области спектра.

Лазеры на квантовых ямах

Как было выяснено в разд. 17.2 и 17.3, прибор на квантовых ямах, схема которого изображена на рис. 17.45, *а*, имеет существенно лучшие рабочие характеристики, чем прибор с двойной гетероструктурой. Выигрыш получается благодаря малой толщине отдельной квантовой ямы, которая обычно меньше 10 нм (сравните с ≈100 нм для диода с двойной гетероструктурой и ≈2 мкм для старомодного диода на гомопереходе).

Сравнение зависимостей пикового коэффициента усиления γ_p от плотности тока J для полупроводниковых лазеров с единичной квантовой ямой (ЕКЯ) и объемной двойной гетероструктурой (ДГ) приведено на рис. 17.46. Лазер с квантовой ямой имеет намного меньшее значение J_Γ (плотность тока, при которой достигается прозрачность), хотя его усиление насыщается на более низком уровне.

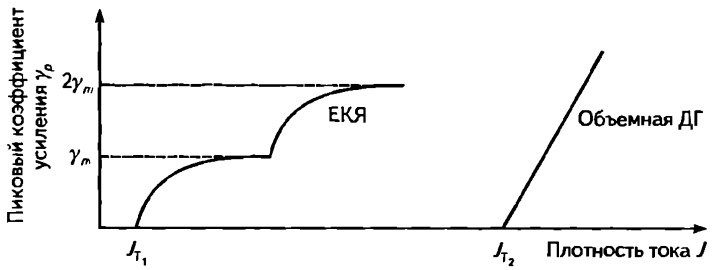


Рис. 17.46. Зависимость пикового коэффициента усиления γ_p от плотности тока J для полупроводниковых лазеров с единичной квантовой ямой (ЕКЯ) и объемной двойной гетероструктурой (ДГ). Пиковый коэффициент усиления для лазера с ЕКЯ резко растет, а затем насыщается при значениях, кратных γ_m [см. (17.42)]

Лазер на КЯ имеет следующие преимущества по сравнению со своим конкурентом на ДГ:

- меньшая пороговая плотность тока;
- больший внешний дифференциальный квантовый выход;
- больший коэффициент преобразования мощности;
- меньшая ширина контура усиления;
- меньшая ширина лазерных мод;
- меньшее время отклика, позволяющее использовать большие частоты модуляции;
- меньшая зависимость от температуры.

Лазеры на множественных квантовых ямах

Лазер на множественных квантовых ямах (МКЯ) (рис. 17.47) имеет больший коэффициент усиления, чем лазер с единичной квантовой ямой (ЕКЯ). Действительно, коэффициент усиления МКЯ-лазера с N квантовыми ямами в N раз больше, чем с каждой из них в отдельности. Для объективного сравнения рабочих характеристик этих двух нриборов, однако, необходимо, чтобы в обоих из них накачка была одинакова. Рассмотрим отдельную квантовую яму, в которой за счет инжекции обеснечена избыточная концентрация носителей Δn и никовый коэффициент усиления γ_p.

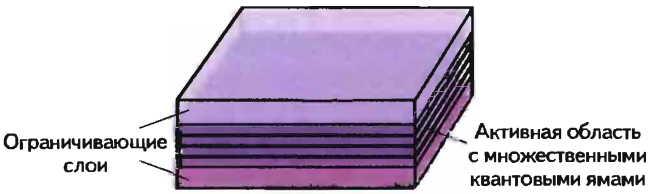


Рис. 17.47. Схема активной области лазера с множественными квантовыми ямами. Ограничивающие слои удерживают носители заряда в области квантовых ям

В структуре с МКЯ в каждую из N ям будет инжектировано только $\Delta n/N$ носителей. Однако, из-за нелинейной зависимости усиления от Δn коэффициент усиления каждой ямы равен $\xi\gamma_p/N$, где ξ может быть как меньше, так и больше единицы, в зависимости от рабочего режима. Тогда нолное усиление лазера с МКЯ равно $N(\xi\gamma_p/N) = \xi\gamma_p$. Получается, что ЕКЯ в типичном случае имеет преимущество нри низких плотностях тока, в то время как МКЯ обычно работают лучше при высоких плотностях тока, но фактор выигрыша меньше, чем N.

Лазеры с напряженными слоями

Приложение механического нанряжения может оказать ноложительное воздействие на работу лазерных диодов, хотя этот факт нротиворечит интуиции. Лазеры с нанряженными слоями могут иметь улучшенные свойства и работать на длинах волн, недоступных с номощью изменения состава нолунроводниковых соединений. Квантово-размерные лазеры с нанряженными слоями были изготовлены из нолупроводниковых материалов III—V групп в различных конфигурациях. Вместо полного согласования решеток с ограничиваюшими слоями активный слой умышленно нодбирается с другой постоянной решетки. Если он достаточно тонок, то расстояния между его атомами могут нодстраиваться под окружающие слои, в процессе чего возникают механические напряжения (если активный слой слишком толстый, он не приспосабливается к окружению, и материал будет содержать дислокации). Например, нанряженный активный слой InGaAs в лазере на AlGaAs/InGaAs с квантовыми ямами имеет значительно большую собственную постоянную решетки, чем у ограничивающих слоев AlGaAs. Поэтому тонкий слой InGaAs испытывает двуосное сжатие в плоскости слоя, а межатомные расстояния в нанравлении, перпендикулярном к слою, увеличиваются.

Напряжение сжатия вносит в зонную структуру три существенных изменения:

1) оно увеличивает ширину запрещенной зоны E_g;

2) оно снимает вырождение при $k = 0$ между зонами легких и тяжелых дырок;

3) оно делает валентные зоны анизотропными, так что наивысшая зона имеет малую эффективную массу в направлении, параллельном плоскости слоя, и большую — в перпендикулярном направлении.

Все это может значительно улучшить работу лазеров. Во-первых, лазерная длина волны перестаивается из-за зависимости E_g от напряжения. Во-вторых, пороговая плотность тока может быть уменьшена за счет напряжения. Для достижения инверсии заселенностей необходимо, чтобы расстояние между квазиуровнями Ферми было больше, чем ширина запрещенной зоны, т. е. $E_{f_c} - E_{f_v} > E_g$ [см. (16.44)]. Уменьшение массы дырок позволяет E_{f_v} легче онускаться в валентную зону, обеспечивая таким образом выполнение этого условия при более низких значениях тока инжекции.

Лазеры на одной и нескольких квантовых проволоках

Квантовые проволоки (см. подразд. 16.1.7) также могут служить активной областью полупроводникового лазера, как показано на рис. 17.45, *б*. Лазеры на квантовых проволоках включают наборы квантовых проволок, как изображено на рис. 17.48. В принципе лазеры на квантовых проволоках имеют более узкие линии, чем лазеры на квантовых ямах, благодаря более тесному удержанию носителей. Однако технология изготовления структур на основе квантовых проволок из элементов III—V групп далеко отстает от таковой для структур на основе квантовых ям, что частично объясняется трудностью создания наборов проволок, расположенных достаточно тесно. Отстают и рабочие характеристики лазеров на квантовых проволоках.

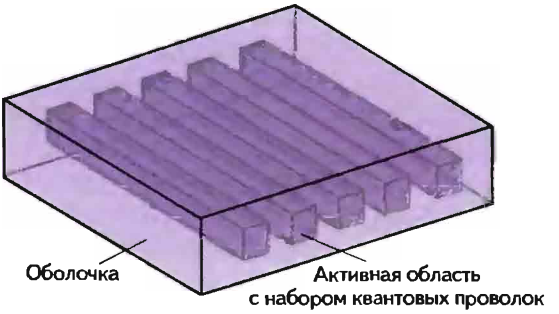


Рис. 17.48. Схема активной области лазера на множестве квантовых проволок. Свет обычно излучается во всех направлениях, хотя лазерное излучение можно ограничить торцевыми поверхностями с применением подходящего резонатора

Пример 17.11

Рабочие характеристики лазеров на квантовых проволоках и квантовых ямах

Активный слой из пяти квантовых проволок длиной 1 мм и толщиной 23 нм из InGaAsP, погруженных в оболочку из InP на расстоянии 80 нм друг от

друга, работает при комнатной темнературе как лазер ненрерывного действия на длине волны $\lambda_0 \approx 1550$ нм. Пороговый ток, пороговая плотность тока, внешний дифференциальный квантовый выход и коэффициент нреобразования мощности найдены равными $i_{\text{п.пл}} = 140$ мА, $J_{\text{п.пл}} = 800$ А/см, $\eta_d = 40\,\%$ и $\eta_c = 2\,\%$ соответственно[2]. Однако из-за малого объема активной области и существенных оптических нотерь рабочие характеристики этого лазера устунают таковым для лазера на квантовых ямах, изготовленного на том же чипе: $i_{\text{п.пл}} = 100$ мА, $J_{\text{п.пл}} = 500$ А/см, $\eta_d = 50\,\%$ и $\eta_c = 6\,\%$.

Лазеры на одной и нескольких квантовых точках

Кваитовые точки, называемые также **квантовыми ящиками** и иногда **нанокристаллами**, обычно имеют форму кубов, сфер или пирамид. Их тиничные размеры составляют 1—10 нм (куб с ребром 10 нм содержит около 40 000 атомов GaAs). Удержание носителей в активной области можно обеспечить, окружив квантовую точку оболочкой из нолупроводника с большей шириной запрещенной зоны или погрузив ее в стекло или полимер. На рис. 17.11, *в* изображена схема лазера на квантовой точке.

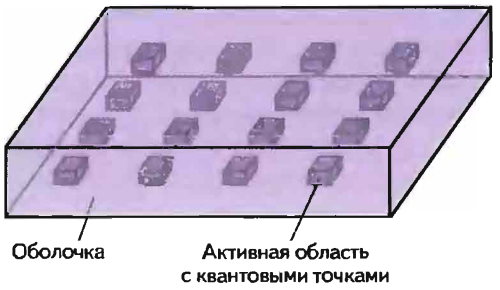


Рис. 17.49. Схема активной области лазера на квантовых точках, которая часто состоит из нескольких слоев. Каждый слой содержит самоупорядоченное множество квантовых точек. Типичные размеры квантовых точек при самоупорядочении попадают в интервал 10—50 нм

Энергетическими уровнями квантовой точки являются уровни ее экситонов. Хотя уровни дискретные в результате тесного удержания носителей, энергии сильно зависят от размера точки. Как впечатляюще показано на рис. 13.12, энергия фотолюминесцентных фотонов растет с убыванием размеров точки, носкольку для удержания возбуждения нолупроводника в меньшем объеме требуется большая энергия (см. нодразд. 13.1.3). Такая перестраиваемость является благоприятным качеством при иснользовании квантовых точек в качестве активной среды. Набор квантовых точек, каждая из которых вносит свой вклад, может обеспечить достаточно высокий уровень лазерной мощности. Поскольку квантовые точки часто самоорганизуются в нростанственно-упорядоченные структуры, из них нетрудно создать лазер на сверхрешетке из квантовых точек, значительное число которых находится в активной зоне, как показано на рис. 17.49.

[2] См. *Yagi H., Sano T., Ohira K., Plumwongrot D., Maruyama T., Haque A., Tamura S., Arai S.* GaInAsP/InP Partially Strain-Compensated Multiple-Quantum-Wire Lasers Fabricated by Dry Etching and Regrowth Processes. Japanese Journal of Applied Physics. Vol. 43, 2004. P. 3401—3409.

Структуры из квантовых точек обеспечивают гибкость при проектировании приборов; они, например, могут использоваться для генерации широкополосного света или служить оптическими усилителями.

Пример 17.12

Кремниевая фотоника квантовых точек

Удержание носителей в квантовой точке означает уменьшение неопределенности их координаты Δx. Поскольку $\Delta x \Delta k \geq 1/2$, в соответствии с формулой (A.13) приложения А, это сопровождается соответствующим увеличением неопределенности волнового числа Δk. Рост Δk устраняет необходимость участия фонона в радиационной рекомбинации. В этом смысле использование квантовых точек аналогично включению азотных примесных центров в непрямозонный GaP, благодаря чему становится реальностью СИД на GaP:N (см. подразд. 17.1.3). Поэтому благодаря малым размерам квантовых точек непрямозонные полупроводники имеют гораздо больший квантовый выход, чем их объемные аналоги. Пассивация поверхности также увеличивает скорость радиационных переходов за счет поверхностно-локализованных экситонов. В результате излучение света наночастицами кремния, а также пористым кремнием и германием приобретает практическое значение. Предпринимаются попытки получения лазерной генерации в кремниевых наноструктурах.

Квантово-каскадные лазеры

Все рассмотренные до сих пор полупроводниковые лазеры работают за счет радиационной электронно-дырочной рекомбинации. Получение света представляет собой однофотонный процесс с участием двух носителей заряда: рекомбинация электрона в зоне проводимости с дыркой в валентной зоне порождает фотон. В **квантово-каскадном лазере (ККЛ)**, напротив, используется лишь один носитель заряда — электрон, но каждый электрон порождает много фотонов. ККЛ, следовательно, является однополярным, а не биполярным устройством. Квантово-каскадные лазеры содержат последовательность квантовых ям, спроектированных и смещенных таким образом, что электрон, инжектированный в зону проводимости, совершает каскад радиационных переходов между подзонами по мере прохождения через прибор. ККЛ, возможно, является одним из первых примеров **зонной инженерии**.

Квантово-каскадные лазеры, как правило, имеют активные области либо на квантовых ямах, либо на сверхрешетках. Как показано на рис. 17.50, лазер с квантовыми ямами имеет ряд ступеней, каждая из которых включает инжектор электронов *n*-тина и активную область с квантовыми ямами из собственного полупроводника. Инжектор содержит набор ям переменной ширины, разделенных тонкими барьерами и образующих сверхрешетку, энергетический спектр которой состоит из разрешенных мини-зон, разделенных запрещенными минизонами (см. подразд. 13.1.3). Число состояний в мини-зоне равно числу квантовых ям. При наличии смещения электроны инжектируются посред-

ством резонансного туннелирования из нижнего (основного) состояния мини-зоны, обозначенного как уровень 3, на верхний лазерный уровень в активной зоне с квантовыми ямами, обозначенный как уровень 2. При вынужденном переходе между подзонами 1 и 2 испускается фотон с частотой $\nu = E_{12}/h$, что показано красной стрелкой (см. подразд. 16.2.4). Затем за счет рассеяния на фононах электрон без излучения релаксирует на уровень 0, после чего входит в мини-зону инжектора следующей ступени в результате резонансного туннелирования. На этой ступени процесс повторяется и испускается еще один фотон. Типичный ККЛ содержит 20—100 ступеней, так что при прохождении каждого электрона через прибор генерируется значительное число фотонов.

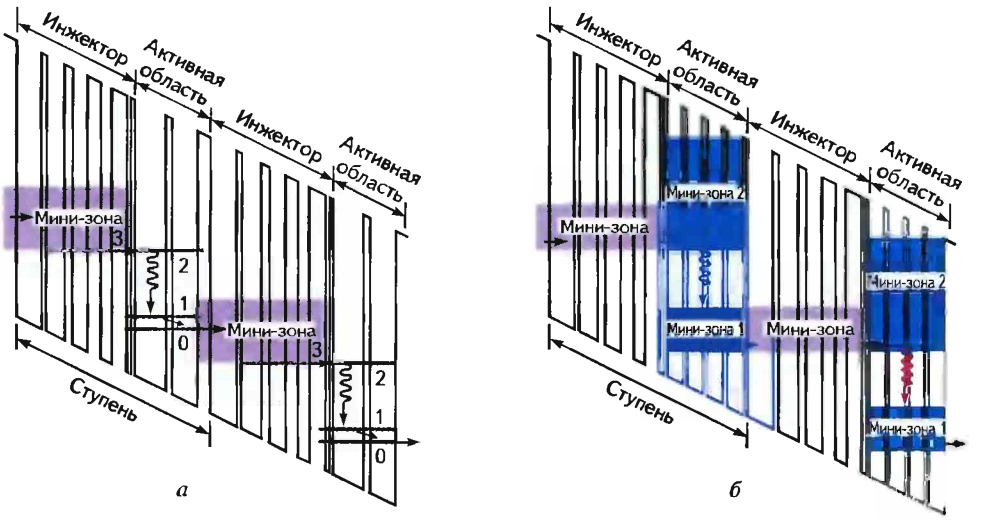


Рис. 17.50. Схемы двух ступеней ККЛ с квантовыми ямами в активной зоне (*а*) и со сверхрешеткой в активной зоне (*б*). ККЛ обычно имеют от 20 до 100 ступеней: полная длина типичного прибора составляет 0,5—3 мм, а ширина от 5 до 20 мкм. Квантово-каскадные лазеры часто делаются с квантовыми ямами из AlInAs/InGaAs, согласованного по постоянной решетке с подложкой из InP, или с квантовыми ямами из AlGaAs/GaAs с использованием молекулярно-пучковой эпитаксии или химического осаждения из паровой фазы методом разложения металлоорганических соединений. Из других материалов представляют интерес AlAsSb/InGaAs, AlGaN/GaN и SiGe

Ввиду использования переходов между подзонами действие ККЛ с квантовыми ямами близко напоминает работу атомного лазера. Как видно из рис. 17.50, *а*, уровень 2 не совпадает с мини-зоной следующей ступени, поэтому он имеет достаточно большое время жизни ($\tau_2 \approx 1$ пс) и, следовательно, накапливает заселенность. Уровень 1, напротив, не сохраняет заселенность, поскольку безызлучательный переход на уровень 0 происходит быстро, а за ним следует туннелирование на следующую ступень ($\tau_1 \approx 0{,}1$ пс). Таким образом, активная зона с квантовыми ямами ведет себя как четырехуровневая лазерная система, в которой инверсия заселенностей достигается на переходе $2 \rightarrow 1$ (см. подразд. 14.2.2).

ККЛ со сверхрешеткой, показанный на рис. 17.50, *б*, отличается от ККЛ с квантовыми ямами тем, что вынужденное излучение происходит при переходе между дном верхней и потолком нижней мини-зоны в активной области, которая в этом случае выполнена в виде сверхрешетки (см. подразд. 16.2.4). Частота лазера определяется шириной запрещенной мини-зоны. Такая структура в общем случае лучше подходит для генерации когерентного света на больших длинах волн ($\lambda_0 \geq 10$ мкм), поскольку совпадение между инжекторной и активной областями менее критично. Более того, можно использовать большие токи возбуждения и инверсия заселенностей достигается легче из-за быстрой релаксации в нижней мини-зоне лазерного перехода. Еще одним вариантом активной области ККЛ является так называемая связанно-свободная схема, где в лазерной генерации участвуют переходы с дискретного верхнего уровня в мини-зону сверхрешетки. Такая схема сочетает эффективную инжекцию электронов на верхний лазерный уровень, присущую ККЛ с квантовыми ямами, с быстрым опустошением нижнего лазерного уровня, характерным для ККЛ на сверхрешетках, что понижает порог и увеличивает мощность.

ККЛ могут работать в громадном диапазоне длин волн от 2 до 70 мкм в средней и дальней ИК-области, что достигается выбором ширины ям, которая, в свою очередь, определяет энергетические уровни подзон и мини-зон (см. упражнение 16.5). Обратная связь с помощью внешнего резонатора в сочетании с поворотной решеткой (см. подразд. 15.2.4) обеспечивает менее грубую перестройку длины волны в пределах около 10 % центральной длины волны (≈1 мкм при центральной длине волны в 10 мкм). Тонкая подстройка длины волны (≈0,1 мкм при центральной длине волны в 10 мкм) может быть достигнута путем изменения тока инжекции и/или температуры, которая меняет эффективный показатель преломления — это, в свою очередь, меняет оптическую длину пути в резонаторе и, таким образом, длину волны излучения. Одномодовый режим достигается при снабжении прибора элементом распределенной обратной связи.

ККЛ среднего ИК-диапазона работают при комнатной температуре и излучают от нескольких до сотен мини-ватт в непрерывном режиме[3]. В них достигается высокий внешний квантовый выход и низкий порог, поскольку один носитель может порождать много фотонов; эти приборы выдерживают сильные токи, поскольку нет необходимости делать их из материалов с узкой запрещенной зоной. ККЛ можно модулировать с очень высокой частотой, в них можно получить режим синхронизации мод, дающий генерацию оптических импульсов длительностью несколько фемтосекунд. Их также можно использовать как источники терагерцового излучения мощностью в десятки миливатт в непрерывном режиме при условии охлаждения. Генерацию от ККЛ удается получить на длинах волн до 150 мкм в дальней ИК-области ($\nu = 2$ ТГц).

---

[3] Хотя распространенные лазерные диоды на солях свинца могут работать на длинах волн до ≈30 мкм, они имеют ряд недостатков, отсутствующих у ККЛ:

1) уровни мощности ограничены несколькими милливаттами;

2) длина волны излучения перестраивается лишь в небольших пределах;

3) они нуждаются в охлаждении.

Разработано множество других конструкций ККЛ, включая приборы на сверхрешетках, в которых инжекторная область исключена, приборы, в которых свет направляется в виде поверхностно-плазмонных мод, так что генерация длинных волн достигается без необходимости в толстых диэлектрических волноводах, приборы, работающие одновременно на многих длинах волн, приборы, генерирующие суперконтинуум за счет различной ширины квантовых ям в активных областях различных ступеней, и, наконец, ВКР-лазеры с инжекционной накачкой от квантово-каскадного лазера, интегрированного в ту же структуру.

Развитие ККЛ в сочетании с приемниками среднего ИК-диапазона, работающими при комнатной температуре, таких как фотоэлектрические батареи из HgCdTe и микроболометры из VO_x (см. разд. 18.5), открывает двери к изобилию научных, промышленных и военных приложений в среднем и дальнем ИК-диапазоне. Поскольку эти диапазоны соответствуют колебательно-вращательным переходам молекул (см. подразд. 13.1.2), приложения включают анализ газов на микроэлементы, химическое зондирование, изотопный анализ и инфракрасную спектроскопию. ККЛ представляются также для безопасной для зрения беспроводной оптической связи в среднем ИК-диапазоне.

17.4.2. Лазеры с микрорезонаторами

Квантовое ограничение, рассмотренное в разд. 17.4, относилось к *носителям заряда*, свобода которых ограничивалась пространственной областью с размерами порядка *длины волны де Бройля* электрона (для теплового электрона в GaAs $\lambda \approx 50$ нм). В отличие от этого **лазеры с микрорезонаторами**, рассматриваемые в данном разделе, включают ограничение свободы *фотонов* пространственной областью, размеры которой порядка *оптической длины волны* ($\lambda_0 \approx 1$ мкм $\gg \lambda$). *Микрорезонаторы* (microresonators) — это резонаторы, у которых один или несколько пространственных размеров порядка нескольких длин волн света или менее $d \approx \lambda_0$. Для микрорезонаторов, все размеры которых малы, в англоязычной литературе применяется термин *microcavities* (микрополости), однако оба термина часто употребляются как взаимозаменяемые, и, в соответствии со сложившейся русскоязычной терминологией, мы будем всегда использовать первый из них.

Ограничение свободы фотонов и носителей — разные свойства приборов фотоники. Можно создать *микрорезонаторный* лазер, активная зона которого *не подвержена* квантово-размерным эффектам (например, микрорезонатор, содержащий простой *p*—*n*-гомопереход), или лазер с *большим* резонатором, у которого активная область *является квантово-размерной* (например, квантово-каскадный лазер). Однако на практике большинство лазеров с микрорезонаторами имеют в активной области квантово-размерные структуры. Несколько типичных примеров лазеров с микрорезонаторами обсуждается в подразд. 17.4.3.

Примерами лазеров с микрорезонаторами, ограничивающими свет на волновых масштабах в различном числе измерений, являются представленные на рис. 17.51 структуры в виде микростолбика, микродиска и микросферы. Эти и другие микрорезонаторы описаны в подразд. 10.4.2 и 10.4.3.

В лазерных диодах с большими резонаторами ($d \gg \lambda$) межмодовые расстояния малы во всех направлениях — пространства и плотность числа возможных резонансных частот $M(\nu)$ можно определять с помощью континуального приближения (см. разд. 10.3). Полная вероятность спонтанного излучения ($с^{-1}$) зависит от плотности числа мод $M(\nu)$, в которые атом может излучать в соответствии с (13.22). В лазерах с большими резонаторами, как и в свободном пространстве, плотность числа мод принимает квадратичный вид $M(\nu) = 8\pi\nu^2/c^3$ в соответствии с (10.82). Это дает большое число мод для спонтанного испускания, однако спонтанное излучение в моды, отличные от лазерной, представляет собой потери энергии после того, как в лазерной моде инициировано вынужденное излучение. В действительности для типичного лазерного диода с торцевым излучателем доля β спонтанного излучения, вносимая в данную лазерную моду, очень мала ($\beta \approx 10^{-5}$). Таким образом, ток, инжектируемый в лазер с большим резонатором в пороговом режиме, по существу расходуется на компенсацию потерь энергии из-за спонтанного излучения, а не на наращивание вынужденного излучения.

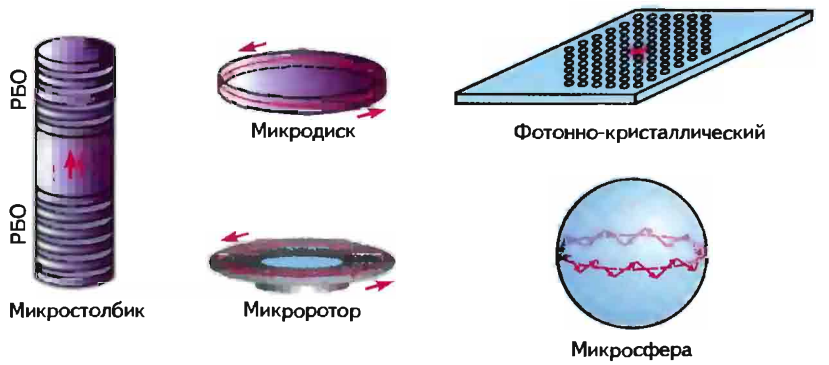


Рис. 17.51. Лазеры с микрорезонаторами ограничивают свет в областях размером порядка длины волны в различных измерениях. Дефект в двумерном фотонном кристалле создает резонатор, который удерживает свет. Аналогичными структурами с квантовым ограничением свободы носителей являются квантовая яма, квантовая проволока и квантовая точка

Однако плотность числа мод $M(\nu)$ можно существенно уменьшить за счет использования микрорезонатора, как обсуждалось в разд. 10.4. Собственные моды микрорезонатора могут иметь большие межмодовые расстояния в одном или нескольких направлениях k-пространства, так что в обширных областях спектра частот мод вообще нет. Это сокращение особенно сильно в трехмерных микрорезонаторах (микрополостях), где межмодовые расстояния велики во всех направлениях k-пространства, что приводит к образованию разреженного дискретного набора мод (см. рис. 10.26). Возможность управления модовым окружением важна в связи со спонтанным излучением. Помещение источника в такое окружение подавляет спонтанное испускание в несуществую-

щие моды и перенаправляет его в немногие имеющиеся моды. Излучение света в отдельные моды высокодобротного микрорезонатора малого объема может быть значительно увеличено по сравнению с излучением в обычную оптическую моду из-за эффекта Перселла, как описано в подразд. 13.3.5. Лазеры с микрорезонаторами проектируются с расчетом на максимальное использование возможности такого подавления и усиления. Модификация плотности числа мод, обеспечиваемая микрорезонатором, может увеличить коэффициент вклада спонтанного излучения β на несколько порядков величины, снижая таким образом пороговый ток лазера $i_{\text{п.пл}}$ на соизмеримую величину (т. е. микроамперы вместо миллиампер).

Хотя лазеры с микрорезонаторами большей частью полупроводниковые, они могут изготавливаться также из таких материалов, как органические красители, кварц, легированный редкоземельными элементами и органические полимеры.

Резюме

Микрорезонаторные лазеры имеют ряд преимуществ перед своими обычными аналогами:

- меньшие размеры;
- более низкий порог генерации;
- меньшую спектральную ширину;
- меньшую пространственную ширину;
- высокую эффективность.

Однако малые размеры означают низкую мощность излучения.

Пример 17.13

Единичные фотоны от квантовой точки по требованию

Отдельные квантовые точки могут служить источниками единичных фотонов, испускаемых по требованию, если их возбудить оптическим или электрическим путем. Этот процесс можно конструктивно реализовать в нескольких вариантах. Например, отдельная квантовая точка InAs, помещенная в полость резонатора-микростолбика (см. рис. 17.51), генерирует спонтанно испущенные фотоны преимущественно в направлении наружу через волоконно-оптический соединитель на верхней грани прибора. Фактор Перселла усиления спонтанного излучения (13.58) велик, поскольку микростолбик имеет малый объем полости и высокую добротность Q (см. табл. 10.1). Так как квантовая точка способна одновременно излучать только один фотон, свет получается антигруппированным и сжатым по числу фотонов; дисперсия числа отсчетов при этом меньше, чем среднее значение (см. подразд. 12.3.2). В другом варианте наблюдается электролюминесценция от отдельной квантовой точки в p—i—n-переходе. Прилагаются активные усилия по повышению надежности процесса генерации одиночных фотонов. Помимо других приложений, надежный источник одиночных фотонов найдет применение в квантовой криптографии.

17.4.3. Материалы и структуры устройств

Полупроводниковые лазеры изготавливаются в ошеломляющем разнообразии видов. Они работают на длинах волн, область которых простирается от среднего ультрафиолета до дальнего инфракрасного диапазона, а выходные мощности могут быть от нановатт до киловатт (для матриц лазерных диодов). По существу все полупроводниковые лазеры, используемые в настоящее время, имеют активные области с квантово-размерными структурами. Мы вначале рассмотрим обычные лазерные диоды, как правило, имеющие большие резонаторы, а затем перейдем к полупроводниковым лазерам с микрорезонаторами. В частности, мы изучим поверхностно-излучающие лазеры с вертикальным резонатором (VCSEL — vertical-cavity surface-emitting laser) и фотонно-кристаллические лазеры, которые становятся все более важными.

Обычные лазерные диоды

Лазерные диоды с торцевым излучателем используются в огромном множестве приложений — от потребительских товаров, таких как DVD-плейеры и лазерные принтеры, до оптоволоконных систем дальней связи. Они служат высокоэффективными источниками оптической накачки для оптоволоконных усилителей, волоконных и твердотельных лазеров.

Материалы и структуры большинства обычных лазерных диодов близко напоминают материалы и структуры светоизлучающих диодов (см. подразд. 17.1.3). В диапазоне от ближнего инфракрасного до среднего ультрафиолетового обычно используются прямозонные трех- и четырехкомпонентные материалы, поскольку их ширина запрещенной зоны может перестраиваться за счет изменения состава. Особенно важными материалами, как и в случае СИД, являются AlInGaN, AlInGaP, InGaAs и InGaAsP. Излучающие с торца приборы имеют типичную длину $l \approx 500$ мкм и ширину $w \approx 2$ мкм. Часто встречающимися длинами волн излучения лазерных диодов являются: 635—650 нм для лазерных указок, DVD и систем ближней связи на пластиковых волокнах; 785 нм для CD; 850 нм для систем ближней связи и 1300—1600 нм для дальней связи. Другие длины волн доступных лазерных диодов — это 375, 405, 440, 670 и 830 нм. Хотя лазерные диоды на солях свинца могут работать на длинах волн вплоть до ≈30 мкм, они страдают множеством недостатков, и квантово-каскадные лазеры в среднем и дальнем ИК-диапазоне намного лучше.

Лазерные диоды с торцевым излучателем могут работать на одной продольной и одной поперечной моде. Однако они часто используются в качестве источника с множеством пространственных мод в приложениях, требующих высокой мощности, таких как накачка волоконных оптических усилителей, многооболочечных волоконных лазеров и твердотельных лазеров (см. подразд. 15.3.1). Многомодовые ЛД с торцевым излучателем могут давать оптическую мощность более 5 Вт для одиночно-полосковых приборов с шириной ≈50 мкм и до нескольких киловатт для стопы многополосковых решеток. Часто они присоединяются к оптоволокну для обеспечения эффективной доставки излучения. Обычно для

накачки твердотельных лазеров используются длины волн 808 нм для ванадата иттрия и иттрий-алюминиевого граната, легированных неодимом, и 980 нм для кварцевого волокна, легированного эрбием. Типичные коэффициенты преобразования мощности составляют $\eta_c \approx 45\,\%$, но достигаются и значения более 75 %.

Лазеры с распределенной обратной связью (РОС) с гребенчатым волноводом или скрытой гетероструктурой демонстрируют высокую надежность в множестве приложений, поэтому мы рассмотрим их следующими в порядке очереди. Широко используются также другие типы лазерных диодов, включая ЛД с широким контактом, клиновидные ЛД и ЛД с РБО.

Лазеры с гребенчатым волноводом

Лазерный диод с гребенчатым волноводом (ГВ) работает на одной пространственной моде. В типичном случае он генерирует в ближнем ИК-диапазоне и применяется в спектроскопии и метрологии. Гребенчатый волновод обеспечивает направление света вдоль узкой полосы и латеральное ограничение за счет локализации тока инжекции в активной области под гребнем. Лазерные диоды с ГВ обычно делаются в виде структуры Фабри—Перо со сколотыми торцами и могут обеспечивать до нескольких сот милливатт мощности. Прибор длиной 500 мкм, показанный на рис. 17.52, имеет активную область, содержащую шесть квантовых ям из InGaAsP толщиной 7 нм, деформированных сжатием, между которыми находятся барьеры из InGaAsP толщиной 10 нм, деформированные растяжением. Данный лазерный диод имеет пороговый ток $i_{\text{п.пл}} = 15$ мА, внешний квантовый выход $\eta_d = 0{,}33$, дифференциальную токовую чувствительность $\mathcal{R}_d = 0{,}26$ Вт/А и выходную мощность около 20 мВт.

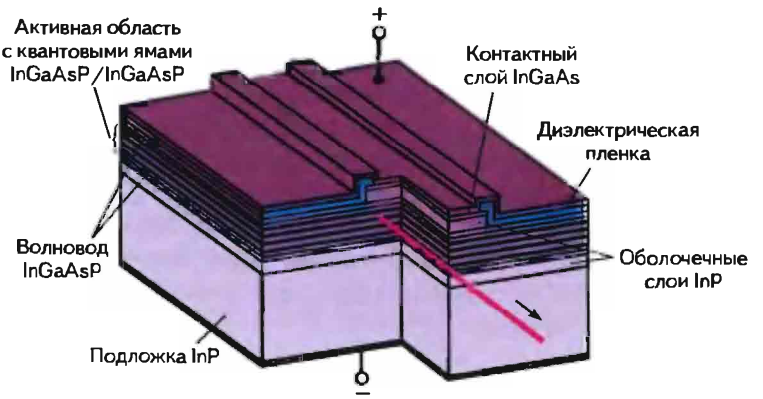


Рис. 17.52. Схема лазерного диода с напряженными квантовыми ямами на InGaAsP/InGaAsP и гребенчатым волноводом, работающего на 1550 нм. Ватт-амперная характеристика приведена на рис. 17.38, *а*

РОС-лазеры со скрытой гетероструктурой

Как показано на рис. 17.53, чередующиеся *p*- и *n*-слои позволяют току течь только вблизи активной области данного прибора со скрытой гетероструктурой, навязывая, таким образом, латеральное ограничение свободы носителей. Диэлектрическая пленка обеспечивает волновод за счет усиления.

Распределенная обратная связь осуществляется с помощью решетки в виде гофрированного слоя, прилегающей к активной области и служащей распределенным отражателем (см. подразд. 17.3.3). Конструкция прибора допускает интеграцию в микросхему. Такие лазеры дают хорошее усиление при умеренных уровнях тока и выдают мощность более 1 Вт в одной пространственной моде. Типичные значения порогового тока и дифференциальной токовой чувствительности $i_{\text{п.п}} < 10$ мА и $\mathcal{R}_d \approx 0{,}4$ Вт/А, соответственно. Эти приборы с торцевым излучателем дают узкие спектральные линии, что критично для эффективной работы систем связи с мультиплексированием по длине волны в диапазоне 1300—1600 нм.

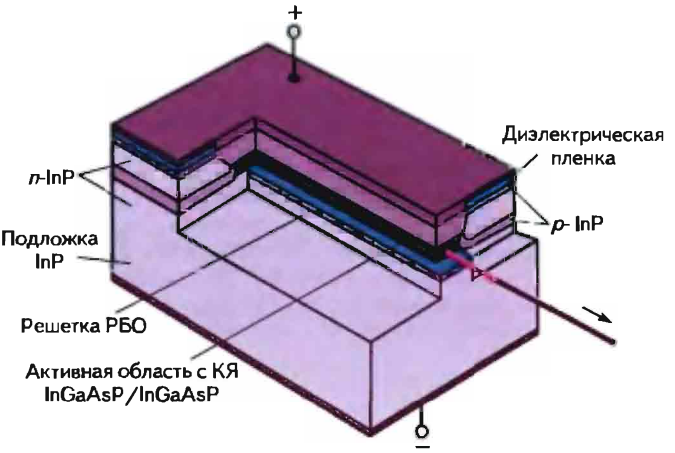


Рис. 17.53. Лазер на скрытой гетероструктуре с квантовыми ямами с распределенной обратной связью, используемый для оптоволоконной связи в диапазоне длин волн 1300—1600 нм

Поверхностно-излучающие лазеры с вертикальным резонатором (VCSEL)

Из лазеров с микрорезонаторами наиболее распространенными являются **поверхностно-излучающие лазеры с вертикальным резонатором** (VCSEL). Эти приборы устроены так, что свет выходит через верхнюю грань одномерного планарного микрорезонатора перпендикулярно активному слою. Они обычно работают в видимой и ближней ИК-области и могут иметь различный диаметр вплоть до 1 мкм.

Пример лазера VCSEL большой площади показан на рис. 17.54. Этот прибор имеет активную область с квантовыми ямами на GaAs/InGaAs и работает на длине волны 995 нм. Резонатор образован решетками Брэгга с большим коэффициентом отражения. Отражаясь между ними, свет многократно проходит через активный слой. Ключевым элементом высокоэффективного лазера VCSEL является диэлектрическая пленка, которая локализует инжекцию носителей и ограничивает оптическую моду в поперечном направлении.

Спектральная интенсивность, оптическая мощность и угловое распределение излучения этого лазера показаны на рис. 17.55.

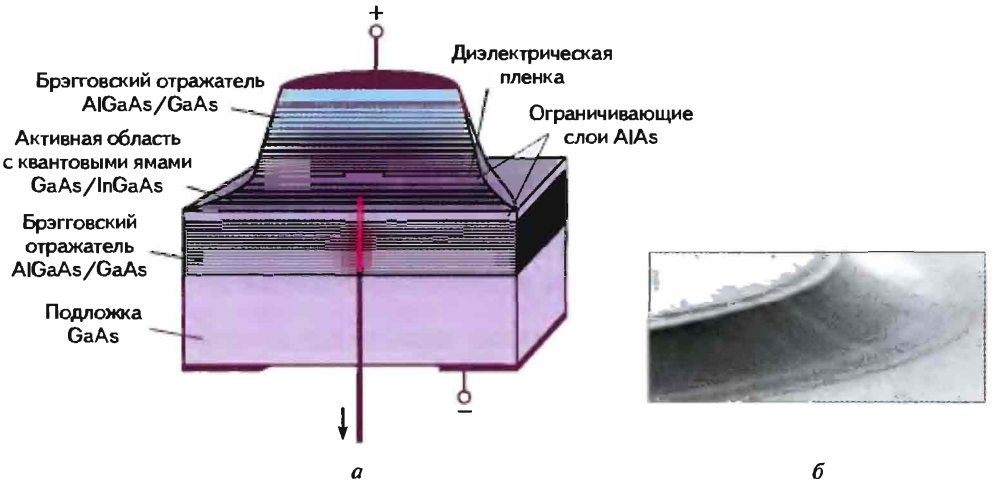


Рис. 17.54. Схема лазера VCSEL большой площади (диаметр 320 мкм) с квантовыми ямами на GaAs/InGaAs, работающего на длине волны 995 нм (*а*). Мезаструктура, полученная методом травления, на которой видны *p*-контакт, РБО *p*-типа и активная область (*б*) (заимствовано из *Miller M.*, *Grabherr M.*, *King R.*, *Jäger R.*, *Michalzik R.*, *Ebeling K.J.* IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, 2001. P. 210—216, Fig. 2 © IEEE)

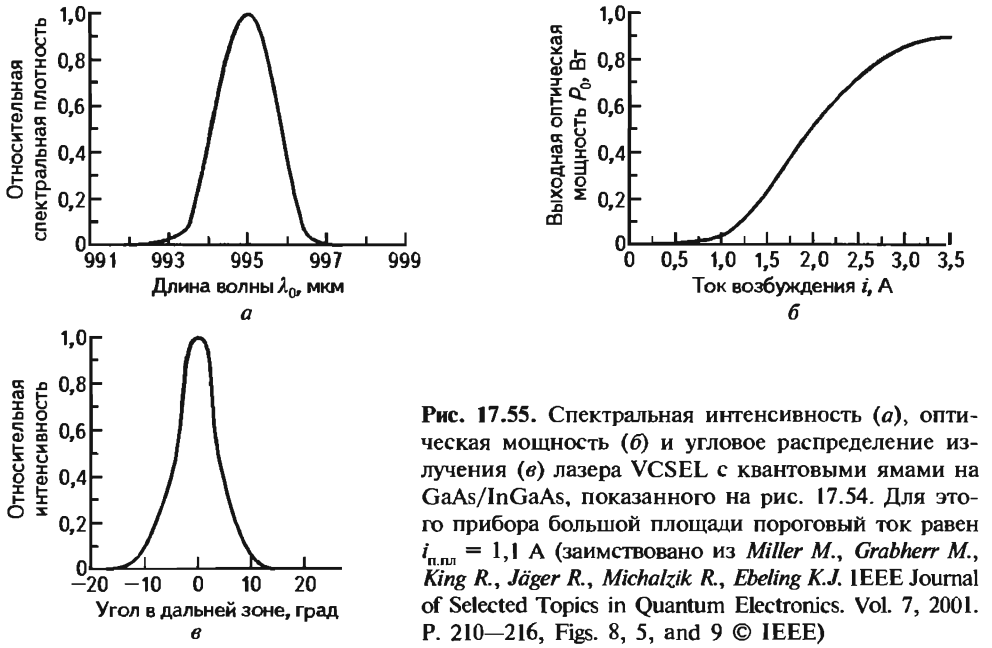


Рис. 17.55. Спектральная интенсивность (*а*), оптическая мощность (*б*) и угловое распределение излучения (*в*) лазера VCSEL с квантовыми ямами на GaAs/InGaAs, показанного на рис. 17.54. Для этого прибора большой площади пороговый ток равен $i_{\text{п.пл}} = 1{,}1$ А (заимствовано из *Miller M.*, *Grabherr M.*, *King R.*, *Jäger R.*, *Michalzik R.*, *Ebeling K.J.* IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, 2001. P. 210—216, Figs. 8, 5, and 9 © IEEE)

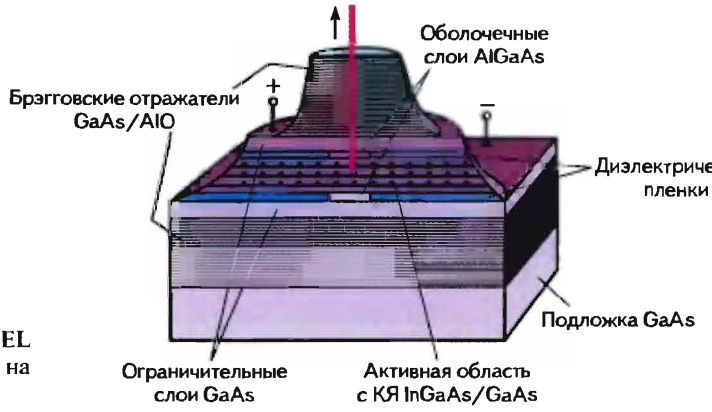
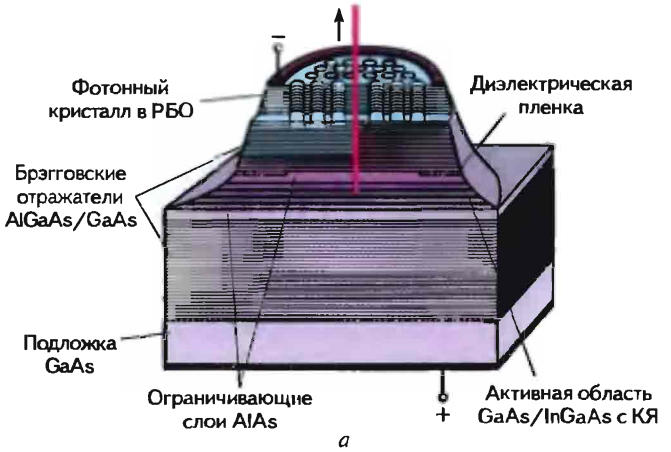
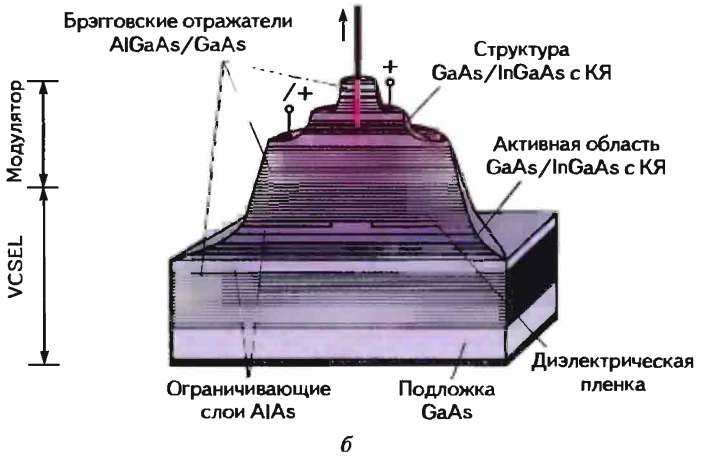


Рис. 17.56. Лазер VCSEL с активной областью на квантовых точках



а



б

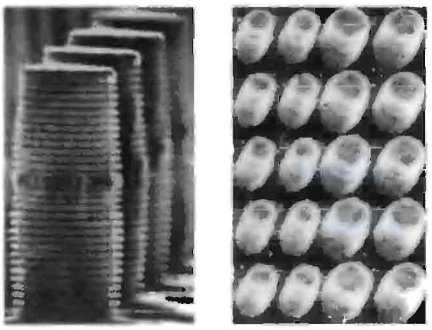
Рис. 17.57. Вариации на тему VCSEL:

а — VCSEL с фотонным кристаллом для поперечного контроля моды; *б* — VCSEL с монолитно встроенным электропоглощающим модулятором

Поскольку толщина активной области составляет всего несколько десятков нанометров, усиление за один проход обычно мало (доли процента). Высокий коэффициент усиления активной среды и коэффициент отражения зеркал поэтому являются решающими для лазеров VCSEL. Типичный РБО, используемый в VCSEL, состоит из десятков слоев. Кроме того, требуются специальные меры по предотвращению нагрева. Пороговые токи у лазеров VCSEL малой площади составляют несколько мкА, а коэффициент преобразования мощности достигает 70 %.

Хотя обычно активная область VCSEL содержит квантовые ямы, можно изготовить VCSEL также на квантовых точках, как показано на рис. 17.56. Лазеры VCSEL принимают невероятное разнообразие видов (рис. 17.57) и могут включать вспомогательные элементы, такие как фотонные кристаллы для поперечного контроля моды, связанные резонаторы и интегрированные модуляторы, что увеличивает скорости прямой модуляции до 40 Гбит/с.

Особенно важно, что лазеры VCSEL допускают высокую плотность упаковки на пластине и могут легко изготавливаться в виде плотных матриц. Одним из первых примеров было изготовление матрицы примерно из миллиона мельчайших VCSEL на квантовых ямах из InGaAs (диаметр ≈2 мкм, высота ≈5,5 мкм) с электрической накачкой и длиной волны генерации вблизи 970 нм на одной пластине GaAs площадью 1 см^2. Этот прибор имел пороговый ток $i_{\text{п.пл}} \approx 100$ мкА при $T = 300$ °K в непрерывном режиме. Полученное с помощью сканирующего электронного микроскопа изображение малой части этой матрицы показано на рис. 17.58. VCSEL-матрицы можно изготавливать из элементов, имеющих заланное распределение лазерных частот.



а *б*

Рис. 17.58. Полученное с помощью сканирующего электронного микроскопа изображение одной из первых матриц, состоящей из $In_{0,2}Ga_{0,8}As$-лазеров на квантовых ямах с вертикальным резонатором и электрической накачкой на пластине из GaAs. Диаметр отдельного лазера от 1 до 5 мкм. Микрорезонаторы образованы брэгговскими отражателями из AlAs/GaAs:

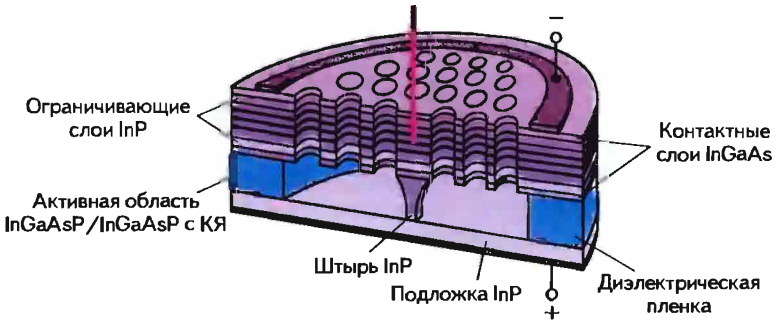
а — AlAs преимущественно вытравливается из брэгговских отражателей, поэтому ясно видны диски из GaAs, скрепленные в центре оставшимся AlAs; *б* — вид сверху на небольшую часть матрицы. Выходные пучки кругового сечения легко вводятся в оптическое волокно (любезно предоставлено Джеком Л. Джуэллом, Picolight Incorporated)

Фотонно-кристаллические лазеры с микрорезонаторами

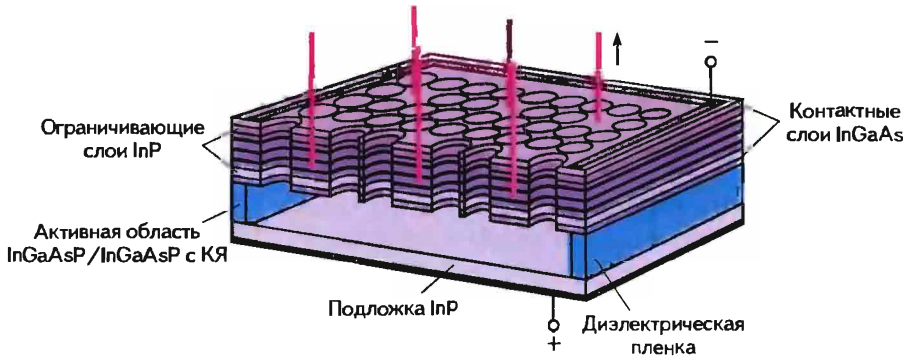
Микрорезонаторы, состоящие из дефектов в двумерных фотонных кристаллах, в сочетании с миниатюрными квантово-размерными источниками света открывают возможность создания беспороговых лазеров с размерами порядка длины волны и матриц из них. Такие приборы имеют длину

волны и пространственное распределение моды дефекта, которые можно перестраивать, а также допускают высокие скорости прямой модуляции. Источником света может служить квантовая яма или квантовая точка.

На рис 17.59, *а* и *б* показаны, соответственно, один такой прибор и матрица из когерентно связанных приборов такого типа. Прибор, схема которого показана на рис. 17.59, *а*, представляет собой одномодовый фотонно-кристаллический лазер, работающий при комнатной температуре[4]. Его электрическая накачка производится через субмикронный штырь, пороговый ток равен 260 мкА. Активная область включает шесть квантовых ям из InGaAsP с компенсацией за счет напряжения; генерация происходит на длине волны $\lambda_0 = 1520$ нм. Выходная мощность равна 2 нВт при токе 0,5 мА, а оценка дифференциальной токовой чувствительности дает $\mathcal{R} = 10^{-5}$. Добротность и модовый объем равны $Q \approx 2500$ и $V \approx 6 \cdot 10^{-2}$ мкм^3 соответственно.



а



б

Рис. 17.59. Фотонно-кристаллический лазер на квантовых ямах InGaAsP/InGaAsP (*а*). Штырь из InP имеет высоту 1 мкм и служит электрическим контактом. Матрица когерентно связанных фотонно-кристаллических лазеров на квантовых ямах (*б*)

Когда ширина линии излучения $\Delta\nu$ меньше ширины электромагнитной моды $\delta\nu$, спонтанное излучение в микрорезонаторах с высокой добротностью

[4] *Park H.-G., Kim S.-H., Kwon S.-H., Yu Y.-G., Yang J.-K., Baek J.-H., Kim S.-B., Lee Y.-H.* Electrically Driven Single-Cell Photonic Crystal Laser. *Science.* Vol. 305, 2004. P. 1444—1447.

может усиливаться за счет эффекта Перселла (см. подразд. 13.3.5 и рис. 13.23). Фактор Перселла для данного прибора составляет

$$\frac{3}{4}\pi^2\frac{\lambda^3}{V}Q \approx 400.$$

Нановаттный уровень мощности отдельного прибора можно существенно повысить за счет использования матриц связанных микрорезонаторов. Прибор, показанный на рис. 17.59, *б*, — фотонно-кристаллический лазер с **матрицей микрорезонаторов** (называемый также **лазером с матрицей нанорезонаторов**), имеющий четыре квантовые ямы InGaAsP/InP и длину волны излучения $\lambda_0 = 1534$ нм [5]. Каждый из $9 \times 9 = 81$ резонаторов, образующих матрицу, занимает площадь 1,5 мкм^2, а площадь матрицы составляет ≈15 мкм^2.

Прибор имеет оптическую накачку от импульсного диодного лазера на 808 нм, излучение которого фокусируется в пятно, размеры которого равны размерам матрицы. Пиковая пороговая мощность для матрицы связанных резонаторов составляет ≈2,5 мВт, а коэффициент связи спонтанного излучения $\beta \approx 0{,}1$. Наблюдалась пиковая выходная мощность 12 мкВт в одномодовом режиме. Порог генерации растет с увеличением числа связанных резонаторов, однако эффективность прибора растет быстрее. Такие приборы допускают очень высокие скорости модуляции.

Рекомендуемая литература

КНИГИ И СТАТЬИ ПО ЛАЗЕРНЫМ УСИЛИТЕЛЯМ И ЛАЗЕРАМ

См. также литературу к гл. 14 и 15.

КНИГИ И СТАТЬИ ПО ФИЗИКЕ ПОЛУПРОВОДНИКОВ, ПОЛУПРОВОДНИКОВЫМ ПРИБОРАМ И НАНОСТРУКТУРАМ

См. также литературу к гл. 16.

КНИГИ ПО СИД И ЛАЗЕРНЫМ ДИОДАМ

Schubert E.F. Light-Emitting Diodes. Cambridge University Press, 2nd ed. 2006.

Müllen K., Scherf U., eds. Organic Light Emitting Devices: Synthesis, Properties and Applications. Wiley-VCH, 2006.

Alfano R.R., ed. The Supercontinuum Laser Source. Springer-Verlag, 1989; 2nd ed. 2006.

Kafafi Z.H., ed. Organic Electroluminescence. CRC Press, 2005.

Sands D. Diode Lasers. Institute of Physics, 2005.

Ohtsubo J. Semiconductor Lasers: Stability. Instability and Chaos, Springer-Verlag, 2005.

Numai T. Fundamentals of Semiconductor Lasers. Springer-Verlag, 2004.

Suhara T. Semiconductor Laser Fundamentals. Marcel Dekker, 2004.

Ye C. Tunable External Cavity Diode Lasers. World Scientific, 2004.

Choi H.K., ed. Long- Wavelength Infrared Semiconductor Lasers. Wiley, 2004.

[5] *Altug H., Vučkovic2 J.* Photonic Crystal Nanocavity Array Laser. Optics Express. Vol.13, 2006. P 8819—8828.

Zappe H. Laser Diode Microsystems. Springer-Verlag, 2004.

Shur M.S., Žukauskas A., eds. UV Solid-State Light Emitters and Detectors. NATO Science Series II: Mathematics, Physics and Chemistry. Vol. 144, Springer-Verlag, 2004.

Shinar J., ed. Organic Light-Emitting Devices: A Survey. Springer-Verlag, 2004.

Risk W.P., Gosnell T.R., Nurmikko A.V. Compact Blue-Green Lasers. Cambridge University Press, 2003.

Sorokina I.T., Vodopyanov K.L., eds. Solid-State Mid-Infrared Laser Sources. Springer-Verlag, 2003.

Yu S.F. Analysis and Design of Vertical Cavity Suiface Emitting Lasers. Wiley, 2003.

Ghafouri-Shiraz H. Distributed Feedback Laser Diodes and Optical Tunable Filters. Wiley, 2003.

Li H., Iga K., eds. Vertical-Cavity Suiface-Emitting Laser Devices. Springer-Verlag, 2003.

Gehrig E., Hess O. Spatio-Temporal Dynamics and Quantum Fluctuations in Semiconductor Lasers. Springer-Verlag, 2003.

Ustinov V.M., Zhukov A.E., Egorov A.Yu., Maleev N.A. Quantum Dot Lasers. Oxford University Press, 2003.

Žukauskas A., Shur M.S., Gaska R. Introduction to Solid-State Lighting. Wiley, 2002.

Kim J., Somani S., Yamamoto Y. Nonclassical Light from Semiconductor Lasers and LEDs. Springer-Verlag, 2001.

Nakamura S., Pearton S., Fasol G. The Blue Laser Diode: The Complete Story. Springer-Verlag, 2nd ed. 2000.

Nakamura S., Chichibu S.F., eds. Introduction to Nitride Semiconductor Blue Lasers and Light Emitting Diodes. Taylor & Francis, 2000.

Diehl R., ed. High-Power Diode Lasers: Fundamentals, Technology, Applications. Springer-Verlag, 2000.

Kapon E., ed. Semiconductor Lasers. Academic Press, 1999.

Wilmsen C., Temkin H., Coldren L.A., eds. Vertical-Cavity Surface-Emitting Lasers: Design, Fabrication, Characterization, and Applications. Cambridge University Press, 1999.

Carroll J., Whiteaway J., Plumb D. Distributed Feedback Semiconductor Lasers. Institution of Engineering and Technology (London), 1998.

Loehr J.P. Physics of Strained Quantum Well Lasers. Kluwer, 1998.

Amann M.-C., Buus J. Tunable Laser Diodes. Artech, 1998.

Morthier G., Vankwikelberge P. Handbook of Distributed Feedback Laser Diodes. Artech, 1997.

Chow W.W., Koch S.W., Sargent III M. Semiconductor-Laser Physics. Springer-Verlag, 1994; corrected ed. 1997.

Willardson R.K., Weber E.R., eds. Semiconductors and Semimetals. Vol. 48. High-Brightness Light Emitting Diodes. G.B. Stringfellow and M.G. Craford, eds. Academic Press, 1997.

Coldren L.A., Corzine S.W. Diode Lasers and Photonic Integrated Circuits. Wiley, 1995.

Ikegami T., Sudo S., Sakai Y. Frequency Stabilization of Semiconductor Laser Diodes. Artech, 1995.

Botez D., Scifres D.R., eds. Diode Laser Arrays. Cambridge University Press, 1994.

Carlson N.W. Monolithic Diode-Laser Arrays. Springer-Verlag, 1994.

Suto K., Nishizawa J.-I. Semiconductor Raman Lasers. Artech, 1994.

Agrawal G.P., Dutta N.K. Semiconductor Lasers. Van Nostrand Reinhold, 2nd ed. 1993.

Ohtsu M. Highly Coherent Semiconductor Lasers. Artech, 1992.

Yamamoto Y., ed. Coherence, Amplification, and Quantum Effects in Semiconductor Lasers. Wiley, 1991.

Willardson R.K., Beer A.C., eds. Semiconductors and Semimetals. Vol. 22: Lightwave Communications Technology. W.T. Tsang, ed. Part B: Semiconductor Injection Lasers, I. Academic Press, 1985.

Willardson R.K., Beer A.C., eds. Semiconductors and Semimetals. Vol. 22: Lightwave Communications Technology. W.T. Tsang, ed. Part C: Semiconductor Injection Lasers, II and Light Emitting Diodes. Academic Press, 1985.

Casey, Jr. H.C., Panish M.B. Heterostructure Lasers, Part B, Materials and Operating Characteristics. Academic Press, 1978.

КНИГИ ПО ОПТОЭЛЕКТРОНИКЕ

Krier A., ed. Mid-infrared Semiconductor Optoelectronics. Springer-Verlag, 2006.

Dakin J.P., Brown R.G.W., eds. Handbook of Optoelectronics. Vol. 1 and 2. CRC Press, 2006.

Liu J.-M. Photonic Devices. Cambridge University Press, 2005.

Parker M.A. Physics of Optoelectronics. Taylor & Francis, 2005.

Razeghi M., Henini M., eds. Optoelectronic Devices: III-Nitrides. Elsevier, 2005.

Piprek J. Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation. Academic Press 2003.

Pearsall T.P. Photonics Essentials: An Introduction with Experiments. McGraw-Hill 2003.

Rosencher E., Vinter B. Optoelectronics. Cambridge University Press, 2002.

Kasap S.O. Optoelectronics and Photonics: Principles and Practices. Prentice Hall, 2001.

Wilson J., Hawkes J.F.B. Optoelectronics. Prentice Hall, 3rd ed. 1998.

Bhattacharya P. Semiconductor Optoelectronic Devices. Prentice Hall, 2nd ed. 1996.

Chuang S.L. Physics of Optoelectronic Devices. Wiley 1995.

Gowar J. Optical Communication Systems (Optoelectronics). Prentice Hall, 2nd ed. 1993.

КНИГИ И СТАТЬИ ПО КРЕМНИЕВОЙ ФОТОНИКЕ

Walters R.J., Kalkman J., Polman A., Atwater H.A., de Dood M.J.A. Photoluminescence Quantum Efficiency of Dense Silicon Nanocrystal Ensembles in SiO_2. Physical Review B. Vol. 73, 2006. 132302.

Makarova M., Vuckovic J., Sanda H., Nishi Y. Silicon-Based Photonic Crystal Nanocavity Light Emitters. Applied Physics Letters. Vol. 89, 2006. 221101.

Jalali B., Raghunathan V., Dimitropoulos D., Boyraz O. Raman-Based Silicon Photonics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, 2006. P. 412—421.

Paniccia M., Koehl S. The Silicon Solution. IEEE Spectrum. Vol. 42, № 10, 2005. P 38—43.

Pavesi L., Lockwood D.J., eds. Silicon Photonics. Springer-Verlag, 2004.

Reed G.T., Knights A.P. Silicon Photonics: An Introduction. Wiley, 2004.

Pavesi L., Gaponenko S., Dal Negro L., eds. Towards the First Silicon Laser. NATO Science. Series II: Mathematics, Physics and Chemistry. Vol. 93. Kluwer, 2003.

Pavesi L., Dal Negro L., Mazzoleni C., Franzo1 G., Priolo F. Optical Gain in Silicon Nanocrystals. Nature. Vol. 408, 2000. P. 440—444.

Zimmermann H. Integrated Silicon Optoelectronics. Springer-Verlag, 2000.

Issue on silicon-based optoelectronics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 4, № 6, 1998.

СТАТЬИ

Issue on nanophotonics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 6, 2006.

Sun Y., Giebink N.C., Kanno H., Ma B., Thompson M.E., Forrest S.R. Management of Singlet and Triplet Excitons for Efficient White Organic Light-Emitting Devices. Nature. Vol. 440, 2006. P. 908—912.

Faist J. Continuous-Wave, Room-Temperature Quantum Cascade Lasers. Optics & Photonics News. Vol. 17, № 5, 2006. P. 32—36.

Holonyak, Jr. N., Feng M. The Transistor Laser. IEEE Spectrum. Vol. 43, № 2, 2006. P. 50—55.

Altug H., Englund D., Vučković J. Ultrafast Photonic Crystal Nanocavity Laser. Nature Physics. Vol. 2, 2006. P. 484—488.

Narendran N. The Solid-State Lighting Revolution. Physics World. Vol. 18, № 7, 2005. P. 25—29.

Snee P.T., Chan Y., Nocera D.G., Bawendi M.G. Whispering-Gallery-Mode Lasing from a Semiconductor Nanocrystal/Microsphere Resonator Composite. Advanced Materials. Vol. 17, 2005. P. 1131—1136.

Issue on semiconductor lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 11, № 5, 2005.

Issue on organic light-emitting diodes. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 10, № 1, 2004.

Howard W.E. Better Displays with Organic Films. Scientific American. Vol. 290, № 2, 2004. P. 76—81.

Park H.-G., Kim S.-H., Kwon S.-H., Ju Y.-G., Yang J.-K., Back J.-H., Kim S.-B., Lee Y.-H. Electrically Driven Single-Cell Photonic Crystal Laser. Science. Vol. 305, 2004. P. 1444—1447.

Issue on semiconductor lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 9, № 5, 2003.

Feature issue on mid-infrared quantum-cascade lasers. IEEE Journal of Quantum Electronics. Vol. 38, № 6, 2002.

Issue on high-efficiency light-emitting diodes. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, № 2, 2002.

Capasso F., Gmachl C., Sivco D.L., Cho A.Y. Quantum Cascade Lasers. Physics Today. Vol. 55, № 5, 2002. P. 34—40.

Craford M.G., Holonyak, Jr. N., Kish, Jr. F.A. In Pursuit of the Ultimate Lamp. Scientific American. Vol. 284, № 2, 2001. P. 62—67.

Issue on semiconductor lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, № 2, 2001.

Huang M.H., Mao S., Feick H., Van H., Wu Y., Kind H., Weber E., Russo R., Yang P. Room-Temperature Ultraviolet Nanowire Nanolasers. Science. Vol. 292, 2001. P. 1897—1899.

Michler P., Kiraz A., Becher C., Schoenfeld W.V., Petroff P.M., Zhang L., Hu E., Imamoglu A. A Quantum Dot Single-Photon Turnstile Device. Science. Vol. 290, 2000. P. 2282—2285.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Issue on semiconductor lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 5, № 3, 1999.

Painter O., Lee R.K., Scherer A., Yariv A., O'Brien J.D., Dapkus P.D., Kim I. Two-Dimensional Photonic Band-Gap Defect Mode Laser. Science. Vol. 284, 1999. P. 1819—1821.

Chang-Hasnain C.J. VCSELs: Advances and Future Prospects. Optics & Photonics News. Vol. 9, № 5, 1998. P. 34—39.

Little G.R., ed. Selected Papers on Fundamentals of Optoelectronics. SPIE Optical Engineering Press (Milestone Series. Vol. 90), 1994.

Coleman J.J., ed. Selected Papers on Semiconductor Diode Lasers. SPIE Optical Engineering Press (Milestone Series. Vol. 50), 1992.

Jewell J. Surface-Emitting Lasers: A New Breed. Physics World. Vol. 3, № 7, 1990. P. 28—30.

Capasso F., Datta S. Quantum Electron Devices. Physics Today. Vol. 43, № 2, 1990. P. 74—82.

Saul R.H., Lee T.P., Burrus C.A. Light-Emitting-Diode Device Design. In R.K. Willardson and A.C. Beer, eds. Semiconductors and Semimetals. Vol. 22: Lightwave Communications Technology. W.T. Tsang, ed. Part C: Semiconductor Injection Lasers, II and Light Emitting Diodes. Academic Press, 1985.

ИСТОРИЧЕСКИЕ РАБОТЫ

Alferov Zh.I. Double Heterostructure Concept and its Applications in Physics. Electronics and Technology. In G. Ekspong, ed. Nobel Lectures, Physics 1996—2000. World Scientific, 2002.

Kroemer H. Quasi-Electric Fields and Band Offsets: Teaching Electrons New Tricks. In G. Ekspong, ed. Nobel Lectures, Physics 1996—2000. World Scientific, 2002.

Faist J., Capasso F., Sivco D.L., Sirtori C., Hutchinson A.L., Cho A.Y. Quantum Cascade Laser. Science. Vol. 264, 1994. P. 553—556.

Jewell J.L., Scherer A., McCall S.L., Lee Y.H., Walker S., Harbison J.P., Florez L.T. Low-Threshold Electrically Pumped Vertical-Cavity Surface-Emitting Microlasers. Electronics Letters. Vol. 25, 1989. P. 1123—1124.

Dupuis R.D. An Introduction to the Development of the Semiconductor Laser. IEEE Journal of Quantum Electronics. Vol. QE-23, 1987. P. 651—657.

Basov N.G. Quantum Electronics at the P.N. Lebedev Physics Institute of the Academy of Sciences of the USSR (FIAN). Soviet Physics-Uspekhi. Vol. 29. 1986 P. 179—185 [Успехи физических наук. Т. 148, 1986. С. 313—324].

Butler J.K., ed. Semiconductor Injection Lasers. IEEE Press, 1980.

Loebner E.E. Subhistories of the Light Emitting Diode. IEEE Transactions on Electron Devices. Vol. ED-23, 1976. P. 675—699.

Basov N.G. Semiconductor Lasers. In Nobel Lectures in Physics, 1963—1970. Elsevier, 1972.

Kazarinov R.F., Suris R.A. Amplification of Electromagnetic Waves in a Semiconductor Superlattice. Soviet Physics-Semiconductors. Vol. 5, 1971. P. 707—709.

Esaki L., Tsu R. Superlattice and Negative Differential Conductivity in Semiconductors. IBM Journal of Research and Development. Vol. 14, 1970. P. 61—65.

Pankove J.I., Berkeyheiser J.E. A Light Source Modulated at Microwave Frequencies. Proceedings of the IRE. Vol. 50, 1962. P 1976—1977.

Quist T.M., Rediker R.H., Keyes R.J., Krag W.E., Lax B., McWhorter A.L., Zeiger H.J. Semiconductor Maser of GaAs. Applied Physics Letters. Vol. 1, 1962. P. 91—92.

Holonyak, Jr. N. Bevacqua S.F. Coherent (Visible) Light Emission from $Ga(As_{1-x}P_x)$ Junctions. Applied Physics Letters. Vol. 1, 1962. P. 82—83.

Nathan M.I., Dumke W.P., Bums G., Dill, Jr. F.H., Lasher G. Stimulated Emission of Radiation from GaAs *p—n*-Junctions. Applied Physics Letters. Vol. 1, 1962. P. 62—64.

Hall R.N., Fenner G.E., Kingsley J.D., Soltys T.J., Carlson R.O. Coherent Light Emission from GaAs Junctions. Physical Review Letters. Vol. 9, 1962. P. 366—368.

Keyes R.J., Quist T.M. Recombination Radiation Emitted by Gallium Arsenide. Proceedings of the IRE. Vol. 50, 1962. P. 1822—1823.

Basov N.G., Krokhin O.N., Popov Yu.M. Production of Negative-Temperature States in *p—n*-Junctions of Degenerate Semiconductors. Soviet Physics-JETP. Vol. 13, 1961. P. 1320—1321 [Журнал экспериментальной и теоретической физики. Т. 40, 1961. С. 1879—1880].

Bernard M.G.A., Duraffourg G. Laser Conditions in Semiconductors. Physica Status Solidi. Vol. 1, 1961. P. 699—703.

John von Neumann, in unpublished calculations sent to Edward Teller in September 1953, showed that it was possible, in principle, to upset the equilibrium concentration of carriers in a semiconductor and thereby obtain light amplification by stimulated emission, e.g., via the recombination of electrons and holes injected into a *p—n*-junction [see *von Neumann J.* Notes on the Photon-Disequilibrium-Amplification Scheme (JvN). Sept. 16, 1953. IEEE Journal of Quantum Electronics. Vol. QE-23, 1987. P. 658—673].

Round H.J. A Note on Carborundum. Electrical World. Vol. 49,1907. P. 309.

Задачи

К РАЗДЕЛУ 17.1

1. **Спектральные ширины СИД.** Рассмотрим семь спектров СИД, показанных на рис. 17.14 и 17.60, а именно те, центральные длины волн которых равны λ_0 = 0,37; 0,53; 0,64; 0,91; 1,30; 1,93 и 2,25 мкм. Оцените по графикам спектральные ширины по высоте половины максимума в нанометрах, герцах и электронвольтах. Сравните ваши оценки с результатами расчета по формулам, приведенным в упражнении 17.3. Оцените уширение из-за неоднородности состава для спектра СИД с центром на λ_0 = 0,53 мкм в нанометрах, герцах и электронвольтах.

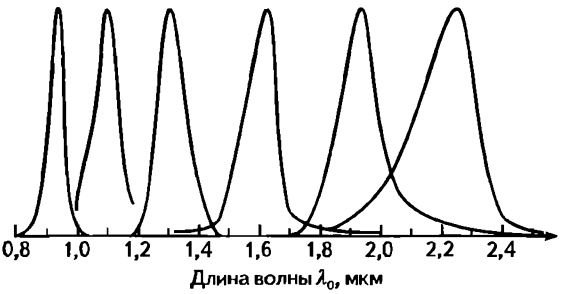


Рис. 17.60. Спектральные интенсивности в зависимости от длины волны для СИД на InGaAsP, работающих в ближнем ИК-диапазоне спектра. Спектры нормированы на одинаковое значение максимальной интенсивности. В общем случае спектральная ширина растет как λ_p^2 в соответствии с (17.29)

2. **Коэффициент вывода излучения из СИД.** Выведите выражение для коэффициента вывода неполяризованного внутри СИД света с учетом угловой зависимости френелевского отражения на границе полупроводник—воздух.

3. **Ввод света из СИД в оптическое волокно.** Рассчитайте долю оптической мощности, испущенной СИД, вводимую в оптическое волокно со ступенчатым профилем показателя преломления, имеющее числовую аппретуру NA = 0,1 в воздухе и показатель преломления сердцевины 1,46 (см. разд. 9.1). Считайте, что СИД имеет плоскую поверхность, показатель преломления n = 3,6 и угловое распределение мощности, пропорциональное $\cos^4 \theta$. Излучающая поверхность СИД плотно прижата к торцу волокна, а площадь излучения меньше, чем площадь сердцевины волокна.

К РАЗДЕЛУ 17.2

1. **Ширина полосы полупроводникового оптического усилителя.** С использованием данных рис. 17.26, *а* постройте график зависимости полной ширины полосы усилителя на InGaAsP от концентрации инжектируемых носителей

Δn. Найдите линейную аппроксимацию для этой зависимости и постройте график зависимости коэффициента усиления прибора от ширины полосы.

2. **Пиковый коэффициент усиления полупроводникового оптического усилителя при $T = 0$ °K.**

а. Покажите, что пиковое значение γ_p коэффициента усиления $\gamma_0(\nu)$ при $T = 0$ °K приходится на $\nu = (E_{f_c} - E_{f_v})/h$.

б. Получите аналитическое выражение для пикового коэффициента усиления γ_p как функции концентрации инжектируемых носителей Δn при $T = 0$ °K.

в. Постройте график γ_p от Δn для усилителя на InGaAsP ($\lambda_0 = 1300$ нм, $n = 3{,}5$, $\tau_{рад} = 2{,}5$ нс, $m_c = 0{,}06m_0$, $m_v = 0{,}4m_0$) для значений Δn в диапазоне от $1\ 10^{18}$ до $2 \cdot 10^{18}$ см$^{-3}$.

г. Сравните результаты с данными рис. 17.26, *б*.

3. **Коэффициент усиления полупроводникового оптического усилителя на GaAs.** ПОУ на GaAs p-типа, работающий при $T = 300$ °K ($E_g \approx 1{,}40$ эВ, $m_c = 0{,}07m_0$, $m_v = 0{,}50m_0$), с показателем преломления $n = 3{,}6$ легирован так ($p_0 = 1{,}2 \cdot 10^{18}$), что время жизни радиационной рекомбинации равно $\tau_{рад} \approx 2$ нс.

а. Считая известной стационарную концентрацию инжектируемых носителей Δn (которая определяется скоростью инжекции R и полным временем рекомбинации τ), с помощью (17.32)—(17.34) найдите коэффициент усиления $\gamma_0(\nu)$ как функцию энергии фотона $h\nu$, предполагая, что $T = 0$ °K.

б. Проделайте те же вычисления при $T = 300$ °K с помощью компьютера.

в. Постройте графики зависимости коэффициента усиления от Δn в обоих случаях.

г. Определите коэффициент потерь α и концентрацию Δn_T, соответствующую прозрачности среды, с использованием линейной аппроксимации.

д. Постройте график полной ширины полосы усиления (в герцах, нанометрах и электронвольтах) как функции Δn в обоих случаях.

е. Сравните результаты с кривыми коэффициента усиления и его пикового значения, показанными на рис. 17.61.

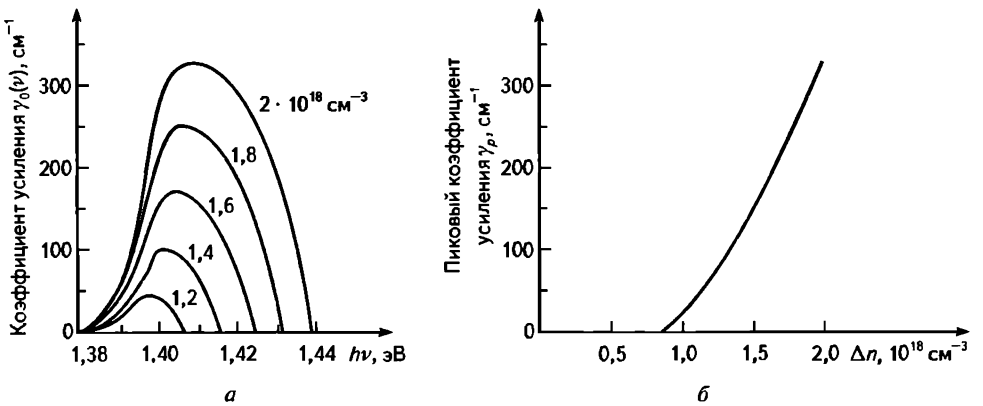


Рис. 17.61. Коэффициент усиления (*а*) и его пиковое значение (*б*) для ПОУ на GaAs (Заимствовано из *Panish M.B.* Heterstructure Injection Lasers. Proceedings of the IEEE. Vol. 64, 1976. P. 1512—1540, Fig. 4 © IEEE.)

4. **Уменьшение ширины запрещенной зоны из-за хвостов зон.** Уменьшение ширины запрещенной зоны за счет состояний, принадлежащих к хвостам зон, в InGaAsP и GaAs можно выразить эмпирической формулой

$$\Delta E_g\,(\text{эВ}) \approx \left(-1{,}6 \cdot 10^{-8}\right)\left(p^{1/3} + n^{1/3}\right),$$

где n- и p-концентрации носителей (см$^{-3}$), обусловленные легированием, инжекцией или обоими факторами.

а. Для InGaAsP p-типа и GaAs определите концентрацию p, при которой ширина запрещенной зоны уменьшается примерно на 0,02 эВ.

б. Для собственного InGaAsP и GaAs определите плотность числа инжектируемых носителей Δn, при которой ширина запрещенной зоны уменьшается примерно на 0,02 эВ. Считайте n_i пренебрежимо малым.

в. Рассчитайте $E_g + \Delta E_g$ и сравните результат с энергией, при которой коэффициент усиления на рис. 17.61, *а* равен нулю с низкочастотной стороны.

5. **Усиление и ширина полосы усилителя.** GaAs имеет собственную концентрацию носителей $n_i = 1{,}8 \cdot 10^6$ см$^{-3}$, рекомбинационное время жизни $\tau = 5$ нс, ширину запрещенной зоны $E_g = 1{,}42$ эВ, эффективную массу электрона $m_c = 0{,}07m_0$ и эффективную массу дырки $m_v = 0{,}50m_0$. Предположим, что $T = 0$ °K.

а. Определите центральную частоту, ширину полосы и максимальное полное усиление в пределах ширины полосы для усилителя на GaAs длиной $d = 200$ мкм, шириной $w = 10$ мкм и толщиной $l = 2$ мкм, через который пропускается ток 1 мА.

б. Определите число голосовых сообщений, которые поддерживает данная ширина полосы, если считать, что передача каждого их них занимает полосу в 4 кГц.

в. Определите битовую скорость передачи через усилитель, если каждый голосовой канал требует 64 кБит/с.

6. **Сечение перехода.** Определите сечение перехода $\sigma(\nu)$ для GaAs как функцию Δn при $T = 0$ °K. Плотность вероятности вынужденного излучения или поглощения есть $\phi\sigma(\nu)$, где ϕ — плотность потока фотонов. Почему сечение перехода менее полезно для полупроводниковых оптических усилителей, чем для других типов оптических усилителей?

7. **Профиль усиления.** Рассмотрим усилитель на InGaAsP, работающий на длине волны 1550 нм ($n = 3{,}5$), конфигурация которого показана на рис. 17.29, с одинаковыми антиотражательными покрытиями входной и выходной граней. Рассчитайте максимальный допустимый коэффициент отражения каждой из граней, если изменения профиля усиления из-за частотной зависимости пропускания резонатора Фабри—Перо не должны превышать 10 % [см. (7.34)].

К РАЗДЕЛУ 17.3

1. **Зависимость выходной мощности от показателя преломления.** В выражении (17.53) для выходного потока фотонов Φ_0 распознайте члены, зависящие от показателя преломления кристалла.

2. **Продольные моды.** В диод на InGaAsP с шириной запрещенной зоны $E_g = 0{,}91$ эВ и показателем преломления $n = 3{,}5$ инжектируется такой ток, что разность уровней Ферми $E_{f_c} - E_{f_v} = 0{,}96$ эВ. Определите максимальное число продольных мод, на которых возможна генерация, если длина резонатора $d = 250$ мкм, а потери отсутствуют.

3. **Минимальное усиление, необходимое для генерации.** Кристалл InGaAsP длиной 500 мкм работает на длине волны, на которой его показатель преломления $n = 3{,}5$. Пренебрегая рассеянием и другими потерями, определите коэффициент усиления, требуемый для компенсации потерь при отражении от границ кристалла.

4. **Межмодовое расстояние при показателе преломления, зависящем от длины волны.** Частотное расстояние между модами лазерного диода осложняется тем фактом, что показатель преломления зависит от длины волны [т. е. $n = n(\lambda_0)$]. Лазерный диод длиной 430 мкм генерирует на центральной длине волны $\lambda_c = 650$ нм. В пределах ширины полосы излучения $n(\lambda_0)$ можно приближенно считать линейной [т. е. $n(\lambda_0) = n_0 - a(\lambda_0 - \lambda_c)$, где $n_0 = n(\lambda_c) = 3{,}4$ и $a = \mathrm{d}n/\mathrm{d}\lambda_0$].

   а. Расстояние между лазерными модами длиной волны вблизи λ_c оказалось равным $\Delta\lambda \approx 0$, 12 нм. Объясните, почему это не соответствует обычному межмодовому расстоянию $\nu_F = c/2d$.

   б. Найдите оценку для a.

   в. Объясните явление затягивания мод в газовом лазере и сравните его с описанным выше эффектом в полупроводниковых лазерах.

ГЛАВА 18 ПОЛУПРОВОДНИКОВЫЕ ДЕТЕКТОРЫ ФОТОНОВ



Генрих Герц (1857—1894) открыл в 1887 г. фотоэлектрический эффект, природа которого была объяснена Эйнштейном в 1905 г.



Симон Дени Пуассон (1781—1840) разработал фундаментальное распределение вероятностей, описывающее шум фотоприемника.

Фотоприемник — прибор, измеряющий поток фотонов или оптическую мощность посредством преобразования энергии поглощенных фотонов в измеряемую форму. Существует два основных класса широко используемых фотоприемников: **фотоэлектрические приемники** и **тепловые приемники.**

1. Действие фотоэлектрических приемников основано на **фотоэлектрическом эффекте**, называемом также **фотоэффектом**. Поглощение фотонов материалом вызывает переходы электронов на вышележащие уровни энергии, что приводит к появлению подвижных носителей заряда. Под действием электрического поля эти носители приходят в движение, и возникает электрический ток, который можно измерить. Фотоэффект бывает двух видов: внешний и внутренний. **Внешний фотоэффект** включает **фотоэлектронную эмиссию**, при которой возбужденные светом электроны выходят из материала наружу. При **внутреннем фотоэффекте** возникает **фотопроводимость**, когда возбужденные носители остаются внутри материала и служат для увеличения электропроводности.

2. Тепловые детекторы работают посредством преобразования энергии фотонов в теплоту. Ввиду большого времени, необходимого чтобы вызвать измене-

ние температуры, тепловые приемники неэффективны и медленны в сравнении с фотоэлектрическими приемниками. Однако последние достижения в изготовлении и миниатюризации резко улучшили свойства матриц из тепловых приемников и теперь они являются жизнеспособными конкурентами в приложениях, связанных с формированием изображений в среднем инфракрасном диапазоне.

О данной главе

Эта глава посвящена изучению различных фотоэлектрических приемников, находящих применение в фотонике. В разд. 18.1 мы начинаем с обсуждения внешнего и внутреннего фотоэффекта и вводим несколько важных свойств фотодетекторов, включая квантовую эффективность, токовую чувствительность и время отклика. В разд. 18.2—18.4 наше внимание направлено на три типа полупроводниковых фотоприемников, основанных на внутреннем фотоэффекте: фотопроводниках, фотодиодах и лавинных фотодиодах. Матричные приемники, формирующие электронные версии оптических изображений, рассмотрены в разд. 18.5.

Для оценки рабочих характеристик полупроводниковых фотодетекторов в различных приложениях важно понять их шумовые свойства, которые приводятся в разд. 18.6. Шум в выходной цепи фотоэлектрического приемника порождается несколькими источниками: квантовый характер самого света (фотонный шум), преобразование фотонов в носители заряда (фотоэлектронный шум), генерация вторичных носителей при внутреннем усилении (шум усиления), а также шумы в цепи приемника. Мы также кратко обсудим работу аналоговых и цифровых оптических приемников.

18.1. ФОТОПРИЕМНИКИ

18.1.1. Внешний и внутренний фотоэффект

Фотоэлектронная эмиссия

Если энергия фотонов, которыми облучается материал в вакууме, достаточно велика, возбужденные электроны могут преодолеть потенциальный барьер на поверхности материала и стать свободными электронами в вакууме. Этот процесс, называемый **фотоэлектронной эмиссией**, показан на рис. 18.1, *а* для металла. Падающий фотон с энергией $h\nu$ высвобождает электрон из частично заполненной зоны проводимости. Закон сохранения энергии требует, чтобы электроны, испускаемые с уровней ниже уровня Ферми, где их много, имели максимальную кинетическую энергию

$$E_{\max} = h\nu - W, \tag{18.1}$$

где **работа выхода** W равна разности энергий между уровнем вакуума и уровнем Ферми в металле. Уравнение (18.1) известно как уравнение Эйнштейна

для фотоэмиссии. Электрон может приобрести максимальную кинетическую энергию, описываемую уравнением (18.1) только в том случае, если он находился на уровне Ферми; удаление электрона с более глубоко лежащих уровней требует дополнительной энергии для перевода на уровень Ферми, что уменьшает кинетическую энергию свободного электрона. Наименьшая работа выхода для металла (Cs) составляет около 2 эВ, так что оптические приемники на основе внешнего фотоэффекта применимы в видимой и ультрафиолетовой областях спектра.

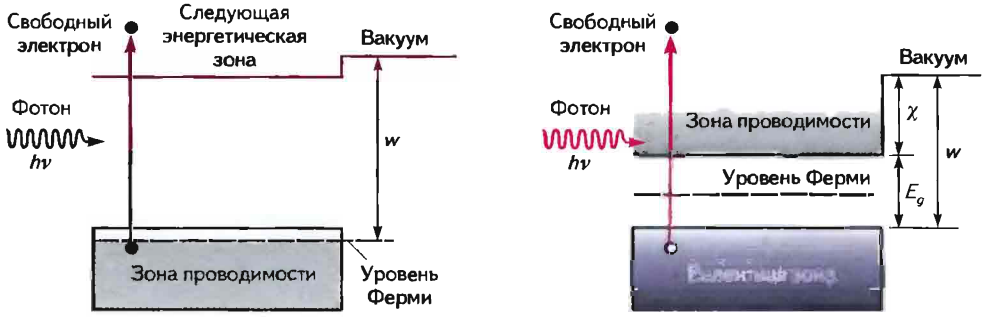


Рис. 18.1. Фотоэлектронная эмиссия из металла (*а*) и из собственного полупроводника (*б*). Ширина запрещенной зоны и электронное сродство материала обозначены через E_g и χ соответственно, а W — работа выхода. Все три величины обычно выражаются в электронвольт

Фотоэлектронная эмиссия из собственного полупроводника схематически изображена на рис. 18.1, *б*. Фотоэлектроны обычно высвобождаются из валентной зоны, где электронов много. Формула, аналогичная (18.1), имеет вид:

$$E_{\max} = h\nu - W = h\nu - \left(E_g + \chi\right), \qquad (18.2)$$

где E_g — ширина запрещенной зоны; χ — сродство вещества к электрону (разность энергий между уровнем вакуума и дном зоны проводимости). Энергия $E_g + \chi$ для некоторых материалов (например, соединение NaKCsSb, содержащее несколько щелочных металлов, на основе которого делаются так называемые S-20-фотокатоды) может снижаться до 1,4 эВ. Поэтому полупроводниковые детекторы с внешним фотоэффектом могут работать не только в видимой и ультрафиолетовой, но и в ближней ИК-области.

Более того, были разработаны полупроводники **с отрицательным сродством к электрону (ОСЭ)**, у которых край зоны проводимости лежит выше уровня вакуума в объеме материала, так что электрону нужно только преодолеть E_g для того, чтобы фотоэлектронная эмиссия могла произойти (наложение тонкого слоя материала *n*-типа или металла на материал *p*-типа может изогнуть зоны на поверхности материала так, что дно зоны проводимости на самом деле лежит ниже уровня вакуума). Поэтому ОСЭ-приемники, например GaAs, покрытый Cs,

чувствительны к несколько большим длинам волн, а также имеют более низкий темновой ток и лучшую квантовую эффективность. Фотокатоды из неоднородных материалов или оксидов, такие как фотокатод S-1-типа, также могут использоваться в ближнем ИК-диапазоне, но лишь до длин волн ≈1 мкм.

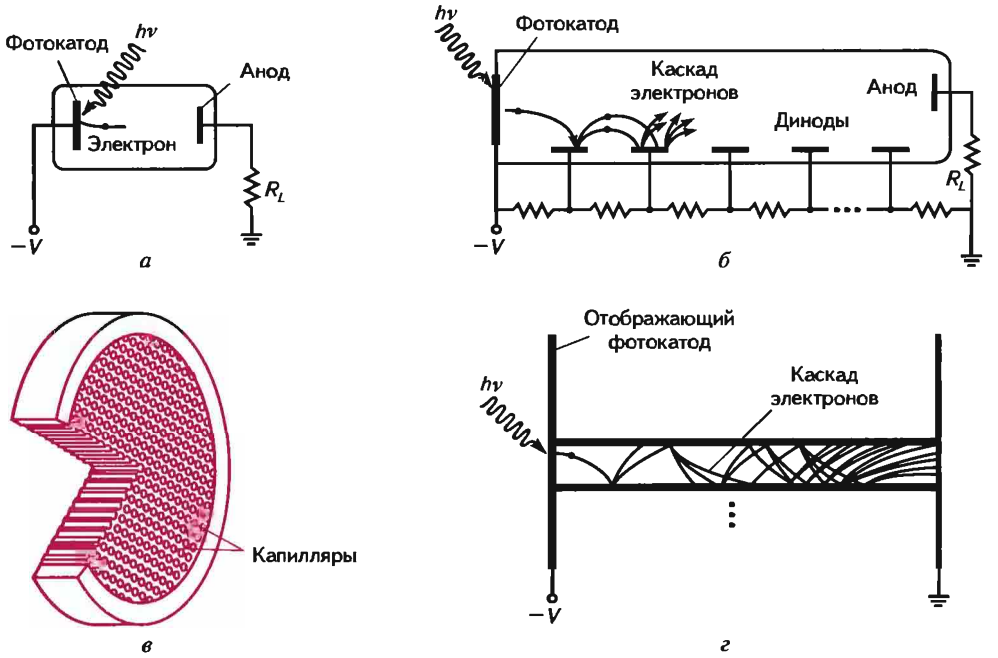


Рис. 18.2. Вакуумный фотодиод с фотокатодом, работающим в режиме отражения (*а*). Умножение числа электронов в трубке фотоумножителя с полупрозрачным фотокатодом, работающим в режиме пропускания (*б*). Вид микроканальной пластины в разрезе (*в*). Умножение числа электронов в одном капилляре микроканальной пластины (*г*)

В простейшем виде фотоприемники на основе фотоэлектронной эмиссии выполняются в виде вакуумных трубок, называемых **вакуумными фотодиодами** или **электровакуумными фотоэлементами**. Электроны испускаются с поверхности фотоэмиссионного материала, называемого **фотокатодом,** и движутся к электроду (аноду) с более высоким электрическим потенциалом. Фотокатод может быть непрозрачным и работать в режиме отражения (рис. 18.2, *а*) или быть полупрозрачным и работать в режиме пропускания (рис. 18.2, *б*). В результате переноса электронов между катодом и анодом в цепи возникает ток, пропорциональный потоку фотонов и **называемый фототоком**. Испущенные электроны могут также создавать каскады электронов в процессе **вторичной эмиссии**. Это происходит, когда испущенные электроны ударяются о другие специально размещенные на их пути полупроводниковые или цезиево-оксидные пластины, называемые **динодами**, потенциалы которых последовательно увели-

чиваются. В результате фототок можно увеличить в 10^8 раз. Этот прибор, показанный на рис. 18.2, *б*, называется **фотоэлектронным умножителем (ФЭУ)**. Фотоэлектронный умножитель можно использовать для детектирования единичных фотонов; при этом он имеет широкий динамический диапазон, однако он объемист и требует высокого напряжения для питания трубки.

Прибор для получения изображений, основанный на этом принципе, — это **микроканальная пластина**. Она состоит из матрицы, образованной миллионами капилляров (с внутренним диаметром ≈10 мкм), выполненных в стеклянной пластине толщиной ≈1 мм. Обе поверхности пластины покрыты тонкой металлической пленкой, чтобы служить электродами, на которые подается напряжение (рис. 18.2, *в*). Внутренние стенки каждого капилляра покрыты материалом, который испускает вторичные электроны и служит распределенным динодом, умножая фототок, созданный в этом месте (рис. 18.2, *г*). Локальный поток фотонов слабого изображения, таким образом, преобразуется в интенсивный поток электронов, который может быть непосредственно измерен. Более того, электронный поток может быть снова преобразован в оптическое изображение (усиленное) с помощью нанесения на задний электрод фосфора, который испускает свет за счет катодолюминесценции (см. подразд. 13.5.1); эта комбинация называется **усилителем изображения**.

Фотопроводимость

Большинство современных фотоприемников работает на основе внутреннего фотоэффекта, при котором возбужденные светом носители заряда (электроны и дырки) остаются внутри образца. Регистрация света на основе фотопроводимости опирается непосредственно на индуцированное светом повышение электропроводности материала. Например, поглощение фотона собственным полупроводником имеет своим результатом *генерацию* свободного электрона, перешедшего из валентной зоны в зону проводимости (рис. 18.3). Одновременно в валентной зоне появляется дырка. Приложение электрического поля к материалу приводит к *переносу* обоих носителей и, следовательно, к возникновению электрического тока в цепи.

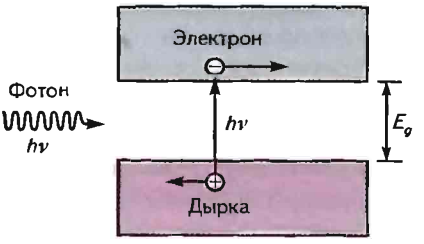


Рис. 18.3. Фотогенерация электрона и дырки в полупроводнике

Полупроводниковый **фотодиодный** приемник представляет собой структуру с *p*—*n*-переходом, также работающую на основе внутреннего фотоэффекта. Фотоны, поглощаемые в обедненном слое, *генерируют* электроны и дырки, которые подвергаются действию локального электрического поля внутри этого слоя. Два типа носителей дрейфуют в противоположных направлениях. Процесс *переноса* индуцирует электрический ток во внешней цепи.

В некоторых фотоприемниках имеется механизм внутреннего усиления, так что фототок может быть усилен за счет умножения числа носителей внут-

ри приемника и становится более доступным для измерения. Если достаточно сильно увеличить электрическое поле внутри обедненного слоя путем большого обратного смещения перехода, то генерируемые электроны и дырки будут сами иметь достаточную энергию, чтобы высвобождать дополнительные электроны и дырки внутри этого слоя посредством процесса, называемого ударной ионизацией. Приборы, в которых происходит такое внутреннее усиление, называются лавинными фотодиодами (ЛФД). ЛФД может использоваться как альтернатива лазерному усилителю, который усиливает оптический сигнал перед его регистрацией, или в сочетании с ним (см. гл. 14 и 17). Однако каждый из упомянутых механизмов усиления вносит свой собственный вид шума.

Таким образом, действие полупроводниковых фотоэлектрических приемников с усилением включает три основных процесса.

1. *Генерация*: поглощение фотона генерирует свободные носители.

2. *Перенос*: приложенное электрическое поле заставляет эти носители двигаться, что вызывает появление тока в цепи

3. *Усиление*: в лавинных фотодиодах большое электрическое поле сообщает свободным носителям достаточную энергию для того, чтобы, в свою очередь, освобождать дополнительные носители путем ударной ионизации. Процесс внутреннего усиления, таким образом, повышает токовую чувствительность приемника.

18.1.2. Общие свойства

Всем полупроводниковым фотоприемникам присущи некоторые общие свойства. Прежде чем изучать детали отдельных фотоприемников, представляющих интерес для фотоники, мы изучим квантовую эффективность, токовую чувствительность и время отклика фотоэлектрических приемников с общей точки зрения.

Полупроводниковые детекторы фотонов и полупроводниковые источники фотонов являются приборами обратного действия. Детекторы служат для того, чтобы преобразовывать поток фотонов на входе в электрический ток на выходе, а источники — наоборот. При изготовлении приборов обоих типов часто используются одни и те же материалы (см. гл. 16). У всех количественных параметров, обсуждаемых в данном разделе, есть свои аналоги в случае источников света (см. гл. 17).

Квантовая эффективность

Квантовой эффективностью η ($0 \le \eta \le 1$) фотоприемника называется вероятность того, что падающий на прибор фотон создаст пару носителей, вносящих вклад в ток приемника. Когда число падающих фотонов, как обычно, велико, η превращается в отношение потока генерируемых электронно-дырочных пар, вносящих вклад в ток детектора, к потоку падающих фотонов.

Не все падающие фотоны генерируют электронно-дырочные пары, поскольку не все они поглощаются средой. Как показано на рис. 18.4, некоторые фотоны отражаются поверхностью детектора, а другие не успевают поглотиться ввиду

недостаточной толщины слоя материала (скорость поглощения фотонов в полупроводниковых материалах была рассмотрена в подразд. 16.2.3). Кроме того, некоторые электронно-дырочные пары, рожденные вблизи поверхности, быстро рекомбинируют благодаря изобилию поверхностных центров рекомбинации, и следовательно, не могут вносить вклада в ток приемника.

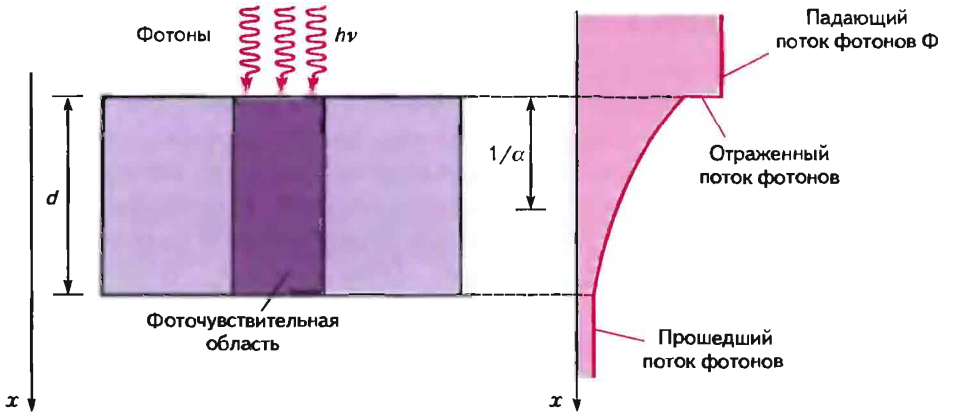


Рис. 18.4. Влияние поверхностного отражения и неполного поглощения на квантовую эффективность приемника η

Таким образом, квантовую эффективность можно выразить как

$$\eta = (1-\mathcal{R})\zeta\left[1-\exp(-\alpha d)\right], \tag{18.3}$$

Квантовая эффективность

где $\mathcal{R}$ — оптический коэффициент отражения на поверхности; ζ — доля электронно-дырочных пар, которые вносят вклад в ток приемника; α — коэффициент поглощения материала (см$^{-1}$), обсуждавшийся в подразд. 16.2.3; d — толщина фотоприемника. Выражение (18.3) представляет собой произведение трех сомножителей.

1. Первый множитель, $(1-\mathcal{R})$, описывает эффект отражения на поверхности прибора. Отражение можно уменьшить, например, применяя антиотражательные покрытия. Некоторые определения квантовой эффективности η не учитывают отражение, которое в этом случае должно учитываться отдельно.

2. Второй множитель есть доля электронно-дырочных пар, которые успешно избегают рекомбинации на поверхности материала и дают вклад в полезный фототок. Поверхностную рекомбинацию можно уменьшить при тщательном выращивании материала и проектировании прибора.

3. Третий множитель,

$$\frac{\int_0^d e^{-\alpha x}\mathrm{d}x}{\int_0^\infty e^{-\alpha x}\mathrm{d}x} = \left[1-\exp(-\alpha d)\right],$$

представляет долю потока фотонов, поглощенных в объеме материала. Чтобы этот множитель был максимальным, толщина прибора d должна быть достаточно большой.

Конечно, возникают дополнительные потери, если свет недостаточно хорошо сфокусирован на активную зону приемника.

Зависимость квантовой эффективности от длины волны

Квантовая эффективность η является функцией длины волны в основном из-за спектральной зависимости коэффициента поглощения α (см. рис. 16.3). Таким образом, свойства полупроводникового материала определяют спектральное окно, в котором η велико. Для достаточно больших значений длины волны λ_0 в свободном пространстве η мало, поскольку поглощение не может происходить при $\lambda_0 \geq \lambda_g = hc_0/E_g$ (т. е. когда энергия фотона меньше ширины запрещенной зоны и материал прозрачен). Длина волны λ_g, соответствующая ширине запрещенной зоны, является, таким образом, **красной границей** полупроводникового материала. Типичные значения E_g и λ_g представлены в табл. 16.2 и показаны на рис. 16.7 и 16.8 для большинства полупроводниковых материалов, представляющих интерес для фотоники. Для достаточно малых значений λ_0 η также убывает, поскольку большинство фотонов в этом случае поглощается вблизи поверхности прибора (например, при $\alpha = 10^4$ см$^{-1}$ большая часть света поглощается на расстоянии $1/\alpha = 1$ мкм). Рекомбинационное время жизни вблизи поверхности гораздо короче, поэтому фотоносители рекомбинируют до того, как будут собраны.

Фотоприемники с резонаторами

Квантовую эффективность η можно повысить путем построения приемника, в котором свет взаимодействует с фоточувствительным материалом, многократно проходя через него. Это эквивалентно увеличению толщины приемника d, что увеличивает поглощение и уменьшает бесполезно прошедший поток фотонов. На практике этого можно достичь, помещая фотоприемник в полость резонатора, удерживающего свет и повышающего квантовую эффективность.

Токовая чувствительность

Токовая чувствительность фотоприемника связывает электрический ток i_p, протекающий через цепь прибора, с оптической мощностью падающего света P. Если бы каждый фотон генерировал в приборе пару фотоносителей, то поток фотонов Ф (число фотонов в секунду) производил бы электронный поток Ф (число электронов в секунду) в цепи фотоприемника, что соответствовало бы току короткого замыкания прибора $i_p = e\Phi$. Таким образом, оптическая мощность $P = h\nu\Phi$ (ватт) на частоте ν вызывала бы электрический ток $i_p = eP/h\nu$.

Однако, поскольку доля фотонов, производящих регистрируемые электроны, равна η, а не единице, электрический ток равен

$$i_p = \eta e \Phi = \frac{\eta e P}{h\nu} \equiv \Re P. \tag{18.4}$$

Коэффициент пропорциональности $\Re = i_p/P$ между электрическим током и оптической мощностью имеет размерность А/Вт и называется токовой чувствительностью фотоприемника:

$$\boxed{\Re = \frac{\eta e}{h\nu} = \eta \frac{\lambda_0}{1,24}.} \tag{18.5}$$

Токовая чувствительность фотоприемника (А/Вт; λ_0 в мкм)

Важно отличать токовую чувствительность фотоприемника (А/Вт) от чувствительности светоизлучающего диода (Вт/А), определяемой выражением (17.28).

Токовая чувствительность пропорциональна квантовой эффективности η и длине волны λ_0 в свободном пространстве, как очевидно из (18.5) и рис. 18.5. Оценка порядка величины токовой чувствительности получается, если положить $\eta = 1$ и $\lambda_0 = 1,24$ мкм в (18.5), тогда $\Re = 1$ А/Вт $= 1$ нА/нВт.

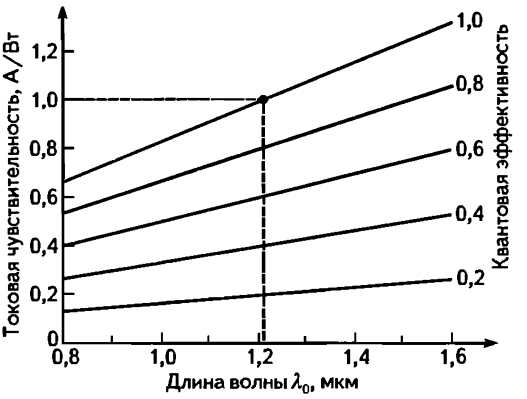


Рис. 18.5. Токовая чувствительность $\Re$ (А/Вт) в зависимости от длины волны l_0 при различных значениях квантовой эффективности η. При $\eta = 1$ $\Re = 1$ А/Вт для $\lambda_0 = 1,24$ мкм

Пропорциональность $\Re$ длине волны λ_0 возникает от того, что токовая чувствительность определяется на основе оптической мощности, тогда как большинство фотоприемников генерируют токи, пропорциональные потоку фотонов Ф. Для данного потока

$$\Phi = \frac{P}{h\nu} = \frac{P\lambda_0}{hc_0}$$

(что соответствует данному току фотоприемника i_p), произведение $P\lambda_0$ фиксировано, так что возрастание λ_0 требует соответственного уменьшения P, а это влечет за собой повышение токовой чувствительности. Некоторые тепловые приемники окликаются на оптическую мошность, а не на поток фотонов, и у них $\Re$ не зависит от λ_0.

Область возрастания $\Re$ с увеличением λ_0 ограничена, поскольку в игру вступает зависимость η от λ_0 в области как больших, так и малых длин волн. Токовая чувствительность падает также при подаче на приемник чрезмерно больших оптических мощностей. Это явление, известное как насыщение приемника, ограничивает **линейный динамический диапазон** приемника, который определяется как область, в которой приемник откликается на мощность падающего света линейным образом.

Приборы с усилением

Представленные выше формулы основаны на предположении, что каждая пара носителей заряда, созданная фотоэффектом (фотоносителей), создает заряд e в цепи фотоприемника. Однако многие приборы создают в цепи заряд q, отличный от e. Такие приборы называются фотоприемниками с усилением. Усиление G определяется как среднее число электронов в цепи, генерируемых одной парой фотоносителей

$$G \equiv \frac{q}{e}. \tag{18.6}$$

Оно может быть как больше, так и меньше единицы, в чем мы убедимся впоследствии.

В присутствии усиления формулы для фототока (18.4) и токовой чувствительности (18.5) нуждаются в модификации. Подставляя $q = Ge$ вместо e в эти выражения, получаем соответственно:

$$\boxed{i_p = \eta q \Phi = \eta G e \Phi = \frac{\eta G e P}{h\nu};} \tag{18.7}$$

Фототок при усилении

$$\boxed{\Re = \frac{\eta G e}{h\nu} = \eta G \frac{\lambda_0}{1{,}24}.} \tag{18.8}$$

Токовая чувствительность фотоприемника (А/Вт; λ_0 в мкм)

Усиление прибора G нужно отличать от его эффективности η — вероятности того, что падающий фотон произведет детектируемую пару фотоносителей. Другие полезные количественные характеристики работы фотоприемника, такие как отношение сигнал—шум и обнаружительная способность, отложим до обсуждения шумовых свойств в разд. 18.6.

Время отклика

Время дрейфа

Постоянное электрическое поле E, приложенное к полупроводнику (или металлу), вызывает ускорение свободных носителей заряда. При этом они испытывают частые столкновения с ионами решетки, совершающими тепловые колебания вблизи своих равновесных положений, а также с дефектами

кристаллической решетки, которые связаны с примесными ионами. Благодаря столкновениям носители испытывают случайные торможения, так что в среднем их движение происходит с некоторой постоянной скоростью, а не с постоянным ускорением. Средняя скорость носителя дается выражением $v = a\tau_{ст}$, где $a = eE/m$ — ускорение, вызываемое электрическим полем; $\tau_{ст}$ — среднее время между столкновениями, которое играет роль времени релаксации. В результате носители дрейфуют в направлении поля со средней **дрейфовой скоростью** $v = e\tau_{ст}E/m$, которую удобно записать в виде

$$v = \mu E, \tag{18.9}$$

где $\mu = e\tau_{ст}/m$ — **подвижность** носителей заряда.

Движение носителей в фотоприемнике создает ток в его внешней цепи. Для определения величины тока $i(t)$ рассмотрим электронно-дырочную пару, созданную (например, путем поглощения фотона) в произвольной точке x полупроводникового материала длиной w, к которому приложено напряжение V, как показано на рис. 18.6, *а*. Ограничимся рассмотрением движения в направлении x и используем энергетические соображения. Если носитель заряда Q (дырка с зарядом $Q = e$ или электрон с зарядом $Q = -e$) проходит расстояние dx за время dt под действием электрического поля напряженностью $E = V/w$, то совершается работа

$$-QE\,dx = -Q\frac{V}{w}dx.$$

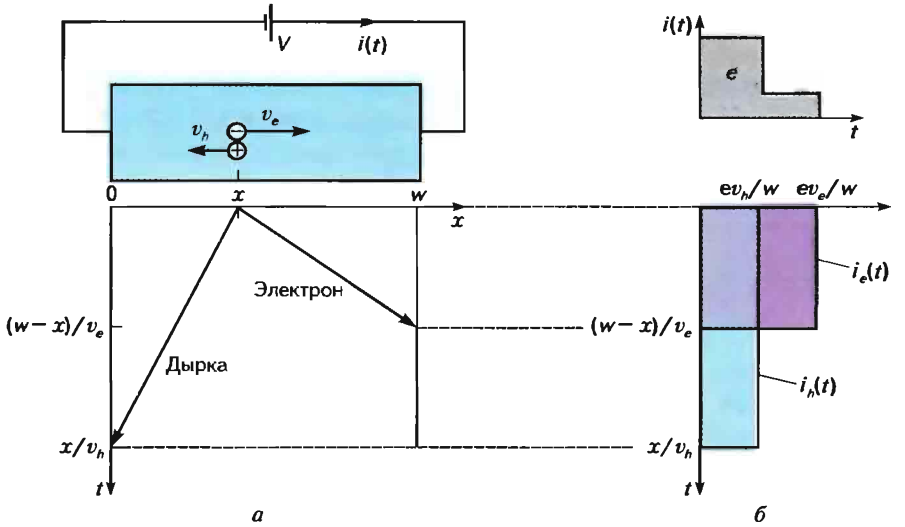


Рис. 18.6. Электронно-дырочная пара рождается в точке x (*а*). Дырка дрейфует влево со скоростью v_h, а электрон вправо со скоростью v_e. Процесс прекращается, когда носители достигают краев материала. Дырочный ток $i_h(t)$, электронный ток $i_e(t)$ и полный ток $i(t)$, индуцированный в цепи (*б*). Полный заряд, индуцированный в цепи на одну пару носителей, равен e

Эта работа должна равняться энергии, полученной от внешней цепи $i(t)V\mathrm{d}t$. Таким образом,

$$i(t)V\,\mathrm{d}t = -Q\frac{V}{w}\,\mathrm{d}x,$$

откуда

$$i(t) = -\frac{Q}{w}\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{Q}{w}v(t).$$

Следовательно, носитель движущийся с дрейфовой скоростью $v(t)$ в направлении x создает ток во внешней цепи, описываемый **теоремой Рамо**:

$$\boxed{i(t) = -\frac{Q}{w}v(t).} \qquad (18.10)$$

Теорема Рамо

Предполагая, что дырка движется со скоростью v_h влево, а электрон со скоростью v_e вправо, находим из (18.10), что дырочный ток равен $i_h = -e(-v_h)/w$, а электронный $i_e = -(-e)v_e/w$, как показано на рис. 18.6, *б*. Каждый носитель вносит вклад в ток до тех пор, пока он движется. Если носители продолжают свое движение до тех пор, пока не достигнут краев материала, то движение дырки продолжается в течение времени x/v_h, а электрона — в течение $(w-x)/v_e$ (см. рис. 18.6, *а*). В полупроводниках v_e обычно больше, чем v_h, так что полная длительность отклика равна x/v_h. Указанная конечная длительность тока обусловлена **временем дрейфа** носителей и является важным фактором, ограничивающим быстродействие всех полупроводниковых фотоприемников.

Можно возразить, что заряд, индуцируемый во внешней цепи при рождении электронно-дырочной пары в результате поглощения фотона материалом фотодетектора, должен быть равен $2e$, поскольку имеется два носителя. На самом деле этот заряд равен e, как показывает вычисление полного заряда q, индуцированного во внешней цепи, как суммы площадей под i_e и i_h:

$$q = e\frac{v_h}{w}\frac{x}{v_h} + e\frac{v_e}{w}\frac{w-x}{v_e} = e\left(\frac{x}{w} + \frac{w-x}{w}\right) = e. \qquad (18.11)$$

Этот результат не зависит от положения точки, в которой рождается электронно-дырочная пара.

Время дрейфа носителей еще больше, если электронно-дырочные пары равномерно рождаются по всему объему материала, как показано на рис. 18.7. При $v_h < v_e$ полная длительность отклика из-за конечного времени дрейфа носителей равна w/v_h, а не x/v_h. Это происходит потому, что равномерное освещение продуцирует пары носителей во всех точках, включая точку $x = w$, из которой дыркам приходится преодолевать наибольшее расстояние, чтобы достичь границы среды $x = 0$, где они, наконец, рекомбинируют.

Итак, теорема Рамо показывает, что доставка заряда во внешнюю цепь благодаря движению носителей в материале фотоприемника происходит не мгно-

венно, а в течение конечного промежутка времени. Это выглядит так, словно движение заряженных частиц внутри материала медленно вытягивает заряд из провода с одной стороны прибора и медленно вталкивает в провод с другой стороны, так что прохождение каждого заряда по внешней цепи оказывается растянутым по времени.

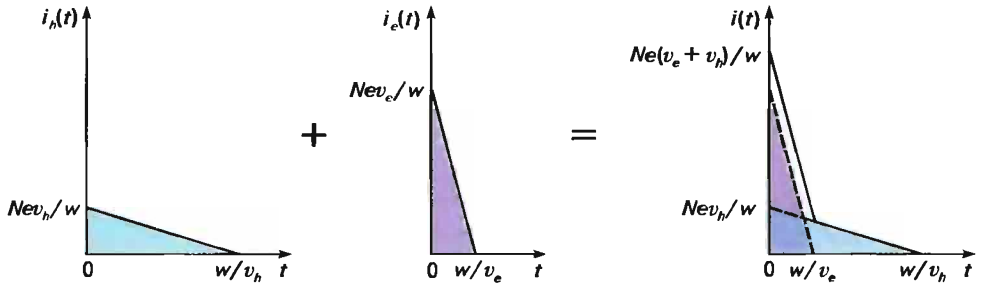


Рис. 18.7. Дырочный ток $i_h(t)$, электронный ток $i_e(t)$ и полный ток $i(t)$, индуцируемый в цепи рождением электронно-дырочных пар при поглощении N фотонов, равномерно распределенных между 0 и w (см. задачу 4 к разд. 18.1). Хвост полного тока возникает из-за движения дырок. Полный ток $i(t)$ можно рассматривать как функцию отклика на импульсное воздействие (см. приложение Б, разд. Б.1) для однородно освещаемого приемника с конечным временем дрейфа носителей

Закон Ома

Если вместо отдельного точечного заряда Q имеется непрерывное распределение заряда с плотностью ρ, то полный заряд, содержащийся в материале фотоприемника, равен ρAw, где A — площадь поперечного сечения (см. рис. 18.6, *а*). Тогда уравнение (18.10) дает

$$i(t) = -\frac{\rho Aw}{w} v(t) = -\rho A v(t),$$

так что плотность тока в направлении x:

$$J(t) = -\frac{i(t)}{A} = \rho v(t).$$

Хорошо известная векторная форма этого выражения

$$\boxed{J = \rho v.} \qquad (18.12)$$

Плотность тока

Комбинируя выражения (18.12) и (18.9), получаем $J = \sigma E$, где σ — удельная электропроводность среды,

$$\sigma = \rho\mu = e\rho\frac{\tau_{ст}}{m} = Ne^2\frac{\tau_{ст}}{m}, \qquad (18.13)$$

где N — число носителей в единице объема (см. подразд. 5.5.4). В более общем случае удельная электропроводность представляет собой тензор σ, и векторная версия этого выражения является **законом Ома**:

$$\boxed{J = \sigma \mathcal{E}.} \tag{18.14}$$

Закон Ома

Для однородного проводника с площадью поперечного сечения A и длиной w $J = \sigma E$ можно записать как

$$i = \frac{\sigma A}{w} Ew = \frac{\sigma A}{w} V = GV = \frac{V}{R},$$

где G и R — проводимость и сопротивление материала соответственно. В этом случае закон Ома принимает свою наиболее известную форму

$$V = iR. \tag{18.15}$$

Постоянная времени *RC*-цепи

Сопротивление R и емкость C фотоприемника, последовательно включенные в цепь, порождают еще одно конечное время отклика, связанное с постоянной времени образующейся RC-цепи, $\tau_{RC} = RC$. Сочетание емкости и сопротивления интегрирует ток на выходе приемника и удлиняет функцию отклика на импульсное воздействие. При наличии конечного времени дрейфа носителей и переходного процесса в RC-цепи полная функция отклика на импульсное воздействие определяется сверткой тока $i(t)$, показанного на рис. 18.7, с экспоненциальной функцией $(1/RC)\exp(-t/RC)$ (см. разд. Б.1).

Стоит заметить, что фотоприемники различных типов могут иметь свои собственные ограничения скорости отклика, которые мы рассмотрим по мере возникновения необходимости. Наконец, отметим, что фотоприемники с данной структурой и материалом часто имеют постоянное произведение ширины полосы на усиление. Повышение усиления ведет к сужению полосы и наоборот. Необходимость компромисса между чувствительностью и частотным откликом — следствие конечного времени, необходимого для процесса усиления.

18.2. ФОТОПРОВОДНИКИ

При поглощении фотонов полупроводником подвижные носители заряда *генерируются* (в идеальном случае одна электронно-дырочная пара на каждый фотон). Удельная электропроводность материала σ растет пропорционально потоку фотонов Ф. Электрическое поле, приложенное к материалу от внешнего источника напряжения, вызывает перенос электронов и дырок. В свою очередь, это приводит к возникновению измеряемого электрического

тока во внешней цепи, как показано на рис. 18.8, *а*. Фотопроводящие приемники действуют на основе измерения либо фототока i_p, который пропорционален потоку фотонов Ф, либо падения напряжения на сопротивлении нагрузки R, включенном в цепь последовательно.

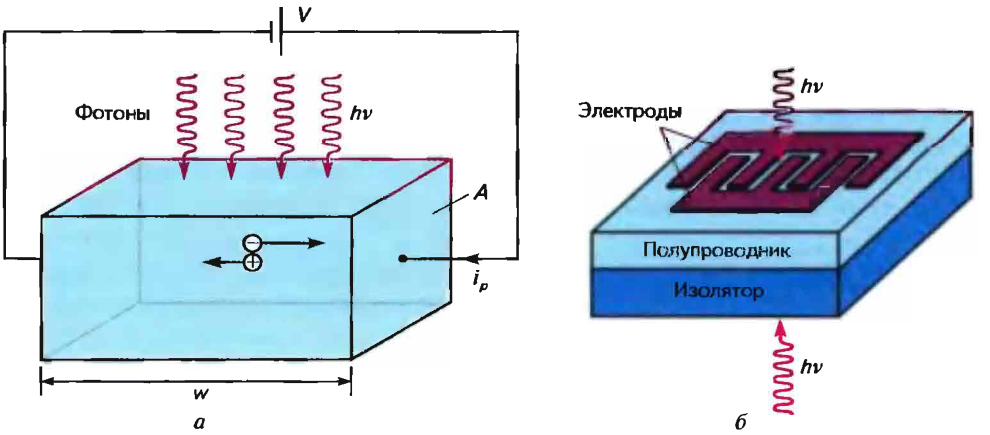


Рис. 18.8. Фотопроводящий приемник (*а*). Генерируемые светом пары носителей движутся под действием приложенного напряжения V, создавая фототок i_p, пропорциональный падающему потоку фотонов Ф. Взаимопроникающая структура электродов предназначена для минимизации времени дрейфа носителей с целью расширения частотной полосы прибора (*б*)

18.2.1. Собственные полупроводники

Если энергия фотона больше ширины запрещенной зоны, то фотоны поглощаются посредством межзонных переходов. Прибор на основе фотопроводимости может иметь форму слоя или тонкой пленки. Анодный и катодный контакты часто размещают на одной стороне прибора в виде близко смыкающихся структур для обеспечения максимального доступа света в материал и уменьшения дрейфового времени (рис. 18.8, *б*). Свет может поступать с нижней стороны прибора, если изолирующая подложка имеет достаточную ширину запрещенной зоны, чтобы обеспечить прозрачность.

Прирост удельной электропроводности, вызванный потоком фотонов Ф (число фотонов в секунду), освещающим полупроводник объема wA (см. рис. 18.8), рассчитывается следующим образом. Доля η падающего потока фотонов поглощается и рождает избыточные пары электрон—дырка. Скорость R генерации пар (на единицу объема) составляет $R = \eta\Phi/wA$. Если τ — время релаксации избыточных носителей, то электроны теряются со скоростью $\Delta n/\tau$, где Δn — концентрация электронов (см. гл. 16). В стационарном режиме обе скорости одинаковы, $R = \Delta n/\tau$, так что $\Delta n = \eta\tau\Phi/wA$. Рост концентрации носителей Δn сопровождается ростом плотности заряда $\Delta\rho = e\Delta n$ и,

следовательно, в соответствии с (18.13) ростом удельной электропроводности $\Delta\sigma = \Delta\rho\mu = e\Delta n\mu$, так что

$$\Delta\sigma = \frac{\eta e\tau(\mu_e + \mu_h)}{wA}\Phi, \tag{18.16}$$

где μ_e и μ_h — подвижность электронов и дырок соответственно. В соответствии с (18.16) рост удельной электропроводности пропорционален потоку фотонов.

Согласно закону Ома (18.14), плотность фототока $J_p = \Delta\sigma E$. Объединяя это с (18.16) и (18.9), откуда $v_e = \mu_e E$ и $v_h = \mu_h E$, получим

$$J_p = \left[\eta e\tau\frac{v_e + v_h}{wA}\right]\Phi,$$

что соответствует электрическому току

$$i_p = AJ_p = \left[\eta e\tau\frac{v_e + v_h}{w}\right]\Phi.$$

Если $v_e \ll v_h$ и формула записана через время дрейфа электрона через образец $\tau_e = w/v_e$, мы получаем

$$i_p \approx \eta\frac{\tau}{\tau_e}e\Phi. \tag{18.17}$$

Сравнение с (18.7) показывает, что отношение τ/τ_e в (18.17) соответствует усилению приемника G по причинам, к выяснению которых мы приступаем.

Усиление

Токовая чувствительность фотопроводника с усилением дается формулой (18.8). Упрощенно можно сказать, что прибор демонстрирует внутреннее усиление по той причине, что рекомбинационное время жизни отличается от времени дрейфа. Предположим, что электроны движутся быстрее дырок (см. рис. 18.8) и времена рекомбинации очень велики. При переносе электронов и дырок к противоположным сторонам фотопроводника электроны завершают свой путь быстрее, чем дырки. Условие непрерывности тока требует, чтобы во внешней цепи немедленно появился другой электрон, входящий в прибор из левого провода. Этот новый электрон быстро движется вправо, снова завершая свой путь до того, как дырка достигнет левого края. Процесс продолжается до тех пор, пока не произойдет рекомбинация электрона и дырки.

Таким образом, поглощение одного фотона может вызывать многократное прохождение электронов через внешнюю цепь. Ожидаемое число проходов, которое может сделать электрон до того, как процесс прервется, равно

$$G = \frac{\tau}{\tau_e}, \tag{18.18}$$

где τ — рекомбинационное время жизни избыточных носителей; τ_e — время дрейфа электрона через образец. Заряд, сообщенный внешней цепи одной электронно-дырочной, парой тогда составит $q = Ge > e$, так что прибор демонстрирует усиление.

В другом предельном случае рекомбинационное время жизни достаточно мало, так что носители рекомбинируют до того, как достигнут краев материала. Это может случиться, если имеется изобилие носителей противоположного типа, с которыми можно рекомбинировать. В этом случае $\tau < \tau_e$ и усиление меньше единицы, так что в среднем каждая пара носителей вносит только долю электронного заряда во внешнюю цепь. Конечно, заряд сохраняется, и множество пар носителей сообщает внешней цепи заряд, равный целому числу электронных зарядов.

Таким образом, усиление фотопроводника $G = \tau/\tau_e$ можно интерпретировать как долю длины образца, проходимую в среднем возбужденным носителем до того, как он исчезнет в процессе рекомбинации. Время дрейфа τ_e определяется длиной прибора и приложенным напряжением по формуле (18.9) и $\tau_e = w/v_e$; типичные значения $w = 1$ мм и $v_e = 10^7$ см/с дают $\tau_e \approx 10^{-8}$ с. Рекомбинационное время жизни может иметь значение от 10 до многих секунд в зависимости от фотопроводящего материала и примесей [см. (16.24)]. Поэтому G может лежать в очень широком интервале от значений, меньших единицы, до значений, много больших единицы. Это зависит от материала, размеров прибора и приложенного напряжения. Однако усиление фотопроводника, как правило, не может превышать 10^6 в силу ограничений, накладываемых пространственной ограниченностью тока, ударной ионизацией и диэлектрическим пробоем.

Спектральный отклик

Спектральная чувствительность фотопроводников определяется в основном зависимостью ηRC от длины волны, что обсуждалось в подразд. 18.1.2. Разные полупроводники имеют разные красные границы (см., например, табл. 16.2). В режиме фотопроводимости могут использоваться одно-, двух- и трехкомпонентные полупроводники. Фотопроводящие приемники (в отличие от фотоэмиссионных) работают в инфракрасной области на межзонных переходах. Однако работа на длинах волн более 2 мкм обычно требует охлаждения приборов для минимизации теплового возбуждения электронов в зону проводимости в этих материалах с малой шириной запрещенной зоны.

Время отклика

Время отклика фотопроводящего приемника, конечно, ограничено соображениями, касающимися времени дрейфа носителей и постоянной времени RC-цепи, которые обсуждались в подразд. 18.1.2. Время отклика, связанное с переносом носителей, примерно равно времени рекомбинации τ, так что связанная с ним ширина полосы B обратно пропорциональна τ. По-

скольку в соответствии с (18.18) усиление G прямо пропорционально τ, рост τ служит для увеличения усиления, что желательно, но одновременно уменьшает ширину полосы, что нежелательно. Таким образом, произведение GB усиления на ширину полосы оказывается приблизительно не зависящим от τ, типичные значения GB достигают $\approx 10^9$

18.2.2. Примесные материалы

Фотопроводимость достигается на более длинных волнах при использовании легированных полупроводников. Подвижные носители заряда могут генерироваться посредством поглощения фотонов примесными центрами, уровни энергии которых лежат внутри запрещенной зоны. Процесс может идти по одному из двух путей:

1) падающий фотон взаимодействует со связанным электроном донорного центра, который переходит в зону проводимости и оставляет после себя связанную дырку;

2) падающий электрон взаимодействует со связанной дыркой акцепторного центра, в результате чего она переходит в валентную зону, оставляя после себя связанный электрон, как показано на рис. 16.43, *б*.

Донорные и акцепторные уровни в запрещенной зоне примесного полупроводника могут иметь очень низкие энергии активации E_A, в силу чего длина волны $\lambda_A = hc_0/E_A$ красной границы материала может быть весьма большой. Такие приемники необходимо охлаждать во избежание теплового возбуждения, для чего часто используется жидкий гелий при 4 °K. Типичные значения E_A и λ_A приведены в табл. 18.1 для ряда примесных полупроводников.

Таблица 18.1. **Некоторые примесные полупроводниковые материалы со своей энергией активации и длиной волны красной границы**

Полупроводник: примесь	E_A, эВ	λ_A, мкм
Ge : Hg	0,088	14
Ge : Cu	0,041	30
Ge : Zn	0,033	38
Ge : Ga	0,010	115
Si : B	0,044	23

Спектральные характеристики нескольких примесных полупроводниковых материалов приведены на рис. 18.9. Токовая чувствительность растет почти линейно с увеличением λ_0 в соответствии с (18.8), имеет максимум, немного не доходя красной границы, и затем резко падает. Квантовая эффективность таких приемников может быть весьма высокой (например, $\eta \approx 0{,}5$ для Ge : Cu),

хотя усиление может быть низким при обычных условиях работы (например, $G \approx 0{,}03$ для Ge : Cu).

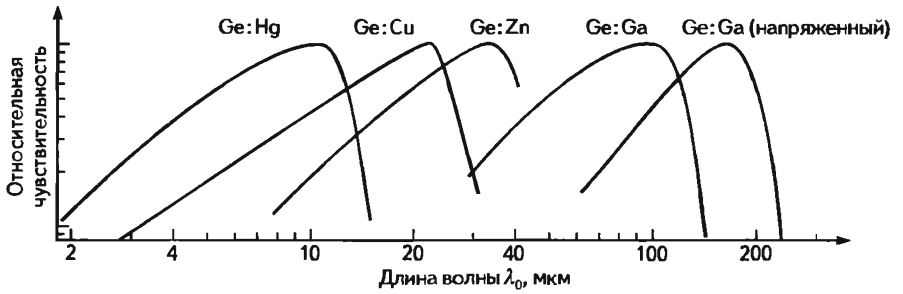


Рис. 18.9. Относительная чувствительность в зависимости от длины волны λ_0 (мкм) для пяти различных материалов на основе германия, отличающихся примесями и используемых как фотопроводящие приемники инфракрасного света

18.2.3. Гетероструктуры

Должным образом конфигурированные гетероструктуры могут служить хорошими фотопроводящими приемниками. Примером является **инфракрасный фотоприемник на квантовых ямах** (quantum-well infrared photodetector, QWIP). Падающий инфракрасный фотон переводит электрон с дискретного уровня связанного состояния в квантовой яме в непрерывный спектр, таким образом создавая подвижный носитель заряда, увеличивающий удельную электропроводность материала (рис. 18.10).

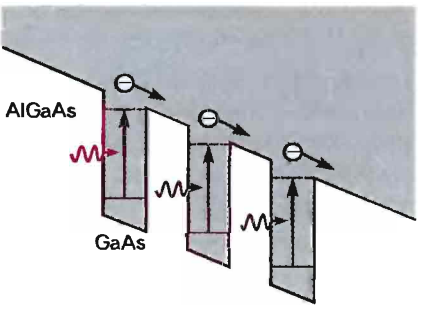


Рис. 18.10. Генерация подвижных носителей заряда при поглощении фотонов в QWIP. Прибор сконструирован таким образом, что в каждой яме имеется единственный дискретный уровень энергии, обеспечивающий чувствительность в некоторой спектральной полосе. У показанного приемника барьеры состоят из AlGaAs, а ямы из GaAs. В ямах находятся электроны на дискретных уровнях энергии. QWIP, изготовленные из полупроводников III—V групп, обеспечивают высокую чувствительность на длинах волн от среднего до дальнего инфракрасного диапазона ($\lambda_0 \approx 4$—20 мкм) и высокое быстродействие, но требуют охлаждения

Инфракрасный фотоприемник на квантовых точках (quantum-dot infrared photodetector, QDIP) — вариация на ту же тему; он может применяться для многоволновой регистрации инфракрасного излучения за счет переходов между подзонами.

18.3. ФОТОДИОДЫ

18.3.1. *p—n*-Фотодиод

Как у фотопроводников, действие **фотодиодных приемников** основано на фотогенерации подвижных носителей заряда. Фотодиод представляет собой *p—n*-переход (см. подразд. 16.1.5), обратный ток которого возрастает при поглощении фотонов. Хотя *p—n*- и *p—i—n*-фотодиоды обычно имеют большее быстродействие, чем фотопроводники, они не обладают усилением.

Рассмотрим обратносмещенный *p—n*-переход в условиях освещения, как показано на рис. 18.11. Фотоны поглощаются везде с коэффициентом поглощения α. Как только поглощается фотон, *генерируется* электронно-дырочная пара. Но лишь там, где присутствует электрическое поле, носители заряда *переносятся* в определенном направлении. Поскольку в *p—n*-переходе электрическое поле имеется только в обедненном слое, именно здесь желательна фотогенерация носителей.

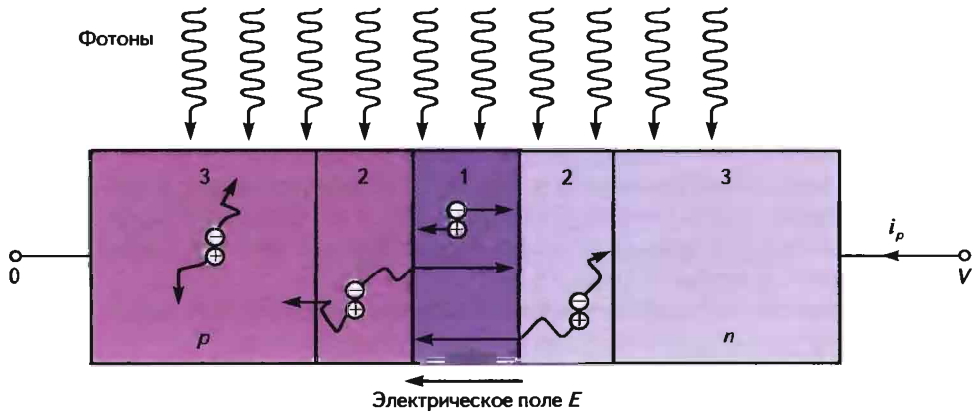


Рис. 18.11. Фотоны, облучающие идеализированный фотодиодный приемник с обратносмещенным *p—n*-переходом. Области дрейфа и диффузии обозначены цифрами 1 и 2 соответственно

Однако имеется три возможные области, в которых может происходить генерация электронно-дырочных пар.

1. Электроны и дырки, генерируемые в обедненном слое (область 1), быстро дрейфуют в противоположных направлениях под действием сильного электрического поля. Поскольку электрическое поле всегда указывает направление $n \rightarrow p$, электроны движутся в сторону *n*-области, а дырки в сторону *p*-области. В результате фототок, создаваемый во внешней цепи, всегда имеет обратное направление (от *n*- к *p*-области). Каждая пара носителей генерирует во внешней цепи импульс электрического тока с площадью e ($G = 1$), поскольку в обедненном слое рекомбинация не происходит.

2. Электроны и дырки, генерируемые вне обедненного слоя (область 3), не могут направленно переноситься ввиду отсутствия электрического поля. Они случайным образом блуждают, пока не аннигилируют в результате рекомбинации. Они не вносят сигнала во внешнюю электрическую цепь.

3. Электроны и дырки, генерируемые вне обедненного слоя, но в непосредственной близости от него (область 2), имеют шанс попасть в обедненный слой в результате диффузии. Электрон, приходящий со стороны *p*-области, быстро переносится через переход и вносит заряд *e* во внешнюю цепь. Дырка, приходящая со стороны *n*-области, дает такой же эффект.

Фотодиоды изготавливаются из многих полупроводниковых материалов, перечисленных в табл. 16.2, а также двух-, трех- и четырехкомпонентных полупроводников, таких как SiC, InGaAs и InGaAsP. Приборы часто делаются таким образом, что свет падает на *p*—*n*-переход по нормали, а не параллельно переходу. В этом случае дополнительный ток из-за диффузии носителей увеличивает η, однако этот эффект уравновешивается уменьшением толщины материала, понижающим η.

Время отклика

Время переноса носителей, дрейфующих через обедненный слой (w_d/v_e для электронов и w_d/v_h для дырок), и время отклика *RC*-цепи играют в формировании времени отклика фотодиода роль, описанную в подразд. 18.1.2. Результирующий ток в цепи показан на рис. 18.6, *б* для электронно-дырочной пары, созданной в положении *x*, и на рис. 18.7 для однородной генерации электронно-дырочных пар.

У фотодиодов имеется дополнительный вклад в время отклика, происходящий от диффузии. Носители, образованные вне обедненного слоя, но достаточно близко к нему, диффундируют в него за конечное время. Это достаточно медленный процесс по сравнению с дрейфом. Максимальное время, в течение которого диффузия может продолжаться, равно времени жизни носителей (τ_p для электронов в *p*-области и τ_n для дырок в *n*-области). Влияние диффузии можно уменьшить за счет использования *p*—*i*—*n*-диодов, как мы увидим впоследствии.

Тем не менее фотодиоды обычно имеют большее быстродействие, чем фотопроводники, так как сильное поле в обедненном слое сообщает полученным в результате фотогенерации носителям большую скорость. Более того, фотодиоды не подвержены действию многих эффектов захвата носителей, присущих фотопроводникам.

Смещение

Как электронный прибор, фотодиод имеет вольт-амперную характеристику, определяемую соотношением

$$i = i_s\left[\exp\left(\frac{eV}{k_B T}\right) - 1\right] - i_p \qquad (18.19)$$

и изображенную на рис. 18.12. Это обычная вольт-амперная характеристика p—n-перехода [см. (16.32)] с добавленным фототоком $-i_p$, пропорциональным потоку фотонов.

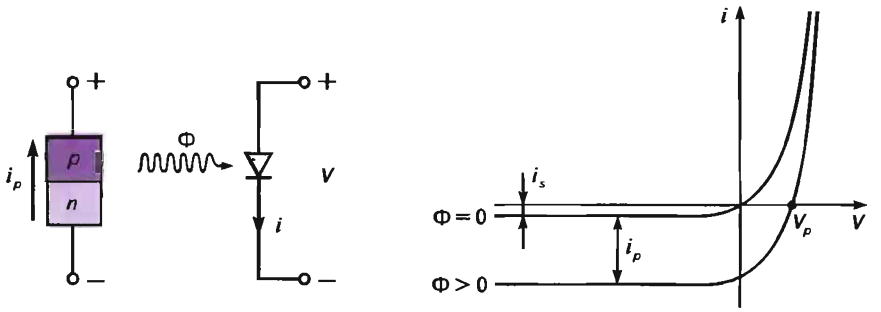


Рис. 18.12. Общая схема фотодиода и его вольт-амперная характеристика

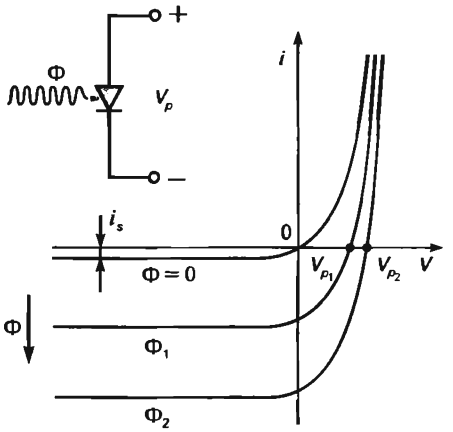


Рис. 18.13. Фотогальванический режим работы фотодиода

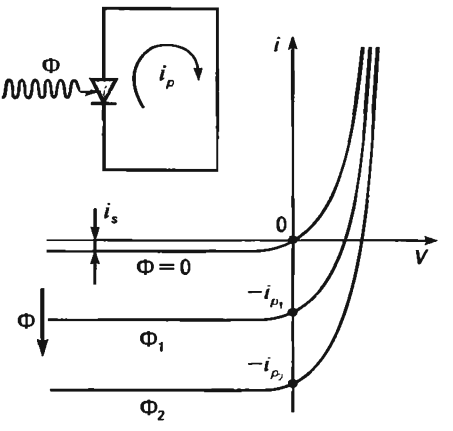


Рис. 18.14. Короткозамкнутый режим работы фотодиода

Существует три классических режима работы фотодиода: режим холостого хода (фотогальванический), короткозамкнутый и обратносмещенный (фотопроводящий). В режиме холостого хода (рис. 18.13) свет генерирует электронно-дырочные пары в обедненном слое. Дополнительные электроны, высвобожденные с n-стороны слоя, рекомбинируют с дырками с p-стороны и наоборот. В результате повышается электрическое поле, что создает на приборе фотоЭДС, величина которой растет с увеличением потока фотонов. Этот режим используется, например, в солнечных батареях. Чувствительность фотогальванического диода измеряется в вольтах на ватт, а не в амперах на ватт (вольтовая чувствительность). Короткозамкнутый режим ($V = 0$) показан на рис. 18.14. Ток короткого замыкания представляет собой просто фототок i_p.

Наконец, фотодиод может работать в режиме обратного смещения, или «фотопроводящем» режиме, как показано на рис. 18.15, *а*. Если в цепь последовательно включается сопротивление нагрузки, прибор работает, как показано на рис. 18.15, *б*.

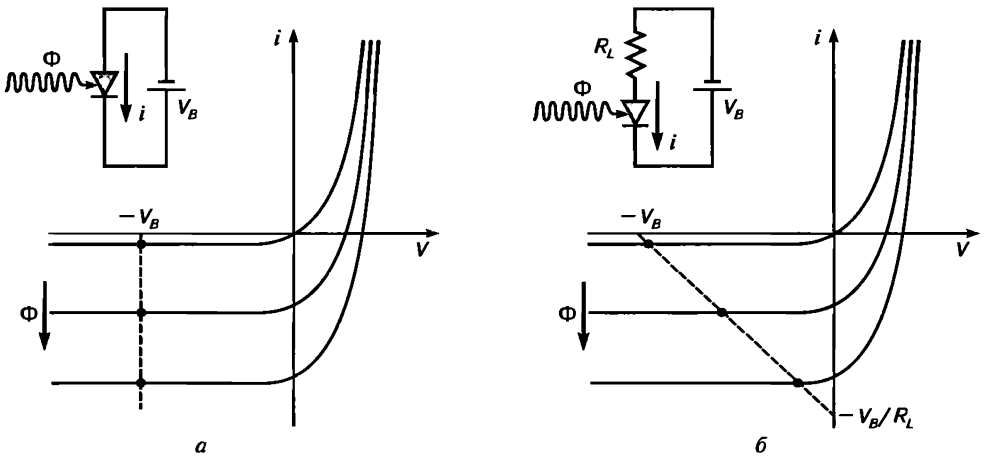


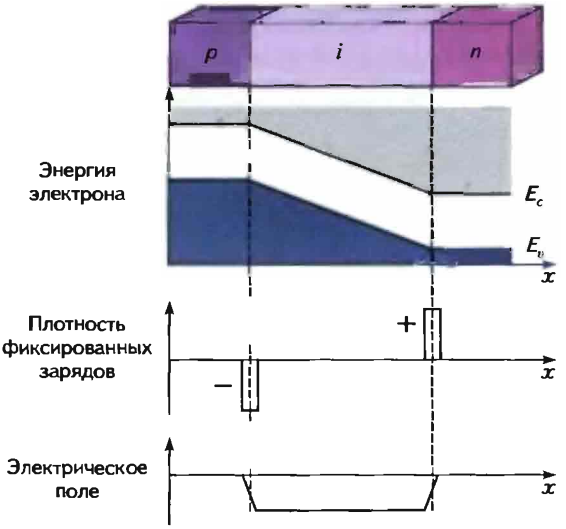
Рис. 18.15. Обратносмещенный режим работы фотодиода без сопротивления нагрузки (*а*) и с сопротивлением нагрузки (*б*). Рабочая точка лежит на штриховой линии

18.3.2. *p—i—n*-Фотодиод

Как приемник, *p—i—n*-фотодиод имеет ряд преимуществ перед *p—n*-фотодиодом. *p—i—n*-фотодиод представляет собой *p—n*-переход со слоем собственного (обычно слабо легированного) полупроводника между *p*- и *n*-слоями (см. подразд. 16.1.5). Он может работать при различных условиях смещения, обсуждавшихся в предыдущем разделе. Схема энергетических зон, распределение заряда и электрического поля в обратносмещенном *p—i—n*-фотодиоде показаны на рис. 18.16. Эта структура служит для расширения ширины области, в которой существует электрическое поле, в сущности, для расширения обедненного слоя.

Фотодиоды с *p—i—n*-структурой имеют следующие преимущества.

- Увеличение ширины обедненного слоя (в котором генерируемые носители переносятся за счет дрейфа) увеличивает площадь, с которой собирается свет.
- Увеличение ширины обедненного слоя уменьшает емкость перехода и, следовательно, *RC*-постоянную времени. С другой стороны, время дрейфа увеличивается с ростом толщины обедненного слоя.
- Уменьшение отношения длины диффузии к дрейфовой длине приводит к тому, что большая часть генерируемого тока переносится более быстрым дрейфовым процессом.



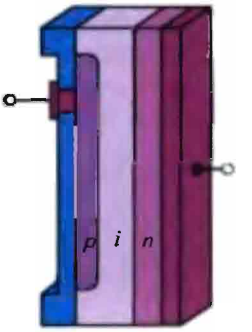


Рис. 18.16. Структура p—i—n-фотодиода, схема энергетических зон, распределение заряда и электрического поля. Прибор может освещаться либо перпендикулярно, либо параллельно плоскости перехода

На рис. 18.17 сравнивается токовая чувствительность идеального p—i—n-фотодиода ($\eta = 1$) и двух промышленных приборов. Максимальная токовая чувствительность приходится на длину волны, меньшую, чем соответствующая ширине запрещенной зоны. Это происходит потому, что Si — непрямозонный полупроводник. Поэтому переходы с поглощением фотонов в типичном случае происходят из состояний валентной зоны в состояния зоны проводимости, лежащие достаточно высоко над дном зоны (см. рис. 16.37).

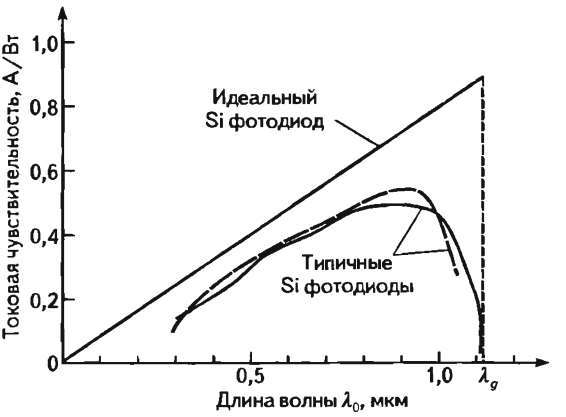


Рис. 18.17. Зависимость токовой чувствительности от длины волны (мкм) для идеального и двух серийно выпускаемых кремниевых p—i—n-фотодиодов. Квантовая эффективность тщательно спроектированного кремниевого фотодиода с антиотражательными покрытиями может приближаться к единице

18.3.3. Гетероструктуры

Гетероструктурные фотодиоды, образованные двумя полупроводниками с различной шириной запрещенной зоны, могут обладать преимуществами перед диодами на p—n-переходах из одного материала. Например, гете-

роперехода, включающий материал с большой шириной запрещенной зоны ($E_g > h\nu$), позволяет уменьшить потери на поглощение вне обедненного слоя за счет прозрачности этого материала. В этом случае материал с большой шириной запрещенной зоны играет роль **приемного окна**. Использование различных материалов позволяет также создавать приборы с большей гибкостью. Наибольший интерес представляют несколько материалов (см. рис. 16.7 и 16.8).

- $Al_xGa_{1-x}As/GaAs$ (решетка AlGaAs согласуется с подложкой GaAs) используется в области длин волн от 0,7 до 0,87 мкм.
- $In_xGa_{1-x}As/InP$ (решетка InGaAs согласуется с подложкой InP) можно перестраивать за счет состава в диапазоне длин волн 1300—1600 нм, представляющем интерес для оптоволоконной связи (см. подразд. 24.1.4). Типичный p—i—n-фотодиод, работающий на 1550 нм, имеет квантовую эффективность $\eta \approx 0{,}75$ и токовую чувствительность $\Re \approx 0{,}9$ А/Вт.
- $Hg_xCd_{1-x}Te/CdTe$ — материал, широко используемый в средней инфракрасной области спектра. Это объясняется тем, что HgTe и CdTe имеют почти одинаковые постоянные решетки и поэтому могут быть согласованы по постоянной решетки почти при любом составе. Данный материал имеет перестраиваемую за счет состава ширину запрещенной зоны, что позволяет работать в диапазоне длин волн от 3 до 17 мкм. Список приложений включает приборы ночного зрения, тепловизоры и длинноволновую оптическую связь.
- Четырехкомпонентные материалы, такие как $In_{1-x}Ga_xAs_{1-y}P_y/InP$ и $Ga_{1-x}Al_xAs_ySb/GaSb$, применяемые в области от 0,92 до 1,7 мкм, представляют интерес, так как четвертый материал обеспечивает дополнительную степень свободы, позволяющую осуществлять согласование постоянных решетки при различных значениях E_g, определяемых составом соединения.

Фотодиоды на барьерах Шоттки

Фотодиоды на структурах «металл—полупроводник», называемые также **фотодиодами на барьерах Шоттки**, содержат гетеропереходы между металлом и полупроводником. Тонкая полупрозрачная металлическая пленка используется вместо слоя p-типа (или n-типа) в фотодиоде на p—n-переходе. Эта тонкая пленка часто делается из сплава металла с полупроводником, который ведет себя как металл. Структура с барьером Шоттки и ее энергетические зоны схематически показаны на рис. 18.18.

Фотодиоды на барьерах Шоттки применяются по многим причинам:

- Не все полупроводники могут иметь оба типа, и n, и p; диоды на барьерах Шоттки особенно важны при использовании таких материалов.
- Полупроводники, используемые для приема видимого и ультрафиолетового излучения с энергией фотона намного больше ширины запрещенной зоны, имеют большой коэффициент поглощения. Это ведет к значительной поверхностной рекомбинации и снижению квантовой эффективности. Переход «металл—полупроводник» имеет обедненный слой непосредственно на поверхности, что резко сокращает поверхностную рекомбинацию.

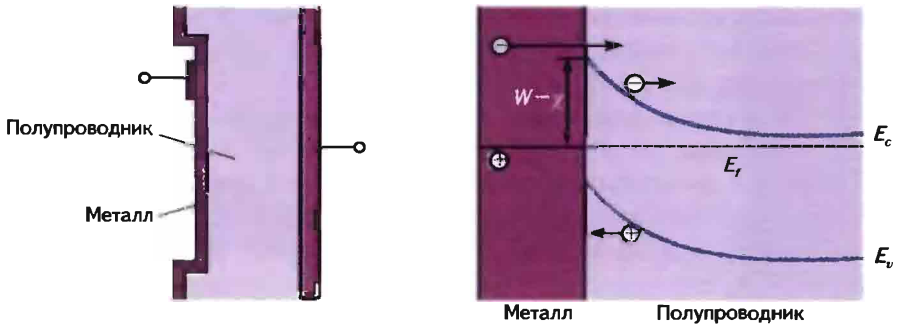


Рис. 18.18. Структура (*а*) и схема (*б*) энергетических зон фотодиода с барьером Шоттки, образованного нанесением металла на полупроводник *n*-типа. Такие фотоприемники чувствительны к фотонам, энергия которых больше, чем высота барьера Шоттки, $h\nu > W - \chi$. Фотодиоды Шоттки можно изготавливать из многих материалов, таких как Au на Si *n*-типа (для видимого диапазона) и силицид платины (PtSi) на Si *p*-типа (для широкой области длин волн от ультрафиолетового до инфракрасного диапазона)

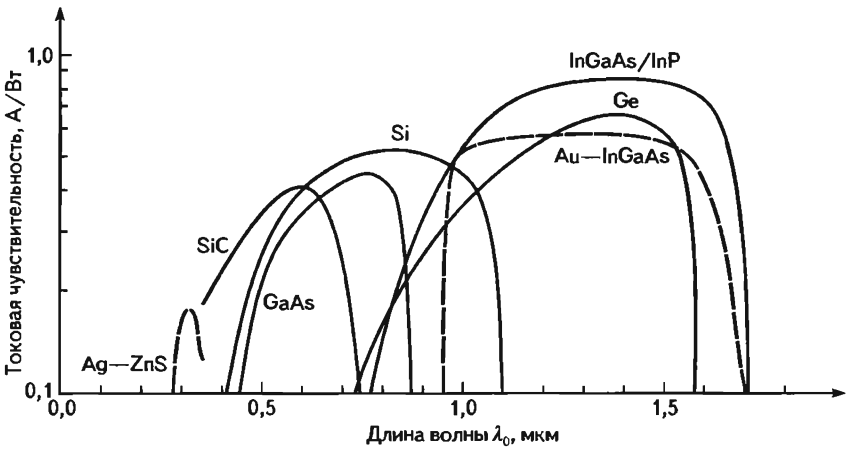


Рис. 18.19. Токовая чувствительность $\Re$ в зависимости от длины волны λ_0 (мкм) для ряда *p—i—n*-фотодиодов (сплошные линии) и фотодиодов на барьерах Шоттки (штриховые линии). Для трех- и четырехкомпонентных полупроводников длина волны максимального отклика зависит от состава. Типичные времена отклика достигают десятков пикосекунд, что соответствует ширине полосы ≈50 ГГц

• Скорость отклика *p—n*- и *p—i—n*-фотодиодов частично ограничена медленным током диффузии фотоносителей, генерируемых вне обедненного слоя, но в непосредственной близости от его границ. Один из способов снижения этого нежелательного поглощения состоит в уменьшении толщины одного из слоев, образующих переход. Однако это должно быть достигнуто без существенного увеличения последовательного сопротивления прибора, так как по-

следнее нежелательным образом увеличивает постоянную времени *RC*-цепи. В структурах с барьером Шоттки это достигается за счет низкого сопротивления металла. Более того, структуры с барьером Шоттки работают на основных носителях и, следовательно, имеют исходно быстрый отклик и большую ширину рабочей полосы. Легко достигаются пикосекундные времена отклика, что соответствует ширине полосы ≈100 ГГц.

Типичные кривые токовой чувствительности для нескольких фотодиодов с *p*—*i*—*n*-переходами и барьерами Шоттки показаны на рис. 18.19.

18.4. ЛАВИННЫЕ ФОТОДИОДЫ

Лавинный фотодиод (ЛФД) работает на основе преобразования каждого детектируемого фотона в *каскад* пар подвижных носителей заряда. В этом случае слабый свет способен вызвать электрический ток, достаточно большой для последующей регистрации электронной аппаратурой. Прибор устроен как фотодиод с сильным обратным смещением, в котором электрическое поле перехода велико. Тогда носители заряда могут приобретать достаточную энергию, чтобы возбуждать новые носители в процессе **ударной ионизации**.

18.4.1. Принципы действия

История жизни типичной электронно-дырочной пары в обедненной области ЛФД изображена на рис. 18.20. Фотон поглощается в точке 1, создавая электронно-дырочную пару (электрон в зоне проводимости и дырку в валентной зоне). Электрон ускоряется в сильном электрическом поле, увеличивая свою энергию относительно дна зоны проводимости. Процесс ускорения постоянно прерывается случайными столкновениями с решеткой, в которых электрон теряет часть приобретенной им энергии. Эти конкурирующие процессы приводят к достижению средней установившейся скорости. Если в любой момент в течение этого процесса электрону удастся накопить энергию, превышающую E_g, он получает возможность генерации второй электронно-дырочной пары посредством ударной ионизации (скажем, в точке 2). Тогда в электрическом поле ускоряются уже два электрона, и каждый из них может быть источником последующей ударной ионизации. Дырки, генерируемые в точках 1 и 2, также ускоряются, двигаясь влево. Каждая из них также имеет шанс вызвать ударную ионизацию, если они приобретут достаточную энергию, и породить новую электронно-дырочную пару, инициированную дыркой (например, в точке 3).

Коэффициент ионизации

Способность электронов и дырок вызывать ударную ионизацию характеризуется коэффициентами ионизации α_e и α_h. Эти величины представляют собой вероятности ионизации на единицу длины ($см^{-1}$); обратные ве-

личины — $1/\alpha_e$ и $1/\alpha_h$ — средние длины пробега между двумя последовательными актами ионизации. Коэффициенты ионизации растут с увеличением напряженности электрического поля внутри обедненного слоя (поскольку именно оно обеспечивает ускорение) и падают с ростом температуры (поскольку участившиеся столкновения с решеткой мешают носителям приобрести энергию, достаточную для ионизации). Упрощенная теория, представленная ниже, предполагает, что α_e и α_h постоянны. Однако для уменьшения шумов выгодно разрабатывать приборы, у которых коэффициенты ионизации зависят от положения и истории носителя определенным образом, как обсуждается в подразд. 18.6.2.

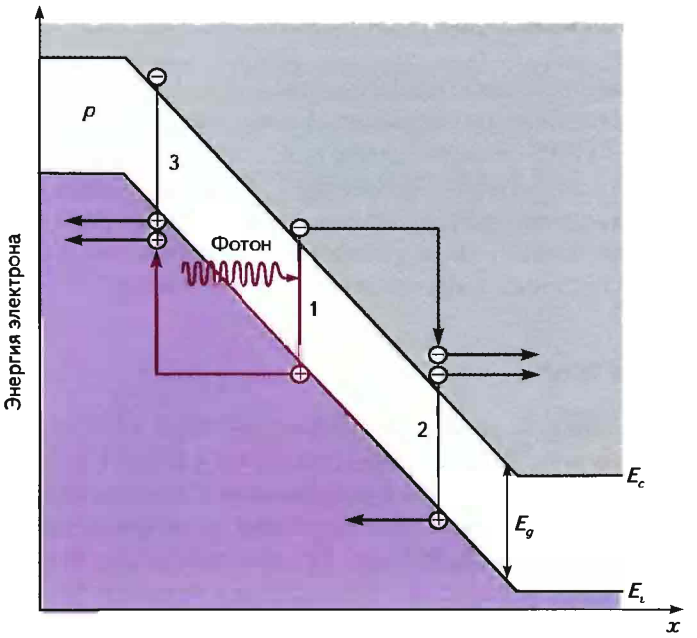


Рис. 18.20. Схематическое изображение процесса умножения носителей в ЛФД

Важный параметр, характеризующий действие ЛФД, — **отношение ионизирующих способностей**, определяемое как отношение коэффициентов ионизации

$$k = \frac{\alpha_h}{\alpha_e}. \tag{18.20}$$

Когда дырки не производят заметной ионизации, т. е. когда $\alpha_h \ll \alpha_e$ и $k \ll 1$, основная ионизация производится электронами. Процесс образования лавины на рис. 18.20 тогда идет в основном слева направо (т. е. от p- к n-области прибора). Он прекращается некоторое время спустя, когда все электроны достигают границы обедненного слоя с n-областью. С другой стороны, если вклад электронов и дырок в ионизацию примерно одинаков ($k \approx 1$), то дырки, движущиеся влево, создают электроны, которые движутся вправо, в свою очередь, генерируя дырки, движущиеся влево, и так далее. Получается беско-

нечный замк-нутый процесс. Хотя этот процесс обратной связи увеличивает усиление прибора (полный заряд, генерируемый в цепи на одну пару фотоносителей, q/e), он, тем не менее, нежелателен по ряду причин.

- Он занимает время и поэтому сужает полосу частот прибора.
- Он хаотичен и поэтому повышает уровень шумов.
- Он может быть неустойчивым и приводить к лавинному пробою.

Следовательно, желательно изготавливать ЛФД из материалов, допускающих ударную ионизацию только одним типом носителей заряда (либо электронами, либо дырками). Если, например, электроны имеют более высокий коэффициент ионизации, то оптимальное поведение достигается при инжекции электрона рожденной фотоном пары вблизи p-границы обедненного слоя и использовании материала с минимально возможным значением k. Если инжектируются дырки, то дырка из рожденной фотоном пары должна инжектироваться вблизи n-границы обедненного слоя, а k — иметь максимальное возможное значение. Идеальные условия умножения числа носителей одного типа соответствуют $k = 0$ или ∞.

Конструкция

Как и у любых фотодиодов, геометрия ЛФД должна обеспечивать максимальное поглощение фотонов, например, принимая форму p—i—n-структуры. С другой стороны, область размножения носителей должна быть тонкой, чтобы минимизировать вероятность возникновения локализованных неконтролируемых лавин (нестабильностей или микроплазменных сгустков), порождаемых сильным электрическим полем. Большая однородность электрического поля может быть достигнута в тонком слое.

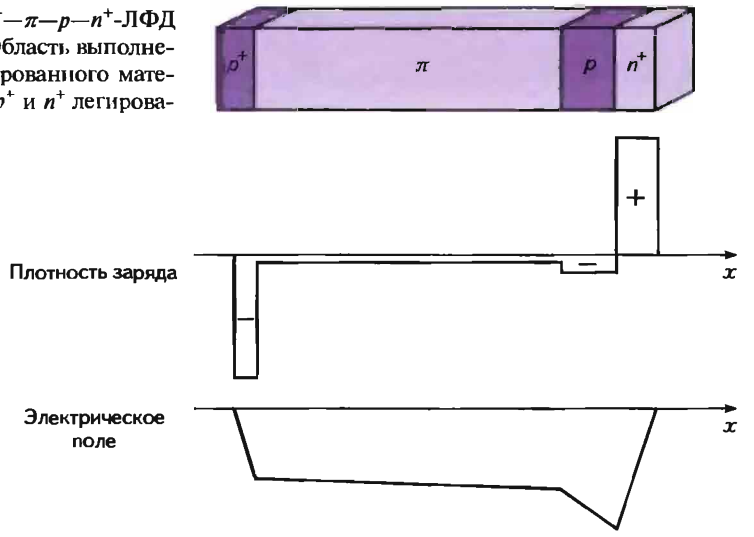


Рис. 18.21. Структура p^+—π—p—n^+-ЛФД с проникновением. π-Область выполнена из очень слабо легированного материала p-типа. Области p^+ и n^+ легированы сильно

Два эти конфликтующих требования стимулируют создание таких конструкций ЛФД, у которых области поглощения и размножения разделены. Структуры такого типа известны как **ЛФД с разделением поглощения и умножения**. Их действие наиболее просто понять на примере материала с $k \approx 0$, например, Si. Фотоны поглощаются в обширной собственной или слабо легированной области. Фотоэлектроны дрейфуют через эту область под действием умеренного электрического поля, а затем входят в тонкий умножающий слой с сильным электрическим полем, где происходит лавинообразование. Структура **ЛФД с проникновением** (проколом), изображенная на рис. 18.21, реализует этот принцип. Поглощение фотонов происходит в широкой области π. Электроны дрейфуют через эту область в тонкий p—n^+-переход, где испытывают действие достаточно сильного электрического поля, чтобы вызвать лавинообразование. Напряжение обратного смещения, приложенное к прибору, достаточно велико для того, чтобы обедненный слой охватывал p- и π-области и проникал в контактный p^+-слой.

18.4.2. Усиление и токовая чувствительность

В качестве прелюдии к определению усиления ЛФД с обоими типами носителей, вызывающими ионизацию, рассмотрим сначала более простую задачу размножения носителей одного типа (электронов), т. е. $\alpha_h = 0$, $k = 0$. Пусть J — плотность электрического тока, переносимого электронами, в точке x, как показано на рис. 18.22. В пределах расстояния dx в среднем ток увеличивается на величину

$$dJ_e(x) = \alpha_e J_e(x)\,dx, \tag{18.21}$$

откуда получаем дифференциальное уравнение

$$\frac{dJ_e}{dx} = \alpha_e J_e(x), \tag{18.22}$$

решение которого — экспоненциальная функция $J_e(x) = J_e(0)\exp(\alpha_e x)$. Тогда усиление $G = J_e(w)/J_e(0)$ равно

$$G = \exp(\alpha_e w). \tag{18.23}$$

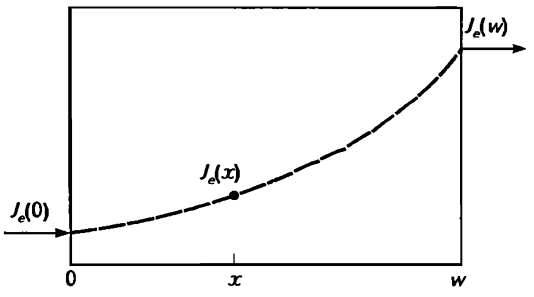


Рис. 18.22. Экспоненциальный рост плотности электрического тока в ЛФД с одним типом носителей

Плотность электрического тока возрастает экспоненциально с ростом произведения коэффициента ионизации α_e и толщины умножающего слоя w. Результат аналогичен усилению в лазере [см. (14.7)].

Проблема умножения числа носителей обоих типов требует знания плотностей тока электронов $J_e(x)$ и дырок $J_h(x)$. Предполагается, что в умножающий слой инжектируются только электроны. Однако поскольку ионизация дырками тоже производит электроны, рост $J_e(x)$ определяется дифференциальным уравнением

$$\frac{dJ_e}{dx} = \alpha_e J_e(x) + \alpha_h J_h(x). \tag{18.24}$$

Вследствие нейтральности среды $dJ_e/dx = -dJ_h/dx$, так что сумма $J_e(x) + J_h(x)$ должна оставаться постоянной для всех x в стационарном режиме. Это ясно из рис. 18.23: полное число носителей заряда, пересекающих любую плоскость, одинаково, независимо от положения.

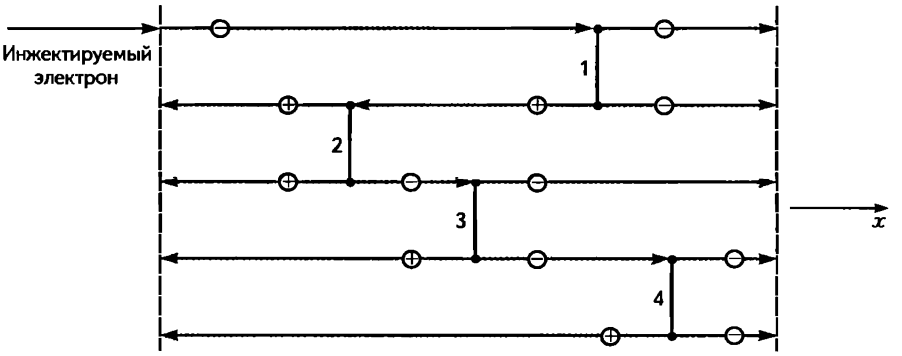


Рис. 18.23. Постоянство суммы плотностей электронного и дырочного токов в поперечной плоскости при любом x. Для иллюстрации показаны четыре акта ударной ионизации и пять пересечений каждой плоскости электронами и дырками

Поскольку предполагается, что дырки не инжектируются при $x = w$, $J_h(w) = 0$, так что

$$J_e(x) + J_h(x) = J_e(w), \tag{18.25}$$

как показано на рис. 18.24. Таким образом, плотность дырочного тока $J_h(x)$ можно исключить из (18.24) и получить

$$\frac{dJ_e}{dx} = (\alpha_e - \alpha_h) J_e(x) + \alpha_h J_e(w). \tag{18.26}$$

Это дифференциальное уравнение первого порядка легко решается относительно усиления

$$G = \frac{J_e(w)}{J_e(0)}.$$

Для $\alpha_e \neq \alpha_h$ получается

$$G = \frac{\alpha_e - \alpha_h}{\alpha_e \exp\left[-(\alpha_e - \alpha_h)w\right] - \alpha_h},$$

откуда получаем

$$\boxed{G = \frac{1 - k}{\exp\left[-(1-k)\alpha_e w\right] - k}}. \qquad (18.27)$$

Усиление ЛФД

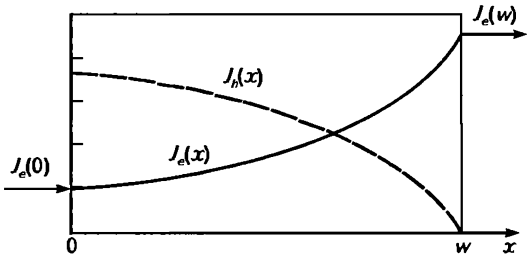


Рис. 18.24. Рост электронного и дырочного токов в результате лавинного размножения носителей

Результат (18.23) для усиления при умножении числа однотипных носителей с его экспоненциальным ростом воспроизводится при $k = 0$. Когда $k = \infty$ усиление остается равным единице, так как инжектируются только электроны, а они не размножаются в данной среде. При $k = 1$ соотношение (18.27) приводит к неопределенности и усиление надо получать непосредственно из (18.26). Тогда получается результат

$$G = \frac{1}{1 - \alpha_e w}.$$

При $\alpha_e w = 1$ достигается неустойчивость. Зависимость усиления от $\alpha_e w$ для нескольких значений отношения ионизирующих способностей k представлена на рис. 18.25. Токовая токовая чувствительность $\mathfrak{R}$ получается с использованием (18.27) в общем соотношении (18.8).

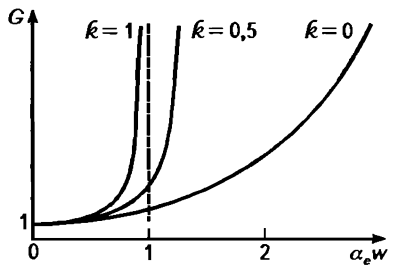


Рис. 18.25. Рост усиления G с увеличением ширины умножающего слоя для нескольких значений отношения ионизирующих способностей k в предположении чисто электронной инжекции

Представляющие интерес материалы тесно связаны с материалами, используемыми в p—i—n-фотодиодах (см. рис. 18.19) с дополнительным требованием, чтобы они имели наименьшее (для электронной инжекции) или наи-

большее (для дырочной инжекции) отношение ионизирующих способностей k. Кремниевые ЛФД имеют $k \approx 0{,}1$—$0{,}2$, однако можно изготовить и кремниевые приборы, у которых k достигает 0,006, что обеспечивает превосходные рабочие характеристики на длинах волн 700—900 нм. ЛФД на InGaAs часто применяются на телекоммуникационных длинах волн (1300—1600 нм), несмотря на более высокие значения k, так как они обеспечивают высокую токовую чувствительность (см. рис. 18.19) и умеренный уровень шума. Электрические поля $\approx 10^5$ В/см, соответствующие напряжению на приборе в десятки вольт, инициируют лавинный механизм. По мере роста напряжения обратного смещения растет усиление и темновой ток, как видно из рис. 18.26.

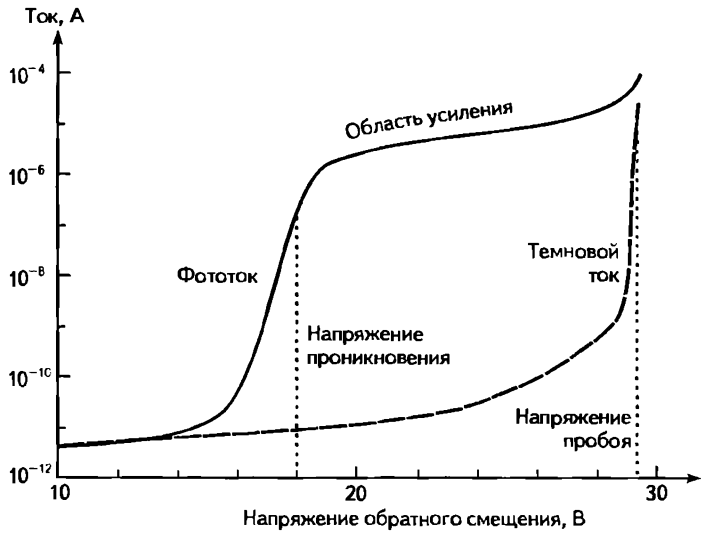


Рис. 18.26. Вольт-амперная характеристика InGaAs ЛФД с раздельным поглощением и умножением. Прибор работает при напряжениях обратного смещения, лежащих между напряжением, обеспечивающим режим проникновения, и напряжением пробоя

Пример 18.1

Усиление в ЛФД на InGaAs

По причине наличия внутреннего усиления лавинные фотодиоды на InGaAs широко применяются в качестве фотоприемников для систем оптоволоконной связи, работающих на телекоммуникационных длинах волн 1300—1600 нм (см. подразд. 14.1.4). Эти приборы обычно работают при напряжениях, лежащих между режимом проникновения и пробоем. Оптимальное усиление $G \approx 10$, а типичный темновой ток $\approx 10^{-11}$ А. Приборы на материалах II—VI групп (например, HgCdTe) и IV—VI групп (например, PbSnTe) находят применение на более длинных волнах.

18.4.3. Время отклика

Кроме обычных дрейфовых, диффузионных и *RC*-эффектов, вносящих вклад в время отклика фотодиодов, у ЛФД имеется дополнительный фактор замедления отклика из-за конечного **времени нарастания лавины**. Формирование времени отклика в ЛФД с двумя типами лавинно размножающихся носителей иллюстрируется рис. 18.27, изображающим историю фотоэлектрона, рожденного вблизи края области поглощения (точка 1). Дрейфуя с установившейся скоростью v_e, электрон достигает области размножения (точка 2) за время w_d/v_e. В этой области электрон также движется со средней скоростью v_e. Посредством ударной ионизации электрон генерирует новые электронно-дырочные пары, скажем, в точках 3 и 4. Дырки дрейфуют в противоположном направлении со своей установившейся скоростью v_h. Дырки также могут индуцировать электронно-дырочные пары посредством ударной ионизации, как, например, в точках 5 и 6. Образовавшиеся носители могут сами вызывать рождение электронно-дырочных пар, таким образом возникает обратная связь. Процесс обрывается, когда последняя дырка, рожденная в области размножения носителей (точка 7), покидает ее и проходит через область дрейфа до точки 8. Полное время τ всего процесса (от точки 1 до точки 8) равно сумме времен дрейфа (от 1 до 2 и от 7 до 8) и времени размножения, обозначаемого τ_m:

$$\tau = \frac{w_d}{v_e} + \frac{w_d}{v_h} + \tau_m. \tag{18.28}$$

В силу случайности процесса лавинного размножения время τ_m является случайным. В частном случае $k = 0$ (дырки не размножаются) максимальное значение τ_m, как легко увидеть из рис. 18.27, равно

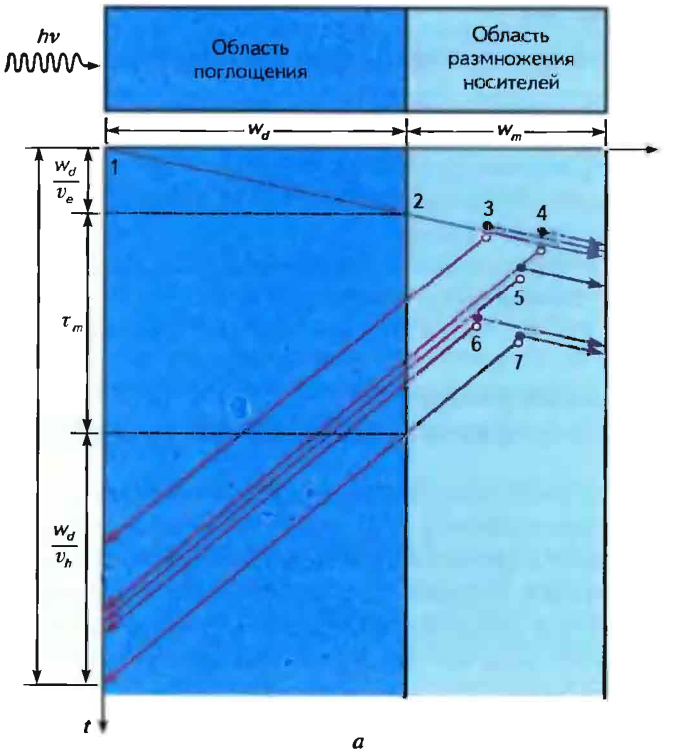
$$\tau_m = \frac{w_m}{v_e} + \frac{w_m}{v_h}. \tag{18.29}$$

При большом усилении для электронной инжекции с $0 < k < 1$ порядок величины среднего значения τ_m получается увеличением первого члена в (18.29) на множитель Gk,

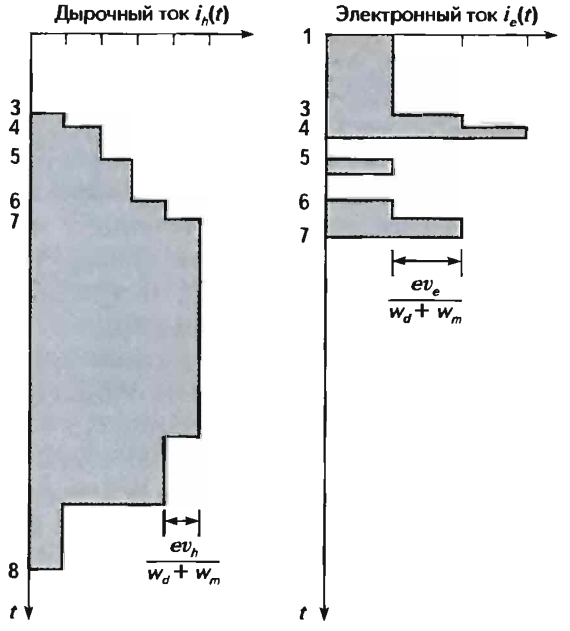
$$\tau_m \approx \frac{Gkw_m}{v_e} + \frac{w_m}{v_h}. \tag{18.30}$$

Более строгая теория весьма сложна.

Рис. 18.27. Отслеживание процесса формирования времени жизни лавины в ЛФД с помощью графика «координата—время» (*а*). Синими линиями представлены электроны, а красными дырки. Электроны движутся вправо со скоростью v_e, а дырки влево со скоростью v_h. В области размножения генерируются вторичные электронно-дырочные пары. Носители перестают двигаться, достигая края материала. Дырочный ток $i_h(t)$ и электронный ток $i_e(t)$, индуцируемые в цепи (*б*). Каждая пара носителей индуцирует заряд e в цепи. Полный индуцированный заряд q, представляющий собой площадь под графиком зависимости суммарного тока $i_h(t) + i_e(t)$ от времени t, равен Ge. Эта схема — обобщение рис. 18.6, где показана история одной электронно-дырочной пары


hν
Область поглощения
Область размножения носителей
w_d
w_m
$\frac{w_d}{v_e}$
τ_m
$\frac{w_d}{v_h}$
1
2
3
4
5
6
7
t

а


Дырочный ток $i_h(t)$
Электронный ток $i_e(t)$
3
4
5
6
7
8
1
$\frac{ev_h}{w_d + w_m}$
$\frac{ev_e}{w_d + w_m}$
t

б

Пример 18.2

Время формирования лавины в кремниевом ЛФД

Рассмотрим кремниевый ЛФД с $w_d = 50$ мкм, $w_m = 0{,}5$ мкм, $v_e = 10^7$см/с, $v_h = 5 \cdot 10^6$ см/с, $G = 100$ и $k = 0{,}1$. Формула (18.29) дает $\tau_m = 5 + 10 = 15$ пс, после чего из (18.28) получаем $\tau = 1020$ пс $= 1{,}02$ нс. С другой стороны, (18.30) дает $\tau_m = 60$ пс, так что из (18.28) получается $\tau = 1065$ пс $= 1{,}07$ нс. Для p—i—n-фотодиода с такими же значениями w_d, v_e и v_h время переходного процесса $w_d/v_e + w_d/v_h \approx 1$ нс. Эти результаты не сильно отличаются, потому что в кремниевом приборе τ_m весьма мало.

18.4.4. Лавинные диоды для регистрации одиночных фотонов (SPAD)

Способность детектировать и пересчитывать отдельные фотоны важна во многих приложениях (например, получение изображений, лазерная локация искусственных спутников и дальняя космическая лазерная связь). В режиме **счета фотонов** подавляется шум усиления и внешней цепи, так как отклик преобразуется в двоичную форму. Счет фотонов может быть достигнут с помощью **лавинного диода для регистрации одиночных фотонов** (SPAD), известного также как **лавинный фотодиод в режиме счетчика Гейгера.** Этот прибор представляет собой ЛФД, смещенный таким образом, что приход одиночного фотона взывает лавинный пробой, создавая большой импульс тока, означающий приход фотона. Каждый импульс тока нужно оборвать, приготовляя прибор к регистрации следующего фотона. Это достигается либо пассивными, либо активными средствами; последнее сложнее, но обеспечивает существенно более высокую максимальную частоту регистрации фотонов.

Кремниевые SPAD работают в видимой и ближней инфракрасной области спектра ($\lambda_0 = 400$—1000 нм) и обладают высокой эффективностью ($\eta \approx 75\,\%$), низкой темновой скоростью счета (≈75 отсчетов/с) и субнаносекундным временным разрешением (≈100 пс). В полосе волоконно-оптической связи ($\lambda_0 = 1300$—1600 нм) наиболее подходящими являются гетероструктуры InGaAs/InP, однако их рабочие характеристики намного менее впечатляющи, чем у коротковолновых приборов: $\eta \approx 20\,\%$, темновая скорость счета ≈5000 отсчетов/с, временное разрешение ≈500 пс. В этой области иногда также используются Ge и Si—Ge. На еще больших длинах волн ($\lambda_0 > 4$ мкм) использовались приборы с поглощающим слоем InAsSb и умножающим слоем AlGaAsSb на подложке из GaSb. Приборы на GaN и SiC нашли применение в ультрафиолетовом диапазоне. Особенностью SiC является стойкость к высоким температурам и агрессивному окружению. Во всех случаях для SPAD важен компромисс между эффективностью и шириной полосы.

Счет фотонов может быть достигнут также с помощью **сверхпроводящих детекторов одиночных фотонов** (SSPD), которые являются широкополосными,

малошумящими и быстродействующими, хотя нуждаются в охлаждении. Приход фотона вызывает в этих приборах появление локального нарушения сверхпроводимости, что вызывает внешний отклик, сигнализирующий о событии.

18.5. МАТРИЧНЫЕ ДЕТЕКТОРЫ

Отдельный фотоприемник регистрирует поток падающих фотонов как функцию времени. Матрица из большого числа фотоприемников может одновременно регистрировать зависящие от времени потоки фотонов в множестве пространственных точек. **Матричные детекторы**, таким образом, позволяют формировать электронные версии оптических изображений. Один из типов матричных детекторов — микроканальная пластина (см. рис. 18.2, *в*) — уже обсуждался нами ранее.

Технологии современной микроэлектроники позволяют изготавливать много типов матриц. Они содержат большое число фотодетектирующих элементов, называемых **пикселами**, представляющих собой фотопроводники, фотодиоды, лавинные фотодиоды или тепловые приемники, такие как болометры. Двумерная матрица фоточувствительных элементов, предназначенная для записи электронной версии изображения в фокальной плоскости изображающей системы, известна как **матрица видеопреобразователя**. В **гибридной** матрице видеопреобразователя накопление сигнала и его обработка происходит в слое, лежащем непосредственно под матрицей фоточувствительных элементов.

Для передачи сигналов с фотоприемников существует две основные технологии считывающих схем: технология **приборов с зарядовой связью** (ПЗС) и технология **комплементарных металлооксидных полупроводников** (КМОП).

Материалы и структуры

Матричные детекторы бывают многих типов, как видно из следующих примеров.

- Матрицы микроболометров часто применяются в системах термического формирования изображений. Падающие фотоны вызывают повышение температуры освещаемых элементов. Сопутствующее нагреву изменение электрического сопротивления записывается внешней схемой. Такие устройства работают при температуре окружающей среды и стали играть важную роль в последние годы, поскольку удалось достичь резкого повышения их разрешающей способности и чувствительности. Матрицы микроболометров из окиси ванадия имеют сотни тысяч пикселов размером около 25 мкм каждый и чувствительны к излучению в средней части инфракрасного диапазона. Эти приборы широко используются в военных и промышленных приложениях.
- Фотопроводящие матрицы, как правило, используются в среднем инфракрасном диапазоне. Фотон, энергия которого превышает ширину запрещенной зоны полупроводника, такого как InSb или HgCdTe, создает электронно-дырочную пару, вносящую вклад в проводимость материала.

• Матрицы из несобственных полупроводников, таких как Ge : Ga, применяются в видеопреобразователях дальнего ИК-диапазона. Фотон переводит электрон донорной примеси в зону проводимости (или дырку акцепторной примеси в валентную зону), внося вклад в проводимость материала.

• Фотоприемники на квантовых ямах используются в мегапиксельных матрицах видеопреобразователей. Фотон обеспечивает достаточную энергию для выхода электрона из квантовой ямы, что дает вклад в проводимость материала. Изображения в длинноволновой и средневолновой частях ИК-диапазона получаются при помощи элементов на основе GaAs/AlGaAs и GaAs/InGaAs/AlGaAs соответственно. С помощью узкополосных спектральных фильтров можно получить сотни полос.

• Матрицы, изготовленные из *p*—*i*—*n*-диодов на составных полупроводниках, таких как InGaAs и HgCdTe, используются в видимой и инфракрасной областях. Фотон, энергия которого больше ширины запрещенной зоны, рождает электронно-дырочную пару, дающую вклад в ток диода.

• Фотодиодные элементы с барьерами Шоттки на переходах металл—полупроводник применяются в матрицах видеопреобразователей многофункциональных видеокамер. Фотон, энергия которого больше барьера Шоттки, создает электронно-дырочную пару, вносящую вклад в ток диода. Для получения изображений в множестве спектральных областей можно использовать PtSi, поскольку этот материал чувствителен к широкому диапазону длин волн, простирающемуся от ближнего ультрафиолета до примерно 6 мкм в среднем ИК-диапазоне. Несмотря на низкий квантовый выход в инфракрасном диапазоне, PtSi находит широкое применение из-за легкости изготовления и высокой стабильности.

• Лавинные диоды на *p*—*n*-переходах с областью умножения также используются в матричных фотодетекторах. Фотон, энергия которого больше ширины запрещенной зоны, создает электронно-дырочную пару, которая входит в усиливающую область полупроводника, обеспечивающую усиление. Возникающий субнаносекундный электрический импульс имеет амплитуду в несколько вольт, достаточный для прямого переключения цифровой схемы КМОП, что устраняет необходимость аналого-цифрового преобразования.

• В лавинных детекторах одиночных фотонов (SPAD) на обратно-смещенных *p*—*n*-переходах используются области умножения, работающие в режиме счетчиков Гейгера. Фотон, энергия которого превышает ширину запрещенной зоны, создает электронно-дырочную пару, которая, входя в область полупроводника с сильным полем, вызывает лавинный пробой и, следовательно, генерацию большого импульса тока.

• Фоточувствительные матрицы могут также использоваться как гетеродинные детекторы; коэффициент передачи преобразователя обеспечивается гетеродином (см. разд. 24.5).

Схемы считывания данных

Для передачи сигналов с фотодетекторов на дисплей камеры или выход системы используется два основных вида схем считывания данных: прибо-

ры с зарядовой связью (ПЗС) и комплементарные структуры металл—оксид—полупроводник (КМОП-структуры).

1. *ПЗС-технология.* В ПЗС заряд, создаваемый отдельным элементом детектора, передается в скрытый ПЗС-канал в заданный момент времени. Затем через этот канал заряд последовательно передается от одного положения детектора к другому, пока не достигнет угла микросхемы, где считывается. Множество электродных структур и синхронизирующих элементов разработано для периодического считывания заряда, накопленного каждым элементом, и генерации потока данных, представляющих изображение.

2. *КМОП-технология.* Технология комплементарных структур металл—оксид—полупроводник (КМОП) широко применяется при изготовлении электронных приборов и интегральных схем. Благодаря низкому энергопотреблению и сравнительной дешевизне, эта технология стимулировала массовое производство матричных видеопреобразователей. Каждый элемент детектора соединен с несколькими полевыми МОП-транзисторами, которые усиливают и считывают сигнал детектора. В отличие от последовательного считывания в ПЗС, считывание с каждого элемента матрицы детекторов производится индивидуально.

18.6. ШУМ В ФОТОДЕТЕКТОРАХ

Фотодетектор обладает чувствительностью к потоку фотонов (или оптической мощности). В соответствии с (18.4) поток фотонов Ф (оптическая мощность $P = h\nu\Phi$) порождает пропорциональный электрический ток $i_p = \eta e\Phi = \Re P$. Однако в действительности электрический ток, генерируемый в приборе, является случайной величиной i, колеблющейся вниз и вверх относительно среднего значения $\bar{i} \equiv i_p = \eta e\Phi = \Re P$. Флуктуации тока i, рассматриваемые как шум, характеризуются стандартным отклонением тока σ_i, где $\sigma_i^2 = \langle (i - \bar{i})^2 \rangle$. Для тока с нулевым средним значением ($\bar{i} = 0$) стандартное отклонение сводится к среднеквадратичному значению тока $\sigma_i^2 = \langle (i)^2 \rangle^{1/2}$.

В процессе детектирования фотонов задействовано множество источников шума:

- *Фотонный шум.* Наиболее фундаментальный источник шума связан со случайным приходом отдельных фотонов на приемник, который обычно описывается статистикой Пуассона, как обсуждалось в разд. 12.2.
- *Фотоэлектронный шум.* В детекторе фотонов с квантовым выходом $\eta < 1$ каждый фотон генерирует электронно-дырочную пару с вероятностью η и не генерирует с вероятностью $1 - \eta$. Поскольку этот процесс случайный, он является источником шума.
- *Шум усиления.* Процесс усиления, обеспечивающий внутреннее усиление в некоторых фотодетекторах, таких как фотопроводники и ЛФД, является стохастическим. Каждый детектируемый фотон генерирует случайное число носителей G со средним значением $\bar{G}$. Флуктуации усиления зависят от природы его механизма.
- *Шум приемной схемы.* Различные компоненты электрической схемы оптического приемника, такие как резисторы и транзисторы, вносят вклад в шум приемной схемы.

Эти четыре источника шума схематически иллюстрирует рис. 18.28. Средний сигнал, попадающий на детектор (входной оптический сигнал), обладает собственным фотонным шумом. Фотоэффект преобразует фотоны в фотоэлектроны. В ходе этого процесса средний сигнал уменьшается на множитель η (квантовую эффективность). Соответствующий фотоэлектронный шум также уменьшается, но в меньшее число раз, чем сигнал; таким образом, отношение сигнал—шум у фотоэлектронного сигнала ниже, чем у входного фотонного сигнала. Шумы схемы вносятся в детектируемый сигнал. Если присутствует механизм усиления, он усиливает и сигнал, и шум. Кроме того, он вносит собственный шум усиления. В конце концов шум схемы вносится в точке токосъема.

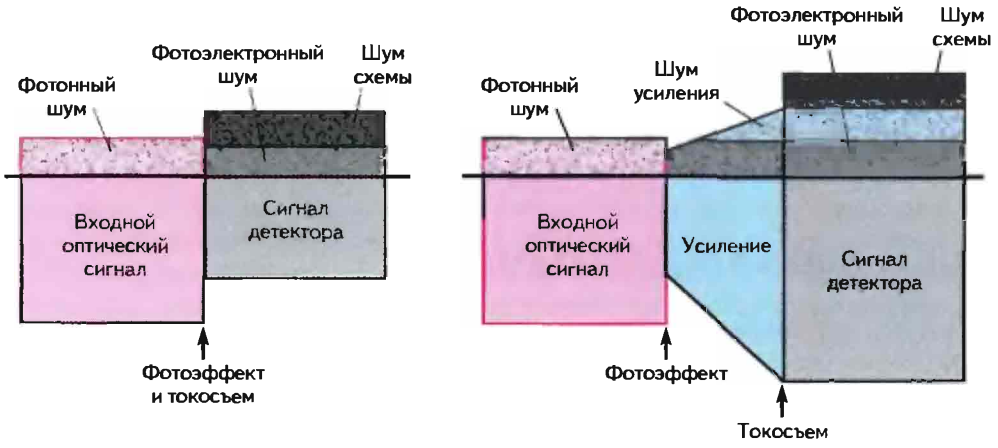


Рис. 18.28. Входной сигнал и сигнал детектора вместе с различными источниками шума для фотодетектора без усиления, такого как p—i—n-фотодиод (*а*), и фотодетектора с усилением, такого как лавинный фотодиод (*б*)

Как составная часть системы передачи информации оптический приемник можно характеризовать следующими рабочими параметрами:

- **Отношение сигнал—шум** (ОСШ) случайной величины определяется как отношение квадрата среднего значения к дисперсии. Таким образом, ОСШ для тока равно $\bar{i}^2/\sigma_i^2$, а для числа фотонов $\bar{n}^2/\sigma_n^2$.
- **Минимальный обнаружимый сигнал** определяется как средний сигнал, при котором ОСШ равно единице.
- **Коэффициент избыточного шума** F случайной величины определяется как отношение среднего квадрата к квадрату среднего. Так, коэффициент избыточного шума для усиления G фотодетектора равен $F = \langle G^2\rangle/\langle G\rangle^2$.
- **Частота появления ошибочных битов** определяется как вероятность ошибки на каждый бит у цифрового оптического приемника.
- **Обнаружительная способность приемника** определяется как сигнал, соответствующий заданной величине отношения сигнал—шум, ОСШ = ОСШ_0.

В то время как минимальный обнаружимый сигнал соответствует обнаружительной способности приемника, обеспечивающей $ОСШ_0 = 1$, для гарантии заданного уровня точности часто используются более высокие уровни $ОСШ_0$ (например, $ОСШ_0 = 10—10^3$, что соответствует 10—30 дБ). Для цифровой системы обнаружительная способность приемника определяется как минимальная оптическая энергия (или соответствующее среднее число фотонов) на один бит, которая требуется для достижения заданной частоты появления ошибочных битов, которая часто устанавливается равной 10^{-9}

Мы начнем с вывода выражений для отношения сигнал—шум у фотодетектора с указанными четырьмя источниками шума. К другим источникам шума, которые мы не будем явно рассматривать, относятся фоновый шум и шум темнового тока. **Фоновый шум** — это фотонный шум, связанный с попаданием на приемник света от посторонних источников, не являющихся предметом изучения, например, света Солнца и звезд. Фоновый шум особенно вреден в системах детектирования, работающих в среднем и дальнем инфракрасном диапазоне, из-за обильного теплового излучения, испускаемого на этих длинах волн окружающими предметами при комнатной температуре (см. рис. 13.31). Фотодетекторы генерируют также **шум темнового тока**, который, как следует из его названия, имеется даже в отсутствие падающего света. Шум темнового тока является результатом случайного теплового рождения электронно-дырочных пар за счет туннелирования. Игнорируются также ток утечки и $1/f$-шум.

18.6.1. Фотоэлектронный шум

Фотонный шум

Как описано в разд. 12.2, поток фотонов, ассоциированный с фиксированной оптической мощностью P, по своей природе является неопределенным. Средний поток фотонов равен $\Phi = P/h\nu$, однако эта величина флуктуирует случайным образом в соответствии с вероятностным законом, зависящим от природы источника света. Таким образом, число фотонов n, зарегистрированных за промежуток времени T, является случайной величиной со средним значением $\bar{n} = \Phi T$. У света от идеального лазера или теплового источника со спектральной шириной, намного превышающей $1/T$, число фотонов подчиняется распределению Пуассона, для которого $\sigma_n^2 = \bar{n}$. Таким образом, флуктуации, связанные со средним числом 100 фотонов, приводят к тому, что действительное число фотонов лежит приблизительно в интервале 100 ± 10.

Отношение сигнал—шум для числа фотонов $ОСШ = \bar{n}^2/\sigma_n^2$ тогда равно

$$\boxed{ОСШ = \bar{n},} \tag{18.31}$$

Отношение сигнал—шум для числа фотонов

а минимальное обнаружимое число фотонов равно $\bar{n} = 1$. Если время наблюдения $T = 1$ мкс, а длина волны $\lambda_0 = 1{,}24$ мкм, то это эквивалентно мини-

мальной обнаружимой мощности 0,16 пВт. Обнаружительная способность приемника при $ОСШ_0 = 10^3$ (30 дБ) составляет 1000 фотонов. Если интервал времени $T = 10$ нс, это эквивалентно обнаружительной способности 10^{11} фотонов/с или, по мощности, 16 пВт при $\lambda_0 = 1,24$ мкм.

Фотоэлектронный шум

Фотон, падающий на фотодетектор с квантовым выходом η, генерирует фотособытие (т. е. создаст электронно-дырочную пару или высвобождает электрон) с вероятностью η и не генерирует с вероятностью $1 - \eta$. Предполагается, что фотособытия выбираются случайным образом из потока фотонов. Средний падающий поток фотонов Ф (число фотонов в секунду) порождает средний поток фотоэлектронов $\eta\Phi$ (число фотоэлектронов в секунду). Число фотоэлектронов m, зарегистрированных за время T, является случайной величиной со средним значением

$$\bar{m} = \eta\bar{n}, \tag{18.32}$$

где $\bar{n} = \Phi T$ — среднее число падающих фотонов за один и тот же интервал времени T. Если число фотонов имеет распределение Пуассона, то таким же будет распределение числа фотоэлектронов. Это можно подтвердить рассуждением, параллельным проведенному в подразд. 12.2.4. Следовательно, дисперсия числа фотоэлектронов равна $\bar{m}$, так что

$$\sigma_m^2 = \bar{m} = \eta\bar{n}. \tag{18.33}$$

Ясно, что фотоэлектронный шум не аддитивен по отношению к фотонному шуму.

Таким образом, изначальным фундаментальным источником шума, с которым необходимо бороться при использовании света для передачи сигнала, является случайность числа фотонов, благодаря которой отношение сигнал—шум для числа фотоэлектронов составляет

$$\boxed{ОСШ = \bar{m} = \eta\bar{n}.} \tag{18.34}$$

Отношение сигнал—шум для числа фотоэлектронов

Минимальное обнаружимое число фотоэлектронов оставляет $\bar{m} = \eta\bar{n} = 1$ фотоэлектрон, приходящийся на $1/\eta$ фотонов. Обнаружительная способность приемника для $ОСШ_0 = 10^3$ составляет 1000 фотоэлектронов, или $1000/\eta$ фотонов.

Шум фототока

Исследуем теперь свойства электрического тока $i(t)$, индуцированного в цепи случайным потоком фотонов со средним значением $\eta\Phi$. В наше рассмотрение включаются эффекты фотонного шума, фотоэлектронного шума и характерный временной отклик детектора и цепи (фильтрация). Каждая электронно-дырочная пара генерирует импульс электрического тока с зарядом (площадью под кривой) e и длительностью τ_p о внешней цепи фотоде-

тектора (рис. 18.29). Таким образом, поток фотонов, падающих на фотодетектор, порождает поток импульсов тока, которые, складываясь друг с другом, образуют фототок $i(t)$. Случайный поток фотонов преобразуется в флуктуирующий электрический ток. Если падающие фотоны распределены по Пуассону, то эти флуктуации называются **дробовым шумом**. В более общем случае для детектора с усилением G генерируемый в каждом импульсе заряд равен $q = Ge$.

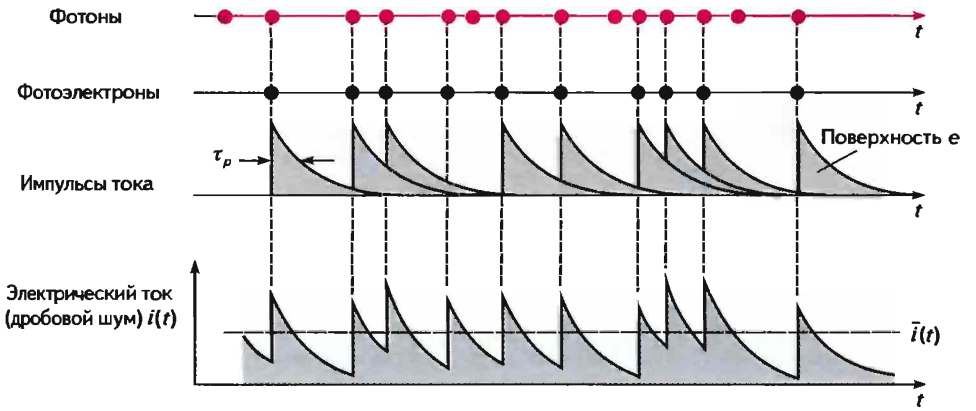


Рис. 18.29. Фототок, индуцируемый в цепи фотодетектора, представляет собой суперпозицию импульсов тока, каждый из которых связан с регистрацией фотона. На данном рисунке отдельные импульсы представляют собой ступенчатые функции с экспоненциальным затуханием, однако они могут иметь произвольную форму (см., например, рис. 18.6, *б* и 18.7)

Прежде чем приступить к аналитическому описанию свойств фототока $i(t)$, рассмотрим задачу в упрощенной постановке. Рассмотрим поток фотонов Ф, падающий на фотоэлектрический приемник с квантовым выходом η. Пусть случайное число m фотоэлектронов, сосчитанных на протяжении характеристического временного интервала $T = 1/2B$ (время разрешения цепи), генерирует фототок $i(t)$, где t — момент времени, непосредственно следующий за интервалом T. Для прямоугольных импульсов тока длительности T случайные величины тока и числа фотоэлектронов связаны формулой $i = (e/T)m$. Следовательно, среднее значение и дисперсия фототока выражаются как

$$\bar{i} = \frac{e}{T}\bar{m}; \tag{18.35}$$

$$\sigma_i^2 = \left(\frac{e}{T}\right)^2 \sigma_m^2, \tag{18.36}$$

где $\bar{m}$ — среднее число фотоэлектронов, собранных за время $T = 1/2B$,

$$\bar{m} = \eta\Phi T = \frac{\eta\Phi}{2B}.$$

Подставляя $\sigma_m^2 = \bar{m}$ для распределения Пуассона, получаем для среднего значения и дисперсии фототока выражения

$$\boxed{\bar{i} = e\eta\Phi;} \qquad (18.37)$$

Среднее значение фототока

$$\boxed{\sigma_i^2 = 2e\bar{i}B.} \qquad (18.38)$$

Дисперсия фототока

Отсюда следует, что отношение сигнал—шум для фототока, ОСШ $= \bar{i}^2/\sigma_i^2$ выражается как

$$\boxed{\text{ОСШ} = \frac{\bar{i}}{2eB} = \frac{\eta\Phi}{2B} = \bar{m}.} \qquad (18.39)$$

Отношение сигнал—шум для фототока

Отношение сигнал—шум для тока прямо пропорционально потоку фотонов Ф и обратно пропорционально ширине полосы электрической цепи B. Как и ожидалось, результат идентичен отношению сигнал—шум для числа фотоэлектронов $\bar{m}$, поскольку цепь не вносит дополнительной случайности.

Пример 18.3

ОСШ и обнаружительная способность приемника

Для $\bar{i} = 10$ нА и $B = 100$ МГц $\sigma_I \approx 0{,}57$ нА, что соответствует отношению сигнал—шум ОСШ $= 310$, или 25 дБ. Среднее число 310 фотоэлектронов детектируется за каждый интервал времени $T = 1/2B = 5$ нс. Минимальный обнаружимый поток фотонов есть $\Phi = 2B/\eta_0$, а обнаружительная способность приемника при ОСШ $= 10^3$ равна $\Phi = 1000 \cdot (2B/\eta) = 2 \ 10^{11}/\eta$ фотонов/с.

Вывод среднего значения и дисперсии фототока

Перейдем к доказательству (18.37) и (18.38) в общем случае. Предположим, что фотособытие, генерируемое в момент $t = 0$, производит электрический импульс $h(t)$ с площадью e во внешней цепи. Тогда фотособытие, генерируемое в момент времени t_1, производит сдвинутый импульс $h(t - t_1)$. Разделим временную ось на бесконечно малые интервалы Δt, так что вероятность p того, что фотособытие произойдет в течение такого интервала, равна $p = \eta\Phi\Delta t$. Электрический ток i в момент времени t записывается как

$$i(t) = \sum_l X_l h(t - l\Delta t), \qquad (18.40)$$

где X_l принимает значение 1 с вероятностью p и 0 с вероятностью $1 - p$. Переменные $\{X_l\}$ независимы.

Среднее значение X_l равно $0 \cdot (1 - p) + 1 \quad p = p$. Среднее от квадрата этой величины равно

$$\langle X_l^2 \rangle = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

Среднее значение произведения $X_l X_k$ равно p^2 при $l \neq k$ и p при $l = k$.

Теперь среднее и среднеквадратичное значения $i(t)$ можно найти с помощью выражений

$$\bar{i} = \langle i \rangle = \sum_l ph(t - l\Delta t); \tag{18.41}$$

$$\begin{aligned} \langle i^2 \rangle &= \sum_l \sum_k \langle X_l X_k \rangle h(t - l\Delta t) h(t - k\Delta t) = \\ &= \sum_{l \neq k} p^2 h(t - l\Delta t) h(t - k\Delta t) + \sum_l ph^2 (t - l\Delta t). \end{aligned} \tag{18.42}$$

В результате подстановки $p = \eta\Phi\Delta t$ и перехода к пределу при $\Delta t \to 0$, когда все суммы становятся интегралами, формулы (18.41) и (18.42) дают соответственно:

$$\bar{i} = \eta\Phi \int_0^\infty h(t)\,\mathrm{d}t = e\eta\Phi; \tag{18.43}$$

$$\langle i^2 \rangle = (e\eta\Phi)^2 + \eta\Phi \int_0^\infty h^2(t)\,\mathrm{d}t. \tag{18.44}$$

Следовательно,

$$\sigma_i^2 = \langle i^2 \rangle - \langle i \rangle^2 = \eta\Phi \int_0^\infty h^2(t)\,\mathrm{d}t. \tag{18.45}$$

Вводя определение

$$B = \frac{1}{2e^2} \int_0^\infty h^2(t)\,\mathrm{d}t = \frac{\int_0^\infty h^2(t)\,\mathrm{d}t}{\left[\int_0^\infty h^2(t)\,\mathrm{d}t\right]^2}, \tag{18.46}$$

мы, наконец, получаем (18.37) и (18.38).

Параметр B, определенный формулой (18.46), представляет собой ширину полосы устройства/цепи. Это легко проверяется, если заметить, что Фурье-образ функции $h(t)$ есть функция передачи $H(\nu)$. Площадь под $h(t)$ есть просто

$H(0) = e$. По теореме Парсеваля [см. (А.7)] площадь под $h^2(t)$ равна площади под симметричной функцией $|H(\nu)|^2$, так что

$$B = \int_0^\infty \left|\frac{H(\nu)}{H(0)}\right|^2 d\nu. \tag{18.47}$$

Таким образом, величина B представляет собой ширину спектра мощности для функции $|H(\nu)|$ (т. е. ширину полосы пропускания прибора/цепи) в соответствии с (А.17). Например, если $H(\nu) = 1$ при $-\nu_c < \nu < \nu_c$ и 0 в остальных случаях, то (18.47) дает $B = \nu_c$.

Эти соотношения применимы ко всем фотоэлектрическим детектирующим приборам без усиления (т. е. фотоэлементам и фотодиодам на переходах). Использование формул требует знания ширины полосы прибора, цепи смещения и усилителя; B определяется путем подстановки передаточной функции всей системы в (18.47).

18.6.2. Шум усиления

Среднее значение и дисперсия фототока в приборе с фиксированным (детерминированным) усилением G определяется с помощью замены e на $q = Ge$ в (18.37) и (18.38), что приводит к выражениям

$$\bar{i} = eG\eta\Phi = \frac{eG\eta P}{h\nu}; \tag{18.48}$$

$$\sigma_i^2 = 2eG\bar{i}B = 2e^2G^2\eta B\Phi. \tag{18.49}$$

Отношение сигнал—шум, в соответствии с (18.39), становится равным

$$\text{ОСШ} = \frac{\bar{i}}{2eGB} = \frac{\eta\Phi}{2B} = \bar{m}. \tag{18.50}$$

Оно не зависит от G, потому что детерминированное усиление не вносит дополнительной случайности; среднее значение тока $\bar{i}$ и его среднеквадратичное значение σ_i содержат один и тот же множитель G.

Этот простой результат неприменим, когда усиление само является случайным, как в случае фотоумножителей, фотопроводников и лавинных фотодиодов. Вывод выражения для среднего значения и дисперсии фототока, проведенный в предыдущем разделе, нуждается в модификации. В частности, электрический ток (18.40) теперь нужно записывать как

$$i(t) = \sum_l X_l G_l h(t - l\Delta t), \tag{18.51}$$

где, как и раньше, X_l принимает значение 1 с вероятностью $p = \eta\Phi\Delta t$ и 0 с вероятностью $1 - p$. Теперь в расчет включены случайные числа G_l, описы-

вающие усиление, сообщаемое фотоносителю, зарегистрированному в течение l-го временного промежутка, как показано на рис. 18.30.

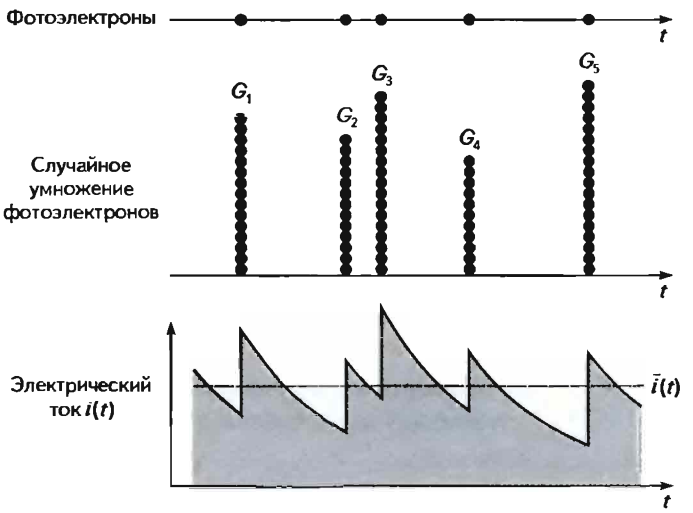


Рис. 18.30. В фотодетекторе с усилением каждое фотособытие генерирует случайное число G_l носителей, что порождает импульс электрического тока площадью eG_l. Полный электрический ток в цепи детектора $i(t)$ — суперпозиция этих импульсов

Если случайная переменная G_l имеет среднее значение $\langle G\rangle = \bar{G}$ и средний квадрат $\langle G^2\rangle$, то анализ, аналогичный проведенному в (18.40)—(18.47), даст

$$\boxed{\bar{i} = e\bar{G}\eta\Phi;} \qquad (18.52)$$

Среднее значение фототока (случайное усиление)

$$\boxed{\sigma_i^2 = 2e\bar{G}\bar{i}BF,} \qquad (18.53)$$

Дисперсия фототока (случайное усиление)

где **коэффициент избыточного шума** F определяется как

$$\boxed{F = \frac{\langle G^2\rangle}{\langle G\rangle^2},} \qquad (18.54)$$

Коэффициент избыточного шума

Коэффициент избыточного шума связан с дисперсией усиления σ_G^2 соотношением

$$F = 1 + \frac{\sigma_G^2}{\langle G\rangle^2}.$$

В частном случае детерминированного усиления $\sigma_G^2 = 0$ и $F = 1$, в результате чего (18.53) сводится к (18.49). Когда усиление случайно, $\sigma_G^2 > 0$ и $F > 1$; обе эти

величины растут с увеличением флуктуаций усиления. Результирующий электрический ток i тогда демонстрирует флуктуации бо́льшие, чем дробовой шум.

При наличии случайного усиления отношение сигнал—шум для тока $\bar{i}^2/\sigma_i^2$ принимает вид

$$\text{ОСШ} = \frac{\bar{i}}{2e\bar{G}BF} = \frac{\eta\Phi/2B}{F} = \frac{\bar{m}}{F}, \tag{18.55}$$

Отношение сигнал—шум (случайное усиление)

где $\bar{m}$ — среднее число фотоэлектронов, собранных за время $T = 1/2B$. Это меньше, чем ОСШ при детерминированном усилении на множитель F; уменьшение связано со случайностью усиления.

Пример 18.4

Коэффициент избыточного шума трубки фотоэлектронного умножителя (ФЭУ)

Трубка ФЭУ работает на основе умножения электронов посредством вторичной электронной эмиссии на ее динодах. Для типичного прибора случайность усиления, связанная с этим процессом, дает коэффициент избыточного шума, равный $F \approx 1{,}2$. Поскольку

$$F = 1 + \frac{\sigma_G^2}{\langle G \rangle^2},$$

ОСШ для усиления составляет $1/(F - 1) \approx 5$. Если ФЭУ имеет среднее значение усиления $\bar{G} = 10^8$, то стандартное отклонение для флуктуаций усиления есть $\sigma_G = 10^8/\sqrt{5}$.

Коэффициент избыточного шума для ЛФД

Когда фотоэлектроны инжектируются на краю однородной области усиления обычного ЛФД, усиление прибора дается формулой (18.27). Оно зависит от коэффициента ионизации для электрона α_e и ионизационного отношения $k = \alpha_h/\alpha_e$, а также от ширины области усиления w. С помощью аналогичного (но более сложного) анализа, учитывающего случайность процесса усиления, можно получить выражение для среднеквадратичного усиления $\langle G^2 \rangle$ и, следовательно, для коэффициента избыточного шума F в (18.54). Этот, более общий, вывод приводит к выражению для среднего усиления $\bar{G}$, которое идентично (18.27).

Расчеты показывают, что коэффициент избыточного шума F можно связать со средним усилением и ионизационным отношением:

$$F = k\bar{G} + (1-k)\left(2 - \frac{1}{\bar{G}}\right). \tag{18.56}$$

Коэффициент избыточного шума (обычный ЛФД)

Графики, соответствующие этой формуле, показаны на рис. 18.31 с $\bar{k}$ в качестве параметра.

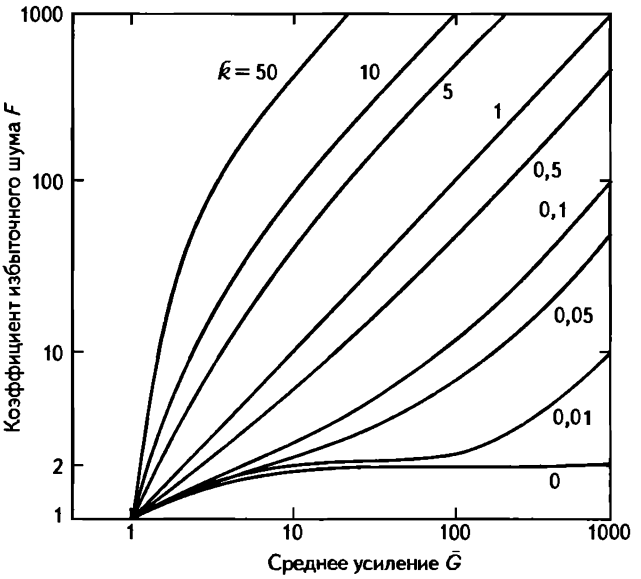


Рис. 18.31. Коэффициент избыточного шума F для обычного ЛФД с однородной областью умножения при инжекции электронов как функция среднего усиления $\bar{G}$ при различных значениях ионизационного отношения $\bar{k}$. При инжекции дырок $\bar{k}$ заменяется на $1/\bar{k}$

Уравнение (18.56) справедливо, когда электроны инжектируются с края области умножения, однако как электроны, так и дырки имеют способность инициировать ударную ионизацию. Если инжектируются только дырки, можно пользоваться теми же выражениями, заменив в них $\bar{k}$ на $1/\bar{k}$. Шум усиления минимизируется путем инжекции носителей с более высоким коэффициентом ионизации, а также путем изготовления структуры, в которой $\bar{k}$ минимально, если инжектируются электроны, и максимально, если инжектируются дырки. Таким образом, коэффициенты ионизации для двух типов носителей должны различаться как можно больше. Говорят, что уравнение (18.56) справедливо в условиях **двойного усиления при одинарном возбуждении**, имея в виду, что инжектируется один тип носителей, а способностью к ударной ионизации обладают оба типа. Если электроны и дырки инжектируются одновременно, полный результат является суммой двух частных результатов.

Шум усиления, вносимый обычным ЛФД, берет начало от двух источников:

1) случайность расположения точек, в которых происходят акты ионизации;

2) процесс обратной связи, связанный с тем фактом, что оба типа носителей могут производить ударную ионизацию.

Первый фактор возникновения шума присутствует даже тогда, когда умножается число носителей лишь одного типа. Он порождает минимальный коэффициент избыточного шума $F = 2$ при больших значениях среднего усиления $\bar{G}$, что очевидно, если положить $\bar{k} = 0$ и сделать $\bar{G}$ бóльшим в (18.56). Второй источник шума — обратная связь — потенциально более вреден, поскольку он может вызывать намного больший рост F.

Пример 18.5

Коэффициент избыточного шума для кремниевого ЛФД

Кремниевый ЛФД с использованием электронной инжекции имеет среднее усиление $\bar{G} = 100$ и ионизационное отношение $k = 0{,}1$. Уравнение (18.56) дает $F = 11{,}8$, так что среднее значение детектируемого тока увеличивается в 100 раз, тогда как отношение сигнал—шум сокращается в 11,8 раза. Однако, при наличии шума схемы, использование ЛФД позволяет увеличить общее ОСШ, как будет продемонстрировано ниже.

ЛФД с коэффициентами ионизации, обладающими памятью

Вновь образованный носитель заряда может вызвать ударную ионизацию только после прохождения некоторого расстояния внутри области умножения, чтобы приобрести от ускоряющего поля достаточную энергию. Это расстояние называется **мертвым пространством**. Следовательно, коэффициенты ионизации на самом деле не являются независимыми от положения и истории носителя, как это предполагается в теории обычного ЛФД. Мертвое пространство организует локализацию областей, в которых может происходить ударная ионизация. В свою очередь, это усиливает упорядоченность процесса генерации носителей и приводит к снижению шума усиления. Сказанное в особенности верно, когда область умножения очень тонкая ($w < 400$ нм) и число умножений мало.

Дальнейшее снижение шума может быть достигнуто за счет должного управления энергией носителей.

• *Эффекты начальной энергии.* Носители, движущиеся в поле со специально созданным градиентом до входа в область умножения, могут приобрести значительную кинетическую энергию, за счет чего уменьшается начальное мертвое пространство в области умножения и происходит регуляризация дальнейшей цепочки ударных ионизаций.

• *Эффекты пороговой энергии ударной ионизации.* Прибор можно спроектировать таким образом, чтобы носитель, пересекающий область умножения, встретил внезапное изменение пороговой энергии ионизации при переходе из слоя одного материала в слой другого. Носитель, энергия которого была недостаточна для ударной ионизации в первом слое, может вызвать ее во втором.

Локализация процессов ударной ионизации в таких специально спроектированных многослойных структурах позволяет создавать приборы с низким шумом, большим усилением и малым темновым током. Пример диаграммы энергетических зон такого прибора в условиях обратного смещения показан на рис. 18.32. Два тонких умножающих слоя со сравнительно низким порогом ионизации окружают слой с более высоким порогом. Ударная ионизация усиливается на краях двух окружающих слоев и подавляется в центральном слое, который сообщает проходящим носителям энергию. Материалы подбираются так, чтобы ионизация, индуцированная дырками, не проявлялась.

Разработана теория шума ЛФД, включающая указанные эффекты[1]. Она имеет вид рекуррентных соотношений для первого и второго моментов и распределения вероятностей чисел электронов и дырок. Случайные переменные детерминированным образом связаны со случайным усилением. Рекуррентные соотношения сформулированы таким образом, чтобы учесть мертвое пространство, а также эффекты начальной энергии и порога ионизации, описанные выше. Численные решения дают среднее усиление и коэффициент избыточного шума для произвольных размеров мертвого пространства и ширины области умножения. Теория правильно предсказывает результаты измерений, представленные в примере 18.6.

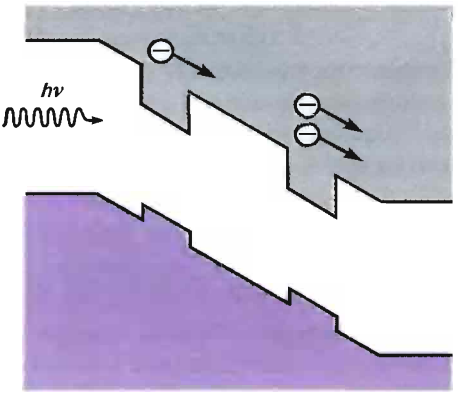


Рис. 18.32. Диаграмма энергетических зон малошумящего гетероструктурного ЛФД в условиях обратного смещения

Пример 18.6

Коэффициент избыточного шума для очень тонкого ЛФД на GaAs

Тонкий гетероструктуный ЛФД, схематически изображенный на рис. 18.32, имеет область умножения в виде двух слоев GaAs толщиной 50 нм, окружающих слой $Al_{0,6}Ga_{0,4}As$ толщиной 85 нм. Измеренный коэффициент избыточного шума равен $F \approx 2{,}5$ при среднем усилении $\bar{G} = 20$. Теоретически предсказываемый формулой (18.56) коэффициент избыточного шума для ЛФД на гомопереходе в объемном GaAs ($k \approx 0{,}75$) равен $F \approx 15{,}5$ при $\bar{G} = 20$ (см. рис. 18.31). Таким образом, измеренные шумовые характеристики гораздо ниже предсказываемых объемной теорией, которая игнорирует мертвое пространство, равно как и эффекты начальной энергии и изменения порога ионизации. Видно, что эти эффекты существенны для понижения уровня шума усиления и должны учитываться при моделировании ЛФД с тонкой умножающей областью. Другие гетероструктурные конфигурации, такие как конфигурация с центрированной ямой, позволяют достичь еще меньших значений F при малых $\bar{G}$.

[1] См. *Hayat M.M., Kwon O.-H., Wang S., Campbell J.C., Saleh B.E.A., Teich M.C.* Boundary Effects on Multiplication Noise in Twin Heterostructure Avalanche Photodiodes: Theory and Experiment, IEEE Transactions on Electron Devices. Vol. 49, 2002. P. 2114—2123.

18.6.3. Шум схемы

Еще одним источником шума является электронная схема, связанная с оптическим приемником. Шум схемы происходит от теплового движения носителей заряда в резисторах и других диссипативных элементах (тепловой шум), а также от флуктуаций носителей заряда в транзисторах, служащих для усиления сигнала.

Тепловой шум

Тепловой шум, называемый также шумом Джонсона или шумом Найквиста, происходит от теплового движения подвижных носителей заряда в резистивных электрических материалах при конечной температуре; это движение порождает случайный электрический ток $i(t)$ даже в отсутствие внешнего источника электрической энергии. Тепловой электрический ток в сопротивлении R есть случайная функция $i(t)$ со средним значением $\langle i(t)\rangle = 0$. Дисперсия тока σ_i^2 совпадает с его среднеквадратичным значением, поскольку среднее исчезает, и растет с увеличением температуры T.

Из соображений статистической механики, представленных в следующем разделе, можно показать, что при температуре T в сопротивлении R возникает случайный электрический ток $i(t)$, характеризуемый спектральной плотностью мощности (см. подразд. 11.1.2)

$$S_i(f) = \frac{4}{R}\frac{hf}{\exp(hf/k_B T) - 1}, \tag{18.57}$$

где f — частота. В области $f \ll k_B T/h$, представляющей основной интерес, поскольку $k_B T/h = 6{,}24$ ТГц при комнатной температуре,

$$\exp\left(\frac{hf}{k_B T}\right) \approx 1 + \frac{hf}{k_B T},$$

так что

$$S_i(f) \approx \frac{4k_B T}{R}. \tag{18.58}$$

Дисперсия электрического тока есть интеграл от спектральной плотности мощности по всем частотам в пределах ширины полосы схемы B, т. е.

$$\sigma_i^2 = \int_0^B S_i(f)\,\mathrm{d}f. \tag{18.59}$$

Для $B \ll k_B T/h$ получаем

$$\boxed{\sigma_i^2 \approx \frac{4k_B TB}{R}}. \tag{18.60}$$

Дисперсия тока теплового шума (сопротивление R)

Таким образом, как показано на рис. 18.33, сопротивление R при температуре T в схеме с шириной полосы B ведет себя как бесшумное сопротивление, параллельно которому подсоединен источник шумового тока с нулевым средним значением и среднеквадратичным значением σ_i, определяемым по формуле (18.60).

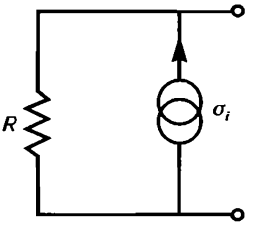


Рис. 18.33. Сопротивление R при температуре T эквивалентно бесшумному сопротивлению, параллельно которому подсоединен источник шумового тока с дисперсией $\sigma_i^2 = \langle i^2 \rangle \approx 4k_B TB/R$, где B — ширина полосы схемы

Пример 18.7

Тепловой шум сопротивления

Сопротивление величиной 1 кОм при $T = 300$ °K в схеме с шириной полосы $B = 100$ МГц создает тепловой шумовой ток со среднеквадратичным значением $\sigma_i \approx 41$ нА.

*Вывод выражения для спектральной плотности мощности теплового шума

Вывод (18.57) основан на демонстрации того, что электрическая мощность, связанная с тепловым шумом в сопротивлении, идентична электромагнитной мощности, излучаемой одномерным черным телом. В множителе $hf/[\exp(hf/k_B T) - 1]$ в формуле (18.57) легко распознать среднюю энергию $\bar{E}$ в электромагнитной моде частоты f (символ ν резервируем для оптических частот) в состоянии теплового равновесия при температуре T [см. (13.66)]. Формулу (18.57) поэтому можно записать как $S_i(f)R = 4\bar{E}$. Электрическая мощность, выделяемая током i при прохождении сопротивления R, равна $\langle i^2 \rangle R = \sigma_i^2 R$, так что $S_i(f)R$ представляет собой плотность электрической мощности (на 1 Гц), выделяемой шумовым током $i(t)$ на сопротивлении R.

Перейдем к демонстрации того, что $4\bar{E}$ есть плотность мощности, излучаемой одномерным черным телом. Как обсуждалось в подразд. 13.4.2, спектральная плотность энергии излучения атомной системы в тепловом равновесии с электромагнитными модами в резонаторе равна $\rho(\nu) = M(\nu)\bar{E}$, где $M(\nu) = 8\pi\nu^2/c^3$ — трехмерная плотность числа мод, а спектральная интенсивность равна $c\rho(\nu)$. Хотя носители заряда в сопротивлении движутся во всех направлениях, только движение в направлении тока в цепи вносит вклад в шумовой ток. Плотность числа мод в одном направлении есть $M(f) = 4/c$ мод/(м · Гц) [см. (10.10)], так что соответствующая плотность энергии есть $\rho(f) = M(f)\bar{E} = 4\bar{E}/c$, а излучаемая плотность мощности составляет $c\rho(\nu) = 4\bar{E}$, что и требовалось доказать.

Параметры шума схемы: оптические приемники, ограниченные по сопротивлению и усилению

Удобно сосредоточить различные источники шума схемы (тепловой шум в сопротивлениях, шум транзисторов и других элементов схемы) в один источник случайного тока i_r на входе приемника, который вызывает эквивалентный полный шум на выходе (рис. 18.34). Среднее значение i_r равно нулю, а дисперсия σ_r^2 зависит от температуры, ширины полосы приемника, параметров схемы и типа прибора.

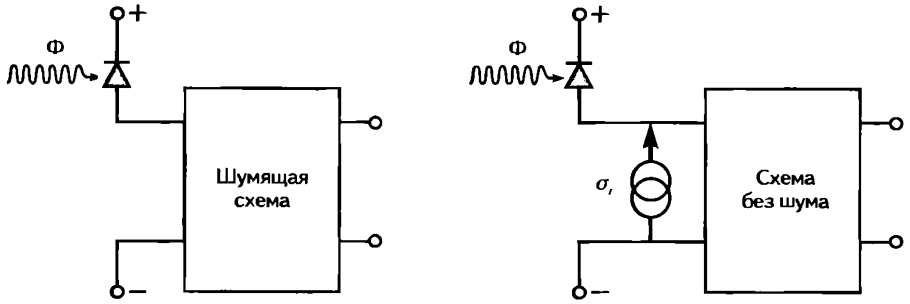


Рис. 18.34. Шумящая схема приемника может быть заменена бесшумной схемой и одним источником случайного тока со среднеквадратичным значением σ_r на входе

Далее, удобно определить безразмерный параметр шума схемы

$$\sigma_q = \frac{\sigma_r T}{e} = \frac{\sigma_r}{2Be}, \tag{18.61}$$

где B — ширина полосы приемника; $T = 1/2B$ — время разрешения. Поскольку σ_r — среднеквадратичное значение шумового тока, σ_r/e представляет собой среднеквадратичный поток электронов (число электронов в секунду) за счет шума схемы, и, следовательно, $\sigma_q = (\sigma_r/e)T$ — среднеквадратичное число шумовых электронов, накопленных за время T. Таким образом, шумовой параметр схемы является показателем ее качества, как станет очевидно в подразд. 18.6.4.

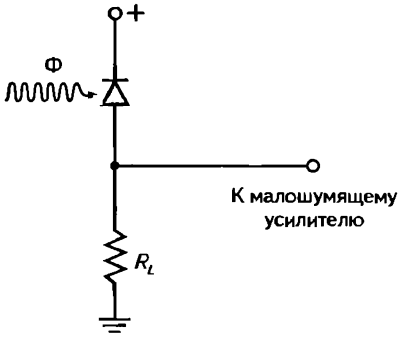


Рис. 18.35. Оптический приемник, ограниченный по сопротивлению

Схема оптического приемника, состоящего из фотодиода, включенного последовательно сопротивлению нагрузки R_L, сигнал с которого подается на усилитель, показана на рис. 18.35. Такой простой приемник называется **ограниченным по сопротивлению**, если шумовой ток схемы из-за теплового шума в сопротивлении значительно превышает вклады всех остальных источников шума. Усилитель в этом случае можно

рассматривать как нешумящий, а среднеквадратичный ток шума схемы равен $\sigma_r^2 = 4k_B TB/R_L$. Параметр шума схемы, определяемый формулой (18.61), в этом случае составляет

$$\sigma_q = \sqrt{\frac{k_B T}{e^2 R_L B}}, \tag{18.62}$$

что обратно пропорционально квадратному корню и B.

Пример 18.8

Параметр шума схемы

При комнатной температуре сопротивление $R_L = 50$ Ом в схеме с шириной полосы $B = 100$ МГц генерирует случайный ток со среднеквадратичным значением $\sigma_r = 0{,}18$ мкА. Это соответствует параметру шума схемы $\sigma_q \approx 5700$.

Приемник с использованием хорошо спроектированного малошумящего усилителя может иметь более низкий параметр шума схемы, чем приемник, ограниченный по сопротивлению. Рассмотрим приемник с усилителем на *полевом транзисторе*. Если шумом на высоком входном сопротивлении усилителя можно пренебречь, то ограничивающим фактором является тепловой шум в канале между истоком и стоком полевого транзистора. При использовании эквалайзера для поддержки высоких частот, ослабляемых емкостным входным импедансом схемы, параметр шума схемы при комнатной температуре для типичных значений параметров элементов схемы оказывается равным

$$\boxed{\sigma_q \approx \frac{\sqrt{B}}{100} \quad (B \text{ выражено в МГц}).} \tag{18.63}$$

Параметр шума схемы (приемник с усилителем на полевом транисторе)

Например, если $B = 100$ МГц, то $\sigma_q = 100$. Это существенно меньше, чем значение параметра шума схемы с ограничением по сопротивлению 50 Ом при той же ширине полосы. Параметр шума схемы σ_q растет с увеличением B из-за влияния эквалайзера[2].

Приемник, в котором используется усилитель на *биполярном транзисторе*, напротив, имеет параметр σ_q, не зависящий от ширины полосы B в широком диапазоне частот. Для ширины полосы от 100 МГц до 2 ГГц в этом случае типичное значение $\sigma_q \approx 500$, при надлежащем выборе транзистора и оптимальном смещении.

[2] Более подробные разъяснения см. в *Personick S.D.* Optical Fiber Transmission Systems. Plenum, 1981, разд. 3.4. Заметим, что в этой работе наш параметр σ_q эквивалентен $Z/2$.

18.6.4. Отношение сигнал—шум и обнаружительная способность приемника

Простейшей мерой качества приема является отношение сигнал—шум. ОСШ для тока на входе в нешумящую схему, показанную на рис. 18.34, есть отношение квадрата среднего тока к сумме дисперсий парциальных источников шума

$$\text{ОСШ} = \frac{\bar{i}^2}{2e\bar{G}\bar{i}BF + \sigma_r^2} = \frac{(e\bar{G}\eta\Phi)^2}{2e^2\bar{G}^2\eta B\Phi F + \sigma_r^2}. \qquad (18.64)$$

Отношение сигнал—шум для оптического приемника

Первый член в знаменателе описывает фотоэлектронный шум и шум усиления детектора [см. (18.53)], а второй — шум внешней схемы. Для детектора без усиления $\bar{G} = 1$ и $F = 1$. Нешумящая схема не меняет отношения сигнал—шум, даже если она обеспечивает усиление.

Упражнение 18.1

Отношение сигнал—шум оптического приемника с ограничением по сопротивлению

Предположим, что в оптическом приемнике, схема которого показана на рис. 18.35, используется идеальный p—i—n-фотодиод ($\eta = 1$), а сопротивление $R_L = 50$ Ом при $T = 300$ °K. Ширина полосы составляет $B = 100$ МГц. При каком значении потока фотонов Ф дисперсия тока фотоэлектронного шума будет равна дисперсии тока теплового шума сопротивления? Какова соответствующая оптическая мощность на длине волны $\lambda_0 = 1550$ нм?

Полезно выразить ОСШ в (18.64) через среднее число детектируемых фотонов $\bar{m}$ за время разрешения приемника $T = 1/2B$:

$$\bar{m} = \eta\Phi T = \frac{\eta\Phi}{2B}, \qquad (18.65)$$

и параметр шума схемы $\sigma_q = \sigma_r/2Be$. В результате получается выражение

$$\text{ОСШ} = \frac{\bar{G}^2\bar{m}^2}{\bar{G}^2F\bar{m} + \sigma_q^2}. \qquad (18.66)$$

Отношение сигнал—шум для оптического приемника

Формула (18.66) имеет простую интерпретацию. Числитель представляет собой квадрат среднего числа умноженных фотоэлектронов, зарегистрированных за время разрешения $T = 1/2B$. Знаменатель есть сумма дисперсий числа

фотоэлектронов и числа электронов шума схемы, собранных за время T. Для фотодиода без усиления $\bar{G} = F = 1$ и (18.66) сводится к выражению

$$\text{ОСШ} = \frac{\bar{m}^2}{\bar{m} + \sigma_q^2}. \qquad (18.67)$$

Отношение сигнал—шум (неусиливающий оптический приемник)

Сравнительная величина $\bar{m}$ и σ_q^2 определяют относительную важность фотоэлектронного шума и шума схемы. Теперь ясно, как параметр σ_q характеризует работу схемы оптического приемника. Например, если $\sigma_q = 100$, то шум схемы преобладает над фотоэлектронным шумом при условии, что среднее число фотоэлектронов, зарегистрированных за время разрешения, не превышает 10 000.

Перейдем к исследованию зависимости ОСШ от потока фотонов Ф, ширины полосы схемы B, параметра шума схемы σ_q и усиления $\bar{G}$. Это даст нам возможность определить, когда применение ЛФД выгодно, а также выбрать подходящий оптический предусилитель для данного потока фотонов. При проведении этого исследования мы будем опираться на выражения (18.64), (18.66) и (18.67) для ОСШ.

Зависимость ОСШ от потока фотонов

Зависимость ОСШ от $\bar{m} = \eta\Phi/2B$ отражает изменения ОСШ от потока фотонов Ф. Рассмотрим вначале случай фотодиода без усиления, когда применимо выражение (18.67).

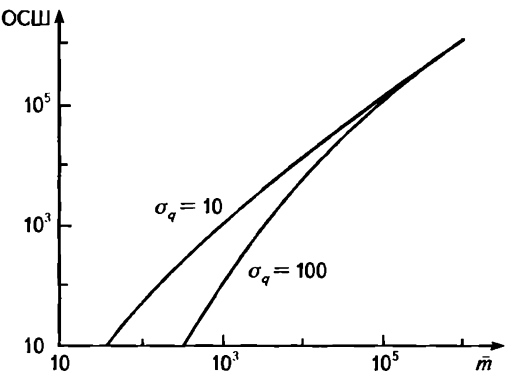


Рис. 18.36. Отношение сигнал—шум (ОСШ) как функция среднего числа фотоэлектронов за время разрешения приемника, $\bar{m} = \eta\Phi/2B$, для фотодиода при двух значениях параметра шума схемы σ_q

Представляют интерес два предельных случая:

1. *Предел шума схемы.* Если Ф достаточно мало, так что $\bar{m} \ll \sigma_q^2$ ($\Phi \ll 2B\sigma_q^2/\eta$), то фотонным шумом можно пренебречь, а шум схемы преобладает, что дает

$$\text{ОСШ} \approx \frac{\bar{m}^2}{\sigma_q^2}. \qquad (18.68)$$

2. *Предел фотонного шума.* Если поток фотонов Ф достаточно велик, так что $\bar{m} \gg \sigma_q^2$ ($\Phi \gg 2B\sigma_q^2/\eta$), то можно пренебречь шумом схемы, откуда

$$\text{ОСШ} \approx \bar{m}. \qquad (18.69)$$

Следовательно, для малых $\bar{m}$ ОСШ пропорционально $\bar{m}^2$ и, следовательно, Φ^2, тогда как для больших $\bar{m}$ ОСШ пропорционально $\bar{m}$ и Ф, как показано на рис. 18.36. Для всех интенсивностей света ОСШ растет с увеличением потока падающих фотонов, т. е. чем больше света, тем лучше работает приемник.

Когда применение ЛФД обеспечивает преимущества

Сравним теперь два приемника, одинаковые во всех отношениях, кроме того, что у одного нет усиления, а другой обеспечивает внутреннее усиление $\bar{G}$ и имеет коэффициент избыточного шума F (например, ЛФД). Для достаточно малых $\bar{m}$ (или фотонных потоков Ф) преобладает шум схемы. В этом случае усиление фототока до значений выше уровня шумов схемы улучшит ОСШ и приемник на ЛФД будет иметь преимущества. Для достаточно большого $\bar{m}$ (или потока фотонов) шум схемы пренебрежимо мал. В этом случае усиление фототока вносит шум усиления, снижая таким образом ОСШ. В этом случае предпочтение следует отдать фотодиоду. Сравнивая (18.66) и (18.67), мы видим, что ОСШ для приемника на ЛФД больше, чем для фотоприемника на фотодиоде, когда

$$\bar{m} < \sigma_q^2 \frac{1 - 1/\bar{G}^2}{F - 1}.$$

Для $\bar{G} \gg 1$ ЛФД имеет преимущества, когда

$$\bar{m} < \frac{\sigma_q^2}{F - 1}.$$

Если это условие не выполняется, использование ЛФД портит, а не улучшает работу приемника. Например, когда σ_q очень мало, из (18.66) очевидно, что для ЛФД ОСШ $= \bar{m}/F$ ниже, чем для фотодиода, у которого ОСШ $= \bar{m}$. График зависимости ОСШ от $\bar{m}$ для двух приемников приведен на рис. 18.37.

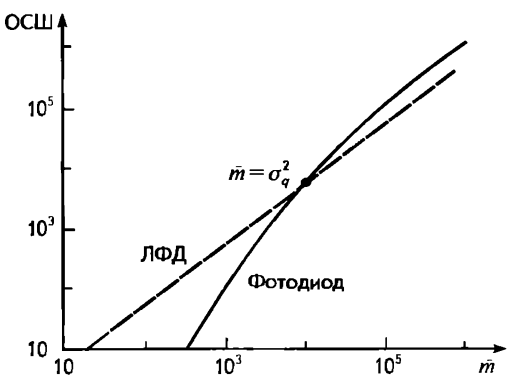


Рис. 18.37. Зависимость ОСШ от $\bar{m} = \eta\Phi/2B$ для приемника на фотодиоде (сплошная линия) и на ЛФД со средним усилением $\bar{G} = 100$ и коэффициентом избыточного шума $F = 2$ (штриховая линия), полученная из (18.66). Параметр шума схемы $\sigma_q = 100$ в обоих случаях. Для малых фотонных потоков (случай ограничения шумами схемы) ЛФД обеспечивает более высокое ОСШ, чем фотодиод. Для больших фотонных потоков (случай ограничения фотонным шумом) приемник на фотодиоде лучше, чем на ЛФД. Переход между двумя областями происходит при $\bar{m} \approx \sigma_q^2/(F - 1) = 10^4$

Зависимость ОСШ от усиления ЛФД

Использование ЛФД выгодно для достаточно малых фотонных потоков, $\bar{m} < \sigma_q^2/(F-1)$. Оптимальное усиление ЛФД определяется с помощью (18.66):

$$\text{ОСШ} = \frac{\bar{G}^2\bar{m}}{\bar{G}^2F + \sigma_q^2/\bar{m}}. \tag{18.70}$$

Коэффициент избыточного шума F сам является функцией $\bar{G}$, как видно из (18.56) для ЛФД большой толщины. Подстановка дает

$$\text{ОСШ} = \frac{\bar{G}^2\bar{m}}{k\bar{G}^3 + (1-k)\left(2\bar{G}^2 - \bar{G}\right)F + \sigma_q^2/\bar{m}}, \tag{18.71}$$

где k — ионизационное отношение ЛФД. График этого соотношения приведен на рис. 18.38 для $\bar{m} = 1000$ и $\sigma_q = 500$. Для ЛФД с размножением одного типа носителей ($k = 0$) ОСШ растет с увеличением усиления и, в конце концов, достигает насыщения. Для ЛФД с размножением двух типов носителей ($k > 0$) ОСШ также растет с увеличением усиления, но достигает максимума при оптимальном усилении, а затем уменьшается в результате резкого роста шума усиления. Таким образом, для ЛФД в общем случае существует оптимальный выбор усиления.

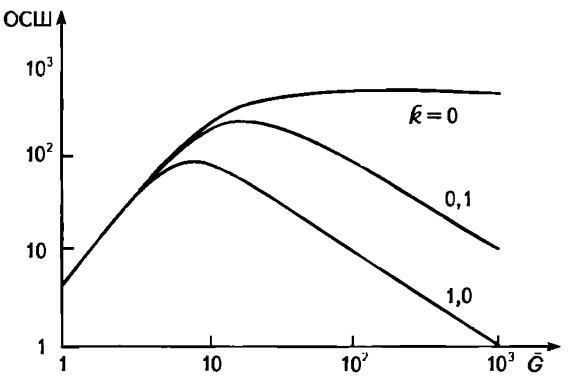


Рис. 18.38. Зависимость ОСШ от среднего усиления $\bar{G}$ ЛФД для различных значений ионизационного отношения k при $\bar{m} = 1000$ и $\sigma_q = 500$

Зависимость ОСШ от ширины полосы приемника

Связь между ОСШ и шириной полосы B явно следует из (18.64). Она определяется зависимостью дисперсии шумового тока схемы σ_r^2 от B. Рассмотрим три приемника.

1. *Приемник с ограничением по сопротивлению* имеет $\sigma_r^2 \propto B$ [см. (18.60)], так что

$$\text{ОСШ} \propto \frac{1}{B}. \tag{18.72}$$

2. Приемник *с усилителем на полевом транзисторе* подчиняется закону $\sigma_r^2 \propto B^{1/2}$ [см. (18.63)], так что $\sigma_r = 2eB\sigma_q \propto B^{3/2}$. Это значит, что зависимость ОСШ от B в (18.64) принимает вид

$$\text{ОСШ} \propto \frac{1}{B + sB^3}, \tag{18.73}$$

где s — постоянная.

3. *Усилитель на биполярном транзисторе* имеет параметр шума схемы σ_q, приближенно не зависящий от B. Таким образом, $\sigma_r \propto B$, так что (18.64) принимает вид

$$\text{ОСШ} \propto \frac{1}{B + s'B^2}, \tag{18.74}$$

где s' — постоянная.

Эти соотношения схематически иллюстрирует рис. 18.39. ОСШ всегда убывает с ростом B. Для сравнительно малых ширин полосы у всех трех приемников ОСШ меняется как $1/B$. Для больших ширин полосы ОСШ для усилителей на полевом и биполярном транзисторах убывают более резко.

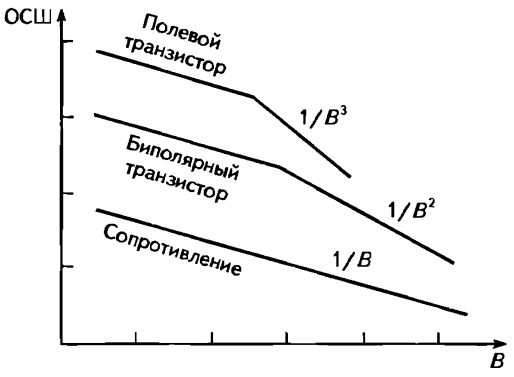


Рис. 18.39. Дважды логарифмический график зависимости ОСШ от ширины полосы B для трех типов приемников

Обнаружительная способность приемника

Обнаружительная способность приемника — это минимальный поток фотонов Φ_0 с соответствующими ему оптической мощностью $P_0 = h\nu\Phi_0$ и средним числом фотоэлектронов $\bar{m}_0 = \eta\Phi_0/2B$, которые требуются для обеспечения заданной величины отношения сигнал—шум ОСШ_0. Величину $\bar{m}_0$ можно определить, решая (18.66) для $\text{ОСШ} = \text{ОСШ}_0$. Мы будем рассматривать только приемник с единичным усилением, оставив более общий случай для упражнения.

Решая квадратное уравнение (18.67) относительно $\bar{m}_0$, получаем

$$\bar{m}_0 = \frac{1}{2}\left[\text{ОСШ}_0 + \sqrt{\text{ОСШ}_0{}^2 + 4\sigma_q^2\text{ОСШ}_0}\right]. \tag{18.75}$$

Возникает два предельных случая:

$$\text{Предел фотонного шума } \left(\sigma_q^2 \ll \frac{\text{ОСШ}_0}{4}\right): \quad \bar{m}_0 = \text{ОСШ}_0. \tag{18.76}$$

$$\text{Предел шума схемы } \left(\sigma_q^2 \gg \frac{\text{ОСШ}_0}{4}\right): \quad \bar{m}_0 = \sqrt{\text{ОСШ}_0}\,\sigma_q. \tag{18.77}$$

Обнаружительная способность приемника

Пример 18.9

Обнаружительная способность приемника

Предположим, что $\text{ОСШ}_0 = 10^4$, что соответствует приемлемому отношению сигнал—шум в 40 дБ. Если параметр шума схемы приемника $\sigma_q = 50$, то приемник ограничен по фотонному шуму, а его обнаружительная способность составляет $\bar{m}_0 = 10\,000$ фотоэлектронов за время разрешения приемника. В более правдоподобной ситуации, для которой $\sigma_q \gg 50$, обнаружительная способность приемника составляет $\approx 100\sigma_q$. Если, например, $\sigma_q = 500$, то обнаружительная способность $\bar{m}_0 = 50\,000$, что соответствует $2B\bar{m}_0 = 10^5 B$ фотоэлектронов в секунду. Обнаружительная способность по оптической мощности

$$P_0 = \frac{2B\bar{m}_0 h\nu}{\eta} = \frac{10^5\, Bh\nu}{\eta}$$

прямо пропорциональна ширине полосы. Если $B = 100$ МГц и $\eta = 0{,}8$, то при $\lambda_0 = 1550$ нм обнаружительная способность приемника составляет $P_0 \approx 1{,}6$ мкВт.

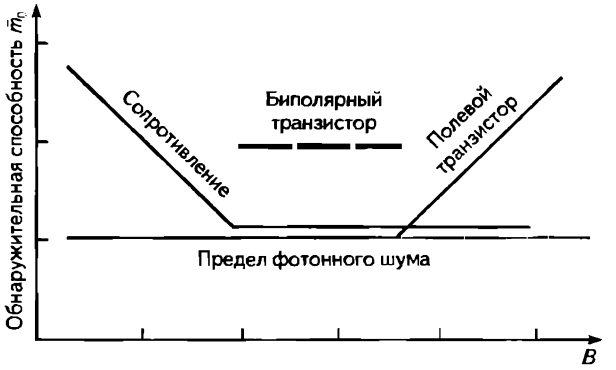


Рис. 18.40. Дважды логарифмический график обнаружительной способности приемника $\bar{m}_0$ (минимального среднего числа фотоэлектронов за время разрешения $T = 1/2B$, гарантирующее заданную величину отношения сигнал—шум ОСШ_0) как функции ширины полосы B для трех типов приемников. Графики приближаются к пределу фотонного шума при значениях B, для которых $\sigma_q^2 \ll \text{ОСШ}_0/4$. В пределе фотонного шума (т. е. когда шумом схемы можно пренебречь) $\bar{m}_0 = \text{ОСШ}_0$ во всех случаях

При использовании (18.75) для определения обнаружительной способности нужно помнить, что параметр шума схемы σ_q, вообще говоря, является функцией ширины полосы B в соответствии с выражениями:

Приемник с ограничением по сопротивлению: $\sigma_q \propto \frac{1}{\sqrt{B}}$;

Усилитель на полевом транзисторе: $\sigma_q \propto \sqrt{B}$;

Усилитель на биполярном транзисторе: σ_q не зависит от B.

Для этих приемников обнаружительная способность $\bar{m}_0$ зависит от ширины полосы B, как показано на рис. 18.40. Оптимальный выбор приемника, следовательно, частично зависит от ширины полосы B.

Упражнение 18.2

Обнаружительная способность приемника на ЛФД

Выведите выражение, аналогичное (18.75), для обнаружительной способности приемника, включающего в себя ЛФД с усилением $\bar{G}$ и коэффициентом избыточного шума F. Покажите, что в пределе пренебрежимо малого шума схемы обнаружительная способность приемника сводится к

$$\bar{m}_0 = F \cdot \text{ОСШ}_0. \tag{18.78}$$

18.6.5. Частота появления ошибочных битов и обнаружительная способность приемника

Обнаружительная способность аналогового приемника была определена в подразд. 18.6.4 как минимальная мощность принимаемого света (или соответствующий поток фотонов), необходимая для достижения предписанного отношения сигнал—шум ОСШ_0. Теперь рассмотрим обнаружительную способность цифрового приемника для систем связи. Для бинарной системы, работающей по принципу «включено—выключено», обнаружительная способность определяется как минимальная оптическая энергия (или соответствующее среднее число фотонов) на один бит, необходимая для достижения предписанной частоты появления ошибочных битов (Bit Error Rate — BER). Вначале определим обнаружительную способность идеального детектора, а затем рассмотрим влияние шума схемы и шума усиления детектора.

Обнаружительная способность идеального оптического приемника

Предположим, что «1» и «0» системы, работающей по принципу «включено—выключено», представлены присутствием либо отсутствием оптической энергии, соответственно, как описано в гл. 24. В течение бита «1» при-

нимается среднее число $\bar{n}$ фотонов. В течение бита «0» фотоны не принимаются. Если оба бита одинаково вероятны, то среднее число фотонов на один бит:

$$\bar{n}_a = \frac{1}{2}\bar{n}.$$

Поскольку в действительности число детектируемых фотонов случайно, имеют место ошибки в идентификации (считывании) бита. Для света, генерируемого светодиодами, вероятность детектирования n фотонов подчиняется распределению Пуассона

$$p(n) = \bar{n}^n \frac{\exp(-\bar{n})}{n!},$$

когда передано среднее число фотонов $\bar{n}$ (см. разд. 12.2). Приемник решает, что передавалась единица, если регистрируется хотя бы один фотон. Следовательно, вероятность p_1 ошибочной регистрации нуля вместо единицы равна вероятности того, что не будет зарегистрирован ни один фотон, т. е.

$$p_1 = p(0) = \exp(-\bar{n}).$$

При передаче бита «0» фотоны отсутствуют, приемник правильно решает, что был передан бит «0», так что $p_0 = 0$. Частота появления ошибочных битов есть среднее двух вероятностей ошибки, BER $= (p_1 + p_0)/2$, откуда

$$\text{BER} = \frac{1}{2}\exp(-\bar{n}) = \frac{1}{2}\exp(-2\bar{n}_a). \qquad (18.79)$$

На рис. 18.41 представлен полулогарифмический график этого соотношения.

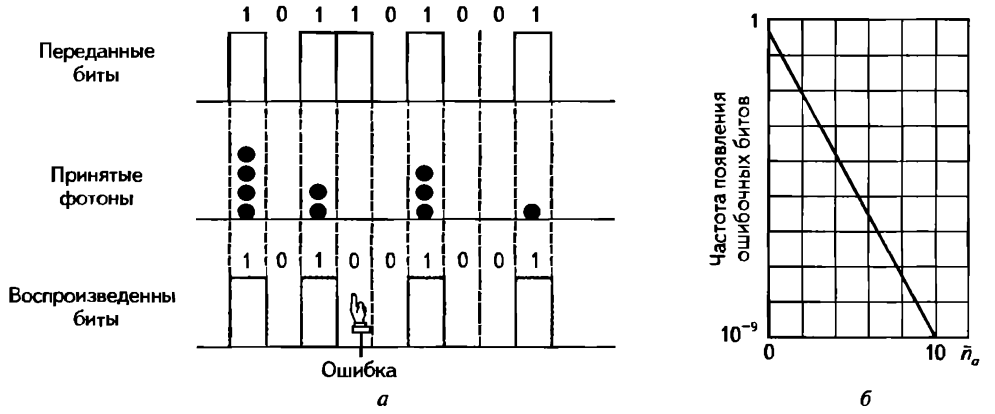


Рис. 18.41. Схематическая иллюстрация возникновения ошибки из-за случайности числа фотонов (*а*). Частота появления ошибочных битов в зависимости от числа фотонов на один бит $\bar{n}_a$ в релейной системе «включено—выключено» с идеальным приемником (*б*)

Обнаружительная способность приемника определяется как среднее число фотонов на один бит, требующееся для достижения некоторого заданного значения BER. В частности, для часто выбираемого значения BER = 10^{-9} формула (18.79) дает $\bar{n}_a \approx 10$ фотонов на бит. Мы приходим к выводу, что:

обнаружительная способность приемника (для частоты появления ошибочных битов BER = 10^{-9}) оптической цифровой коммуникационной системы с использованием идеального приемника равна 10 фотонам на бит.

Упражнение 18.3

Влияние квантового выхода и фонового шума на обнаружительную способность приемника

а. Покажите, что для приемника, детектор которого имеет квантовую эффективность η, но в остальном является идеальным,

$$\mathrm{BER}=\frac{1}{2}\exp\left(-2\eta\bar{n}_a\right),$$

так что обнаружительная способность $\bar{n}_a = 10/\eta$ фотонов на бит, т. е. $\bar{m}_a = \eta\bar{n}_a = 10$ фотоэлектронов на бит.

б. Предполагая, что биты «1» и «0» соответствуют средним числам фотонов $\bar{n}_1 = \bar{n} + \bar{n}_B$ и $\bar{n}_0 = \bar{n}_B$, где $\bar{n}$ — среднее число сигнальных фотонов; $\bar{n}_B$ — средний поток фотонов фона с пуассоновским законом распределения, не зависящий от числа сигнальных фотонов, получите выражение для BER как функции $\bar{n}$ и $\bar{n}_B$. Постройте график зависимости BER от

$$\bar{n}_a = \frac{1}{2}\bar{n}$$

для нескольких значений $\bar{n}_B$. Из этого графика определите обнаружительную способность приемника $\bar{n}_a$ как функцию $\bar{n}_B$.

Указание. Сумма случайных чисел, каждое из которых имеет пуассоновское распределение вероятностей, также распределена по Пуассону.

Обнаружительная способность идеального приемника 10 фотонов на бит применима только к свету с пуассоновским распределением числа фотонов. В принципе обнаружительную способность можно улучшить с использованием света, сжатого по числу фотонов (см. подразд. 12.3.2).

Обнаружительная способность приемника с шумами схемы и усиления

Как объяснено в подразд. 18.6.1, фотодиод преобразует долю η принятых фотонов в электронно-дырочные пары, каждая из которых вносит заряд e в электрический ток внешней цепи. Полный заряд, накопленный за

время T, отводимое на один бит, равен m (в единицах заряда электрона). Это число случайное и имеет пуассоновское распределение со средним значением $\bar{m} = \eta\bar{n}$ и дисперсией $\bar{m}$.

Дополнительный шум вносится схемой фотоприемника в виде случайного электрического тока i_r с гауссовым распределением вероятности, среднее значение которого равно нулю, а дисперсия σ_r^2. За временной интервал T, приходящийся на один бит, накопленный заряд $q = i_r T/e$ (в единицах заряда электрона) имеет среднеквадратичное значение $\sigma_q = \sigma_r T/e$. Параметр σ_q, называемый параметром шума схемы, зависит от ширины полосы приемника B, как описано в подразд. 18.6.3.

Полный накопленный заряд за один бит $s = m + q$ (в единицах заряда электрона) представляет собой сумму пуассоновской случайной переменной m и независимой гауссовой случайной переменной q. Его среднее значение равно сумме средних

$$\mu = \bar{m} = \eta\bar{n}, \tag{18.80}$$

а дисперсия — сумме дисперсий

$$\sigma^2 = \bar{m} + \sigma_q^2. \tag{18.81}$$

Для остаточно больших $\bar{m}$ распределение Пуассона можно аппроксимировать гауссовым распределением, так что полное распределение можно аппроксимировать гауссовым распределением со средним значением μ и дисперсией σ^2. Мы примем такое распределение за основу для последующего анализа.

Для лавинного фотодиода (ЛФД) с усилением $\bar{G}$ среднее число фотоэлектронов усиливается на множитель $\bar{G}$, но при этом вносится дополнительный шум в процессе усиления. Среднее значение полного накопленного заряда s на 1 бит:

$$\mu = \bar{m}\bar{G}, \tag{18.82}$$

а его дисперсия

$$\sigma^2 = \bar{m}\bar{G}^2 F + \sigma_q^2, \tag{18.83}$$

где $F = \langle G^2\rangle/\langle G\rangle^2$ — коэффициент избыточного шума ЛФД (см. подразд. 18.6.2).

Приемник измеряет заряд s, накопленный за каждый бит (например, с помощью интегратора), и сравнивает его с заданным пороговым значением ϑ. Если $s > \vartheta$, выбирается бит «1», в противном случае «0». Вероятности ошибки p_1 и p_0 определяются с помощью двух гауссовых распределений вероятности s, для которых

$$\begin{aligned} &\text{среднее } \mu_0 = 0, \quad \text{дисперсия } \sigma_0^2 = \sigma_q^2 && \text{для бита «0»;} \\ &\text{среднее } \mu_1 = \bar{m}\bar{G}, \quad \text{дисперсия } \sigma_1^2 = \bar{m}\bar{G}^2 F + \sigma_q^2 && \text{для бита «1».} \end{aligned} \tag{18.84}$$

Вероятность p_0 ошибочной регистрации «1» вместо правильного «0» представляет собой интеграл от гауссова распределения вероятности $p(s)$ со средним μ_0 и дисперсией σ_0^2 в пределах от $s = \vartheta$ до $s = \infty$. Вероятность p_1 ошибочной регистрации «0» вместо правильного «1» представляет собой интеграл от гаус-

сова распределения вероятности со средним μ_1 и дисперсией σ_1^2 в пределах от $s = -\infty$ до $s = \vartheta$. Порог выбирается таким, чтобы средняя вероятность ошибки BER = $(p_0 + p_1)/2$ была минимальной.

Данный тип анализа является основой общепринятой теории бинарного детектирования в присутствии гауссова шума. Если μ_0 и σ_0^2, и μ_1 и σ_1^2 — средние значения и дисперсии, связанные с двумя гауссовыми переменными, представляющие биты «0» и «1» соответственно, а σ_0 и σ_1 много меньше $\mu_1 - \mu_0$, то частота появления ошибочных битов приемника с оптимальным порогом приближенно выражается как

$$\mathrm{BER} \approx \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{Q}{\sqrt{2}}\right)\right]. \tag{18.85}$$

Здесь

$$Q = \frac{\mu_1 - \mu_0}{\sigma_0 + \sigma_1}, \tag{18.86}$$

а функция ошибок $\mathrm{erf}(z)$ определяется как

$$\mathrm{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z \exp(-x^2)\,dx. \tag{18.87}$$

Поскольку величина BER = 10^{-9} соответствует $Q \approx 6$, имеем

$$\boxed{\mu_1 - \mu_0 \approx 6(\sigma_0 + \sigma_1).} \tag{18.88}$$

Условие для BER = 10^{-9} (гауссово приближение)

Подставляя (18.84) в (18.88), определяя

$$\bar{m}_a = \frac{1}{2}\bar{m}$$

как среднее число фотоэлектронов на один бит и выполняя несложные преобразования, находим

$$\bar{m}_a \approx 18F + \frac{6\sigma_q}{\bar{G}}. \tag{18.89}$$

Равенство (18.89) связывает обнаружительную способность приемника, выраженную в виде среднего числа $\bar{m}_a$ фотоэлектронов на один бит, необходимого для обеспечения BER = 10^{-9}, с параметрами приемника $\bar{G}$, F и σ_q.

Когда усиление ЛФД достаточно велико, так что $3F\bar{G} \gg \sigma_q$, второе слагаемое (шум схемы) в правой части (18.89) пренебрежимо мало, следовательно

$$\boxed{\bar{m}_a \approx 18F.} \tag{18.90}$$

Обнаружительная способность приемника на ЛФД (без шума схемы)

Таблица 18.2. **Типичные значения обнаружительной способности (в средних числах фотонов на бит) некоторых оптических приемников, работающих с битовой скоростью передачи данных в диапазоне от 1 МБ/с до 2,5 ГБ/с**

Приемник	Обнаружительная способность, фотонов/бит
Идеальный приемник с ограничением по фотонному шуму	10
Кремниевый ЛФД	125
Предусилитель из кварцевого волокна, легированного Er^{3+}/*p*—*i*—*n*-фотодиод InGaAs	215
InGaAs-ЛФД	500
p—*i*—*n*-Фотодиод	6000

Согласно этим расчетам, приемник, имеющий пренебрежимо малый шум схемы, на фотодиоде без усиления ($\bar{G} = 1$ и $F = 1$) имеет обнаружительную способность $\bar{m}_a = 18$ фотоэлектронов на бит. Этот результат отличается от 10 фотоэлектронов на бит для идеального приемника, что было получено ранее (см. упражнение 18.3). Причина этого расхождения состоит в использовании гауссова распределения вместо распределения Пуассона, которое неприменимо для такого малого числа отсчетов. Типичные значения обнаружительной способности некоторых приемников приведены в табл. 18.2. Реальные значения зависят от параметра шума схемы σ_q, который, в свою очередь, зависит от битовой скорости $B_0 \equiv 1/T$.

Рекомендуемая литература

книги

См. также литературу к гл. 16 и 17.

Becker W. Advanced Time-Correlated Single Photon Counting Techniques. Springer-Verlag, 2005.

Shur M.S., *Žukauskas A.*, eds. UV Solid-State Light Emitters and Detectors. NATO Science Series II: Mathematics, Physics and Chemistry. Vol. 144. Springer-Verlag, 2004.

Johnson M. Photodetection and Measurement: Maximizing Performance in Optical Systems. McGraw-Hill, 2003.

Osche G.R. Optical Detection Theory for Laser Applications. Wiley, 2002.

Henini M., *Razeghi M.*, eds., Handbook of Infrared Detection Technologies. Elsevier, 2002.

Jha A.R. Infrared Technology: Applications to Electro-Optics, Photonic Devices, and Sensors. Wiley, 2000.

Howell S.B. Handbook of CCD Astronomy. Cambridge University Press, 2000.

Glass I.S. Handbook of Infrared Astronomy. Cambridge University Press, 2000.

Donati S. Photodetectors: Devices, Circuits, and Applications. Prentice Hall, 1999.

Trishenkov M.A. Detection of Low-Level Optical Signals: Photodetectors, Focal Plane Arrays and Systems. Springer-Verlag, 1997.

Choi K.K. The Physics of Quantum Well Infrared Photodetectors. World Scientific, 1997.

Dereniak E.L., Boreman G.D. Infrared Detectors and Systems. Wiley, 1996.

Kingston R.H. Optical Sources, Detectors, and Systems: Fundamentals and Applications. Academic Press, 1995.

Rieke G.H. Detection of Light: From the Ultraviolet to the Submillimeter. Cambridge University Press, 1994.

Manasreh M.O. Semiconductor Quantum Wells and Superlattices for Long-Wavelength Infrared Detectors. Artech, 1993.

Bube R.H. Photoelectronic Properties of Semiconductors. Cambridge University Press, paperback ed. 1992.

Joshi N.V. Photoconductivity: Art, Science, and Technology. Marcel Dekker, 1990.

Vincent J.D. Fundamentals of Infrared Detector Operation and Testing. Wiley, 1990.

Dennis P.N.J. Photodetectors. Springer-Verlag, 1986.

van der Ziel A. Noise in Solid State Devices and Circuits. Wiley, 1986.

Willardson R.K., Beer A.C., eds. Semiconductors and Semimetals. Vol. 22: Lightwave Communications Technology. W.T. Tsang, ed. Part D: Photodetectors. Academic Press, 1985.

Dereniak E.L., Crowe D.G. Optical Radiation Detectors. Wiley, 1984.

Buckingham M.J. Noise in Electron Devices and Systems. Wiley, 1983.

Boyd R.W. Radiometry and the Detection of Optical Radiation. Wiley, 1983.

Keyes R.J., ed. Optical and Infrared Detectors. Vol. 19: Topics in Applied Physics. Springer-Verlag, 2nd ed. 1980.

Saleh B.E.A. Photoelectron Statistics. Springer-Verlag, 1978.

Rose A. Concepts in Photoconductivity and Allied Problems. Wiley, 1963; Krieger, reissued 1978.

Willardson R.K., Beer A.C., eds. Semiconductors and Semimetals. Vol. 12: Infrared Detectors II. Academic Press, 1977.

Hudson Jr. R.D., Hudson J.W., eds. Benchmark Papers in Optics/2: Infrared Detectors. Dowden, Hutchinson & Ross, 1975.

Willardson R.K., Beer A.C., eds. Semiconductors and Semimetals. Vol. 5: Infrared Detectors. Academic Press, 1970.

Sommer A.H. Photoemissive Materials: Preparation, Properties and Uses. Wiley, 1968; Krieger; reissued 1980.

СТАТЬИ

Special issue on photodetectors. IEEE Lasers & Electro-Optics Society News. Vol. 20, № 5, 2006.

Ramirez D.A., Hayat M.M., Karve G., Campbell J.C., Torres S.N., Saleh B.E.A., Teich M.C. Detection Efficiencies and Generalized Breakdown Probabilities for Nanosecond-Gated Near Infrared Single-Photon Avalanche Photodiodes. IEEE Journal of Quantum Electronics. Vol. 42, 2006. P. 137—145.

Richards P.L., McCreight C.R. Infrared Detectors for Astrophysics. Physics Today. Vol. 58, № 2, 2005. P. 41—47.

Rogalski A. HgCdTe Infrared Detector Material: History, Status and Outlook. Reports on Progress in Physics. Vol. 68, 2005. P. 2267—2336.

Piotrowski J., Rogalski A. Uncooled Long Wavelength Infrared Photon Detectors. Infrared Physics and Technology. Vol. 46, 2004. P. 115—131.

Rogalski A., ed. Selected Papers on Infrared Detectors: Developments. SPIE Optical Engineering Press (Milestone Series. Vol. 179), 2004.

Issue on photodetectors and imaging. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 10, № 4, 2004.

Kang M.G., ed. Selected Papers on CCD and CMOS Imagers. SPIE Optical Engineering Press (Milestone Series. Vol. 177), 2003.

Aull B.F., Loomis A.H., Young D.J., Heinrichs R.M., Felton B.J., Daniels P.J., Landers D.J. Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging. Lincoln Laboratory Journal. Vol. 13, № 2, 2002. P. 335—350.

Hayat M.M., Kwon O.-H., Wang S., Campbell J.C., Saleh B.E.A., Teich M.C. Boundary Effects on Multiplication Noise in Thin Heterostructure Avalanche Photodiodes: Theory and Experiment. IEEE Transactions on Electron Devices. Vol. 49, 2002. P. 2114—2123.

Teich M.C., Saleh B.E.A. Branching Processes in Quantum Electronics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, 2000. P. 1450—1457.

Yuan P., Wang S., Sun X., Zheng X.G., Holmes Jr. A.L., Campbell J.C. Avalanche Photodiodes with an Impact-Ionization-Engineered Multiplication Region. IEEE Photonics Technology Letters. Vol. 12, 2000. P. 1370—1372.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Smith A., ed. Selected Papers on Photon Counting Detectors. SPIE Optical Engineering Press (Milestone Series. Vol. 143), 1998.

Rogalski A., ed. Selected Papers on Semiconductor Infrared Detectors. SPIE Optical Engineering Press (Milestone Series. Vol. 66), 1992.

Joshi N.V., ed. Selected Papers on Photoconductivity. SPIE Optical Engineering Press (Milestone Series. Vol. 56), 1992.

Hayat M.M., Saleh B.E.A., Teich M.C. Effect of Dead Space on Gain and Noise of Double-Carrier-Multiplication Avalanche Photodiodes. IEEE Transactions on Electron Devices. Vol. 39, 1992. P. 546—552.

Teich M.C., Matsuo K., Saleh B.E.A. Excess Noise Factors for Conventional and Superlattice Avalanche Photodiodes and Photomultiplier Tubes. IEEE Journal of Quantum Electronics. Vol. QE-22, 1986. P. 1184—1193.

Sclar N. Properties of Doped Silicon and Germanium Infrared Detectors. Progress in Quantum Electronics. Vol. 9, 1984. P. 149—257.

Chin R., Holonyak Jr. N., Stillman G.E., Tang J.Y., Hess K. Impact Ionization in Multilayered Heterojunction Structures. Electronics Letters. Vol. 16, 1980. P. 467—469.

Bratt P.R. Impurity Germanium and Silicon Infrared Detectors. In Semiconductors and Semimetals. Vol. 12: Infrared Detectors II. R.K. Willardson, A.C. Beer, eds. Academic Press, 1977, P. 33—142.

Melchior H. Demodulation and Photodetection Techniques. In F.T. Arecchi, E.O. Schulz-Dubois, eds. Laser Handbook. Vol. 1. North-Holland, 1972. P. 725—835.

Spicer W.E., Wooten F. Photoemission and Photomultipliers. Proceedings of the IEEE. Vol. 51, 1963. P. 1119—1126.

Shockley W., Pierce J.R. A Theory of Noise for Electron Multipliers. Proceedings of the IRE. Vol. 26, 1938. P. 321—332.

Задачи

К РАЗДЕЛУ 18.1

1. **Влияние отражения на квантовую эффективность.** Определите множитель $1 - \mathcal{R}$ в выражении для квантовой эффективности при нормальном падении и падении под углом 45° неполяризованного светового пучка из воздуха на границу Si, GaAs и InSb (см. разд. 6.2 и табл. 16.5).

2. **Токовая чувствительность**. Найдите максимальную токовую чувствительность идеального (с квантовой эффективностью и усилением, равными единице) полупроводникового фотодетектора, изготовленного из:

а) Si;

б) GaAs;

в) InSb.

3. **Время отклика**. Обратимся к рис. 18.6 и предположим, что фотон генерирует электронно-дырочную пару в положении $x = w/3$, что $v_e = 3v_h$ (в полупроводниках обычно скорость электронов больше, чем дырок) и что носители рекомбинируют на контактах. Для каждого носителя найдите величину токов i_e и i_h и их продолжительность τ_e и τ_h. Выразите результат через e, w и v_e. Проверьте, что полный заряд, индуцированный в цепи, равен e. Для $v_e = 6 \cdot 10^7$ см/с и $w = 10$ мкм изобразите графически временную зависимость токов.

4. **Токовый отклик при равномерном освещении**. Рассмотрим полупроводниковый материал (как на рис. 18.6), облучаемый при $t = 0$ импульсом света, который генерирует N электронно-дырочных пар, равномерно распределенных между 0 и w. Пусть скорость электронов и дырок в материале составляет v_e и v_h соответственно. Покажите, что дырочный ток можно записать как

$$i_h(t) = \begin{cases} -\dfrac{Nev_h^2}{w^2}t + \dfrac{Nev_h}{w}, & 0 \le t \le \dfrac{w}{v_h}; \\ 0, & \text{в остальных случаях,} \end{cases}$$

а электронный ток как

$$i_e(t) = \begin{cases} -\dfrac{Nev_e^2}{w^2}t + \dfrac{Nev_e}{w}, & 0 \le t \le \dfrac{w}{v_e}; \\ 0, & \text{в остальных случаях,} \end{cases}$$

и тогда полный ток равен

$$i(t) = \begin{cases} \dfrac{Ne}{w}\left[(v_h + v_e) - \dfrac{1}{w}(v_h^2 + v_e^2)t\right], & 0 \le t \le \dfrac{w}{v_e}; \\ \dfrac{Nev_h}{w}\left[1 - \dfrac{v_h}{w}t\right], & \dfrac{w}{v_e} \le t \le \dfrac{w}{v_h}. \end{cases}$$

Различные токи показаны на рис. 18.7. Проверьте, что электроны и дырки вносят по заряду $Ne/2$ во внешнюю цепь, так что полный генерируемый заряд равен Ne.

5. **Двухфотонные детекторы**. Рассмотрим пучок фотонов с энергией $h\nu$ и плотностью потока энергии ϕ (фотонов/см с), падающий на полупроводниковый детектор с шириной запрещенной зоны $h\nu \le E_g \le 2h\nu$, так что энергии одного фотона недостаточно для генерации электронно-дырочной пары. Од-

нако такой энергии хватает у двух электронов, поглощаемых в одном акте. Пусть плотность тока, индуцируемого в таком детекторе, выражается как $J_p = \zeta\phi^2$, где ζ — ностоянная. Покажите, что токовая чувствительность для двухфотонного детектора (А/Вт) выражается формулой

$$\Re = \left(\frac{\zeta}{hc_0}\right)^2 \frac{\lambda_0^2 P}{A},$$

где P — оптическая мощность; A — освещаемая площадь детектора. Объясните физически нропорциональность результата величинам λ_0^2 и P/A.

К РАЗДЕЛУ 18.2

1. **Цепь с фотопроводником.** Фотопроводящий детектор часто соединяют последовательно с сопротивлением нагрузки R и источником постоянного напряжения V. Измеряется напряжение V_p на сопротивлении нагрузки. Если нроводимость детектора пропорциональна оптической мощности P, постройте график зависимости V_p от P. При каких условиях эта зависимость линейна?

2. **Фотопроводимость.** Концентрация носителей заряда в собственном Si равна $n_i = 1{,}5 \cdot 10^{10}$ см$^{-3}$, а время рекомбинации $\tau = 10$ мкс. Материал поглощает свет с плотностью оптической мощности 1 мВт/см^3 на длине волны $\lambda_0 = 1$ мкм. Определите процентный рост его проводимости. Квантовая эффективность $\eta = 1/2$.

К РАЗДЕЛУ 18.3

1. **Квантовая эффективность фотодиодного детектора.** Для некоторого p—i—n-фотодиода импульс, содержащий $6 \cdot 10^{12}$ фотонов на длине волны $\lambda_0 = 1550$ мкм, порождает в среднем $2 \cdot 10^{12}$ электронов, собираемых на выводах прибора. Определите квантовую эффективность η и токовую чувствительность $\Re$ фотодиода на данной длине волны.

К РАЗДЕЛУ 18.4

1. **Квантовая эффективность ЛФД.** Обычный ЛФД с усилением $\bar{G} = 20$ работает на длине волны $\lambda_0 = 1550$ мкм. При условии, что его токовая чувствительность на этой длине волны равна $\Re = 12$ А/Вт, рассчитайте его квантовую эффективность η. Каков фототок на выходе нрибора, если на него падает поток $\Phi = 10^{12}$ фотонов/с с той же длиной волны?

2. **Усиление ЛФД.** Покажите, что ЛФД с ионизационным отношением $k = 1$, как для германия, имеет усиление $\bar{G} = 1/(1 - \alpha_e w)$, где α_e — коэффициент ионизации для электронов; w — ширина умножающего слоя.

Замечание. Уравнение (18.27) не дает правильного ответа для усиления нри $k = 1$.

К РАЗДЕЛУ 18.5

1. **Коэффициент избыточного шума для ЛФД с одним типом носителей.** Покажите, что ЛФД с чисто электронной инжекцией и без умножения дырок

($k = 0$) имеет коэффициент избыточного шума $F \approx 2$ для всех заметных значений усиления. С помощью (18.27) нокажите, что среднее усиление в этом случае равно $G = \exp(\alpha_e w)$. Рассчитайте токовую чувствительность кремниевого ЛФД для фотонов с энергией, равной ширине запрещенной зоны E_g, предполагая, что квантовая эффективность $\eta = 0{,}8$, а усиление $\bar{G} = 70$. Найдите коэффициент избыточного шума для кремниевого ЛФД с двумя типами носителей при $k = 0{,}01$. Сравните результат с $F \approx 2$, полученным в пределе одного типа носителей.

2. **Усиление многослойного ЛФД.** С помощью вероятностного закона Бернулли покажите, что среднее усиление многослойного ЛФД, такого как показан на рис. 18.32, $\bar{G} = (1 + P)^l$, где P — вероятность ударной ионизации на каждом шаге; l — число шагов. Покажите, что результат сводится к обычному ЛФД при $P \to 0$ и $l \to \infty$.

3. **Коэффициент избыточного шума для однокаскадного фотоумножителя.** Выведите выражение для коэффициента избыточного шума F однокаскадного фотоумножителя, считая, что число электронов вторичной эмиссии на один первичный электрон распределено по Пуассону со средним значением δ.

4. **Коэффициент избыточного шума для фотопроводящего детектора.** В разд. 18.2 было показано, что усиление фотопроводящего детектора равно $G = \tau/\tau_e$, где τ — рекомбинационное время жизни электронно-дырочной пары; τ_e — время пробега электрона через образец. На самом деле G — случайная величина, поскольку τ можно рассматривать как случайную переменную. Покажите, что экспоненциальная плотность распределения вероятности случайного времени рекомбинации

$$P(t) = \frac{1}{\bar{\tau}} \exp\left(-\frac{\tau}{\bar{\tau}}\right)$$

приводит к коэффициенту избыточного шума $F = 2$. Это подтверждает тот факт, что **шум генерации—рекомбинации** фотопроводника ухудшает ОСШ в два раза.

5. **Ширина полосы *RC*-цепи.** Используя определение ширины полосы, даваемое выражением (18.46), покажите, что цепь с функцией отклика на импульсное воздействие

$$h(t) = \frac{e}{\bar{\tau}} \exp\left(-\frac{t}{\bar{\tau}}\right)$$

имеет ширину полосы $B = 1/4\tau$. Какова ширина полосы RC-цепи? Определите ток теплового шума для сопротивления $R = 1$ кОм при $T = 300$ °K, соединенного с емкостью $C = 5$ пФ.

6. **Отношение сигнал—шум для приемника с ЛФД.** Во сколько раз изменится отношение сигнал—шум приемника на ЛФД со средним усилением $\bar{G} = 100$, если ионизационное отношение k возрастет от $k = 0{,}1$ до $0{,}2$? Считайте шум схемы пренебрежимо малым. Покажите, что если среднее усиление $\bar{G} \gg 1$ и $\bar{G} \gg 2(1 - k)/k$, то ОСШ приблизительно обратно пропорционально $\bar{G}$.

7. **Шум приемника на ЛФД.** Оптический приемник на основе ЛФД имеет следующие параметры: квантовую эффективность $\eta = 0{,}8$; среднее усиление $\bar{G} = 100$; ионизационное отношение $k = 0{,}5$; сопротивление нагрузки $R_L = 1$ кОм; ширину полосы $B = 100$ кГц; ток утечки и темновой ток равен 1 нА. Принимается оптический сигнал мощностью 10 нВт на длине волны $\lambda_0 = 0{,}87$ мкм. Определите среднеквадратичные значения различных шумовых токов и ОСШ. Считайте, что шумовые дисперсии темнового тока и тока утечки подчиняются тому же закону, что и шум фототока, и что приемник ограничен по сопротивлению.

8. **Оптимальное усиление ЛФД.** Приемник на *p—i—n*-фотодиоде имеет отношение дисперсии шума схемы к дисперсии фотоэлектронного шума, равное 100. Если вместо *p—i—n*-фотодиода используется ЛФД с ионизационным отношением $k = 0, 2$, найдите оптимальное среднее оптическое усиление, обеспечивающее максимум отношения сигнал—шум и соответствующее приращение этого отношения.

9. **Обнаружительная способность приемника.** Определите обнаружительную способность приемника (т. е. оптическую мощность, требуемую для достижения ОСШ $= 10^3$) для фотоприемника с квантовой эффективностью $\eta = 0{,}8$ на $\lambda_0 = 1300$ нм в цепи с шириной полосы $B = 100$ МГц при отсутствии шума схемы. Приемник выдает сигнал в виде электрического тока i.

10. **Сравнение шумовых свойств трех фотодетекторов.** Рассмотрим три фотодетектора, соединенных последовательно с сопротивлением нагрузки 50 Ом при 77 °K (температура жидкого азота) для использования в оптической системе с шириной полосы 1 ГГц:

1) *p—i—n*-фотодиод с квантовой эффективностью $\eta = 0{,}9$;

2) ЛФД с квантовой эффективностью $\eta = 0{,}6$, усилением $\bar{G} = 100$ и ионизационным отношением $k = 0$;

3) 10-каскадный фотоумножитель с квантовой эффективностью $\eta = 0{,}3$, общим средним усилением $\bar{G} = 4^{10}$ и дисперсией общего усиления

$$\sigma_G^2 = \frac{\bar{G}^2}{4}.$$

а. Для каждого детектора найдите ОСШ фототока, если детектор освещается потоком фотонов 10^{10} с$^{-1}$.

б. Для каких из трех приборов сигнал является обнаружимым?

11. **Зависимость обнаружительной способности приемника от длины волны.** Обнаружительная способность идеального приемника (с квантовой эффективностью единица и без шумов схемы) на длине волны 870 нм составляет −76 дБм. Какова обнаружительная способность на 1300 нм, если приемник работает с той же скоростью передачи данных?

12. **Частоты появления ошибочных битов.** *p—i—n*-Фотодиод, ограниченный по квантовому шуму (никаких шумов, кроме фотонного шума), с квантовой эффективностью $\eta = 1$ ошибочно принимает присутствующий оптический сигнал на $\lambda_0 = 870$ нм мощностью P (бит 1) за отсутствующий сигнал (бит 0) с вероятностью 10^{-10}. Какова вероятность ошибки при каждом из следующих новых условий:

а) длина волны равна $\lambda_0 = 1300$ нм;

б) мощность в два раза больше, остальное так же;

в) те же условия, но квантовая эффективность $\eta = 0{,}5$;

г) те же условия, но используется идеальный ЛФД с $\eta = 1$ и усилением $G = 100$ (шум усиления отсутствует);

д) как в случае (г), но ЛФД имеет коэффициент избыточного шума $F = 2$.

13. **Обнаружительная способность приемника с амплитудной модуляцией.** Детектор с токовой чувствительностью $\mathfrak{R}$ (А/Вт), шириной полосы B и пренебрежимо малым шумом схемы измеряет модулированную оптическую мощность $P(t) = P_0 + P_s \cos(2pft)$, причем $f < B$. При условии $P_0 \gg P_s$ выведите выражение для минимальной амплитуды модуляции P_s, которую можно измерить с $\text{ОСШ}_0 = 30$ дБ. Каково влияние фоновой мощности P_0 на минимально обнаружимое значение сигнала P_s?

14. **Обнаружительная способность приемника в режиме счета фотонов.** Фотоприемник с квантовой эффективностью $\eta = 0{,}5$ считает фотоэлектроны, приходящие в течение последовательных интервалов времени длительностью $T = 1$ мкс. Определите обнаружительную способность приемника (среднее число фотонов, необходимое для обеспечения ОСШ $= 10^3$) в предположении пуассонова распределения числа фотонов. Считая длину волны света равной $\lambda_0 = 870$ нм, найдите соответствующую оптическую мощность. Если эта мощность принимается, то какова вероятность регистрации нулевых отсчетов детектором?

15. **Фотоумножитель с одним динодом.** Рассмотрим фотоэлектронный умножитель с квантовой эффективностью $\eta = 1$ и единственным динодом. На катод падает свет от гипотетического источника фотонов, обеспечивающего вероятность наблюдения n фотонов за счетное время $T = 1{,}3$ нс, выражаемую как

$$p(n) = \begin{cases} \dfrac{1}{2}, & n = 0, 1; \\ 0, & \text{в остальных случаях.} \end{cases}$$

Когда один электрон ударяется о динод, испускаются и следуют к аноду либо два, либо три вторичных электрона. Распределение усиления $P(G)$ дается выражением

$$P(G) = \begin{cases} \dfrac{1}{3}, & G = 2; \\ \dfrac{2}{3}, & G = 3; \\ 0, & \text{в остальных случаях.} \end{cases}$$

Таким образом, вероятность эмиссии трех электронов вдвое больше, чем двух.

а. Рассчитайте ОСШ для входного числа фотонов и сравните результат с таковым для распределения Пуассона с тем же средним значением.

б. Найдите распределение вероятностей для числа фотоэлектронов $p(m)$ и ОСШ для числа фотоэлектронов.
в. Найдите среднее усиление $\langle G \rangle$ и среднеквадратичное усиление $\langle G^2 \rangle$.
г. Найдите коэффициент избыточного шума F.
д. Найдите средний анодный ток $\overline{i}$ в цепи с шириной полосы $B = 1/2T$.
е. Найдите токовую чувствительность фотоумножителя при длине волны $\lambda_0 = 1550$ нм.
ж. Объясните, почему (18.53) для σ_i^2 неприменимо.

ГЛАВА 19

АКУСТООПТИКА





Сэр Уильям Генри Брэгг (1862—1942) и сэр Уильям Лоуренс Брэгг (1890—1971), отец и сын, получившие в 1915 г. Нобелевскую премию за исследования дифракции света на периодических структурах, таких, как создаваемые звуком.

Показатель преломления оптической среды меняется в присутствии звука. Звук меняет влияние среды на свет, т. е. *звук может управлять светом* (рис. 19.1). Множество полезных устройств основано на **акустоонтическом эффекте**; к ним относятся модуляторы, переключатели, дефлекторы, фильтры, преобразователи частоты и спектроанализаторы.

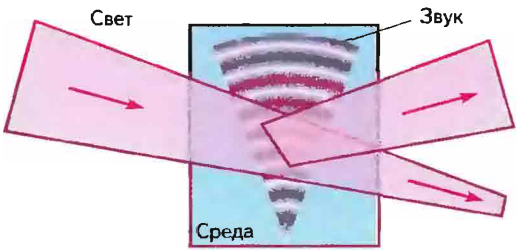


Рис. 19.1. Звук изменяет влияние оптической среды на свет

Звук представляет собой динамическую деформацию, включающую молекулярные колебания, которая принимает форму волн, распространяющихся с

характерной для среды скоростью (скоростью звука). Например, гармоническая плоская волна сжатий и разрежений в газе схематически изображена на рис. 19.2. В тех местах, где среда сжата, плотность выше и показатель преломления больше; там, где среда разрежена, ее плотность и показатель преломления меньше. В твердых телах звук включает в себя колебания молекул вокруг положений их равновесия, меняющие оптическую поляризуемость и, следовательно, показатель преломления.

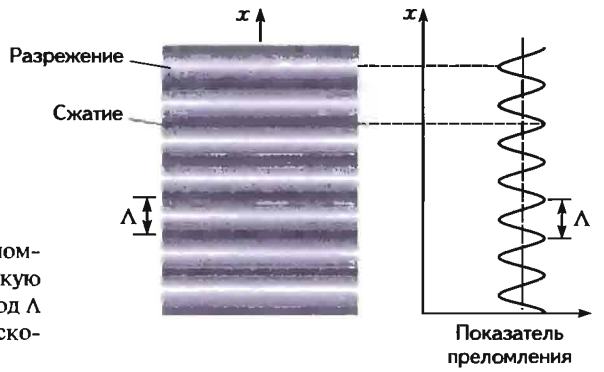


Рис. 19.2. Изменение показателя преломления, сопровождающее гармоническую звуковую волну. Картина имеет период Λ (длину волны звука) и движется со скоростью звука

Акустическая волна создает возмущение показателя преломления в виде волны. Среда приобретает *динамический* градиент показателя преломления, т. е. становится неоднородной средой с плавно-слоистым распределением показателя преломления, зависящим от времени. Акустооптическая теория имеет дело с возмущением показателя преломления, вызванным звуком, и с распространением света сквозь эту возмущенную среду с зависящей от времени неоднородностью.

Распространение света в статических (в нротивоноложность меняющимся во времени) неоднородных (градиентных) средах частично обсуждалось в гл. 1 и 2 (см. разд. 1.3 и подразд. 2.4.3). Поскольку оптические частоты намного выше акустических, изменения показателя преломления в среде, возмущаемой звуком, происходят очень медленно но сравнению с оптическим периодом. Следовательно, имеется два существенно различных временных масштаба — для света и для звука. Поэтому можно использовать адиабатический подход, в котором оптическая задача распространения решается отдельно в каждый момент времени на протяжении сравнительно медленного акустического колебания, нричем материал в каждый такой момент времени рассматривается как статическая (замороженная) неоднородная среда. В этом квазистационарном приближении акустооптика становится оптикой неоднородной среды (обычно нериодической), которая управляется светом.

Простейшим видом взаимодействия света и звука является частичное отражение оптической плоской волны от слоистой структуры из параллельных плоскостей, представляющих изменение показателя преломления, созданное плос-

кой акустической волной (рис. 19.3). Набор параллельных отражателей, разделенных длиной волны звука Λ, будет отражать свет, если угол падения θ удовлетворяет условию Брэгга для конструктивной интерференции

$$\sin\theta = \frac{\lambda}{2\Lambda}, \tag{19.1}$$

Условие Брэгга

где λ — длина волны света в среде (см. упражнение 2.11). Этот вид взаимодействия света со звуком известен как **дифракция Брэгга**, отражение Брэгга или брэгговское рассеяние. Прибор, где используется дифракция Брэгга, называется брэгговским отражателем, брэгговским дефлектором или **ячейкой Брэгга**.

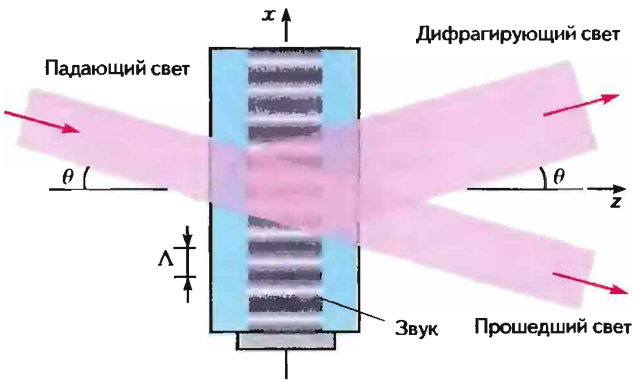


Рис. 19.3. Дифракция Брэгга: плоская акустическая волна действует как частичный отражатель света (делитель пучка), когда угол падения θ удовлетворяет условию Брэгга

О данной главе

Ячейки Брэгга нашли многочисленные приложения в фотонике. Эта глава посвящена их свойствам. В разд. 19.1 представлена простая теория оптики брэгговских отражателей для линейных недиспергирующих сред. Анизотропные свойства среды и поляризационная природа света звука игнорируются. Хотя теория основана на волновой оптике, дана простая квантовая интерпретация ее результатов. В разд. 19.2 обсуждается использование ячеек Брэгга для модуляции и сканирования света. Раздел 19.3 дает краткое введение в эффекты анизотропии и поляризации в акустооптике.

19.1. ВЗАИМОДЕЙСТВИЕ СВЕТА И ЗВУКА

В данном разделе описано влияние скалярной акустической волны на скалярную оптическую волну. Вначале рассмотрим оптические и акустические плоские волны, а затем изучим взаимодействие оптических и акустических пучков.

19.1.1. Дифракция Брэгга

Рассмотрим акустическую плоскую волну, распространяющуюся в среде в направлении x со скоростью v_s, частотой f и длиной волны $\Lambda = v_s/f$. Деформация (относительное смещение) в точке x в момент времени t имеет вид

$$s(x,\ t) = S_0 \cos(\Omega t - qx), \tag{19.2}$$

где S_0 — амплитуда; $\Omega = 2\pi f$ — угловая частота; $q = 2\pi/\Lambda$ — волновое число. Акустическая интенсивность (Вт/м^2):

$$I_s = \frac{1}{2}\rho v_s^2 S_0^2, \tag{19.3}$$

где ρ — плотность среды.

Среда предполагается оптически прозрачной, а показатель преломления в отсутствие звука равен n. Деформация $s(x, t)$ создает пропорциональное возмущение показателя преломления, аналогичное эффекту Поккельса (20.4),

$$\Delta n(x,\ t) = -\frac{1}{2}\mathrm{p}n^3 s(x,\ t), \tag{19.4}$$

где p — феноменологический безразмерный коэффициент, называемый **фотоупругой постоянной** (или коэффициентом оптической чувствительности по деформациям). Знак «минус» указывает на то, что положительная деформация (растяжение) приводит к уменьшению показателя преломления. Как следствие, среда приобретает зависящее от времени неоднородное распределение показателя преломления в виде волны

$$n(x,\ t) = n - \Delta n_0 \cos(\Omega t - qx) \tag{19.5}$$

с амплитудой

$$\Delta n_0 = \frac{1}{2}\mathrm{p}n^3 S_0. \tag{19.6}$$

Поставляя (19.3) в (19.6), находим, что изменение показателя преломления пропорционально квадратному корню из акустической интенсивности

$$\boxed{\Delta n_0 = \sqrt{\frac{1}{2}\mathcal{M} I_s},} \tag{19.7}$$

где $\mathcal{M}$ — материальный параметр, характеризующий эффективность, с которой звук меняет показатель преломления.

$$\boxed{\mathcal{M} = \frac{\mathrm{p}^2 n^6}{\rho v_s^3}.} \tag{19.8}$$

Величина $\mathcal{M}$ является показателем силы акустооптического эффекта в материале.

Пример 19.1

Показатель качества акустооптического эффекта

В особо плотном стекле типа флинт $r = 6{,}3 \cdot 10^3$ кг/м^3, $v_s = 1{,}3$ км/с, $n = 1{,}92$, $p = 0{,}25$, так что $\mathcal{M} = 1{,}67\ \ 10^{-14}$ м^2/Вт. Акустическая волна с интенсивностью 10 Вт/см^2 создаст волну показателя преломления с амплитудой $\Delta n_0 = 2{,}89 \cdot 10^{-5}$.

Рассмотрим теперь оптическую плоскую волну, распространяющуюся в этой среде с частотой ν, угловой частотой $\omega = 2\pi\nu$, длиной волны в свободном пространстве $\lambda_0 = c_0/\nu$, длиной волны в невозмущенной среде $\lambda = \lambda_0/n$, соответствующей волновому числу $k = n\omega/c$, и волновым вектором $\boldsymbol{k}$, лежащим в плоскости x—z и образующим угол θ с осью z, как показано на рис. 19.4.

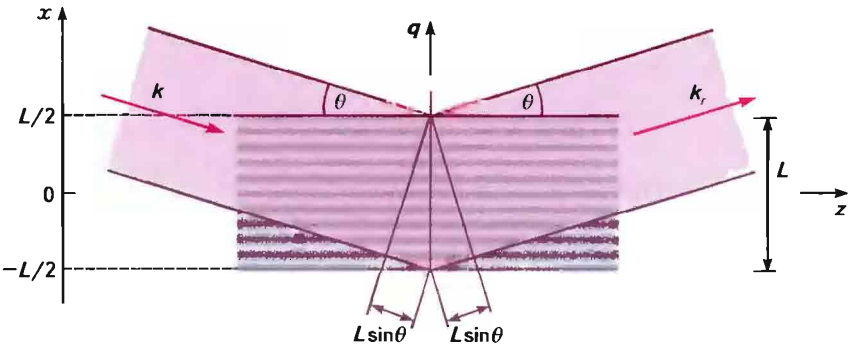


Рис. 19.4. Отражения от слоев неоднородной среды

Поскольку акустическая частота f в типичном случае намного меньше оптической частоты ν (по крайней мере, на пять порядков), для описания взаимодействия света со звуком можно принять адиабатическое приближение, а именно, будем рассматривать показатель преломления как статическую «замороженную» синусоидальную функцию

$$n(x) = n - \Delta n_0 \cos(qx - \varphi), \tag{19.9}$$

где φ — фиксированная фаза. Далее определим отраженный от данной (градиентной) структуры свет и проследим за его медленным изменением во времени, полагая $\varphi = \Omega t$.

Для определения амплитуды отраженной волны разделим среду на бесконечно тонкие плоские слои, перпендикулярные оси x. Падающая оптическая волна частично отражается от каждого слоя из-за изменения показателя преломления. Предположим, что отражательная способность достаточно мала, так что свет, отражаемый от одного слоя, приблизительно сохраняет свою первоначальную амплитуду (не ослабляется) по мере проникновения через последовательные слои в среду.

Если $\Delta r = (\mathrm{d}r/\mathrm{d}x)\Delta x$ — приращение отражательной способности по комплексной амплитуде в бесконечно тонком слое Δx с координатой x, то полная отражательная способность по комплексной амплитуде для полной длины L (см. рис. 19.4) представляет собой сумму малых приращений

$$r = \int_{L/2}^{L/2} e^{j2kx\sin\theta}\frac{\mathrm{d}r}{\mathrm{d}x}\mathrm{d}x. \tag{19.10}$$

Фазовый множитель $e^{j2kx\sin\theta}$ необходим, поскольку волна, отраженная в точке x, проходит путь короче на $2x\sin\theta$, что соответствует фазовому сдвигу $2kx\sin\theta$, относительно волны, отраженной в токс $x = 0$. Волновые числа падающей и отраженной волн полагаются одинаковыми.

Используя (19.9), запишем

$$\frac{\mathrm{d}r}{\mathrm{d}x} = \frac{\mathrm{d}r}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}x} = \frac{\mathrm{d}r}{\mathrm{d}n}q\Delta n_0\sin(qx-\varphi), \tag{19.11}$$

где производная $\mathrm{d}r/\mathrm{d}n$, которую, как мы покажем ниже, можно получить из формул Френеля для отражения, не зависит от x. Теперь подставим (19.11) в (19.10) и используем комплексное представление

$$\sin(qx-\varphi) = \frac{e^{j(qx\ \varphi)} - e^{-j(qx-\varphi)}}{2j},$$

тогда получаем

$$r = jr_0e^{j\varphi}\frac{1}{L}\int_{-L/2}^{L/2} e^{j(2k\sin\theta-q)x}\mathrm{d}x - jr_0e^{-j\varphi}\frac{1}{L}\int_{-L/2}^{L/2} e^{j(2k\sin\theta+q)x}\mathrm{d}x, \tag{19.12}$$

где

$$r_0 = \frac{1}{2}\Delta n_0qL\frac{\mathrm{d}r}{\mathrm{d}n}. \tag{19.13}$$

Выполняя интегрирование в (19.12) и подставляя $\varphi = \Omega t$, получаем

$$r = r_+ + r_-, \tag{19.14}$$

где

$$\boxed{r_\pm = \pm jr_0\,\mathrm{sinc}\left[(2k\sin\theta \pm q)\frac{L}{2\pi}\right]e^{\pm j\Omega t}}; \tag{19.15}$$

Коэффициент отражения по амплитуде

$$\mathrm{sinc}(x) = \frac{\sin\pi x}{\pi x}.$$

По причинам, которые скоро станут ясны, члены r_+ и r_- называются коэффициентами отражения с положительным (отрицательным) сдвигом. Коэффици-

ент отражения с положительным сдвигом r_+ имеет максимальное значение при $2k\sin\theta = q$, а с отрицательным r_- — при $2k\sin\theta = -q$. Если L достаточно велико, эти максимумы острые, так что уже небольшое отклонение от углов $\theta = \pm\arcsin(q/2k)$ делает соответствующий член пренебрежимо малым. Таким образом, только один из двух членов одновременно может иметь заметное значение, зависящее от угла θ. Вначале рассмотрим отражение с положительным сдвигом, $2k\sin\theta \approx q$, когда отражение с отрицательным сдвигом пренебрежимо мало, а затем прокомментируем противоположный случай.

Условие Брэгга

Функция sinc в (19.15) имеет максимальное значение 1, когда ее аргумент равен нулю, т. е. когда $q = 2k\sin\theta$ в случае отражения с положительным сдвигом. Это происходит при $\theta = \theta_B$, где $\theta_B = \arcsin(q/2k)$ — **угол Брэгга**. Поскольку $q = 2\pi/\Lambda$, а $k = 2\pi/\lambda$,

$$\boxed{\sin\theta_B = \frac{\lambda}{2\Lambda}.} \qquad (19.16)$$

Угол Брэгга

Угол Брэгга представляет собой угол, для которого волны, отраженные от плоскостей, разделенных акустической длиной волны Λ, имеют фазовый сдвиг 2π и интерферируют конструктивно [см. упражнение 2.11 и формулу (7.47)].

Пример 19.2

Угол Брэгга

Акустооптическая ячейка изготовлена из стекла флинт, в котором скорость звука равна $v_s = 3$ км/с, а показатель преломления $n = 1{,}95$. Угол Брэгга для отражения оптической волны с вакуумной длиной волны $\lambda_0 = 633$ нм ($\lambda = \lambda_0/n = 325$ нм) от акустической волны с частотой $f = 100$ МГц ($\Lambda = v_s/f = 30$ мкм) равен $\theta_B = 5{,}4$ мрад $\approx 0{,}31°$ Этот угол внутренний (в толще среды). Если ячейка помещена в воздух, то θ_B соответствует внешнему углу $\theta_B' \approx n\theta_B = 0{,}61$. Звуковая волна с частотой в десять раз больше ($f = 1$ ГГц) соответствует углу Брэгга $\theta_B = 3{,}1$.

Условие Брэгга может быть сформулировано также как простое соотношение между волновыми векторами звуковой и оптической волн. Если $\boldsymbol{q} = (q, 0, 0)$, $\boldsymbol{k} = (-k\sin\theta, 0, k\cos\theta)$ и $\boldsymbol{k}_r = (-k\sin\theta, 0, k\cos\theta)$ — компоненты волновых векторов звуковой волны, падающей и отраженной оптических волн соответственно, то условие $q = 2k\sin\theta_B$ эквивалентно векторному соотношению

$$\boxed{\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q},} \qquad (19.17)$$

иллюстрируемое векторной диаграммой на рис. 19.5.

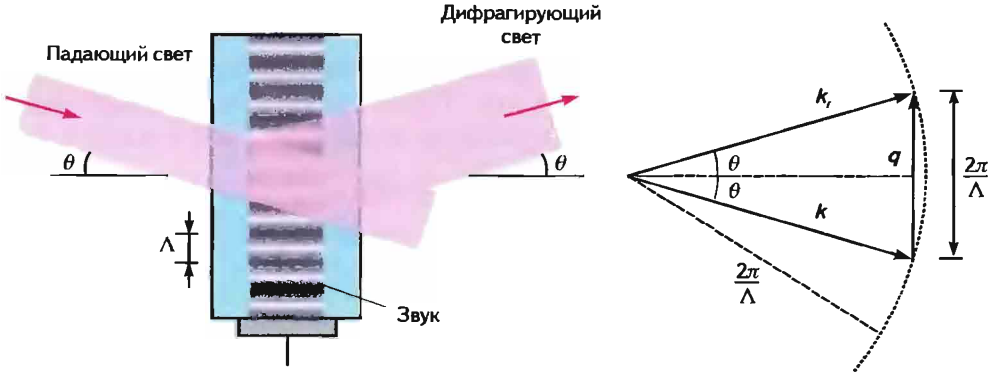


Рис. 19.5. Условие Брэгга $\sin\theta_B = q/2k$ эквивалентно векторному соотношению $\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}$

Допустимое отклонение от условия Брэгга

Зависимость комплексного амплитудного коэффициента отражения от угла θ определяется симметричной функцией

$$\operatorname{sinc}\left(q - 2k\sin\theta\frac{L}{2\pi}\right) = \operatorname{sinc}\left(\sin\theta - \sin\theta_B\frac{2L}{\lambda}\right)$$

в выражении (19.15). Эта функция достигает своего пикового значения при $\theta = \theta_B$ и быстро убывает при небольшом отклонении θ от θ_B. Когда $\sin\theta - \sin\theta_B = \lambda/2L$, функция sinc достигает своего первого нуля и отражение исчезает (рис. 19.6). Поскольку θ_B обычно очень мало, $\sin\theta_B \approx \theta_B$ и отражение исчезает при отклонении от угла Брэгга примерно на $\theta - \theta_B \approx \lambda/2L$. Поскольку L в типичном случае много больше λ, это предельно малая угловая величина. Такое резкое падение отражения при малых отклонениях от угла Брэгга происходит из-за деструктивной интерференции парциальных отраженных волн.

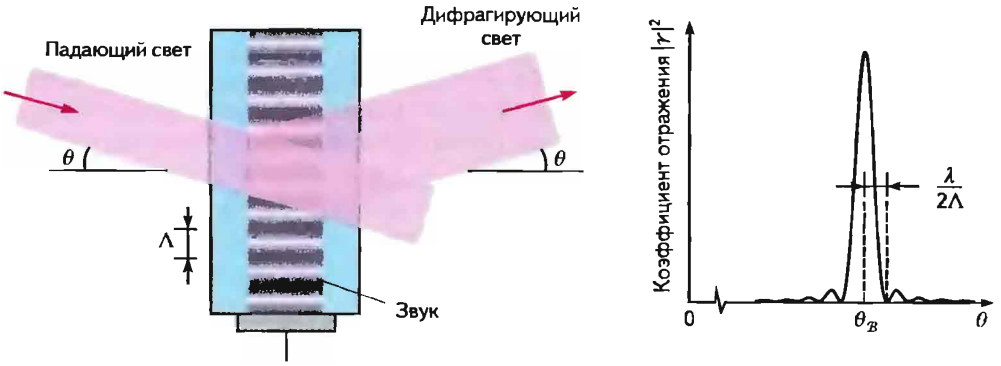


Рис. 19.6. Зависимость коэффициента отражения $|r|^2$ от угла θ. Максимальное отражение происходит под углом Брэгга $\theta_B = \arcsin(\lambda/2\Lambda)$

Доплеровский сдвиг

В соответствии с (19.15) комплексный амплитудный коэффициент отражения r_+ пропорционален $\exp(j\Omega t)$. Так как угловая частота падающего света равна ω [т. е. $E \propto \exp(j\omega t)$], отраженная волна $E_r = r_+ E \propto \exp[j(\omega + \Omega)t]$ имеет угловую частоту

$$\boxed{\omega_r = \omega + \Omega.} \qquad (19.18)$$

Доплеровский сдвиг

Таким образом, процесс отражения сопровождается сдвигом частоты, равным частоте звука. Его можно объяснить как следствие эффекта Доплера (см. упражнение 2.7 и подразд. 13.3.4). Падающий свет отражается от поверхностей, движущихся со скоростью v_s. С учетом доплеровского сдвига его частота становится равной

$$\omega_r = \omega\left(1 + 2v_s \sin\frac{\theta}{c}\right),$$

где $v_s \sin\theta$ — составляющая скорости движения поверхностей в направлении падающей и отраженной волн. С помощью соотношений

$$\sin\theta = \frac{\lambda}{2\Lambda}; \quad v_s = \frac{\Lambda\Omega}{2\pi}; \quad c = \frac{\lambda\omega}{2\pi}$$

получается (19.18). Доплеровский сдвиг равен частоте звука.

Поскольку $\Omega \ll \omega$, частоты падающего и отраженного света приближенно равны (с типичной относительной ошибкой 1 к 10^5). Длина волны этих волн также приблизительно одинакова. При записи (19.10) мы явно использовали это предположение, считая волновое число k одинаковым для обеих волн. При построении векторной диаграммы на рис. 19.5 также предполагалось, что векторы $\boldsymbol{k}_r$ и $\boldsymbol{k}$ имеют примерно одинаковую длину $n\omega/c_0$.

Максимальный коэффициент отражения

Коэффициент отражения $\mathcal{R} = |r_+|^2$ определяется как отношение интенсивности отраженной оптической волны к интенсивности падающей волны. Для угла Брэгга $\theta = \theta_B$ формула (19.15) дает $\mathcal{R} = |r_0|^2$. Подставляя сюда выражение (19.13), находим

$$\mathcal{R} = \frac{1}{4}\Delta n_0^2 q^2 L^2 \left|\frac{dr}{dn}\right|^2 \qquad (19.19)$$

Выражение для производной dr/dn можно получить, если с помощью формул Френеля (см. разд. 6.2) выразить вклады в комплексный амплитудный коэффициент отражения Δr через разность значений показателя преломления Δn между двумя соседними слоями. Для TE (ортогональной) поляризации использует-

ся (6.34) с $n_1 = n + \Delta n$, $n_2 = n$, $\theta_1 = 90° - \theta$, а также закон Снелла $n_1 \sin\theta_1 = n_2 \sin\theta_2$, из которого находится θ_2. Пренебрегая членами второго порядка по Δn, получаем $\Delta r = -\Delta n/2n\sin^2\theta$, так что

$$\frac{dr}{dn} = \frac{-1}{2n\sin^2\theta}. \tag{19.20}$$

Уравнение (6.35) аналогично используется для TM (параллельной) поляризации, что дает

$$\frac{dr}{dn} = \frac{-\cos 2\theta}{2n\sin^2\theta}. \tag{19.21}$$

В большинстве акустооптических приборов θ очень мало и $\cos 2\theta \approx 1$, что делает формулу (19.20) применимой для обеих поляризаций.

Подстановка производной (19.20) в (19.21) и использование условия Брэгга

$$q = 2k\sin\theta = \frac{4\pi n\sin\theta}{\lambda_0}$$

даст

$$\mathcal{R} = \frac{\pi^2}{\lambda_0^2}\left(\frac{L}{\sin\theta}\right)\Delta n_0{}^2. \tag{19.22}$$

С помощью (19.7) приходим к заключению, что коэффициент отражения

$$\boxed{\mathcal{R} = \frac{\pi^2}{2\lambda_0^2}\left(\frac{L}{\sin\theta}\right)^2 \mathcal{M} I_s} \tag{19.23}$$

Коэффициент отражения

пропорционален интенсивности акустической волны I_s, материальному параметру $\mathcal{M}$, определенному выражением (19.8), и квадрату длины наклонного проникновения $L/\sin\theta$ света в среду с акустической волной.

Подставляя $\sin\theta = \lambda/2\Lambda$ в (19.23), получаем

$$\mathcal{R} = 2\pi^2 n^2 \frac{L^2\Lambda^2}{\lambda_0^4}\mathcal{M} I_s \tag{19.24}$$

Таким образом, коэффициент отражения пропорционален λ_0^4 (или прямо пропорционален ω^4). Зависимость эффективности рассеяния от четвертой степени оптической частоты типична для явлений рассеяния света.

Пропорциональность коэффициента отражения интенсивности звука вызывает вопрос: по мере роста интенсивности звука $\mathcal{R}$ может достичь единицы, а затем превысить ее. Отраженная волна оказывается интенсивнее падающей! Этот неприемлемый результат возник из-за нарушения условий применимости нашей приближенной теории. Предполагалось, что отражение от каждого слоя

столь мало, что не ослабляет падающую волну, отражающуюся от последовательных слоев. Очевидно, это предположение неверно, если интенсивность звуковой волны велика. В действительности имеет место процесс насыщения, гарантирующий тот факт, что величина $\mathcal{R}$ никогда не превысит единицы. Более точный анализ (см. подразд. 19.1.2), в котором учитывается ослабление падающей волны при каждом отражении, приводит к следующему выражению для коэффициента отражения:

$$\boxed{\mathcal{R}_e = \sin^2 \sqrt{\mathcal{R}},} \tag{19.25}$$

в котором $\mathcal{R}$ дается приближенным выражением (19.23), а $\mathcal{R}_e$ — точное значение. Данное соотношение иллюстрирует рис. 19.7. Очевидно, что при $\mathcal{R} \ll 1$ $\sin \sqrt{\mathcal{R}} \approx \sqrt{\mathcal{R}}$, так что $\mathcal{R}_e \approx \mathcal{R}$.

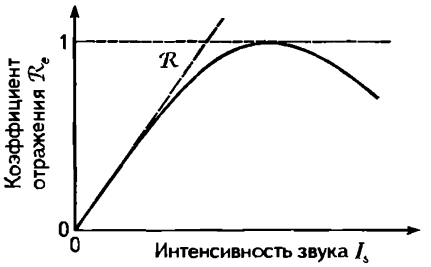


Рис. 19.7. Зависимость коэффициента отражения $\mathcal{R}_e$ брэгговского отражателя от интенсивности звука I_s. Когда I_s мало, $\mathcal{R} \approx \mathcal{R}_e$, что дает линейную функцию от I_s

Пример 19.3

Отражение от брэгговского рефлектора

Ячейка Брэгга изготовлена из очень плотного стекла флинт с материальным параметром $\mathcal{M} = 1{,}67 \ \ 10^{-14}$ м^2/Вт (см. пример 19.1). Если $\lambda_0 = 633$ нм (длина волны He—Ne-лазера), интенсивность звука $I_s = 10$ Вт/см^2, а длина пути света в звуке $L/\sin\theta = 1$ мм, то $\mathcal{R} = 0{,}0206$ и $\mathcal{R}_e = 0.0205$, так что приблизительно 2 % света отражается. Если интенсивность звука увеличить до 100 Вт/см, то $\mathcal{R} = 0{,}206$ и $\mathcal{R}_e = 0, 192$, так что коэффициент отражения увеличивается до 19 %.

Брэгговская дифракция со сдвигом вниз

При другой возможной геометрии брэгговской дифракции $2k\sin\theta = -q$. Это выполняется, когда угол θ отрицателен, т. е. падающая оптическая волна образует острый угол со звуковой волной, как показано на рис. 19.8. В этом случае коэффициент отражения со сдвигом вниз r_-, определяемый формулой (19.15), имеет максимальное значение, тогда как коэффициент отражения со сдвигом вверх r_+ пренебрежимо мал. Комплексный амплитудный коэффициент отражения выражается как

$$r_- = jr_0 e^{-j\Omega t} \tag{19.26}$$

В этом случае частота отраженной волны сдвигается вниз:

$$\omega_s = \omega - \Omega, \tag{19.27}$$

а волновые векторы световой и звуковой волн связаны равенством

$$\boldsymbol{k}_s = \boldsymbol{k} - \boldsymbol{q}, \tag{19.28}$$

которое иллюстрирует рис. 19.8. Равенство (19.28) выражает условие согласования фаз, гарантирующее синфазность волн, отраженных от каждого слоя. Сдвиг частоты вниз (19.27) согласуется с доплеровским сдвигом, поскольку световая и звуковая волны распространяются в одном направлении.

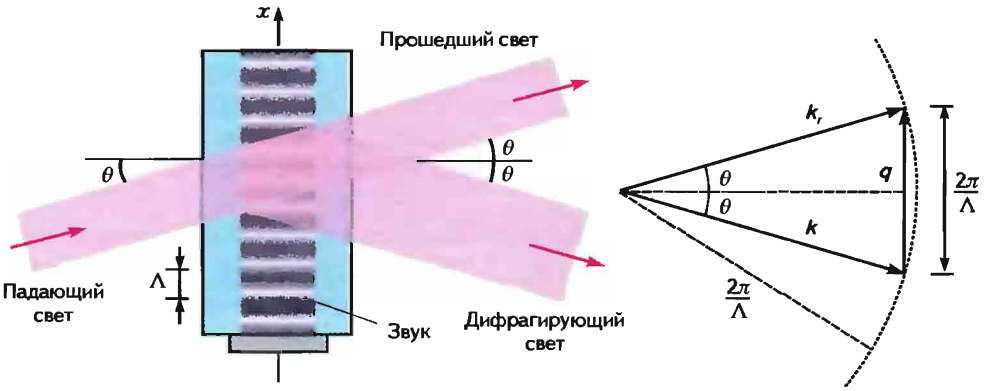


Рис. 19.8. Геометрия отражения света от звука с понижением частоты

Квантовая интерпретация

В соответствии с квантовой теорией света (см. гл. 12) оптическая волны с угловой частотой ω и волновым вектором $\boldsymbol{k}$ рассматривается как поток фотонов, каждый из которых имеет энергию $\hbar\omega$ и импульс $\hbar\boldsymbol{k}$. Аналогично, акустическая волна с угловой частотой Ω и волновым вектором $\boldsymbol{q}$ рассматривается как поток акустических квантов, называемых **фононами**, каждый из которых имеет энергию $\hbar\Omega$ и импульс $\hbar\boldsymbol{q}$.

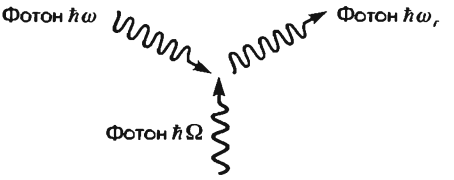


Рис. 19.9. Дифракция Брэгга: фотон комбинирует с фононом и рождает новый фотон с измененной энергией и импульсом

Взаимодействие света и звука происходит, когда фотон и фонон комбинируют с рождением нового фотона с суммарной энергией и импульсом. Падающий фотон с частотой ω и волновым вектором $\boldsymbol{k}$ взаимодействует с фононом частоты Ω с волновым вектором $\boldsymbol{q}$ и рождает новый фотон с частотой ω_r и волновым

вектором $\boldsymbol{k}_r$, как показано на рис. 19.9. Законы сохранения энергии и импульса требуют выполнения равенств $\hbar\omega_r = \hbar\omega + \hbar\Omega$, $\hbar\boldsymbol{k}_r = \hbar\boldsymbol{k} + \hbar\boldsymbol{q}$, откуда сразу следуют формула для доплеровского сдвига $\omega_r = \omega + \Omega$ и условие Брэгга $\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}$.

*19.1.2. Теория связанных волн

Дифракция Брэгга как процесс рассеяния

Как описано в подразд. 5.2.2, распространение света через однородную среду с медленно меняющимся возмущением показателя преломления Δn описывается волновым уравнением

$$\nabla^2\mathcal{E} - \frac{1}{c^2}\frac{\partial^2\mathcal{E}}{\partial t^2} \approx -S, \tag{19.29}$$

где

$$S = -\mu_0\frac{\partial^2\Delta\mathcal{P}}{\partial t^2} = -2\mu_0\varepsilon_0 n\frac{\partial^2}{\partial t^2}(\Delta n\mathcal{E}) \tag{19.30}$$

представляет собой радиационный источник, пропорциональный второй производной от произведения $\Delta n\mathcal{E}$ [см. (5.35)]. При дифракции Брэгга возмущение Δn создается звуковой волной, так что рассеивающий источник зависит как от акустического поля, так и от оптического поля $\mathcal{E}$, включающего как падающую, так и рассеянную волну.

Один из приближенных методов решения такой задачи рассеяния, называемый **первым борновским приближением**, основан на предположении, что источник рассеяния S создается падающей волной (а не действительно существующим полем). Зная источник рассеяния, можно решить волновое уравнение относительно рассеянного поля.

Предполагая, что падающий свет представляет собой плоскую волну

$$\mathcal{E} = \mathrm{Re}\left\{A\exp\left[j(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})\right]\right\} \tag{19.31}$$

и возмущение, вызванное акустической волной, — также плоская волна

$$\Delta n = -\Delta n_0\cos(\Omega t - \boldsymbol{q}\cdot\boldsymbol{r}), \tag{19.32}$$

подставим их в (19.30) и преобразуем члены в произведении $\Delta n\mathcal{E}$, что дает

$$S = -\left(\frac{\Delta n_0}{n}\right)\left(k_r^2\mathrm{Re}\left\{A\exp\left[j(\omega_r t - \boldsymbol{k}_r\cdot\boldsymbol{r})\right]\right\} + k_s^2\mathrm{Re}\left\{A\exp\left[j(\omega_s t - \boldsymbol{k}_s\cdot\boldsymbol{r})\right]\right\}\right), \tag{19.33}$$

где $\omega_r = \omega + \Omega$; $\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}$, $k_r = \omega_r/c$ и $\omega_s = \omega - \Omega$, $\boldsymbol{k}_s = \boldsymbol{k} - \boldsymbol{q}$, $k_s = \omega_s/c$. Таким образом, имеется два источника света с частотами $\omega \pm \mathrm{W}$ и волновыми векторами $\boldsymbol{k} \pm \boldsymbol{q}$, которые могут излучать волны брэгговского отражения со сдвигом частоты вверх или вниз. Отражение со сдвигом вверх происходит, если геометрия такова, что длина вектора $\boldsymbol{k} + \boldsymbol{q}$ равна $\omega_r/c \approx \omega/c$, как легко увидеть из

векторной диаграммы на рис. 19.5. Отражение со сдвигом вниз происходит, если вектор $\boldsymbol{k} - \boldsymbol{q}$ имеет длину $\omega_s/c \approx \omega/c$ (см. рис. 19.8). Очевидно, оба эти условия не могут выполняться одновременно.

Таким образом, мы независимо доказали условие Брэгга и формулу для доплеровского сдвига, используя теорию рассеяния. Формула (19.33) отражает факт пропорциональности интенсивности испускаемого света $\omega_r^4 \approx \omega^4$, так что эффективность рассеяния обратно пропорциональна четвертой степени длины волны. Этот анализ можно продолжить и вывести выражение для коэффициента отражения путем определения интенсивности волны, испускаемой рассеивающим источником (см. задачу 19.5).

Уравнения для связанных волн

Для выхода за рамки первого борновского приближения мы должны учесть вклад рассеянного поля в источник S. Если геометрия эксперимента соответствует брэгговской дифракции с повышением частоты, то поле $\mathcal{E}$ состоит из падающей и брэгговской отраженной волн:

$$\mathcal{E} = \mathrm{Re}\{E\exp(j\omega t)\} + \mathrm{Re}\{E_r\exp(j\omega_r t)\}.$$

С помощью соотношения (19.32) (19.30) дает

$$S = \mathrm{Re}\{S\exp(j\omega t)\} + \mathrm{Re}\{S_r\exp(j\omega_r t)\} + \text{члены с другими частотами,} \quad (19.34)$$

где

$$S = -k^2\left(\frac{\Delta n_0}{n}\right)E_r; \quad S_r = -k_r^2\left(\frac{\Delta n_0}{n}\right)E. \quad (19.35)$$

Сравнивая члены с одинаковыми частотами в обеих частях волнового уравнения

$$\nabla^2\mathcal{E} - \frac{1}{c^2}\frac{\partial^2\mathcal{E}}{\partial t^2} = -S,$$

мы получаем два связанных уравнения Гельмгольца для падающей волны и брэгговской рассеянной волны:

$$(\nabla^2 + k^2)E = -S; \quad (\nabla^2 + k_r^2)E_r = -S_r. \quad (19.36)$$

Эти уравнения вместе с (19.35) можно решить и определить E и E_r.

Рассмотрим, например, случай рассеяния под малым углом ($\theta \ll 1$), когда обе волны распространяются примерно в направлении оси z. Если считать, что $k \approx k_r$, то поля E и E_r записываются как $E = A\exp(-jkz)$ и $E_r = A_r\exp(-jkz)$, где A и A_r — медленно меняющиеся функции z. Используя приближение медленной огибающей

$$(\nabla^2 + k^2)A\exp(-jkz) \approx -j2k\frac{\mathrm{d}A}{\mathrm{d}z}\exp(-jkz)$$

(см. подразд. 2.2.3), получаем из (19.35) и (19.36):

$$\boxed{\frac{dA}{dz} = j\frac{1}{2}\gamma A_r;} \tag{19.37а}$$

$$\boxed{\frac{dA_r}{dz} = j\frac{1}{2}\gamma A,} \tag{19.37б}$$

где

$$\gamma = k\frac{\Delta n_0}{n}. \tag{19.38}$$

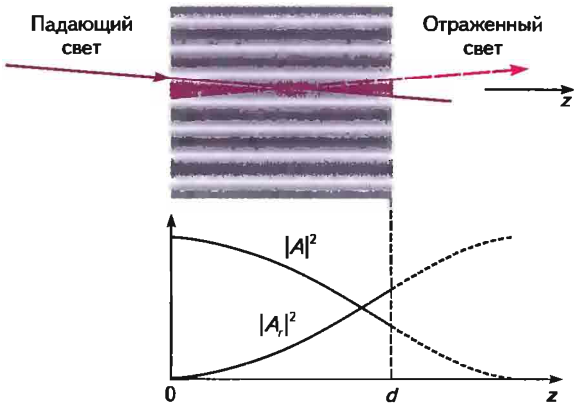


Рис. 19.10. Изменение интенсивности падающей оптической волны (сплошная линия) и интенсивности брэгговской отраженной волны (штриховая линия) в зависимости от расстояния, пройденного в среде с акустической волной

Если ячейка простирается от $z = 0$ до $z = d$, то мы используем граничное условие $A_r(0) = 0$ и обнаруживаем, что уравнения (19.37) имеют гармоническое решение

$$A(z) = A(0)\cos\frac{\gamma z}{2}; \tag{19.39а}$$

$$A_r(z) = jA(0)\cos\frac{\gamma z}{2}. \tag{19.39б}$$

Эти уравнения описывают возникновение отраженной волны и спад падающей волны, как показано на рис. 19.10. Следовательно, коэффициент отражения

$$\mathcal{R}_e = \frac{|A_r(d)|^2}{|A(0)|^2}$$

выражается как $\mathcal{R}_e = \sin^2(\gamma d/2)$, так что $\mathcal{R}_e = \sin^2\sqrt{\mathcal{R}}$, где $\mathcal{R} = (\gamma d/2)^2$.

Используя (19.38), получаем

$$\mathcal{R} = \frac{\pi^2}{\lambda_0^2}\Delta n_0^2 d^2$$

Это в точности повторяет выражение для коэффициента отражения слабого звука (19.22) при $d = L/\sin\theta$.

19.1.3. Брэгговская дифракция пучков

До этих пор было показано, что оптическая *плоская волна* с волновым вектором $\boldsymbol{k}$, взаимодействуя с акустической *плоской волной* с волновым вектором $\boldsymbol{q}$, порождает оптическую плоскую волну с волновым вектором $\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}$, если выполняется условие Брэгга (т. е. если угол между $\boldsymbol{k}$ и $\boldsymbol{q}$ таков, что $k_r = |\boldsymbol{k} + \boldsymbol{q}| \approx k = 2\pi/\lambda$). Взаимодействие *пучка* света с *пучком* звука можно объяснить, если пучок рассматривать как суперпозицию плоских волн, распространяющихся в различных направлениях, каждый со своим волновым вектором (см. введение к гл. 4).

Дифракция оптического пучка на плоской акустической волне

Рассмотрим оптический *пучок* шириной D, взаимодействующий с акустической *плоской волной*. В соответствии с принципами Фурье-оптики (см. подразд. 4.3.1) оптический пучок можно разложить по плоским волнам, направления которых заполняют конус с половиной угла при вершине

$$\delta\theta = \frac{\lambda}{D}. \tag{19.40}$$

В определении диаметра D и угла $\delta\theta$ имеется некоторый произвол, и соответствующий постоянный множитель в (19.40) взят равным единице. Если профиль пучка прямоугольный и имеет ширину D, то угловая ширина от максимума до первого нуля равна $\delta\theta = \lambda/D$; для круглого пучка диаметром D $\delta\theta = 1{,}22\lambda/D$; для гауссова пучка с диаметром перетяжки $D = 2W_0$

$$\delta\theta = \frac{\lambda}{\pi W_0} = \frac{2}{\pi}\frac{\lambda}{D} \approx \frac{0{,}64\lambda}{D}$$

[см. (3.20)]. Для простоты мы будем использовать (19.40).

Хотя существует единственный волновой вектор $\boldsymbol{q}$, имеется множество волновых векторов $\boldsymbol{k}$ одинаковой длины $2\pi/\lambda$ внутри конуса с углом $\delta\theta$. Как показывает рис. 19.11, существует только одно направление $\boldsymbol{k}$, для которого выполняется условие Брэгга. Тогда отраженная волна является плоской с единственным волновым вектором $\boldsymbol{k}_r$.

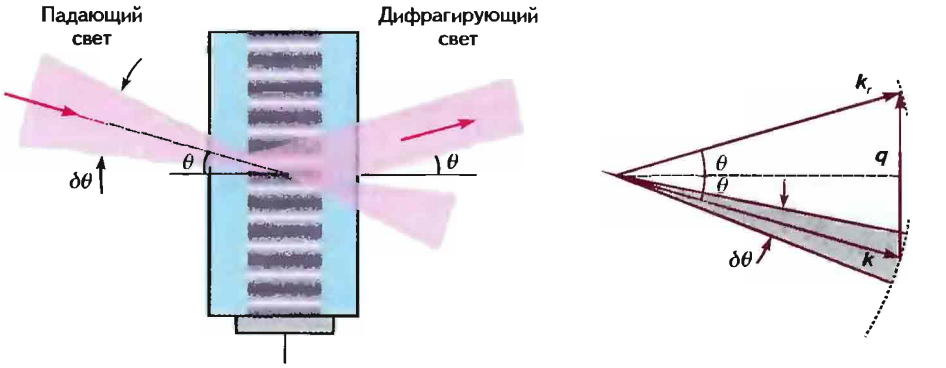


Рис. 19.11. Дифракция оптического *пучка* на акустической *плоской волне*. Существует единственная компонента падающего света в виде плоской волны, которая удовлетворяет условию Брэгга. Дифрагированный свет представляет собой плоскую волну

Дифракция оптического пучка на акустическом пучке

Предположим теперь, что сама акустическая волна является пучком шириной D_s. Если частота звука достаточно велика, чтобы звуковая длина волны была много меньше, чем толщина среды, то звук распространяется в режиме свободных волн (без волноводных эффектов) и имеет свойства, аналогичные свойствам оптических пучков, в частности, обладает угловой расходимостью

$$\delta\theta_s = \frac{\Lambda}{D_s}. \tag{19.41}$$

Это эквивалентно множеству плоских волн, направления распространения которых лежат внутри угла расходимости.

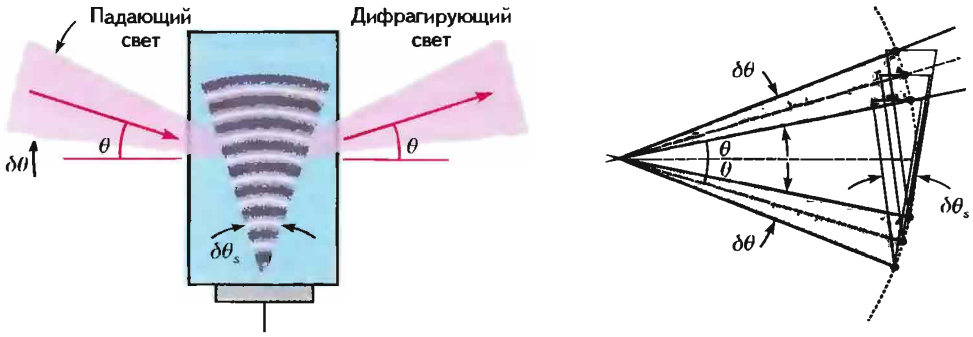


Рис. 19.12. Дифракция оптического *пучка* на акустическом *пучке*

Отражение оптического пучка от акустического можно определить путем поиска подходящих пар оптических и акустических плоских волн, удовлетво-

ряющих условию Брэгга. Сумма отраженных волн составляет отраженный оптический пучок. Существует много волновых векторов $\boldsymbol{k}$ (с одинаковой длиной $2\pi/\lambda$) и много векторов $\boldsymbol{q}$ (с одинаковой длиной $2\pi/\Lambda$), причем только пары векторов, образующие равнобедренные треугольники, вносят вклад в отражение, как показано на рис. 19.12.

Если расходимость акустического пучка больше, чем у оптического ($\delta\theta_s \gg \delta\theta$), и центральные направления обоих пучков удовлетворяют условию Брэгга, то каждая падающая плоская оптическая волна находит себе акустического «партнера», и отраженный световой пучок имеет такую же угловую расходимость $\delta\theta$, что и падающий пучок. Распределение акустической энергии в звуковом пучке в этом случае можно отслеживать как функцию направления с помощью пробного светового пучка со значительно меньшей расходимостью, измеряя отраженный свет в зависимости от угла падения.

Дифракция плоской оптической волны на тонком акустическом пучке: дифракция Рамана—Ната

Поскольку тонкий акустический пучок включает плоские волны, распространяющиеся в множестве направлений, он может вызывать дифракцию света под углами, которые существенно отличаются от брэгговского угла, соответствующего главному направлению пучка. Рассмотрим, например, геометрию эксперимента, показанную на рис. 19.13, в которой падающая плоская оптическая волна перпендикулярна главному направлению тонкого акустического пучка. Условие Брэгга выполняется, если волновой вектор отраженной волны $\boldsymbol{k}_r$ направлен под углами $\pm\theta$, где

$$\sin\frac{\theta}{2} = \frac{\lambda}{2\Lambda}. \tag{19.42}$$

Если угол θ мал, то

$$\sin\frac{\theta}{2} \approx \frac{\theta}{2};$$

$$\boxed{\theta \approx \frac{\lambda}{\Lambda}}. \tag{19.43}$$

Следовательно, падающий пучок отклоняется в одном из двух направлений, образующих углы $\pm\theta$, в зависимости от того, куда бежит звуковая волна: вверх или вниз. Для стоячей акустической волны отражение происходит в обоих направлениях.

Угол $\theta \approx \lambda/\Lambda$ — это угол, на который дифракционная решетка с периодом Λ отклоняет падающую плоскую волну (см. упражнение 2.7). Тонкий акустический пучок фактически модулирует показатель преломления, создавая периодическую структуру с периодом Λ, ограниченную тонким плоским слоем. Таким образом, среда действует как тонкая дифракционная решетка. Дифракция све-

та на этой решетке происходит также в направлении максимумов более высокого порядка, как показано на рис. 19.14, *а*.

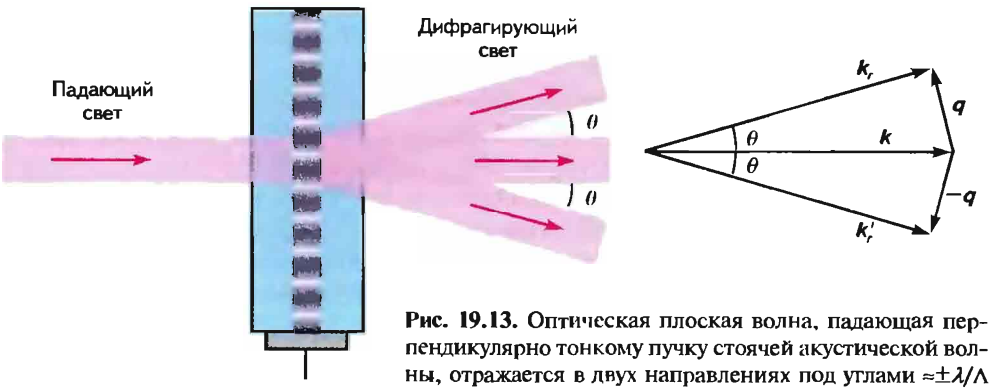


Рис. 19.13. Оптическая плоская волна, падающая перпендикулярно тонкому пучку стоячей акустической волны, отражается в двух направлениях под углами $\approx \pm\lambda/\Lambda$

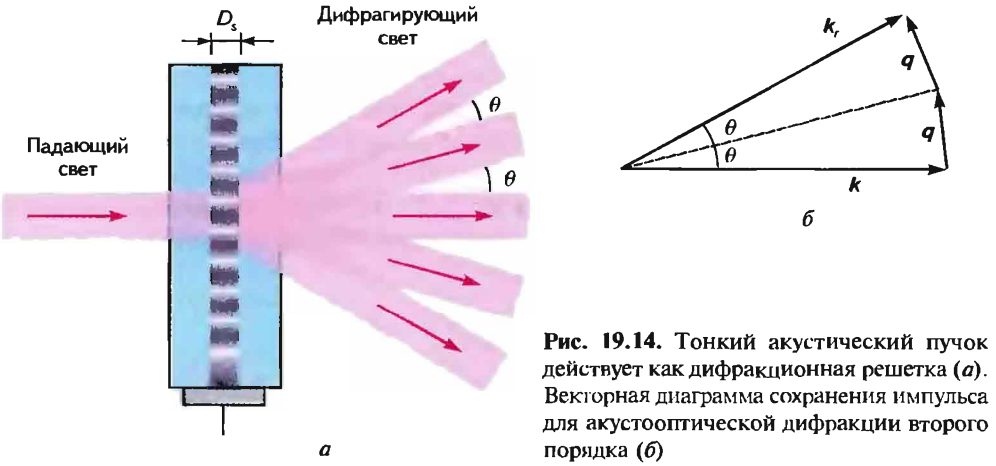


Рис. 19.14. Тонкий акустический пучок действует как дифракционная решетка (*а*). Векторная диаграмма сохранения импульса для акустооптической дифракции второго порядка (*б*)

Дифрагированные волны более высокого порядка, создаваемые фазовой решеткой в направлениях $\pm 2\theta$, $\pm 3\theta$, можно интерпретировать также в рамках квантового описания взаимодействия света и звука. Один падающий фотон комбинирует с двумя фононами (квантовыми акустическими частицами) с образованием фотона отраженной волны второго порядка. Закон сохранения импульса требует выполнения равенства $\boldsymbol{k}_r = \boldsymbol{k} \pm 2\boldsymbol{q}$. Это условие выполняется для геометрии, показанной на рис. 19.14, *б*. Свет, отраженный во втором порядке дифракции, имеет сдвинутую частоту $\omega_r = \omega \pm 2\Omega$. Аналогично можно интерпретировать и более высокие порядки дифракции.

Акустооптическое взаимодействие света с *перпендикулярным тонким* звуковым пучком известно как **рассеяние Рамана—Ната** или **Дебая—Сирса**[1].

[1] Более детальное описание см. в *Born M., Wolf E.* Principles of Optics. Cambridge University Press, 7th expanded and corrected ed. 2002, Chapter 12.

19.2. АКУСТООПТИЧЕСКИЕ УСТРОЙСТВА

19.2.1. Модуляторы

Интенсивность отраженного света в ячейке Брэгга пропорциональна интенсивности звука, если последняя достаточно мала. Используя электрически управляемый акустический преобразователь (рис. 19.15, *а*), можно пропорционально варьировать интенсивность отраженного света. Такой прибор можно использовать как линейный аналог модулятора света.

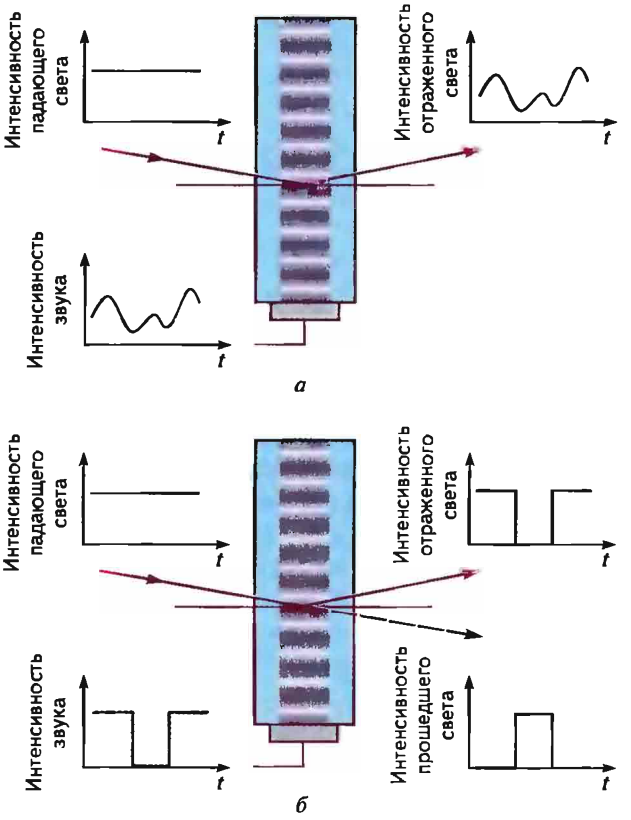


Рис. 19.15. Акустооптический модулятор (*а*). Интенсивность отраженного света пропорциональна интенсивности звука. Акустооптический переключатель (*б*)

С ростом мощности звука, однако, наступает насыщение и можно достичь почти полного отражения (см. рис. 19.7). В этом случае модулятор служит оптическим переключателем, где включение/выключение звука вызывает включение/выключение отраженного света и, соответственно, выключение/включение проходящего света, как показано на рис. 19.15, *б*.

Ширина полосы модуляции

Ширина полосы модулятора определяется максимальной частотой, на которой возможна эффективная модуляция. Когда амплитуда акустической волны, имеющей частоту f_0, меняется во времени в результате модуляции сигналом с шириной полосы B, акустическая волна перестает быть одночастотной гармонической функцией времени; она имеет частотные компоненты в пределах полосы $f_0 \pm B$ с центром на частоте f_0 (рис. 19.16). Как взаимодействует монохроматический свет с такой многочастотной акустической волной и какова максимальная ширина B, при которой акустооптический модулятор может работать?

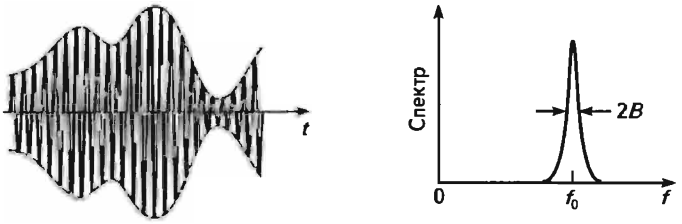


Рис. 19.16. Форма амплитудно-модулированного акустического сигнала и его спектр

Если как падающая оптическая волна, так и акустическая волна являются плоскими, то составляющая звука частоты f соответствует углу Брэгга

$$\theta = \arcsin\frac{\lambda}{2\Lambda} = \arcsin\frac{f\lambda}{2v_s} \approx \frac{\lambda}{2v_s} f, \tag{19.44}$$

который предполагается малым. Для фиксированного угла падения θ падающая монохроматическая оптическая плоская волна с длиной волны l взаимодействует с одной и только одной гармонической компонентой акустической волны, частота которой f удовлетворяет (19.44), как показано на рис. 19.17. Отраженная волна в этом случае является *монохроматической* и имеет частоту $v + f$. Хотя акустическая волна модулированная, отраженная оптическая волна не является таковой. Очевидно, что при таких идеализированных условиях ширина полосы модулятора равна нулю.

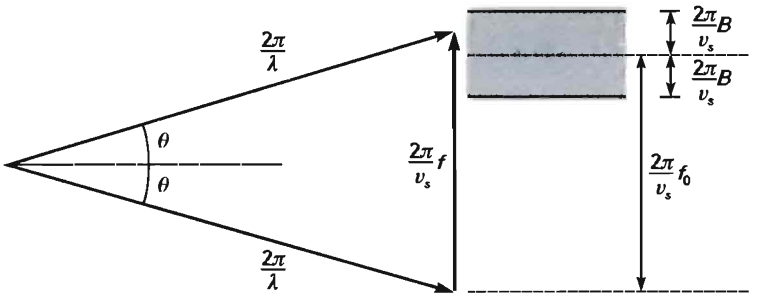


Рис. 19.17. Взаимодействие оптической плоской волны с модулированной (многочастотной) акустической волной. Только одна частотная компонента звука отражает световую волну. Отраженная волна является монохроматической и не модулируется

Для достижения модуляции с шириной полосы B каждая из акустических частотных компонент в полосе $f_0 \pm B$ должна взаимодействовать с падающей световой волной. Следовательно, должны выполняться более мягкие условия. Предположим, что падающий свет — пучок с шириной D и угловой расходимостью $\delta\theta = \lambda/D$, а модулированная акустическая волна плоская. Каждая частотная компонента звука взаимодействует с оптической плоской волной, падающей под подходящим углом Брэгга (рис. 19.18). Частотная полоса $f_0 \pm B$ будет покрываться, если оптический пучок имеет угловую расходимость

$$\theta \approx \frac{(2\pi/v_s)B}{2\pi/\lambda} = \frac{\lambda}{v_s}B. \tag{19.45}$$

В этом случае ширина полосы модулятора составляет

$$B = v_s\frac{\delta\theta}{\lambda} = \frac{v_s}{D}, \tag{19.46}$$

или

$$\boxed{B = \frac{1}{T}; \quad T = \frac{D}{v_s},} \tag{19.47}$$

Ширина полосы

где T — время прохода звука через перетяжку светового пучка. Этого результата следовало ожидать, поскольку для изменения амплитуды звуковой волны во всех точках области взаимодействия звука со светом требуется время T, так что максимальная частота модуляции есть $1/T$ Гц. Для увеличения ширины полосы модулятора световой пучок нужно фокусировать в пятно меньшего диаметра.

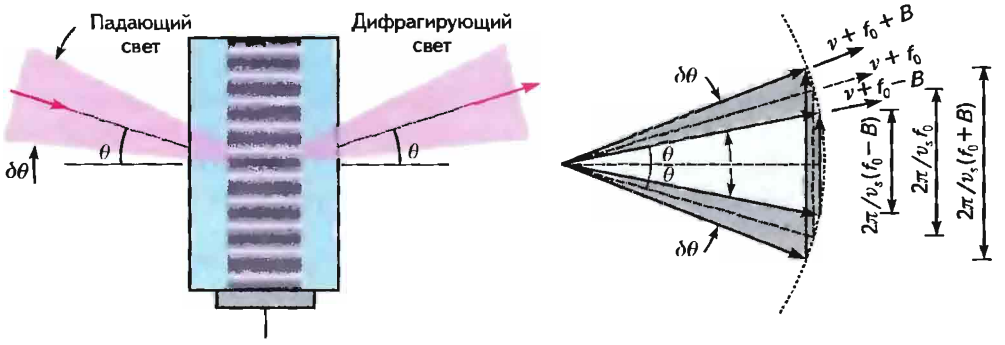


Рис. 19.18. Взаимодействие оптического пучка с угловой расходимостью $\delta\theta$ с акустической плоской волной, частоты которой лежат в полосе $f_0 \pm B$. Существует множество параллельных векторов $\boldsymbol{q}$ различной длины, каждый из которых согласуется с одним из направлений падающего света

Упражнение 19.1

Параметры акустооптических модуляторов

Определите угол Брэгга и максимальную ширину полосы следующих акустооптических модуляторов:

Модулятор 1

Материал:	Плавленый кварц ($n = 1{,}46$; $v_s = 6$ км/с)
Звук:	Частота $f = 50$ МГц
Свет:	He—Ne-лазер, длина волны $\lambda_0 = 633$ нм, угловая расходимость $\delta\theta = 1$ мрад.

Модулятор 2

Материал:	Теллур ($n = 4{,}8$; $v_s = 2,2$ км/с)
Звук:	Частота $f = 100$ МГц
Свет:	CO_2-лазер, длина волны $\lambda_0 = 10{,}6$ мкм, ширина пучка $D = 1$ мм.

19.2.2. Сканеры

Акустооптическая ячейка может использоваться для сканирования света. Основная идея заключена в линейной связи между углом отклонения 2θ и частотой звука f

$$2\theta \approx \frac{\lambda}{v_s} f, \tag{19.48}$$

где θ предполагается достаточно малым, так что $\sin\theta \approx \theta$. Меняя частоту звука f, можно менять угол отклонения света 2θ.

Трудность состоит в том, что θ представляет собой и угол отражения, и угол падения. Для изменения угла отражения нужно одновременно менять и угол падения, и частоту звука. Этого можно достичь путем качания звукового пучка. На рис. 19.19 проиллюстрирован этот принцип. Для изменения частоты звука нужен частотный модулятор (ЧМ). Для качания звукового пучка требуется сложная система, например, образованная сфазированным набором акустических преобразователей (несколько акустических преобразователей, у которых фазы задающих сигналов подобраны так, чтобы полная генерируемая акустическая волна качалась). Угол качания синхронизируется с задающим генератором ЧМ.

Качания звукового пучка можно избежать, если использовать пучок с угловой расходимостью, величина которой равна или больше, чем полный угловой диапазон сканирования. При изменении частоты звука меняется угол Брэгга, и падающая световая волна выбирает только одну плоскую компоненту акустической волны с подходящим направлением. Эффективность такой системы, конечно, будет низкой. Перейдем к изучению некоторых свойств такого устройства.

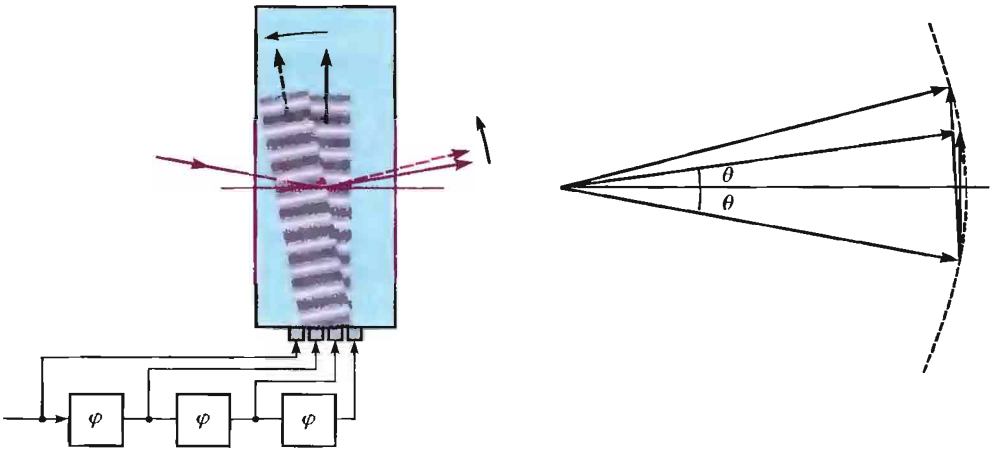


Рис. 19.19. Сканирование посредством изменения частоты и направления звука. Звуковая волна качается благодаря набору преобразователей, задающие сигналы которых различаются по фазе на величину φ

Угол сканирования

Когда частота звука равна f, падающая световая волна взаимодействует с компонентой звука под углом $\theta = (\lambda/2v_s)f$ и отклоняется на угол $2\theta = (\lambda/v_s)f$, как показывает рис. 19.20. При изменении частоты звука от f_0 до $f_0 + B$ угол отклонения 2θ покрывает область сканирования

$$\boxed{\Delta\theta = \frac{\lambda}{v_s} B.} \qquad (19.49)$$

Угол сканирования

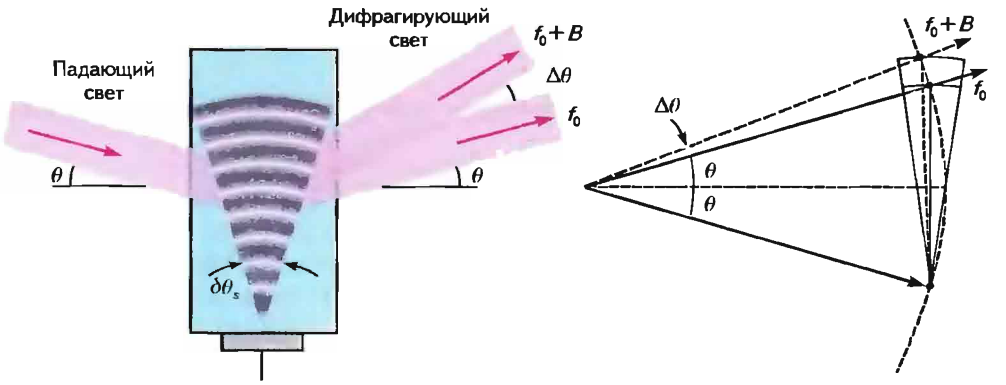


Рис. 19.20. Сканирование оптической волны путем изменения частоты звукового пучка с угловой расходимостью $\delta\theta_s$ в диапазоне частот $f_0 \le f \le f_0 + B$

При этом, конечно, предполагается, что звуковой пучок имеет равную или большую угловую ширину $\delta\theta_s = \Lambda/D_s \geq \Delta\theta$. Поскольку угол сканирования обратно пропорционален скорости звука, его расширение требует применения материалов, в которых скорость звука v_s мала.

Число разрешаемых пятен

Если оптическая волна сама имеет угловую ширину $\delta\theta = \lambda/D$ и $\delta\theta = \delta\theta_s$, то отклоненный пучок также имеет ширину $\delta\theta$. Следовательно, число разрешаемых пятен сканера (число неперекрывающихся угловых ширин в пределах диапазона сканирования) равно

$$N = \frac{\Delta\theta}{\delta\theta} = \frac{(\lambda/v_s)B}{\lambda/D} = \frac{D}{v_s}B, \tag{19.50}$$

или

$$\boxed{N = TB,} \tag{19.51}$$

Число разрешаемых пятен

где B — ширина полосы частотного модулятора, используемого для генерации звука; $T = D/v_s$ — время прохода звука через световой пучок (рис. 19.21).

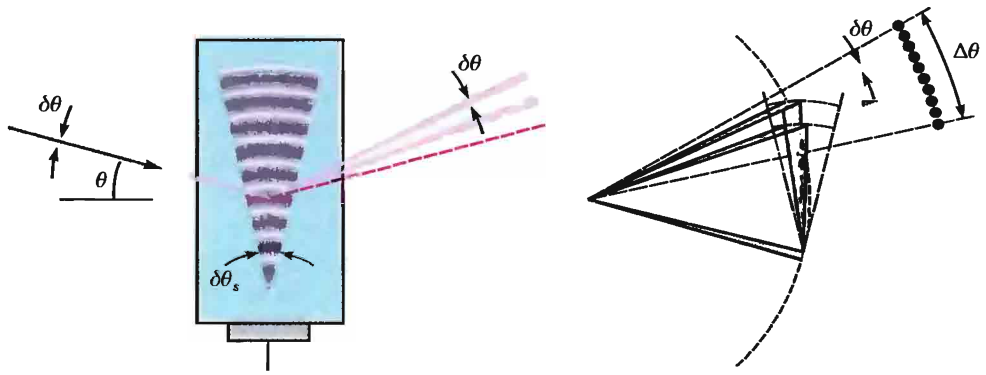


Рис. 19.21. Разрешаемые пятна акустооптического сканера

Таким образом, число разрешаемых пятен равно произведению времени прохода на ширину полосы. Это число представляет число степеней свободы прибора и является важным показателем возможностей сканера. Для увеличения N нужно, чтобы время прохода T было большим. Это требование противоположно тому, которое используется при проектировании акустооптических модуляторов, где ширину полосы модуляции $B = 1/T$ стремятся увеличивать за счет малых T.

Упражнение 19.2

Параметры акустооптического сканера

Акустооптический сканер из плавленого кварца (v_s = 6 км/с, n = 1,46) используется для сканирования пучка He—Ne-лазера (λ_0 = 633 нм). Частота звука сканируется в диапазоне от 40 до 60 МГц. До какой ширины нужно сфокусировать лазерный пучок, чтобы число разрешаемых точек было N = 100? Каков угол сканирования? Какой эффект вызовет использование материала с более низкой скоростью звука, например стекла сорта флинт (v_s = 3, 1 км/с)?

Акустооптический сканер как анализатор спектра

Пропорциональность между углом отклонения и частотой звука можно использовать для создания анализатора акустического спектра. Звуковая волна, содержащая спектр различных частот, отклоняет свет в различных направлениях, причем интенсивность света, отклоненного в данном направлении, пропорциональна мощности компоненты звука на соответствующей частоте (рис. 19.22).

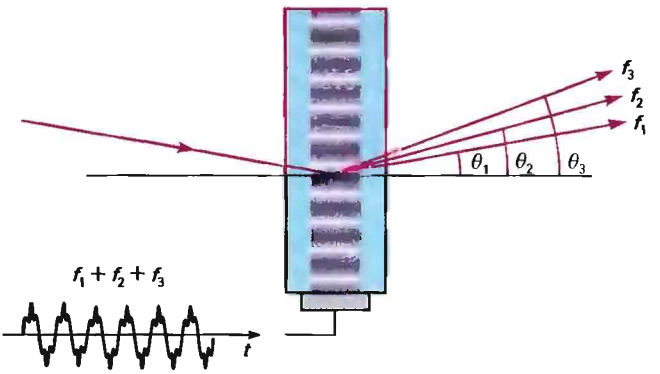


Рис. 19.22. Каждая частотная компонента звуковой волны отклоняет свет в своем направлении. Акустооптическая ячейка служит анализатором акустического спектра

19.2.3. Пространственные переключатели

Акустооптическая ячейка может применяться в качестве пространственного переключателя (см. разд. 23.3), который направляет информацию, несомую одним или несколькими оптическим пучками в одном или более выделенных направлениях. Возможно несколько схем соединения:

- Акустооптическая ячейка, в которой частота акустической волны равна одному из N возможных значений $f_1, f_2, ..., f_N$, отклоняет падающий оптический пучок в одном из N соответствующих направлений $\theta_1, \theta_2, ..., \theta_N$, как показано на рис. 19.23. Прибор направляет один пучок по любому из N направлений.
- С помощью акустической волны, состоящей из двух компонент с частотами f_1 и f_2, можно отразить оптический пучок в двух соответствующих направ-

лениях θ_1 и θ_2 одновременно. Таким образом, один пучок соединяется с любой парой из множества возможных направлений (каналов), как показано на рис. 19.24. Аналогично, с помощью акустической волны, содержащей M частотных компонент, входящий пучок можно отправить одновременно по M направлениям. Примером является анализатор акустического спектра, у которого входящий световой пучок отражается от звуковой волны со спектром, состоящим из M частот. Световой пучок направляется в M точек, причем интенсивность в каждой точке пропорциональна мощности соответствующей частотной компоненты звука.

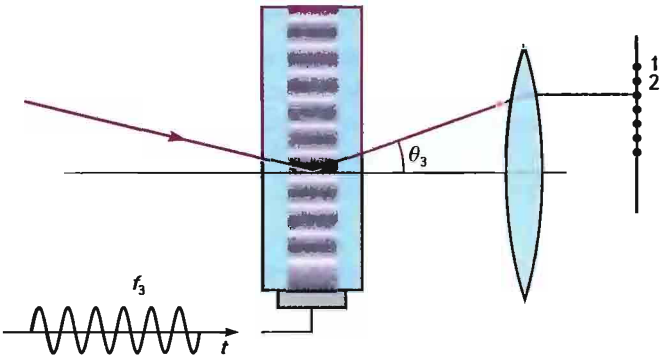


Рис. 19.23. Направление оптического пучка по одному из N каналов. Приложение акустической волны с частотой f_3, например, отклоняет оптический пучок на угол θ_3 и направляет его в точку 3

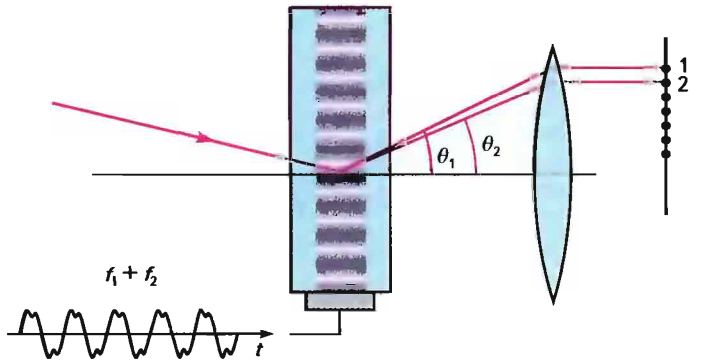


Рис. 19.24. Направление светового пучка одновременно по нескольким каналам

• Длину акустооптической ячейки можно разделить на два участка. В некоторый момент времени в одном участке присутствует акустическая волна частоты f_1, а в другом — частоты f_2. Это можно реализовать путем генерации акустической волны электрическим частотно-манипулируемым сигналом в виде двух импульсов: импульса с частотой f_1, за которым следует другой импульс с частотой f_2, причем длительность каждого импульса равна $T/2$, где $T = W/v_s$ — время прохода звука через ячейку длины W (рис. 19.25). Когда передний фронт акустической волны достигнет конца ячейки, она обрабатывает два входяших оптических пучка, отклоняя верхний из них на угол θ_1, соответствующий f_1, а

нижний на угол θ_2, соответствующий f_2. Получается переключатель, который связывает каждый из двух пучков с любым из множества возможных направлений. Если в каждом участке создать многочастотную звуковую волну, то каждый из оптических пучков будет посылаться в нескольких направлениях одновременно.

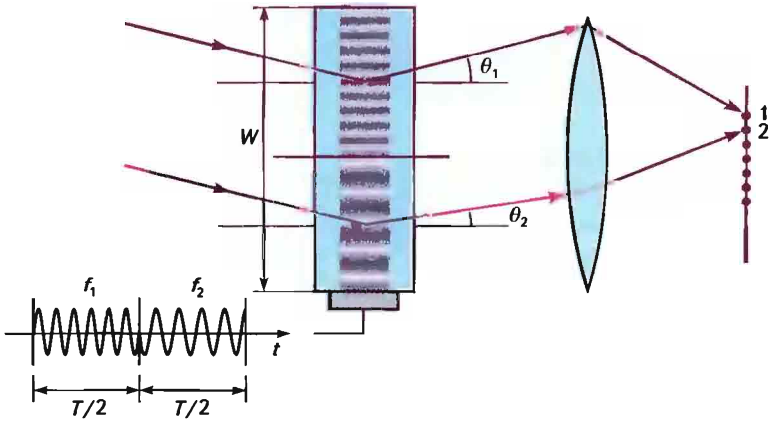
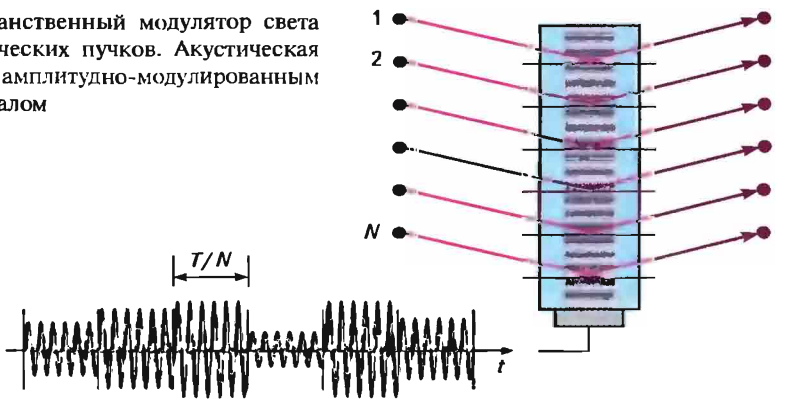


Рис. 19.25. Направление двух световых пучков по набору заданных каналов. Акустическая волна генерируется частотно-манипулируемым сигналом

• Ячейку можно разделить на N участков, каждый из которых несет гармоническую акустическую волну одинаковой частоты f, но с различной амплитудой. В результате получается **пространственный модулятор света** (ПМС), который модулирует интенсивности N входящих пучков (рис. 19.26). Пространственные модуляторы света применяются при обработке оптических сигналов (см. подразд. 20.1.5).

Рис. 19.26. Пространственный модулятор света модулирует N оптических пучков. Акустическая волна возбуждается амплитудно-модулированным электрическим сигналом



• Наиболее общая архитектура переключателя подразумевает деление ячейки на L участков, каждый из которых несет акустическую волну с M частотами.

Прибор действует как ключ прямого доступа, который одновременно связывает каждый из L входящих пучков с M направлениями (рис. 19.27).

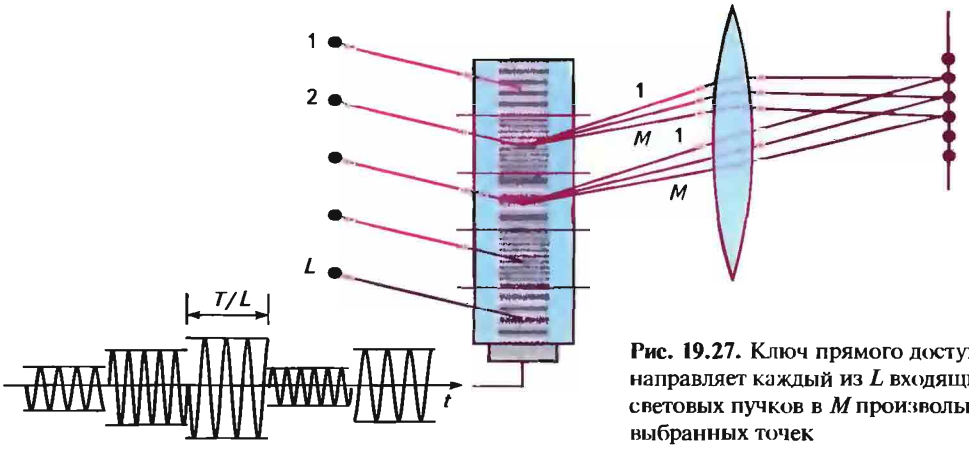


Рис. 19.27. Ключ прямого доступа направляет каждый из L входящих световых пучков в M произвольно выбранных точек

Пропускная способность соединения

Как будет показано впоследствии, существует верхний предел числа соединений, которые можно установить при помощи акустооптического прибора. При использовании акустооптической ячейки для направления L входящих оптических пучков одновременно по максимальному числу M каналов произведение ML не может превышать произведения $N = TB$, где T — время прохода акустической волны через ячейку, а B — ее ширина полосы:

$$\boxed{ML \le N.} \qquad (19.52)$$

Пропускная способность соединения

Эта верхняя граница числа соединений называется пропускной способностью соединения прибора.

В акустооптической ячейке с L участками используется акустическая волна, состоящая из L отрезков, каждый длительностью T/L. Чтобы каждый участок мог посылать пучок в M независимых точек, акустическая волна должна содержать M независимых частотных компонент на каждом своем отрезке. Для сигнала длительностью T/L существует неустранимая неопределенность частоты, равная L/T Гц. Следовательно, частоты M компонент должны отстоять друг от друга, по крайней мере, на величину этой неопределенности. Чтобы M частот поместились в доступную ширину полосы B, необходимо, чтобы $M(L/T) \le B$, откуда $ML \le TB$ и, следовательно, получается (19.52).

Например, одиночный оптический пучок ($L = 1$) можно связать с любой из $N = TB$ точек, но каждый из двух пучков можно связать самое большее с $N/2$ точками, и так далее. Вопрос заключается в том, как представить заданное для данной ячейки произведение $N = TB$ в виде L отрезков времени, на каждом из

которых имеется M независимых частот. Примеры возможного выбора показаны в координатах «время—частота» на рис. 19.28.

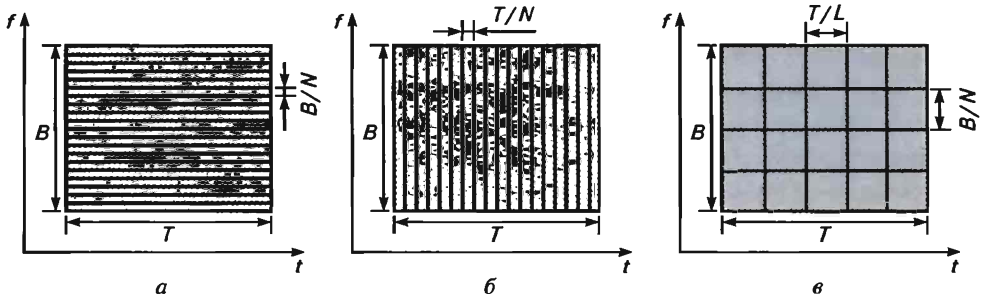


Рис. 19.28. Несколько примеров деления области TB на диаграмме «время—частота» на $N = TB$ частей (в данном примере $N = 20$):

а — **сканер**: один отрезок времени содержит N сегментов частоты; *б* — **пространственный модулятор света**: N отрезков времени, каждый содержит одну частотную компоненту; *в* — **переключатель соединений**: L отрезков времени, каждый содержит $M = N/L$ частотных сегментов (на данной диаграмме $N = 20$, $M = 4$, $L = 5$)

19.2.4. Фильтры, преобразователи частоты и вентили

Акустооптические ячейки имеют ряд других приложений, включающих фильтры, преобразователи частоты и вентили.

Перестраиваемые акустооптические фильтры

Условие Брэгга $\sin\theta = \lambda/2\Lambda$ связывает угол θ, акустическую длину волны Λ и оптическую длину волны λ. Если θ и Λ заданы, то отражение может происходить только для одной длины волны света λ. Это свойство селективности по длине волны можно использовать для частотной фильтрации оптической волны, имеющей широкий спектр. Такой фильтр перестраивается путем изменения угла θ либо частоты звука f.

Упражнение 19.3

Разрешающая сила акустооптического фильтра

Покажите, что спектральная разрешающая сила $\lambda/\Delta\lambda$ акустооптического фильтра равна fT, где f — частота звука; T — время прохода; $\Delta\lambda$ — минимальная разрешимая разность длин волн.

Преобразователи частоты

Оптические преобразователи частоты применяются во многих приложениях фотоники, включая оптическое гетеродинирование, оптические частотные модуляторы и лазерные доплеровские измерители скорости. Акустооптическая ячейка может использоваться как перестраиваемый преобразователь

частоты, поскольку отраженный в соответствии с условиями Брэгга свет сдвинут по частоте (вверх или вниз) на частоту звука. В гетеродинном оптическом приемнике принимаемый сигнал с амплитудной или частотной модуляцией смешивается с когерентной оптической волной от оптического гетеродина с другой частотой. Обе оптические волны дают биения (см. подразд. 2.6.2) и детектируемый сигнал меняется на разностной частоте. Информация об амплитуде и фазе принимаемого сигнала может быть извлечена из детектируемого сигнала (см. разд. 24.5). Акустооптическая ячейка дает практический способ внести частотный сдвиг, необходимый для процесса гетеродинирования.

Оптическая развязка

Оптическая развязка — однонаправленный оптический клапан, который часто используется, чтобы предотвратить попадание отраженного света в исходный источник света (см. подразд. 6.6.3 и 23.1.3). Иногда оптическая развязка применяются в сочетании с полупроводниковыми лазерами, поскольку отраженный свет может участвовать в процессе лазерной генерации создавать шум. Акустооптическая ячейка может работать как оптическая развязка. Если часть сдвинутого по частоте вверх брэгговского дифрагированного света отражается назад зеркалом и проходит обратно в ячейку, как показано на рис. 19.29, она претерпевает повторную брэгговскую дифракцию, сопровождающуюся еще одним сдвигом частоты вверх. Поскольку частота возвращенного света отличается от первоначальной на удвоенный сдвиг, его можно блокировать при помощи фильтра. Даже без фильтра процесс лазерной генерации может быть нечувствителен к свету, сдвинутому по частоте.

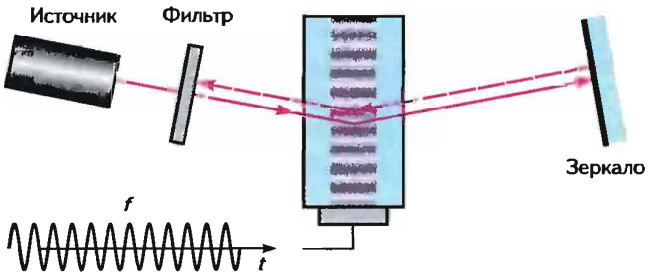


Рис. 19.29. Акустооптическая развязка

*19.3. АКУСТООПТИКА АНИЗОТРОПНЫХ СРЕД

Скалярная теория взаимодействия света и звука обобщается в этом разделе, чтобы учесть анизотропные свойства среды и эффекты поляризации света и звука.

Акустические волны в анизотропных материалах

Акустическая волна — это волна деформации материала. Деформация определяется через смещение молекул относительно их равновесных поло-

жений. Если $\boldsymbol{u} = (u_1, u_2, u_3)$ — вектор смещения молекул в точке $\boldsymbol{x} = (x_1, x_2, x_3)$, то деформация определяется симметричным тензором с компонентами

$$s_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right),$$

где индексы $i, j = 1, 2, 3$ нумеруют координаты (x, y, z). Например, элемент $s_{33} = \partial u_3/\partial x_3$ описывает деформацию растяжения в направлении z (рис. 19.30, *а*), а s_{13} — деформацию сдвига, поскольку $\partial u_1/\partial x_3$ есть относительное смещение в направлении x двух бесконечно близких параллельных плоскостей, перпендикулярных оси z, как показано на рис. 19.30, *б*.

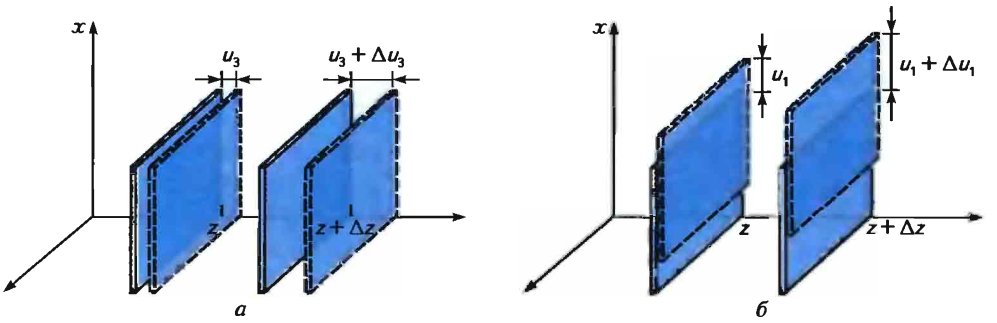


Рис. 19.30. Смещения, связанные с деформацией растяжения и сдвига:
а — деформация растяжения; *б* — сдвиг

Пример 19.4

Продольная акустическая волна

Волна со смещением $u_1 = 0$, $u_2 = 0$, $u_3 = A_0 \sin(\Omega t - qz)$, где A_0 — постоянная, соответствует тензору деформации, у которого исчезают все компоненты, кроме

$$s_{33} = S_0 \cos(\Omega t - qz), \tag{19.53}$$

где $S_0 = -qA_0$. Это волна растяжения вдоль направления z, которая распространяется также вдоль оси z. Поскольку колебания происходят в том же направлении, что и распространение волны, она является продольной.

Пример 19.5

Поперечная акустическая волна

Волна смещения $u_1 = A_0 \sin(\Omega t - qz)$, $u_2 = 0$, $u_3 = 0$ соответствует тензору деформации, все компоненты которого равны нулю, кроме

$$s_{13} = s_{31} = S_0 \cos(\Omega t - qz), \tag{19.54}$$

где $S_0 = -qA_0/2$. Эта волна распространяется в направлении z, а колебания происходят вдоль оси x. Следовательно, это поперечная (сдвиговая) волна.

Скорости продольных и поперечных акустических волн являются характерными для данной среды и, вообще говоря, зависят от направления распространения.

Фотоупругий эффект

Оптические свойства анизотропной среды полностью характеризуются тензором диэлектрической непроницаемости $\eta = \varepsilon_0 \varepsilon^{-1}$ (см. разд. 6.3). По заданному η можно определить эллипсоид показателей преломления и, следовательно, показатель преломления для оптической волны, распространяющейся в произвольном направлении с произвольной поляризацией.

При наличии деформации тензор диэлектрической непроницаемости меняется так, что η_{ij} становится функцией элементов тензора деформации, $\eta_{ij} = \eta_{ij}(s_{kl})$. Эта зависимость называется **фотоупругим эффектом**. Каждую из девяти функций $\eta_{ij}(s_{kl})$ можно разложить по девяти переменным s_{kl} в ряд Тейлора. Оставляя только линейные члены, получаем

$$\eta_{ij}(s_{kl}) \approx \eta_{ij}(0) + \sum_{kl} \mathfrak{p}_{ijkl} s_{kl}, \quad i, j, k, l = 1, 2, 3, \tag{19.55}$$

где $\mathfrak{p}_{ijkl} = \partial \eta_{ij} / \partial s_{kl}$ — постоянные, образующие тензор четвертого ранга, называемый **тензором фотоупругости**.

Поскольку и $\{\eta_{ij}\}$, и $\{s_{kl}\}$ — симметричные тензоры, коэффициенты $\mathfrak{p}_{ijkl}$ инвариантны относительно перестановок i и j, а также k и l. Следовательно, имеется всего шесть, а не девять независимых значений для (i, j) и шесть независимых значений для (k, l). Независимые пары индексов (i, j) обычно обозначаются одним индексом $I = 1, 2, \ldots 6$ (см. табл. 20.1). Независимые пары индексов (k, l) также обозначаются одним индексом $K = 1, 2, \ldots, 6$. В результате тензор четвертого ранга можно записать в виде матрицы $\mathfrak{p}_{IK}$ размером 6×6.

Симметрия кристалла требует обращения в нуль некоторых коэффициентов $\mathfrak{p}_{IK}$ и накладывает связи на другие коэффициенты. Например, для кубического кристалла матрица $\mathfrak{p}_{IK}$ имеет структуру

$$\mathfrak{p}_{IK} = \begin{bmatrix} \mathfrak{p}_{11} & \mathfrak{p}_{12} & \mathfrak{p}_{12} & 0 & 0 & 0 \\ \mathfrak{p}_{12} & \mathfrak{p}_{11} & \mathfrak{p}_{12} & 0 & 0 & 0 \\ \mathfrak{p}_{11} & \mathfrak{p}_{12} & \mathfrak{p}_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathfrak{p}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathfrak{p}_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathfrak{p}_{44} \end{bmatrix}. \tag{19.56}$$

Матрица фотоупругости (кубический кристалл)

Эта матрица применима также к изотропным средам при дополнительном условии $\mathfrak{p}_{44} = (\mathfrak{p}_{11} + \mathfrak{p}_{12})/2$, так что имеется всего два независимых коэффициента.

Пример 19.6

Продольная акустическая волна в кубическом кристалле

Продольная акустическая волна, описанная в примере 19.1, распространяется вдоль одной из осей кубического кристалла с показателем преломления n. Путем подстановки (19.53) и (19.56) в (19.55) находим, что при этом деформация порождает тензор диэлектрической непроницаемости с элементами

$$\eta_{11} = \eta_{22} = \frac{1}{n^2} + p_{12} S_0 \cos(\Omega t - qz); \tag{19.57}$$

$$\eta_{33} = \frac{1}{n^2} + p_{11} S_0 \cos(\Omega t - qz); \tag{19.58}$$

$$\eta_{ij} = 0, \quad i \neq j. \tag{19.59}$$

Таким образом, первоначально оптически изотропный кубический кристалл становится одноосным кристаллом с оптической осью в направлении распространения акустической волны (направление z) и обыкновенным и необыкновенным показателями преломления n_o и $n_н$, которые определяются выражениями

$$\frac{1}{n_o^2} = \frac{1}{n^2} + p_{12} S_0 \cos(\Omega t - qz); \tag{19.60}$$

$$\frac{1}{n_н^2} = \frac{1}{n^2} + p_{11} S_0 \cos(\Omega t - qz). \tag{19.61}$$

Форма эллипсоида показателя преломления периодически меняется в пространстве и времени в виде волны, однако главные оси остаются неизменными (рис. 19.31). Поскольку изменения показателей преломления обычно малы, вторые члены в (19.60) и (19.61) также малы, поэтому можно применить приближенную формулу $(1 + \Delta)^{-1/2} \approx 1 - \Delta/2$, где $|\Delta| \ll 1$, для аппроксимации (19.60) и (19.61):

$$n_o \approx n - \frac{1}{2} n^3 p_{12} S_0 \cos(\Omega t - qz); \tag{19.62}$$

$$n_н \approx n - \frac{1}{2} n^3 p_{11} S_0 \cos(\Omega t - qz). \tag{19.63}$$

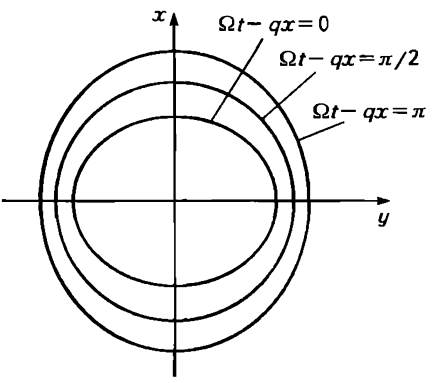


Рис. 19.31. Продольная акустическая волна, распространяющаяся в направлении z в кубическом кристалле, меняет форму эллипсоида показателя преломления от сферы к эллипсоиду вращения, размеры которого синусоидально меняются во времени и координатой z

Упражнение 19.4

Поперечная акустическая волна в кубическом кристалле

Поперечная акустическая волна, описанная в примере 19.2, распространяется вдоль одной из осей кубического кристалла. Покажите, что кристалл становится двуосным с главными занчениями показателя преломления

$$n_1 \approx n - \frac{1}{2} n^3 p_{44} S_0 \cos(\Omega t - qz); \tag{19.64}$$

$$n_2 \approx n; \tag{19.65}$$

$$n_3 \approx n + \frac{1}{2} n^3 p_{44} S_0 \cos(\Omega t - qz). \tag{19.66}$$

В примере 19.3 и упражнении 19.4 акустическая волна меняет главные значения эллипсоида показателя преломления, но не направления главных осей, так что эллипсоид сохраняет свою ориентацию. Очевидно, что это не всегда так. Акустические волны с другими направлениями распространения и поляризациями по отношению к главным осям кристалла приводят к изменению как главных значений показателя преломления, так и направления главных осей.

Дифракция Брэгга

Взаимодействие линейно поляризованной оптической волны с продольной или поперечной акустической волной в анизотропной среде можно описать на основе тех же принципов, которые обсуждались в разд. 19.1. Падающая оптическая волна отражается от акустической волны, если выполняется брэгговское условие конструктивной интерференции. Анализ более сложен по сравнению со скалярной теорией, поскольку падающая и отраженная волны распространяются с различными скоростями и, следовательно, углы падения и отражения не обязаны быть равными.

Условие брэгговской дифракции — это условие сохранения импульса (согласования фаз)

$$\boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}. \tag{19.67}$$

Длины этих волновых векторов равны

$$k = \frac{2\pi}{\lambda_0} n; \quad k_r = \frac{2\pi}{\lambda_0} n_r; \quad q = \frac{2\pi}{\Lambda},$$

где λ_0 и Λ — оптическая и акустическая длина волн; n и n_r — показатели преломлении для падающей и отраженной оптических волн соответственно.

Как показано на рис. 19.32, если θ и θ_r — углы падения и отражения, то векторное равенство (19.67) можно заменить двумя скалярными равенствами, связывающими z- и x-компоненты волновых векторов в плоскости падения:

$$\frac{2\pi}{\lambda_0} n_r \cos\theta_r = \frac{2\pi}{\lambda_0} n \cos\theta; \tag{19.68}$$

$$\frac{2\pi}{\lambda_0} n_r \sin\theta_r + \frac{2\pi}{\lambda_0} n \sin\theta = \frac{2\pi}{\Lambda}, \tag{19.69}$$

откуда

$$n_r \cos\theta_r = n\cos\theta; \tag{19.70а}$$

$$n_r \sin\theta_r + n\sin\theta = \frac{\lambda_0}{\Lambda}. \tag{19.70б}$$

При заданных длинах волн λ_0 и Λ углы θ и θ_r можно определить путем решения уравнений (19.70). Заметим, что n и n_r, вообще говоря, являются функциями θ и θ_r, которые можно определить по эллипсоиду показателя преломления невозмущенного кристалла.

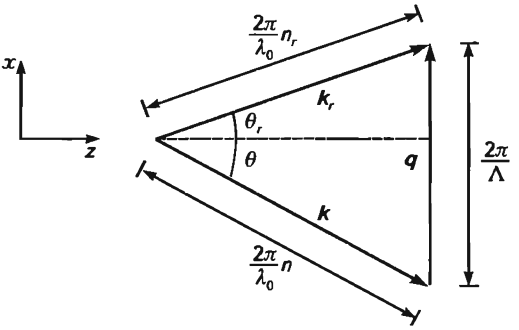


Рис. 19.32. Сохранение импульса (условие согласования фаз или условие Брэгга) в анизотропной среде

Уравнения (19.70) легко решаются, когда акустическая и оптическая волны коллинеарны, так что $\theta = \pm\pi/2$ и $\theta_r = \pi/2$. Знаки «плюс» и «минус» соответствуют отражению назад и вперед, как показано на рис. 19.33. Условия (19.70) тогда сводятся к одному условию

$$n_r \pm n = \frac{\lambda_0}{\Lambda}. \tag{19.71}$$

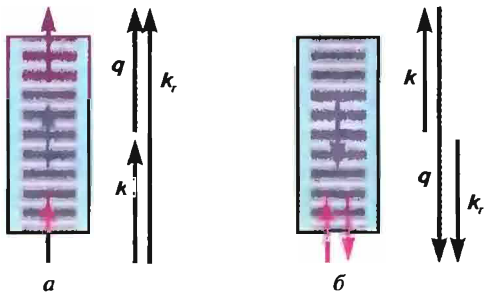


Рис. 19.33. Векторная диаграмма для отражения оптической волны от акустической волны:
а — отражение вперед; *б* — отражение назад

Для рассеяния назад (знак «+») Λ должно быть меньше λ_0, что вряд ли возможно, разве что для очень высокочастотных акустических волн. Для рассеяния вперед (знак «—») падающая и отраженная волны должны иметь различные поляризации, поскольку $n \neq n_r$.

Рекомендуемая литература

КНИГИ

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

Wolfe J.P. Imaging Phonons: Acoustic Wave Propagation in Solids. Cambridge University Press, paperback ed. 2005.

Musgrave M.J.P. Crystal Acoustics. Holden-Day, 1970; Acoustical Society of America, 2003.

Yariv A., Yeh P. Optical Waves in Crystals: Propagation and Control of Laser Radiation. Wiley, 1984; reprinted 2003.

Royer D., Dieulesaint E. Elastic Waves in Solids. Vol. 2: Generation, Acousto-Optic Interaction, Applications. Springer-Verlag, 2000.

Nye J.F. Physical Properties of Crystals: Their Representation by Tensors and Matrices. Oxford University Press, 1957; reprinted with corrections and new material, 2001.

Born M., Wolf E. Principles of Optics. Cambridge University Press, 7th expanded and corrected ed. 2002, Chapter 12.

Korpel A. Acousto-Optics. Marcel Dekker, 1988; 2nd ed. 1997.

Berg N.J., Pellegrino J.M., eds. Acousto-Optic Signal Processing. Marcel Dekker, 1983, 2nd ed., 1996.

Goutzoulis A.P., Pape D.R., eds. Design and Fabrication of Acousto-Optic Devices. Marcel Dekker, 1994.

Gusev V.E., Karabutov A.A. Laser Optoacoustics. American Institute of Physics, 1993.

Scott C. Field Theory of Acousto-Optic Signal Processing Devices. Artech, 1992.

Xu J., Stroud R. Acousto-Optic Devices: Principles, Design and Applications. Wiley, 1992.

Bunkin F.V., Kolomensky A.A., Mikhalevich V.G. Lasers in Acoustics. Harwood, 1991.

Das P.K., De Cusatis C.M. Acousto-Optic Signal Processing: Fundamentals & Applications. Artech, 199].

Tsai C.S. Guided- Wave Acoustooptics: Interactions, Devices and Applications. Springer-Verlag, 1990.

Magdich L.N., Molchanov V.Ya. Acoustooptic Devices and Their Applications. Gordon and Breach, 1989.

Gottlieb M., Ireland C.L.M., Ley J.M. Electro-Optic and Acousto-Optic Scanning and Deflection. Marcel Dekker, 1983.

Narasimhamurty T.S. Photoelastic and Electro-Optic Properties of Crystals. Plenum, 1981.

Nelson D.F. Electric, Optic, and Acoustic Interactions in Dielectrics. Wiley, 1979.

Sapriel J. Acousto-Optics. Wiley, 1979.

Berry M.V. The Diffraction of Light by Ultrasound. Academic Press, 1966.

СТАТЬИ

Korpel A., ed. Selected Papers on Acousto Optics. SPIE Optical Engineering Press (Milestone Series. Vol. 16), 1990.

Special issue on acoustooptic signal processing. Proceedings of the IEEE. Vol. 69, № 1, 1981.

Задачи

К РАЗДЕЛУ 19.1

1. **Дифракция света от различных периодических структур.** Обсудите дифракцию оптической плоской волны с длиной волны λ от следующих периодических структур, отметив в каждом случае геометрическую конфигурацию и частотный сдвиг.

а. Акустическая бегущая волна с длиной волны Λ.

б. Акустическая стоячая волна с длиной волны Λ.

в. Прозрачная градиентная среда с синусоидально меняющимся показателем преломления (с периодом Λ).

г. Слоистая периодическая среда с периодом Λ из параллельных чередующихся слоев двух материалов с различными показателями преломления (см. подразд. 7.1.3).

2. **Дифракция Брэгга как процесс рассеяния.** Падающая оптическая волна с угловой частотой ω, волновым вектором $\boldsymbol{k}$ и комплексной огибающей A взаимодействует со средой, возмущаемой акустической волной с угловой частотой Ω и волновым вектором $\boldsymbol{q}$, создавая источник S, описываемый формулой (19.33). Угол θ соответствует брэгговской дифракции со смещением частоты вверх, так что источник рассеянного света $S = \mathrm{Re}\{S_r(\boldsymbol{r})\exp(j\omega_r t)\}$, где

$$S_r(\boldsymbol{r}) = -\frac{\Delta n_0}{n} k_r^2 A \exp(-j\boldsymbol{k}_r \cdot \boldsymbol{r}); \quad \omega_r = \omega + \Omega; \quad \boldsymbol{k}_r = \boldsymbol{k} + \boldsymbol{q}.$$

Этот источник излучает рассеянное поле E. Считая, что падающая волна не ослабляется в результате акустооптического взаимодействия (первое борновское приближение, т. е. A остается приблизительно постоянным), можно получить поле рассеянного света путем решения уравнения Гельмгольца $\nabla^2 E + k^2 E = -S$. Это уравнение имеет решение в дальней зоне (см. задачу 5 к разд. 21.1)

$$E(\boldsymbol{r}) \approx \frac{\exp(-jkr)}{4\pi r} \int_V S_r(\boldsymbol{r}') \exp(jk\hat{\boldsymbol{r}} \cdot \boldsymbol{r}')\, d\boldsymbol{r}',$$

где $\hat{\boldsymbol{r}}$ — единичный вектор в направлении $\boldsymbol{r}$; $k = 2\pi/\lambda$; V — объем источника. Используйте данное уравнение для вывода выражения для коэффициента отражения акустооптической ячейки при выполнении условия Брэгга. Сравните результат с (19.22).

3. **Условие дифракции Рамана—Ната.** Выведите выражение для максимальной ширины D_s акустического пучка с длиной волны Λ, при которой возможна дифракция Рамана—Ната для света с длиной волны λ (см. рис. 19.13).

4. **Комбинированная акустооптическая и электрооптическая модуляция.** Один конец кристалла ниобата лития ($LiNbO_3$) помещен в микроволновый резонатор с электромагнитным полем частотой 3 ГГц. В результате пьезоэлектрического эффекта (деформации материала электрическим полем) запускается акустическая волна. Свет от He—Ne-лазера ($\lambda_0 = 633$ нм) отражается от акустичес-

кой волны. Показатель преломления равен $n = 2{,}3$, а скорость звука $v_s = 7{,}4$ км/с. Определите угол Брэгга. Поскольку ниобат лития является также электрооптическим материалом, приложенное электрическое поле модулирует показатель преломления, который, в свою очередь, модулирует фазу падающего света. Изобразите спектр отраженного света. Если микроволновое электрическое поле представляет собой импульс малой длительности, изобразите спектр отраженного света в различные моменты времени, выделяя вклады электрооптического и акустооптического эффектов.

К РАЗДЕЛУ 19.2

1. **Акустооптическая модуляция**. Придумайте систему для преобразования монохроматической оптической волны с комплексной волновой функцией $U(t) = A\exp(j\omega t)$ в модулированную волну с комплексной волновой функцией $A\cos(\Omega t)\exp(j\omega t)$ при помощи акустооптической ячейки с акустической волной $s(x, t) = S_0\cos(\Omega t - qx)$

Указание. Рассмотрите случаи брэгговского отражения со сдвигом частоты вверх и вниз.

2. **Брэгговский отражатель, свободный от частотного сдвига**. Спроектируйте акустооптическую систему, отклоняющую свет без внесения частотного сдвига.

Указание. Используйте две брэгговские ячейки.

К РАЗДЕЛУ 19.3

1. **Дифракция Брэгга вперед**. Поперечная акустическая волна с длиной волны Λ распространяется в направлении x в одноосном кристалле с показателями преломления n_o и $n_н$ и оптической осью в направлении z. Выведите выражение для длины волны λ_0 падающего света, распространяющегося в направлении x и поляризованного вдоль z, которая удовлетворяет условию дифракции Брэгга. Какова поляризация волны, отражаемой вперед? Определите Λ, если $\lambda_0 = 633$ нм, $n_н = 2{,}200$ и $n_o = 2{,}286$.

ГЛАВА 20 ЭЛЕКТРООПТИКА



Фридрих Поккельс (1865—1913) описал линейный электрооптический эффект в 1893 г.



Джон Керр (1824—1907) открыл квадратичный электрооптический эффект в 1875 г.

Некоторые прозрачные материалы меняют свои оптические свойства под действием электрического поля. Это является результатом действия сил, которые искажают положение, ориентацию или форму молекул, образующих материал. Электрооптический эффект — это изменение показателя преломления, вызванное наложением постоянного или низкочастотного электрического поля (рис. 20.1). Электрическое поле, приложенное к анизотропному оптическому материалу, меняет его показатели преломления и, таким образом, действие материала на пропускаемый через него поляризованный свет.

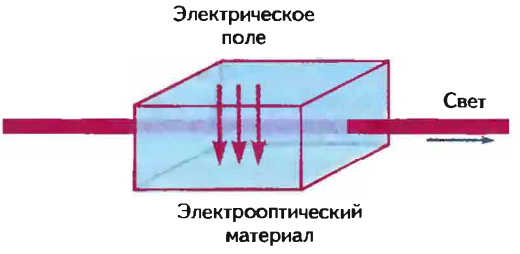


Рис. 20.1. Стационарное электрическое поле, приложенное к электрооптическому материалу, меняет его показатель преломления. Это, в свою очередь, изменяет действие материала на свет, распространяющийся сквозь него. Таким образом, электрическое поле управляет светом

Зависимость показателя преломления от приложенного электрического поля обычно принимает одну из двух следующих форм.

• Показатель преломления меняется пропорционально приложенному электрическому полю — эффект, известный как **линейный электрооптический эффект**, или **эффект Поккельса**.

• Показатель преломления меняется пропорционально квадрату приложенного электрического поля — **квадратичный электрооптический эффект**, или **эффект Керра**.

Изменение показателя преломления в типичном случае мало. Тем не менее фаза оптической волны, распространяющейся сквозь электрооптическую среду, может измениться существенно, если длина пути значительно превышает длину волны света. Например, если показатель преломления увеличивается на 10^{-5} за счет внешнего электрического поля, то оптическая волна, проходящая расстояние в 10^5 длин волн, приобретает дополнительный фазовый сдвиг 2π.

Материалы, показатель преломления которых можно менять за счет приложенного электрического поля, находят применение при производстве электрически управляемых оптических приборов, например:

• линза из материала, показатель которого можно менять, — линза с управляемым фокусным расстоянием;

• призма, способность которой к отклонению луча управляется внешним полем, может использоваться как оптический сканер;

• свет, пропущенный через прозрачную пластинку с управляемым показателем преломления, претерпевает управляемую задержку по фазе. Такая пластинка может использоваться как оптический модулятор фазы;

• анизотропный кристалл, показатели преломления которого можно менять, служит управляемой фазовой пластинкой, которая необходима для изменения поляризационных свойств света;

• фазовая пластинка, помещенная между двумя скрещенными поляризаторами, обеспечивает пропускание света с интенсивностью, зависящей от фазовой задержки между компонентами (см. подразд. 6.6.2). Следовательно, коэффициент пропускания такого прибора управляется электрическим полем, поэтому его можно использовать как оптический модулятор интенсивности или переключатель.

Управляемые элементы такого рода находят важные применения в оптической связи и обработке оптических сигналов.

Кроме сказанного ранее, электрическое поле может менять оптические свойства материала за счет поглощения. Полупроводниковые материалы в нормальном состоянии оптически прозрачны для света, длина волны которого больше, чем длина волны, соответствующая ширине запрещенной зоны (см. подразд. 16.2.2). Однако приложенное электрическое поле может уменьшить ширину запрещенной зоны материала, создавая возможность поглощения и превращая материал из прозрачного в непрозрачный. Этот эффект, называемый **электропоглощением**, применяется в оптических модуляторах и переключателях.

О данной главе

Мы начнем с описания электрооптического эффекта и принципов электрооптической модуляции и сканирования. Это первое знакомство в разд. 20.1 упрощено за счет переноса детального рассмотрения эффектов анизотропии в разд. 20.2.

Раздел 20.3 посвящен электрооптическим свойствам жидких кристаллов. Электрическое поле, действующее на молекулы жидкого кристалла, вызывает изменение их ориентации. Это ведет к изменениям оптических свойств среды, т. е. она демонстрирует электрооптический эффект. Молекулы скрученного нематического жидкого кристалла организованы в спиральную структуру, которая в нормальных условиях действует как вращатель плоскости поляризации. Внешнее электрическое поле можно использовать для устранения спиральной структуры и, следовательно, оптической активности материала. При выключении электрического поля материал восстанавливает свою первоначальную спиральную структуру. Таким образом, прибор работает как динамический вращатель плоскости поляризации. Применение дополнительных фиксированных поляризаторов позволяет использовать такой вращатель поляризации в качестве модулятора интенсивности или переключателя. Такое поведение лежит в основе действия большинства элементов жидкокристаллических дисплеев.

Электрооптические свойства фоторефрактивных сред рассматриваются в разд. 20.4. Это материалы, в которых поглощение света создает внутреннее электрическое поле, которое, в свою очередь, инициирует электрооптический эффект, меняющий оптические свойства среды. В результате оптическими свойствами среды управляет падающий на нее свет. Таким образом, фоторефрактивные приборы позволяют осуществлять *управление света светом.*

Наконец, в разд. 20.5 приводится краткое введение в физику электропоглощения.

20.1. ПРИНЦИПЫ ЭЛЕКТРООПТИКИ

20.1.1. Эффекты Поккельса и Керра

Показатель преломления электрооптической среды является функцией $n(E)$ приложенного стационарного или медленно меняющегося электрического поля E. Функция $n(E)$ слабо зависит от E, так что ее можно разложить в ряд Тейлора в окрестности точки $E = 0$

$$n(E) = n + a_1 E + \frac{1}{2} a_2 E^2 + \ldots, \tag{20.1}$$

где коэффициенты разложения

$$n = n(0); \quad a_1 = \left.\frac{\mathrm{d}n}{\mathrm{d}E}\right|_{E=0}; \quad a_2 = \left.\frac{\mathrm{d}^2 n}{\mathrm{d}E^2}\right|_{E=0}$$

По причинам, которые станут ясны далее, удобно выразить (20.1) через два новых коэффициента

$$\mathfrak{r} = -\frac{2a_1}{n^3}; \quad \sigma = -\frac{a_2}{n^3},$$

называемые электрооптическими коэффициентами, так что

$$n(E) = n - \frac{1}{2}\mathfrak{r}n^3E - \frac{1}{2}\mathfrak{s}n^3E^2 + \ldots . \tag{20.2}$$

Члены второго и высших порядков этого ряда, как правило, на много порядков меньше, чем n. Членами высших порядков можно с уверенностью пренебречь.

Для будущего использования удобно вывести выражение для диэлектрической непроницаемости

$$\eta = \frac{\varepsilon_0}{\varepsilon} = \frac{1}{n^2}$$

электрооптической среды как функции E. Параметр η используется для описания оптических свойств анизотропных сред (см. подразд. 6.3.1). Малое приращение

$$\Delta\eta = \frac{\mathrm{d}\eta}{\mathrm{d}n}\Delta n = -\frac{2}{n^3}\left(-\frac{1}{2}\mathfrak{r}n^3E - \frac{1}{2}\mathfrak{s}n^3E^2\right) = \mathfrak{r}E + \mathfrak{s}E^2,$$

так что

$$\eta(E) \approx \eta + \mathfrak{r}E + \mathfrak{s}E^2, \tag{20.3}$$

где $\eta = \eta(0)$. Следовательно, электрооптические коэффициенты $\mathfrak{r}$ и $\mathfrak{s}$ — это коэффициенты пропорциональности двух слагаемых $\Delta\eta$ величинам E и E^2 соответственно. Этим объясняется целесообразность на первый взгляд странного их определения в формуле (20.2).

Значения коэффициентов $\mathfrak{r}$ и $\mathfrak{s}$ зависят от направления приложенного электрического поля и поляризации света, что будет обсуждаться в разд. 20.2.

Эффект Поккельса

Во многих материалах третий член в (20.2) пренебрежимо мал по сравнению со вторым, в силу чего

$$\boxed{n(E) \approx n - \frac{1}{2}\mathfrak{r}n^3E,} \tag{20.4}$$

Эффект Поккельса

что иллюстрируется рис. 20.2, *а*. Среда в таком случае называется средой (ячейкой) Поккельса. Коэффициент $\mathfrak{r}$ называется **коэффициентом Поккельса**, или линейным электрооптическим коэффициентом. Типичные значения $\mathfrak{r}$ лежат в интервале от 10^{-12} до 10^{-10} м/В (от 1 до 100 пм/В). Например, для $E = 10^6$ В/м

(10 кВ приложено к ячейке толщиной 1 см) член $\varsigma n^3 E/2$ в (20.4) имеет порядок от 10^{-6} до 10^{-4}. Изменения показателя преломления, наведенные электрическим полем, действительно очень малы. Обычно в ячейках Поккельса используются кристаллы $NH_4H_2PO_4$ (ADP), KH_4PO_4 (KDP), $LiNbO_3$, $LiTaO_3$ и CdTe.

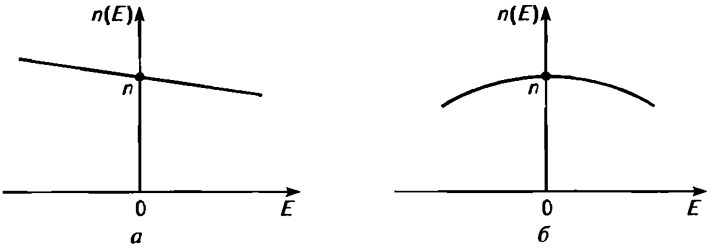


Рис. 20.2. Зависимость показателя преломления от электрического поля: *а* — среда Поккельса; *б* — среда Керра

Эффект Керра

Если материал обладает центром симметрии, как в случае газов, жидкостей и некоторых кристаллов, то $n(E)$ — четная симметричная функция (рис. 20.2, *б*), так как она должна быть инвариантна относительно обращения знака E. Ее первая производная в этом случае обращается в нуль, нулю равен и коэффициент ς, вследствие чего

$$\boxed{n(E) \approx n - \frac{1}{2} s n^3 E^2.} \qquad (20.5)$$

Эффект Керра

В этом случае материал называется керровским (ячейка Керра). Параметр s называется **коэффициентом Керра** s, или квадратичным электрооптическим коэффициентом. Типичные значения s лежат в интервале от 10^{-18} до 10^{-14} м^2/В^2 для кристаллов и от 10^{-22} до 10^{-19} м^2/В^2 для жидкостей. Например, для $E = 10^6$ В/м член $sn^3E^2/2$ в (20.5) имеет порядок от 10^{-6} до 10^{-2} в кристаллах и от 10^{-10} до 10^{-7} в жидкостях.

20.1.2. Электрооптические модуляторы и переключатели

Модуляторы фазы

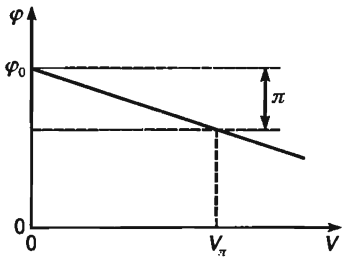
Пучок света, проходящий ячейку Поккельса длиной L, к которой приложено электрическое поле E, претерпевает фазовый сдвиг

$$\varphi = n(E) k_0 L = 2\pi n(E) \frac{L}{\lambda_0},$$

где λ_0 — длина волны света в свободном пространстве. С использованием (20.4) имеем

$$\varphi \approx \varphi_0 - \pi \frac{\varsigma n^3 E L}{\lambda_0}, \qquad (20.6)$$

где $\varphi_0 = 2\pi nL/\lambda_0$. Если электрическое поле получается путем приложения напряжения V к двум граням ячейки, расстояние между которыми d, то $E = V/d$ и (20.6) дает



$$\varphi = \varphi_0 - \pi \frac{V}{V_\pi}, \qquad (20.7)$$

Фазовая модуляция

где

$$V_\pi = \frac{d}{L}\frac{\lambda_0}{\tau n^3}. \qquad (20.8)$$

Полуволновое напряжение

Параметр V_π, называемый **полуволновым напряжением**, представляет собой напряжение, при котором фазовый сдвиг меняется на π. Уравнение (20.7) выражает линейную связь между оптическим сдвигом фазы и напряжением. Таким образом, можно модулировать фазу, изменяя напряжение V, приложенное к среде, через которую пропускается свет. Параметр V_π является важной характеристикой модулятора. Он зависит от свойств материала (n и τ), длины волны λ_0 и отношения геометрических размеров d/L.

Электрическое поле может быть приложено в направлении, перпендикулярном направлению распространения света (поперечный модулятор) или параллельном ему (продольный модулятор). В последнем случае $d = L$ (рис. 20.3). Величина электрооптического коэффициента τ зависит от направлений распространения света и приложенного поля, так как кристалл, вообще говоря, анизотропен (как объясняется в разд. 20.2). Типичные значения полуволнового напряжения лежат в интервале от одного до нескольких киловольт для продольных модуляторов и сотни вольт для поперечных.

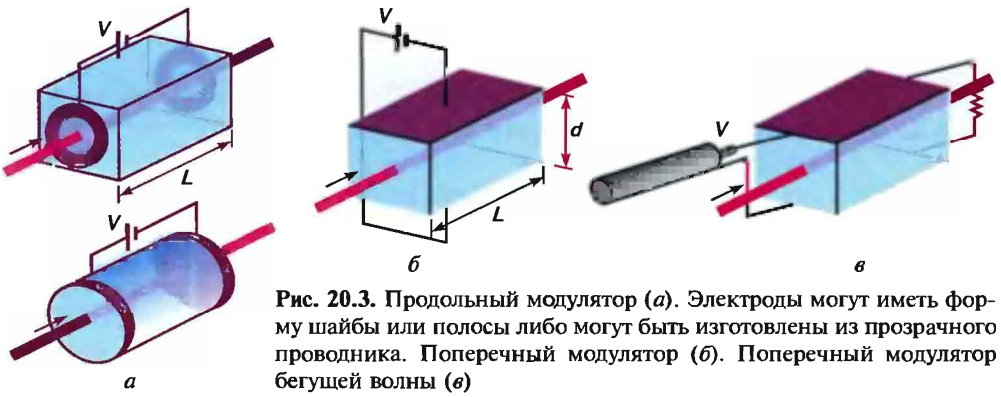


Рис. 20.3. Продольный модулятор (*а*). Электроды могут иметь форму шайбы или полосы либо могут быть изготовлены из прозрачного проводника. Поперечный модулятор (*б*). Поперечный модулятор бегущей волны (*в*)

Скорость работы электрооптического модулятора ограничена электрическими емкостными эффектами и временем прохода света через материал. Если электрическое поле $E(t)$ заметно меняется за время прохода света T, то бегущая оптическая волна будет подвергаться действию электрического поля различной величины в течение времени распространения в кристалле. Модулированная фаза в данный момент времени t тогда будет пропорциональна среднему электрическому полю $E(t)$ за время от $t - T$ до t. В результате ограниченная временем прохода ширина полосы модуляции равна $\approx 1/T$. Один из способов уменьшения этого времени состоит в подаче напряжения V на один конец кристалла, в то время как электроды служат линией передачи, как показано на рис. 20.3, *в*. Если скорость бегущей электрической волны совпадет со скоростью оптической волны, то эффекты, связанные с конечностью времени прохода, в принципе можно исключить. Промышленные модуляторы с конструкциями, показанными на рис. 20.3, как правило, работают на нескольких сотнях мегагерц, однако возможны скорости модуляции и до нескольких гигагерц.

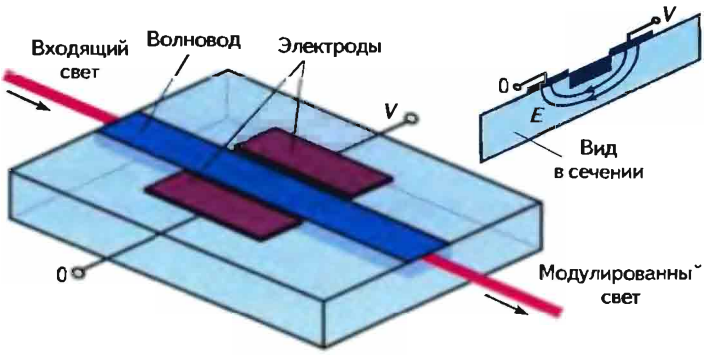


Рис. 20.4. Интегрально-оптический фазовый модулятор с использованием электрооптического эффекта

Электрооптические модуляторы можно изготовить также как интегрально-оптические устройства. Такие приборы работают на более высоких частотах и при более низких напряжениях, чем объемные приборы. Оптический волновод изготавливается на электрооптической подложке (часто из $LiNbO_3$) путем прямой диффузии материала, такого как титан, для увеличения показателя преломления. Электрическое поле прикладывается к волноводу с помощью электродов, как показано на рис. 20.4. Так как конфигурация является поперечной, а ширина волновода много меньше его длины ($d \ll L$), полуволновое напряжение может быть мало и составлять несколько вольт. Такие модуляторы достигают частот, превышающих 100 ГГц. Свет удобно вводить и выводить из модулятора при помощи оптических волокон.

Динамические фазовые пластинки

Анизотропная среда имеет две линейно поляризованные нормальные моды, распространяющиеся с различными скоростями c_0/n_1 и c_0/n_2 (см. подразд. 6.3.2). Если среда проявляет эффект Поккельса, то в присутствии стаци-

онарного электрического поля E оба показателя преломления изменяются в соответствии (20.4), т. е.

$$n_1(E) \approx n_1 - \frac{1}{2}\mathfrak{r}_1 n_1^3 E; \tag{20.9}$$

$$n_2(E) \approx n_2 - \frac{1}{2}\mathfrak{r}_2 n_2^3 E, \tag{20.10}$$

где $\mathfrak{r}_1$ и $\mathfrak{r}_2$ — соответствующе коэффициенты Поккельса (детально эффекты анизотропии исследуются в разд. 20.2). После прохождения расстояния L моды приобретают относительную задержку по фазе, выражаемую формулой

$$\Gamma = k_0\left[n_1(E) - n_2(E)\right]L = k_0(n_1 - n_2)L - \frac{1}{2}k_0\left(\mathfrak{r}_1 n_1^3 - \mathfrak{r}_2 n_2^3\right)EL, \tag{20.11}$$

Если получается путем приложения напряжения между двумя поверхностями среды, разделенных расстоянием, то (20.11) можно записать в компактном виде как

$$\boxed{\Gamma = \Gamma_0 - \pi\frac{V}{V_\pi},} \tag{20.12}$$

Фазовая задержка

где $\Gamma = k_0(n_1 - n_2)L$ — фазовая задержка в отсутствие электрического поля; V_π — напряжение, необходимое для получения фазовой задержки, равной π,

$$\boxed{V_\pi = \frac{d}{L}\frac{\lambda_0}{\mathfrak{r}_1 n_1^3 - \mathfrak{r}_2 n_2^3}.} \tag{20.13}$$

Напряжение задержки в полволны

Формула (20.12) показывает, что фазовая задержка линейно связана с приложенным напряжением. Среда служит электрически управляемой динамической фазовой пластинкой.

Модуляторы интенсивности: использование фазового модулятора в интерферометре

Фазовая задержка сама по себе не влияет на интенсивность светового пучка. Однако фазовый модулятор, помещенный в одно из плеч интерферометра, может действовать как модулятор интенсивности. Рассмотрим, например, интерферометр Маха—Цендера, показанный на рис. 20.5. Если светоделители делят оптическую мощность ровно пополам, то интенсивность, проходящая через выходное плечо интерферометра I_0, связана с интенсивностью падающего света I_i соотношением

$$I_0 = \frac{1}{2}I_i + \frac{1}{2}I_i\cos\varphi = I_i\cos^2\left(\frac{\varphi}{2}\right), \tag{20.14}$$

где $\varphi = \varphi_1 - \varphi_2$ — разность между фазовыми сдвигами, приобретаемыми светом при прохождении двух плеч (см. подразд. 2.5.1). Коэффициент пропускания интерферометра равен

$$\mathcal{T} = \frac{I_0}{I_i} = \cos^2\left(\frac{\varphi}{2}\right).$$

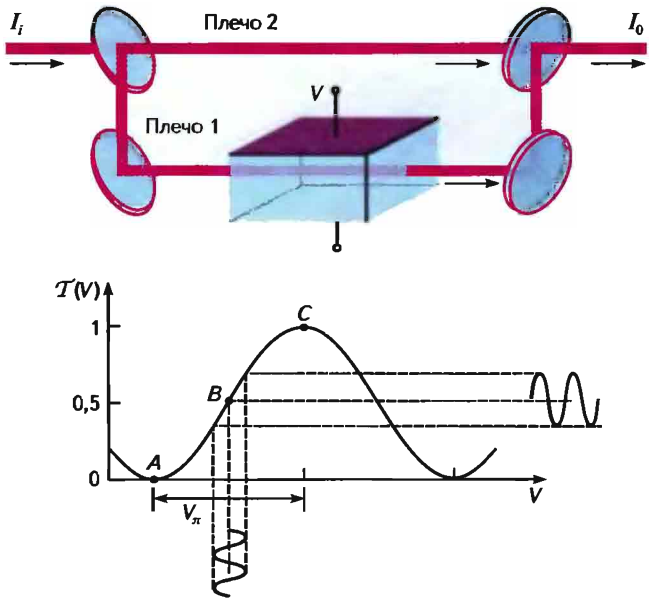


Рис. 20.5. Фазовый модулятор, помещенный в одно из плеч интерферометра Маха—Цендера, может служить модулятором интенсивности. Коэффициент пропускания интерферометра $\mathcal{T}(V) = I_0/I_i$ меняется периодически с ростом приложенного напряжения V. При работе в ограниченной области вблизи точки B прибор ведет себя как линейный модулятор интенсивности. Если переключать V между точками A и C, то прибор действует как оптический переключатель

Из-за присутствия фазового модулятора в плече 1, согласно (20.7), имеем $\varphi_1 = \varphi_{10} - \pi V/V_\pi$, так что φ управляется приложенным напряжением V в соответствии с линейным соотношением $\varphi = \varphi_1 - \varphi_2 = \varphi_0 - \pi V/V_\pi$, где $\varphi_0 = \varphi_{10} - \varphi_2$ зависит от оптической разности хода. Следовательно, коэффициент пропускания прибора является функцией приложенного напряжения V,

$$\mathcal{T}(V) = \cos^2\left(\frac{\varphi_0}{2} - \frac{\pi}{2}\frac{V}{V_\pi}\right). \qquad (20.15)$$

Коэффициент пропускания

График этой функции приведен на рис. 20.5 для произвольной величины φ_0. Прибор может функционировать как линейный модулятор интенсивности при

подборе разности оптических длин пути, для которого $\varphi_0 = \pi/2$, и выборе почти линейного участка вблизи $\mathcal{T} = 0{,}5$. В другом варианте разность оптических длин пути выбирается так, чтобы φ_0 было целым кратным 2π. В этом случае $\mathcal{T}(0) = 1$, а $\mathcal{T}(V_\pi) = 0$, так что модулятор включает и выключает свет при переключении V между 0 и V_π.

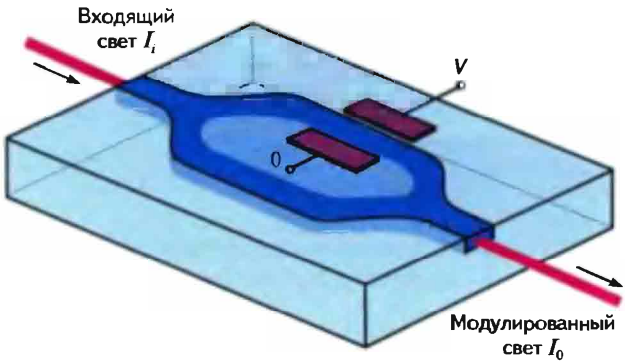


Рис. 20.6. Интегрально-оптический модулятор интенсивности (или оптический переключатель). Интерферометр Маха—Цендера и электрооптический фазовый модулятор выполнены с использованием оптических волноводов, изготовленных из такого материала, как $LiNbO_3$

Модулятор интенсивности Маха—Цендера можно построить также в виде интегрально-оптического устройства. Волноводы размещаются на подложке по схеме, показанной на рис. 20.6. Светоделители реализуются в виде волноводных Y-разветвителей. Ввод и вывод света производится по оптическим волокнам. Промышленные интегрально-оптические модуляторы обычно работают на частотах в несколько гигагерц, однако в отдельных случаях были достигнуты и частоты, превышающие 25 ГГц.

Модуляторы интенсивности: фазовая пластинка между скрещенными поляризаторами

Как описано в подразд. 6.6.3, фазовая пластинка (с фазовой задержкой Γ), помещенная между двумя скрещенными поляризаторами, расположенными под углом 45° к осям пластинки (см. рис. 6.40), обладает коэффициентом пропускания по интенсивности $\mathcal{T} = \sin^2(\Gamma/2)$. Если в качестве пластинки взять ячейку Поккельса, то Γ линейно зависит от приложенного напряжения V в соответствии с формулой (20.12). Тогда коэффициент пропускания прибора является периодической функцией напряжения V

$$\mathcal{T}(V) = \sin^2\left(\frac{\Gamma_0}{2} - \frac{\pi}{2}\frac{V}{V_\pi}\right), \qquad (20.16)$$

Коэффициент пропускания

как показано на рис. 20.7. Изменяя V, можно варьировать коэффициент пропускания между 0 (затвор закрыт) и 1 (затвор открыт). Прибор можно использовать

также как линейный модулятор, если работать в области около $\mathcal{T}(V) = 0{,}5$. Выберем $\Gamma_0 = \pi/2$ и $V \ll V_\pi$, тогда

$$\mathcal{T}(V) = \sin^2\left(\frac{\pi}{4} - \frac{\pi}{2}\frac{V}{V_\pi}\right) \approx \mathcal{T}(0) + \left.\frac{d\mathcal{T}}{dV}\right|_{V=0} V = \frac{1}{2} - \frac{\pi}{2}\frac{V}{V_\pi}, \qquad (20.17)$$

так что $\mathcal{T}(V)$ — линейная функция с наклоном $\pi/2V_\pi$, представляющим чувствительность модулятора. Фазовая задержка может быть установлена либо оптическим путем (снабдив модулятор дополнительным устройством задержки фазы — компенсатором), либо электрическим — путем добавления к V постоянного напряжения смещения.

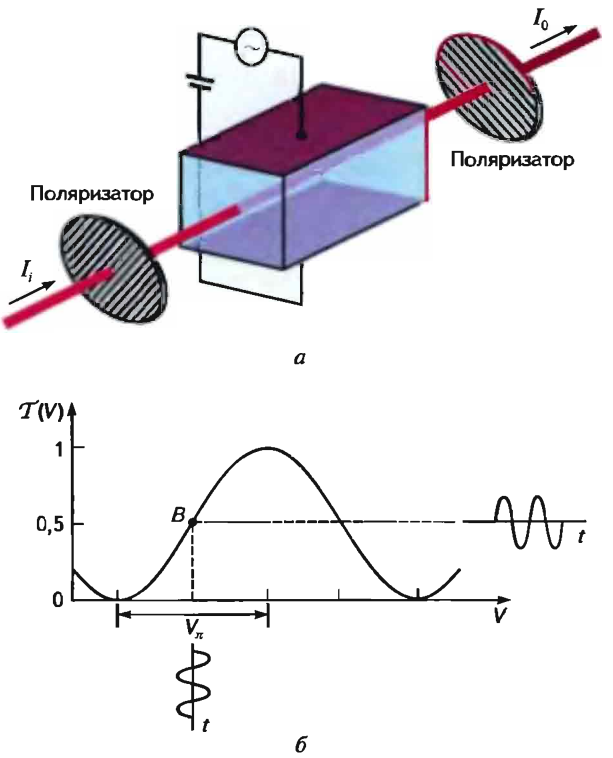


Рис. 20.7. Оптический модулятор интенсивности с ячейкой Поккельса, помещенной между двумя скрещенными поляризаторами (*а*). Оптический коэффициент пропускания в зависимости от приложенного напряжения при произвольном значении; для линейного действия на ячейку накладывается дополнительное напряжение смещения, чтобы сделать *B* рабочей точкой (*б*)

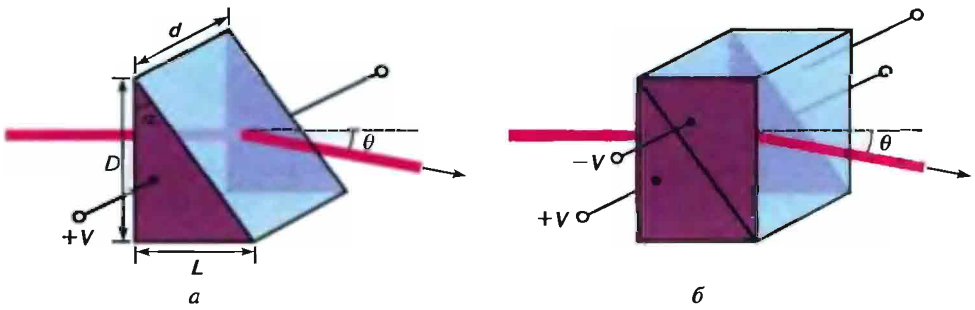
На практике максимальное пропускание модулятора меньше единицы из-за потерь, обусловленных отражением, поглощением и рассеянием. Минимальное пропускание больше нуля из-за разъюстировки направлений распространения и поляризации по отношению к осям кристалла и поляризаторов. Отношение максимального и минимального коэффициентов пропускания называется коэффициентом контрастности. Возможные значения этого коэффициента могут превышать 30 дБ (1000 : 1).

20.1.3. Сканеры

Оптический пучок можно отклонять в динамическом режиме путем использования призмы с электрически управляемым показателем преломления. Угол отклонения, вносимый призмой с малым углом α при вершине и показателем преломления n, равен $\theta \approx (n-1)\alpha$ [см. (1.10)]. Малое приращение показателя преломления Δn, вызванное приложенным электрическим полем E, соответствует малому приращению угла отклонения

$$\Delta\theta = \alpha\Delta n = -\frac{1}{2}\alpha r n^3 E = -\frac{1}{2}\alpha r n^3 \frac{V}{d}. \qquad (20.18)$$

Часто бывает более удобно накладывать электроды треугольной формы, определяя таким образом призму в прямоугольном кристалле. Можно организовать каскад из двух или нескольких призм за счет изменения направления электрического поля, как показано на рис. 20.8, *б*.



а *б*

Рис. 20.8. Электрооптическая призма (*а*). Угол отклонения θ управляется приложенным напряжением. Электрооптическая двойная призма (*б*)

Важным параметром, характеризующим сканер, является разрешение, т. е. число независимых точек сканирования. Оптический пучок ширины D с длиной волны λ_0 имеет угловую расходимость $\delta\theta \approx \lambda_0/D$ [см. (4.40)]. Чтобы минимизировать этот угол, пучок должен быть настолько широким, насколько это возможно; в идеальном случае он должен перекрывать всю призму. Для заданного максимального напряжения V, соответствующего углу сканирования $\Delta\theta$, число независимых точек дается выражением

$$N \approx \frac{|\Delta\theta|}{\delta\theta} = \frac{(1/2)\alpha r n^3 V/d}{(\lambda_0/D)}. \qquad (20.19)$$

Подставляя

$$\alpha = \frac{L}{D}; \quad V_\pi = \frac{d}{L}\frac{\lambda_0}{r n^3},$$

получаем

$$N \approx \frac{V}{2V_\pi}, \tag{20.20}$$

откуда $V \approx 2NV_\pi$. Это обескураживающий результат. Чтобы просканировать N независимых точек, напряжение должно в $2N$ раз превышать свое полуволновое значение. Поскольку V_π обычно велико, создание полезного сканера с $N \gg 1$ требует неприемлемо высоких напряжений. Поэтому наиболее часто используемые сканеры — механические и акустооптические (см. подразд. 19.2.2 и 23.3.2).

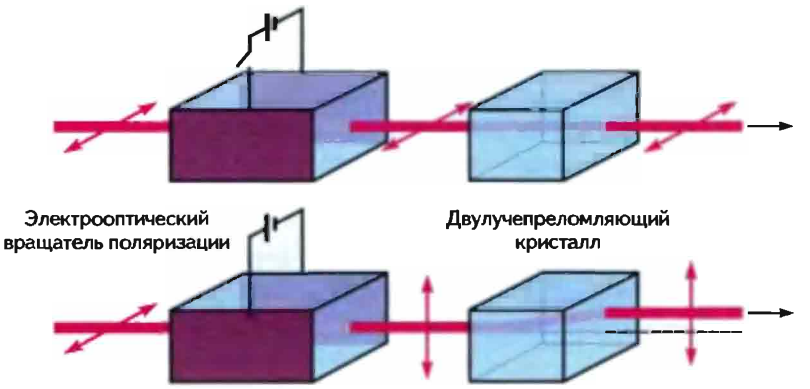


Рис. 20.9. Переключатель положения на основе электрооптической задержки фазы и двулучепреломления

Процесс двойного преломления в анизотропных кристаллах (см. подразд. 6.3.5) вносит поперечный сдвиг падающего пучка параллельно самому себе для одной поляризации и не вносит сдвига для другой. Этот эффект можно использовать для переключения пучка между двумя параллельными позициями посредством переключения его поляризации. Линейно поляризованный оптический пучок сначала пропускается через электрооптическую фазовую пластинку, которая действует как вращатель поляризации, а затем через кристалл. Вращатель управляет поляризацией электрически, определяя, будет ли пучок испытывать поперечное смещение, как показано на рис. 20.9.

20.1.4. Направленные ответвители

Важным приложением электрооптического эффекта является управление связью двух параллельных волноводов в интегрально-оптических устройствах. Электрическое поле может быть использовано для передачи света из одного волновода в другой, так что устройство действует как электрически управляемый направленный ответвитель.

Связь между двумя параллельными одномодовыми планарными волноводами (рис. 20.10, *а*) была изучена нами в подразд. 8.5.2. Было показано, что при распространении света мощности $P_1(z)$ и $P_2(z)$ вдоль волноводов происходит периодический по z обмен энергией между ними. Величина этого обмена управляется двумя параметрами: коэффициентом связи C (который зависит от размеров, длины волны и показателей преломления) и разностью постоянных распространения

$$\Delta\beta = \beta_1 - \beta_2 = \frac{2\pi\Delta n}{\lambda_0},$$

где Δn — разность между показателями преломления волноводов. Если волноводы одинаковы, $\Delta\beta = 0$, то при $P_2(0) = 0$ на расстоянии $z = L_0 = \pi/2C$, которое называется длиной участка связи, или длиной передачи, мощность полностью передается из волновода 1 в волновод 2, т. е. $P_1(L_0) = 0$ и $P_2(L_0) = P_1(0)$, как показано на рис. 20.10, *а*.

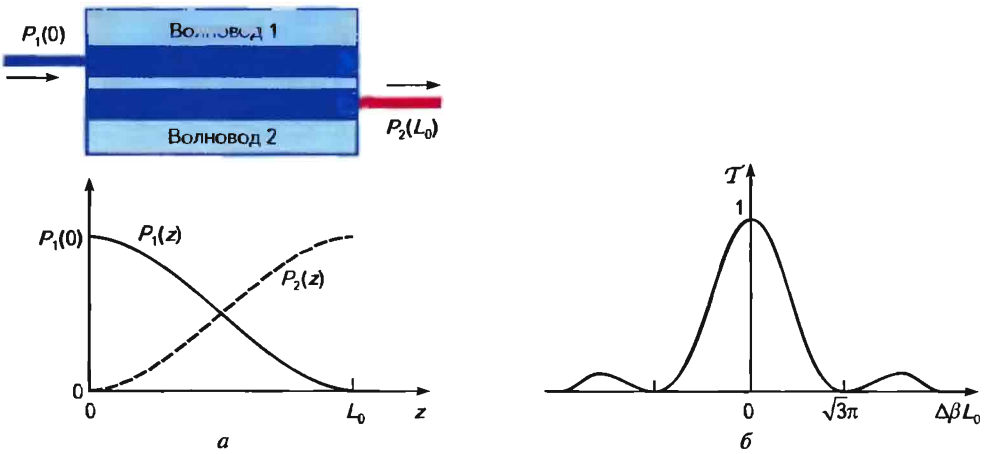


Рис. 20.10. Обмен энергией между двумя параллельными слабо связанными волноводами, которые идентичны и имеют одинаковые постоянные распространения β (*а*). При $z = 0$ вся мощность сосредоточена в волноводе 1. При $z = L_0$ она полностью переносится в волновод 2. Зависимость коэффициента передачи мощности $\mathcal{T} = P_2(L_0)/P_1(0)$ от параметра фазовой расстройки $\Delta\beta L_0$ (*б*)

Для волновода длины L_0 и при $\Delta\beta \neq 0$ коэффициент передачи мощности $\mathcal{T} = P_2(L_0)/P_1(0)$ является функцией фазовой расстройки [см. (8.58а)]

$$\mathcal{T} = \frac{\pi^2}{4}\operatorname{sinc}^2\left[\frac{1}{2}\sqrt{1+\left(\frac{\Delta\beta L_0}{\pi}\right)^2}\right], \qquad (20.21)$$

где

$$\operatorname{sinc}(x) \equiv \frac{\sin\pi x}{\pi x}.$$

На рис. 20.10, *б* показана эта зависимость. Отношение имеет максимальное значение единица при $\Delta\beta L_0 = 0$, убывает с ростом $\Delta\beta L_0$ и обращается в нуль при $\Delta\beta L_0 = \sqrt{3}\pi$. В последнем случае мощность *не передается* в волновод 2.

Зависимость передаваемой мощности от фазовой расстройки является ключом к созданию электрически управляемых направленных ответвителей. Если расстройка $\Delta\beta L_0$ переключается с 0 на $\sqrt{3}\pi$, свет остается в волноводе 1. Электрическое управление величиной $\Delta\beta$ достигается за счет электрооптического эффекта. Электрическое поле, наложенное на один или два идентичных в его отсутствие волновода, меняет показатель преломления на $\Delta n = -n^3 r E/2$, где r — коэффициент Поккельса. Это приводит к сдвигу фазы

$$\Delta\beta L_0 = \Delta n \frac{2\pi L_0}{\lambda_0} = -\frac{\pi}{\lambda_0} n^3 r L_0 E.$$

Типичный электрооптический направленный ответвитель имеет геометрию, показанную на рис. 20.11. Электроды наложены на два волновода, расстояние между которыми равно d. Приложенное напряжение V создает электрическое поле $E \approx V/d$ в одном волноводе и $-V/d$ — в другом, где d — эффективное расстояние, определяемое путем решения электростатической задачи (линии электрического поля идут вниз в одном волноводе и вверх — в другом). Показатель преломления увеличивается в одном волноводе и уменьшается в другом. В результате полная разность показателей преломления составляет $2\Delta n = -n^3 r(V/d)$, что соответствует фазовой расстройке

$$\Delta\beta L_0 = -\frac{2\pi L_0}{\lambda_0} n^3 r \frac{L_0}{d} V,$$

пропорциональной приложенному напряжению V.

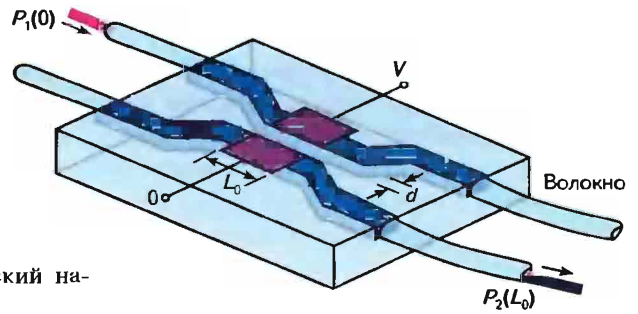


Рис. 20.11. Интегрально-оптический направленный ответвитель

Напряжение V_0, необходимое для переключения оптической мощности, должно обеспечивать выполнение условия $|\Delta\beta L_0| = \sqrt{3}\pi$, т. е.

$$V_0 = \sqrt{3}\frac{d}{L_0}\frac{\lambda_0}{2n^3 r} = \frac{\sqrt{3}}{\pi}\frac{C\lambda_0 d}{n^3 r}, \qquad (20.22)$$

где $L_0 = \pi/2C$; C — коэффициент связи. Оно называется **напряжением переключения.**

Поскольку

$$|\Delta\beta L_0| = \sqrt{3}\pi\frac{V}{V_0},$$

выражение (20.21) дает

$$\boxed{\mathcal{T} = \frac{\pi^2}{4}\operatorname{sinc}^2\left[\frac{1}{2}\sqrt{1+3\left(\frac{V}{V_0}\right)^2}\right].} \qquad (20.23)$$

Эффективность передачи мощности

Это выражение определяет эффективность передачи мощности как функцию приложенного напряжения V; график этой функции показан на рис. 20.12.

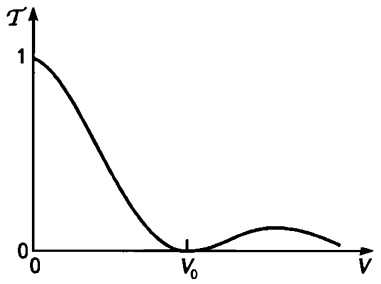


Рис. 20.12. Зависимость коэффициента передачи оптической мощности от приложенного напряжения V. Когда $V = 0$, вся оптическая мощность переходит из волновода 1 в волновод 2, когда $V = V_0$, вся оптическая мощность остается в волноводе 1

Электрооптический направленный ответвитель характеризуется длиной передачи L_0, которая обратно пропорциональна коэффициенту связи C, и напряжением переключения V_0, которое прямо пропорционально C. Таким образом, ключевым параметром является коэффициент C, определяемый геометрией и показателями преломления.

Интегрально-оптические направленные ответвители можно изготавливать, например, путем диффузии титана в подложки из $LiNbO_3$ высокой степени чистоты. Переключающее напряжение V_0, как правило, не превышает 10 В, а рабочая частота переключения может превышать 10 ГГц. Световые пучки фокусируются в пятна размером в несколько микрометров. Концы волновода могут наглухо присоединяться к одномодовым оптическим волокнам, сохраняющим поляризацию (см. подразд. 9.1.3). Дальнейшего увеличения частоты переключения можно достичь с помощью версии данного прибора, работающей в режиме бегущей волны.

Упражнение 20.1

Спектральный отклик эффективности передачи

Уравнение (20.22) отражает тот факт, что переключающее напряжение V_0 пропорционально длине волны. Предположим, что приложенное напряжение $V = V_0$ для некоторой конкретной длины волны $\bar{\lambda}_0$, так что эффективность

передачи мощности $\mathcal{T} = 0$ на $\bar{\lambda}_0$. Пусть теперь длина волны падающего света равна другому значению λ_0. Постройте график зависимости $\mathcal{T}$ от разности $\lambda_0 - \bar{\lambda}_0$. Считайте, что коэффициент связи и показатель преломления приближенно не зависят от длины волны.

20.1.5. Пространственные модуляторы света

Пространственный модулятор света — это прибор, который модулирует интенсивность света так, чтобы в каждой заданной точке она была изменена в заданное число раз (рис. 20.13). Это плоский оптический элемент с управляемым коэффициентом пропускания интенсивности $\mathcal{T}(x, y)$. Интенсивность прошедшего света $I_0(x, y)$ связана с интенсивностью падающего $I_i(x, y)$ формулой $I_0(x, y) = I_i(x, y)\mathcal{T}(x, y)$. Если падающий свет однороден [т. е. $I_i(x, y)$ постоянна], то интенсивность прошедшего света пропорциональна $\mathcal{T}(x, y)$. В этом случае «изображение» $\mathcal{T}(x, y)$ сообщается пропущенному свету, что в значительной степени напоминает «считывание» изображения, хранимого на транспаранте, при его равномерном освещении в слайд-проекторе. Однако в пространственном модуляторе $\mathcal{T}(x, y)$ является управляемым. В электрооптическом модуляторе это управление осуществляется электрическим сигналом.

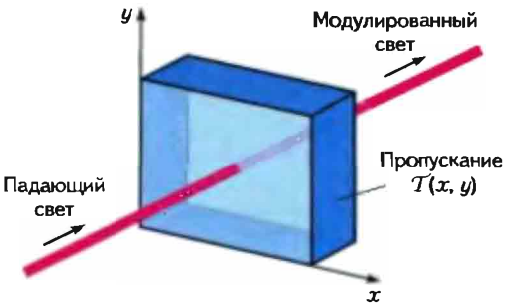


Рис. 20.13. Пространственный модулятор света

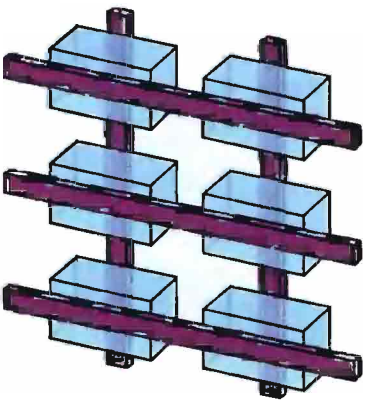


Рис. 20.14. Электрически адресуемая матрица продольных электрооптических модуляторов

Для построения пространственного модулятора света на основе электрооптического эффекта необходимо придумать некоторый механизм для создания электрического поля $E(x, y)$, пропорционального желаемому коэффициенту пропускания $\mathcal{T}(x, y)$ в каждой точке. Это не просто. Один из подходов состоит в размещении матрицы прозрачных электродов на маленьких пластинках электрооптического материала между скрещенными поляризаторами. На каждый электрод подается соответствующее напряжение (рис. 20.14). Напряжение, приложенное к электроду с центром в точке (x_i, y_i), $i = 1, 2, \ldots$, делается пропорциональным желаемому значению $\mathcal{T}(x_i, y_i)$ (см., например, рис. 20.7). Если число электродов достаточно велико,

коэффициент пропускания приближается к $\mathcal{T}(x, y)$. Независимо подавать сигналы на каждый электрод сложно и неудобно, тем не менее мы увидим, что такая схема применяется в жидкокристаллических пространственных модуляторах света для дисплеев, поскольку в этом случае требуемые напряжения малы (см. подразд. 20.3.2).

Электрооптические пространственно-временны́е модуляторы с оптическим управлением

Один из методов оптического управления электрооптическим пространственным модулятором света основан на использовании тонкого слоя фотопроводящего материала для создания электрического поля, управляющего модулятором (рис. 20.15). Проводимость фотопроводящего материала пропорциональна интенсивности падающего света (см. разд. 18.2). При освещении светом с распределением интенсивности $I_W(x, y)$ создается пространственная картина электропроводности $G(x, y) \propto I_W(x, y)$. Фотопроводящий слой помещается между двумя электродами, действующими как конденсатор. Первоначально конденсатор заряжен, и утечка заряда в точке (x, y) пропорциональна локальной проводимости $G(x, y)$. В результате заряд конденсатора уменьшается в тех точках, где проводимость велика. Следовательно, локальное напряжение пропорционально $1/G(x, y)$, а соответствующее электрическое поле $E(x, y) \propto 1/G(x, y) \propto 1/I_W(x, y)$. Если коэффициент пропускания $\mathcal{T}(x, y)$ (или отражения $\mathcal{R}(x, y)$) модулятора пропорционален приложенному полю, то он должен быть обратно пропорционален исходной интенсивности света $I_W(x, y)$.

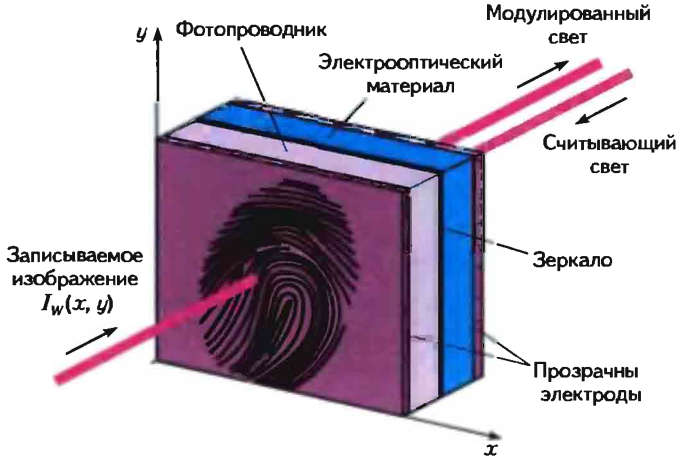


Рис. 20.15. Электрооптический пространственный модулятор света с фотопроводящим материалом, создающим пространственное распределение электрического поля, которое используется для управления электрооптическим материалом

Считывающий оптический модулятор на основе эффекта Поккельса

Остроумное воплощение этого принципа — **поккельсовский считывающий оптический модулятор** (ПСОМ). В одном из конструктивных решений используется кристалл $Bi_{12}SiO_{20}$ (BSO), обладающий необычной комбинацией оптических и электрических свойств:

1) в нем проявляется электрооптический эффект (эффект Поккельса);

2) он является фотопроводящим для синего света, но не для красного;

3) он является хорошим изолятором в темноте. ПСОМ (рис. 20.16) включает в себя тонкую пластину BSO, помещенную между двумя прозрачными электродами.

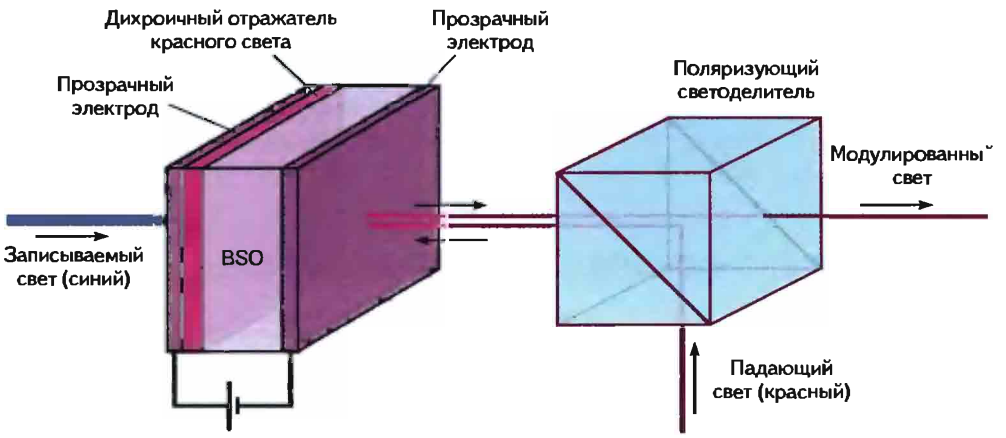


Рис. 20.16. Поккельсовский считывающий оптический модулятор (ПСОМ)

Свет, который нужно промодулировать (считывающий свет), пропускается через поляризатор, входит в слой DSO и отражается дихроичным зеркалом, после чего проходит через второй поляризатор. Зеркало отражает красный свет, но прозрачно для синего света. ПСОМ приводится в действие следующим образом.

• *Зарядка.* Большая разность потенциалов (≈4 кВ) прикладывается к электродам и конденсатор заряжается (без утечки, поскольку кристалл является хорошим изолятором в темноте).

• *Запись.* Интенсивный синий свет с распределением интенсивности $I_W(x, y)$ освещает кристалл. В результате создается пространственное распределение проводимости $G(x, y) \propto I_W(x, y)$, напряжение по поверхности кристалла селективно уменьшается и электрическое поле пропорционально уменьшается в каждой точке, так что $E(x, y) \propto 1/G(x, y) \propto 1/I_W(x, y)$. За счет электрооптического эффекта показатели преломления BSO изменяются, в кристалле создается и записывается пространственная картина изменения показателя преломления $\Delta n(x, y) \propto 1/I_W(x, y)$.

• *Считывание.* Однородный красный свет используется для считывания $\Delta n(x, y)$, как в обычном электрооптическом модуляторе интенсивности (см. рис. 20.7, *а*), при этом поляризующий светоделитель играет роль скрещенных поляризаторов.

• *Стирание.* Распределение показателя преломления стирается однородной вспышкой синего света. Кристалл снова заряжается путем подключения напряжения 4 кВ, и начинается новый цикл.

Преобразователи некогерентного света в когерентный

В оптически управляемом пространственном модуляторе света, таком как ПСОМ, свет, используемый для записи пространственной картины в модулятор, не обязан быть когерентным, поскольку фотопроводящие материалы чувствительны только к интенсивности света. Пространственная оптическая картина (изображение) может записываться с помощью некогерентного света, а считываться когерентным светом. Этот процесс преобразования в реальном времени пространственного распределения естественного некогерентного света в пропорциональное пространственное распределение когерентного света используется в целом ряде приложений, связанных с оптической обработкой данных и изображений.

*20.2. ЭЛЕКТРООПТИКА АНИЗОТРОПНЫХ СРЕД

Основные принципы и приложения электрооптики были представлены в разд. 20.1 в упрощенном виде: эффекты поляризации и анизотропии либо игнорировались, либо рассматривались в самых общих чертах. В данном разделе представлен более полный анализ электрооптики анизотропных сред. Напомним кратко некоторые важные свойства анизотропных сред (см. разд. 6.3).

Оптика кристаллов: краткое напоминание

Оптические свойства анизотропной среды характеризуются геометрическим построением, которое называется **эллипсоидом показателя преломления**

$$\sum_{ij} \eta_{ij} x_i x_j = 1, \quad i, j = 1, 2, 3,$$

где $\eta_{ij} = \eta_{ji}$ — элементы тензора диэлектрической непроницаемости $\eta = \varepsilon_0 \varepsilon^{-1}$. Если оси эллипсоида соответствуют главным осям среды, то его размеры вдоль этих осей равны главным значениям показателя преломления n_1, n_2, n_3 (рис. 20.17):

$$\frac{x_1^2}{n_1^2} + \frac{x_2^2}{n_2^2} + \frac{x_3^2}{n_3^2} = 1.$$

Эллипсоид показателя преломления можно использовать для определения поляризации и показателей преломления n_a и n_b нормальных мод волны, распространяющейся в произвольном направлении в анизотропной среде. Для этого нужно провести плоскость, перпендикулярную направлению распространения и проходящую через центр эллипсоида. Ее пересечение с эллипсоидом дает эллипс, большая и малая полуоси которого равны n_a и n_b, как описано в подразд. 6.3.3.

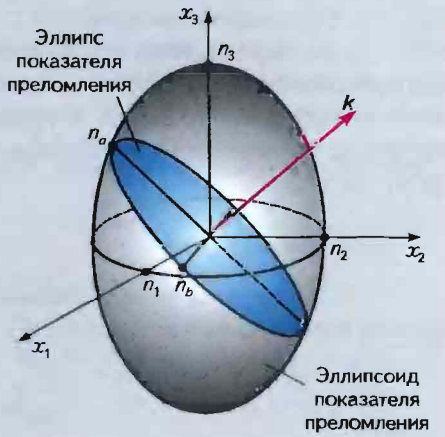


Рис. 20.17. Эллипсоид показателя преломления. Координаты (x_1, x_2, x_3) соответствуют главным осям, а n_1, n_2, n_3 — главные значения показателя преломления. Показатели преломления нормальных мод волны, распространяющейся в направлении k, равны n_a и n_b

20.2.1. Эффекты Поккельса и Керра

Если кристалл поместить в электрическое поле $\boldsymbol{E}$ с компонентами (E_1, E_2, E_3), то элементы тензора η меняются. Каждый из этих девяти элементов η_{ij} становится функцией E_1, E_2, и E_3, т. е. $\eta_{ij} = \eta_{ij}(\boldsymbol{E})$, так что эллипсоид показателя преломления меняет свою форму (рис. 20.18). Зная функции $\eta_{ij}(\boldsymbol{E})$, можно определить эллипсоид показателя преломления и оптические свойства кристалла при произвольно приложенном электрическом поле $\boldsymbol{E}$. Задача в принципе проста, хотя ее решение часто громоздко.



Рис. 20.18. Эллипсоид показателя преломления меняется в результате наложения постоянного электрического поля

Каждый элемент $\eta_{ij}(\boldsymbol{E})$ является функцией трех переменных $\boldsymbol{E} = (E_1, E_2, E_3)$, которая может быть разложена в ряд Тейлора в точке $\boldsymbol{E} = \boldsymbol{0}$:

$$\eta_{ij}(\boldsymbol{E}) = \eta_{ij} + \sum_k r_{ijk} E_k + \sum_{kl} s_{ijkl} E_k E_l, \quad i, j, k, l = 1, 2, 3, \qquad (20.24)$$

где

$$\eta_{ij} = \eta_{ij}(\mathbf{0}); \quad \mathfrak{r}_{ijk} = \frac{\partial \eta_{ij}}{\partial E_k}; \quad \mathfrak{s}_{ijkl} = \frac{1}{2}\frac{\partial^2 \eta_{ij}}{\partial E_k \partial E_l},$$

производные берутся в точке $\boldsymbol{E} = \mathbf{0}$. Формула (20.17) является обобщением (20.3), где $\mathfrak{r}$ заменяется набором из $3^3 = 27$ коэффициентов $\{\mathfrak{r}_{ijk}\}$, а $\mathfrak{s}$ — набором из $3^4 = 81$ коэффициентов $\{\mathfrak{s}_{ijkl}\}$. Коэффициенты $\{\mathfrak{r}_{ijk}\}$ называются линейными электрооптическими коэффициентами (коэффициентами Поккельса). Они образуют тензор третьего ранга. Коэффициенты $\{\mathfrak{s}_{ijkl}\}$ — квадратичные электрооптические коэффициенты (коэффициенты Керра). Они образуют тензор четвертого ранга.

Симметрия

Поскольку тензор η симметричен ($\eta_{ij} = \eta_{ji}$), $\mathfrak{r}$ и $\mathfrak{s}$ инвариантны относительно перестановки индексов i и j, т. е. $\mathfrak{r}_{ijk} = \mathfrak{r}_{jik}$ и $\mathfrak{s}_{ijkl} = \mathfrak{s}_{jikl}$. Коэффициенты

$$\mathfrak{s}_{ijkl} = \frac{1}{2}\frac{\partial^2 \eta_{ij}}{\partial E_k \partial E_l}$$

инвариантны также относительно перестановки k и l (в силу независимости от порядка дифференцирования), так что $\mathfrak{s}_{ijkl} = \mathfrak{s}_{ijlk}$. В силу такой перестановочной симметрии девять комбинаций индексов i и j порождают шесть, а не девять независимых элементов. Такое же уменьшение числа элементов относится к индексам k и l. Следовательно, $\mathfrak{r}_{ijk}$ имеет 6×3 независимых элементов, а $\mathfrak{s}_{ijkl}$ — 6×6 независимых элементов.

Удобно нумеровать пары индексов (i, j), $i, j = 1, 2, 3$, одним индексом $I = 1, 2, \ldots, 6$ в соответствии с табл. 20.1. Пара (k, l) аналогичным образом заменяется одним индексом $K = 1, 2, \ldots, 6$ по тому же правилу. Таким образом, элементы $\mathfrak{r}_{ijk}$ и $\mathfrak{s}_{ijkl}$ заменяются на $\mathfrak{r}_{Ik}$ и $\mathfrak{s}_{IK}$ соответственно. Например, $\mathfrak{r}_{12k}$ обозначается $\mathfrak{r}_{6k}$, $\mathfrak{s}_{1231}$ переименовывается в $\mathfrak{s}_{65}$ и так далее. Итак, тензор третьего ранга $\mathfrak{r}$ заменяется матрицей 6×3, а тензор четвертого ранга $\mathfrak{s}$ — матрицей 6×6.

Таблица 20.1. **Таблица нумерации пар индексов (i, j) одним индексом I***

j/i	1	2	3
1	1	6	5
2	6	2	4
3	5	4	3

* Например, пара $(i, j) = (3, 2)$ нумеруется индексом $I = 4$.

Кристаллическая симметрия

Симметрия кристалла вносит дополнительные ограничения на элементы матриц r и s. Некоторые элементы обращаются в нуль, а другие оказываются равными или имеют противоположные знаки при одинаковой абсолютной величине, или связаны каким-либо другим правилом. Например, для сред с центром симметрии r вообще исчезает, и проявляется только эффект Керра. Список коэффициентов r и s и соотношения симметрии для них для 32 кристаллографических точечных групп можно найти в нескольких книгах из списка литературы для чтения в конце главы. Примеры приведены в табл. 20.2 и 20.3.

Таблица 20.2. **Коэффициенты Поккельса r_{lk} для некоторых кристаллических групп**

$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{41} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{63} \end{bmatrix}$	$\begin{bmatrix} 0 & -r_{22} & r_{13} \\ 0 & r_{22} & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{51} & 0 \\ r_{51} & 0 & 0 \\ -r_{22} & 0 & 0 \end{bmatrix}$
Кубический $\bar{4}3m$ (например, GaAs, CdTe, InAs)	Тетрагональный $\bar{4}2m$ (например, KDP, ADP)	Тригональный $3m$ (например, $LiNbO_3$, $LiTaO_3$)

Таблица 20.3. **Коэффициенты Поккельса s_{lk} для изотропной среды**

$\begin{bmatrix} s_{11} & s_{12} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{11} & s_{12} & 0 & 0 & 0 \\ s_{12} & s_{12} & s_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{44} \end{bmatrix}$	$s_{44} = \frac{1}{2}(s_{11} - s_{12})$

Эффект Поккельса

Для определения оптических свойств анизотропного материала, в котором при наложении электрического поля $\boldsymbol{E}$ имеет место эффект Поккельса, нужно проделать следующее.

1. Найти главные оси и главные значения показателя преломления n_1, n_2 и n_3 в отсутствие $\boldsymbol{E}$.

2. Найти коэффициенты $\{r_{ijk}\}$ из соответствующей матрицы r_{lk}, например, с помощью табл. 20.2, пользуясь правилом соответствия между (i, j) и I, приведенным в табл. 20.1.

3. Определить элементы тензора диэлектрической непроницаемости:

$$\eta_{ij}(E) = \eta_{ij}(0) + \sum_k r_{ijk} E_k, \tag{20.25}$$

где $\eta_{ij}(0)$ — диагональная матрица с элементами $1/n_1^2$, $1/n_2^2$ $1/n_3^2$.

4. Написать уравнение для модифицированного эллипсоида показателя преломления

$$\sum_{ij} \eta_{ij}(E) x_i x_j = 1. \tag{20.26}$$

5. Определить главные оси модифицированного эллипсоида показателя преломления посредством диагонализации и найти соответствующие главные значения $n_1(E)$, $n_2(E)$ и $n_3(E)$.

6. Для данного направления распространения света найти нормальные моды и связанные с ними показатели преломления из найденного эллипсоида показателя преломления.

Пример 20.1

Тригональные 3m-кристаллы ($LiNbO_3$ и $LiTaO_3$)

Тригональные $3m$-кристаллы являются одноосными ($n_1 = n_2 = n_o$, $n_3 = n_н$) с матрицей r, приведенной в таблице (20.2). Пусть $E = (0, 0, E)$, т. е. электрическое поле направлено вдоль оптической оси (см. рис. 20.3). Тогда легко показать, что модифицированный эллипсоид показателя преломления

$$\left(\frac{1}{n_o^2} + r_{13}E\right)\left(x_1^2 + x_2^2\right) + \left(\frac{1}{n_н^2} + r_{33}E\right)x_3^2 = 1. \tag{20.27}$$

Это эллипсоид вращения, главные оси которого не меняются при наложении электрического поля. Обыкновенный $n_o(E)$ и необыкновенный $n_н(E)$ показатели преломления даются выражениями

$$\frac{1}{n_o^2(E)} = \frac{1}{n_o^2} + r_{13}E; \tag{20.28}$$

$$\frac{1}{n_н^2(E)} = \frac{1}{n_н^2} + r_{33}E. \tag{20.29}$$

Ввиду малости членов $r_{13}E$ и $r_{33}E$ в (20.28) и (20.29) используем приближение $(1 + \Delta)^{-1/2} \approx 1 - \Delta/2$, справедливое при $|\Delta| \ll 1$, и получим

$$\boxed{n_o(E) \approx n_o - \frac{1}{2} n_o^3 r_{13} E;} \tag{20.30}$$

$$\boxed{n_н(E) \approx n_н - \frac{1}{2} n_н^3 r_{33} E.} \tag{20.31}$$

Отметим сходство между этими уравнениями и общим уравнением (20.4). Мы приходим к выводу, что при наложении электрического поля на одноосный кристалл вдоль его оптической оси кристалл остается одноосным, его главные оси не меняют направления, как показано на рис. 20.19, однако показатели преломления модифицируются в соответствии с выражениями (20.30) и (20.31).

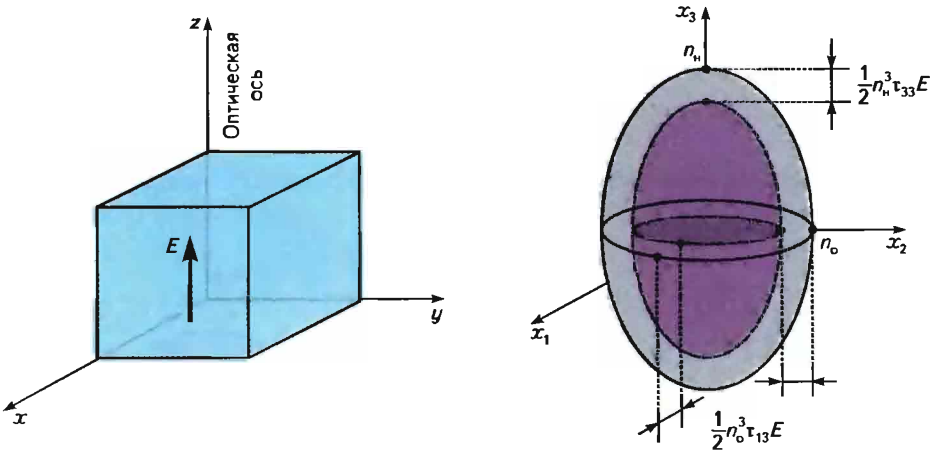


Рис. 20.19. Изменение эллипсоида показателя преломления у тригонального $3m$-кристалла, такого как $LiNbO_3$, в результате наложения постоянного электрического поля вдоль направления оптической оси

Пример 20.2

Тетрагональные $\bar{4}2m$-кристаллы (KDP и ADP)

Выполняя ту же процедуру для этого класса одноосных кристаллов и предполагая, что электрическое поле направлено вдоль оптической оси (рис. 20.20), получаем следующее уравнение для эллипсоида показателя преломления:

$$\frac{x_1^2 + x_2^2}{n_o^2} + \frac{x_3^2}{n_н^2} + 2r_{63}Ex_1x_2 = 1. \qquad (20.32)$$

Измененные главные оси получаются поворотом системы координат на 45° вокруг оси z. Подставляя

$$x_1' = \frac{(x_1 + x_2)}{\sqrt{2}}; \quad x_2' = \frac{(x_1 - x_2)}{\sqrt{2}}; \quad x_3' = x_3$$

в (20.32) и меняя обозначение осей координат на (x_1, x_2, x_3), получаем

$$\frac{x_1^2}{n_1^2(E)} + \frac{x_2^2}{n_2^2(E)} + \frac{x_3^2}{n_3^2(E)} = 1, \qquad (20.33)$$

где

$$\frac{1}{n_1^2(E)} = \frac{1}{n_o^2} + r_{63}E; \quad \frac{1}{n_2^2(E)} = \frac{1}{n_o^2} - r_{63}E; \quad n_3(E) = n_e. \tag{20.34}$$

Используя приближение $(1 + \Delta)^{-1/2} \approx 1 - \Delta/2$, находим

$$n_1(E) \approx n_o - \frac{1}{2}n_o^3 r_{63}E; \tag{20.35}$$

$$n_2(E) \approx n_o + \frac{1}{2}n_o^3 r_{63}E; \tag{20.36}$$

$$n_3(E) = n_н. \tag{20.37}$$

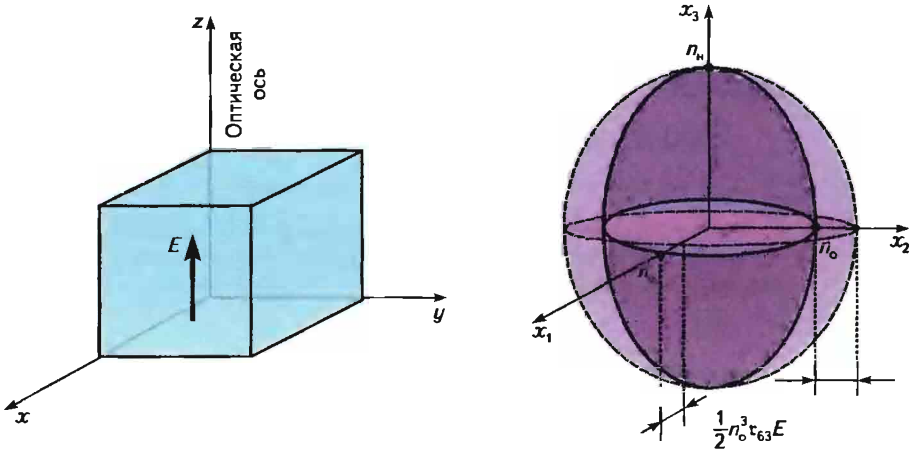


Рис. 20.20. Изменение эллипсоида показателя преломления при наложении постоянного электрического поля E вдоль направления оптической оси в одноосном тетрагональном $\bar{4}2m$-кристалле, таком как KDP

Мы приходим к выводу, что первоначально одноосный кристалл под действием электрического поля, приложенного вдоль оптической оси, ведет себя как двуосный кристалл, что иллюстрируется рис. 20.20.

Пример 20.3

Кубические $\bar{4}3m$-кристаллы (GaAs, CdTe и InAs)

Если электрическое поле приложено вдоль кубической оси материала (ось z на рис. 20.21), то уравнение эллипсоида показателя преломления для этих изотропных кристаллов ($n_1 = n_2 = n_3 = n$) принимает вид

$$\frac{x_1^2 + x_2^2 + x_3^2}{n^2} + 2r_{41}Ex_1x_2 = 1, \tag{20.38}$$

где r_{63} принимает значение r_{41} (см. табл. 20.2). Как и в примере 20.2, новые главные оси повернуты на 45° вокруг оси z, и главные значения показателя преломления оказываются равными

$$n_1(E) \approx n - \frac{1}{2} n^3 r_{41} E; \tag{20.39}$$

$$n_2(E) \approx n + \frac{1}{2} n^3 r_{41} E; \tag{20.40}$$

$$n_3(E) = n. \tag{20.41}$$

Таким образом, при наложении поля изотропный кристалл ведет себя как двуосный (см. рис. 20.21).

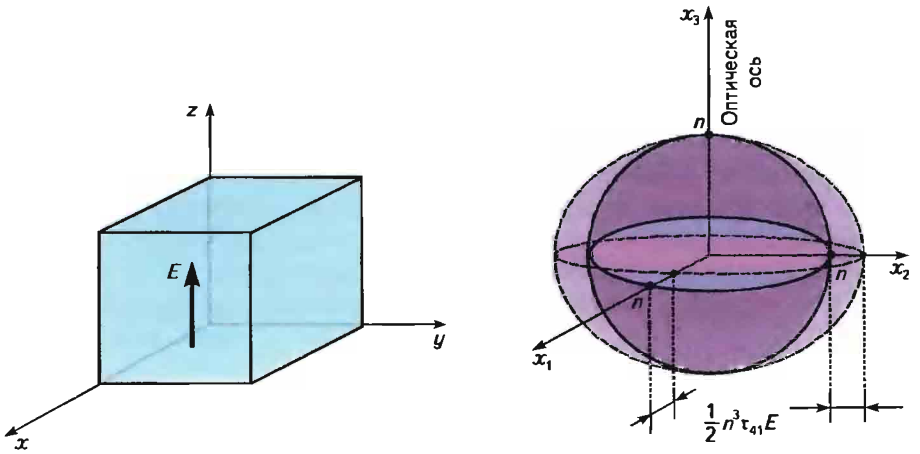


Рис. 20.21. Изменение эллипсоида показателя преломления при наложении постоянного электрического поля вдоль кубической оси в $\bar{4}3m$-кристалле, таком как GaAs

Кубические кристаллы имеют хорошо определенные кристаллические оси, однако по своим оптическим свойствам они изотропны. Наложение постоянного электрического поля нарушает геометрическую симметрию и приводит к появлению оптической анизотропии, как ясно следует из примера 20.3. Для исходно анизотропных материалов, у которых приложенное электрическое поле не меняет направления главных осей, как в примере 20.1, поляризация нормальных мод остается неизменной, но связанные с ними показатели преломления становятся зависящими от E. В этом случае среду удобно использовать в качестве фазового модулятора, фазовой пластинки или модулятора интенсивности в соответствии с общей теорией, приведенной в разд. 20.1. Данный принцип разъясняется далее в подразд. 20.2.2.

Эффект Керра

Оптические свойства керровской среды можно определить с помощью такой же процедуры, какая была использована для сред с эффектом Поккельса, за исключением того, что коэффициенты $\eta_{ij}(E)$ задаются теперь формулой

$$\eta_{ij}(E) = \eta_{ij}(0) + \sum_{kl} s_{ijkl} E_k E_l. \tag{20.42}$$

Пример 20.4

Эффект Керра в изотропной среде

При наложении постоянного электрического поля E вдоль оси z используем коэффициенты Керра s_{IK} из табл. 20.3 для изотропной среды и найдем уравнение для эллипсоида показателя преломления

$$\left(\frac{1}{n^2} + s_{12}E^2\right)\left(x_1^2 + x_2^2\right) + \left(\frac{1}{n^2} + s_{11}E^2\right)x_3^2 = 1. \tag{20.43}$$

Это эллипсоид вращения с осью z, совпадающей с направлением приложенного поля. Главные значения показателя преломления $n_o(E)$ и $n_н(E)$ определяются выражениями

$$\frac{1}{n_o^2(E)} = \frac{1}{n^2} + s_{12}E^2; \tag{20.44}$$

$$n_{н}(E) \approx \frac{1}{n^2} + s_{11}E^2 \tag{20.45}$$

Поскольку стоящие в (20.44) и (20.45) члены малы, снова используем приближение $(1 + \Delta)^{-1/2} \approx 1 - \Delta/2$ и получим

$$\boxed{n_o(E) \approx n - \frac{1}{2}n^3 s_{12}E^2;} \tag{20.46}$$

$$\boxed{n_н(E) \approx n - \frac{1}{2}n^3 s_{11}E^2.} \tag{20.47}$$

Таким образом, постоянно электрическое поле E, приложенное к изначально изотропной среде, заставляет ее вести себя как одноосный кристалл с оптической осью, направленной вдоль поля. Обыкновенный и необыкновенный показатели преломления представляют собой квадратично убывающие функции E.

20.2.2. Модуляторы

Принципы модуляции интенсивности и фазы с помощью электрооптического эффекта были описаны в подразд. 20.1.2. Эффекты анизотропии были учтены лишь в общих чертах. С помощью теории анизотропной среды, представленной в настоящем разделе, общие параметры r и s, использованные в разд. 20.1, теперь можно определить для любого данного кристалла, а также любых направлений электрического поля и распространения света. Мы рассмотрим только модуляторы на эффекте Поккельса, однако такой же подход можно применять и для керровских модуляторов. Для простоты предположим, что направление электрического поля таково, что главные оси кристалла не меняют своего направления в результате модуляции. Предположим также, что направление распространения волны по отношению к этим осям таково, что плоскости поляризации нормальных мод также не меняются под действием электрического поля.

Фазовые модуляторы

Нормальная мода характеризуется показателем преломления

$$n(E) \approx n - \frac{1}{2} r n^3 E,$$

где n и r — соответствующие показатель преломления и коэффициент Поккельса; $E = V/d$ — электрическое поле, получаемое путем приложения напряжения V к точкам на расстоянии d. Волна, проходящая расстояние L, претерпевает фазовый сдвиг

$$\varphi = \varphi_0 - \pi \frac{V}{V_\pi}, \tag{20.48}$$

где

$$\varphi_0 = \frac{2\pi n L}{\lambda_0};$$

$$V_\pi = \frac{d}{L} \frac{\lambda_0}{r n^3} \tag{20.49}$$

полуволновое напряжение. Соответствующие коэффициенты, в общем виде обозначенные n и r, можно легко определить, как показано в следующем примере.

Пример 20.5

Тригональные 3m-кристаллы ($LiNbO_3$ и $LiTaO_3$)

Когда электрическое поле направлено вдоль оптической оси одноосного кристалла такого типа, кристалл остается одноосным с теми же направлениями главных осей (см. рис. 20.19). Главные значения показателя преломления да-

ются выражениями (20.30) и (20.31). Кристалл можно использовать как фазовый модулятор в одной из двух конфигураций.

Продольный модулятор

Если линейно поляризованная оптическая волна распространяется в направлении оптической оси (параллельной электрическому полю), то фазовый модулятор описывается параметрами $n = n_o$, $r = r_{13}$ и $d = L$. Для $LiNbO_3$ $r_{13} = 9{,}6$ пм/В и $n_o = 2{,}3$ при $\lambda_0 = 633$ нм. Тогда формула (20.49) дает $V_\pi = 5{,}41$ кВ для напряжения, необходимого для изменения фазы на π.

Поперечный модулятор

Если волна распространяется вдоль оси x и поляризована вдоль оси z, то соответствующие параметры — это $n = n_н$ и $r = r_{33}$. Ширина d в общем случае не равна длине L. Для $LiNbO_3$ при $\lambda_0 = 633$ нм $r_{33} = 30{,}9$ пм/В и $n_н = 2{,}2$, что дает полуволновое напряжение $V_\pi = 1{,}9(d/L)$ кВ. Если $d/L = 0{,}1$, получаем $V_\pi \approx 190$ В, что существенно ниже, чем полуволновое напряжение для продольного модулятора.

Модуляторы интенсивности

Различие зависимостей показателей преломления двух нормальных мод от приложенного поля в ячейке Поккельса приводит к зависящей от напряжения фазовой задержке

$$\Gamma = \Gamma_0 - \pi \frac{V}{V_\pi}, \tag{20.50}$$

где

$$\Gamma_0 = \frac{2\pi (n_1 - n_2) L}{\lambda_0}; \tag{20.51}$$

$$V_\pi = \frac{(d/L)\lambda_0}{r_1 n_1^3 - r_2 n_2^3}. \tag{20.52}$$

Если ячейка помещена между скрещенными поляризаторами, то система служит модулятором интенсивности (см. подразд. 20.1.2). Нетрудно определить соответствующие показатели преломления n_1 и n_2 и коэффициенты r_1 и r_2, что иллюстрируется следующим примером.

Пример 20.6

Тетрагональные $\bar{4}2m$-кристаллы (KDP и ADP)

Как описано в примере 20.2, при наложении электрического поля вдоль оптической оси одноосного кристалла последний ведет себя как двуосный кристалл. Новые главные оси повернуты на 45° вокруг оптической оси. Рассмотрим продольный модулятор ($d/L = 1$), в котором волна бежит вдоль оптичес-

кой оси. Две нормальные моды имеют показатели преломления, даваемые формулами (20.35) и (20.36). Следовательно, коэффициенты, которые нужно использовать в (20.52), — это $n_1 = n_2 = n_o$, $r_1 = r_{63}$, $r_2 = -r_{63}$ и $d = L$, так что $\Gamma_0 = 0$ и

$$V_\pi = \frac{\lambda_0}{2r_{63}n_o^3}. \tag{20.53}$$

Для KDP при $\lambda_0 = 633$ нм $V_\pi = 8{,}4$ кВ.

Упражнение 20.2

Модулятор интенсивности на основе эффекта Керра

С помощью (20.46) и (20.47) найдите выражение для фазового сдвига φ и фазовой задержки Γ в продольном керровском модуляторе, изготовленном из изотропного материала, в зависимости от приложенного напряжения V. Выведите выражения для полуволновых напряжений V_π в каждом случае.

20.3. ЭЛЕКТРООПТИКА ЖИДКИХ КРИСТАЛЛОВ

Как показано в разд. 6.5, вытянутые молекулы нематических жидких кристаллов стремятся к упорядоченной ориентации, которая меняется, если на материал действуют механические или электрические силы. В силу своей анизотропной природы жидкие кристаллы можно использовать в качестве фазовых пластинок или вращателей поляризации. В присутствии электрического поля ориентация их молекул меняется, так что меняется и их действие на поляризованный свет. Следовательно, жидкие кристаллы можно использовать как электрически управляемые фазовые пластинки, модуляторы и переключатели. Эти устройства особенно полезны в технологии дисплеев.

20.3.1. Фазовые пластинки и модуляторы

Электрические свойства нематических жидких кристаллов

Жидкие кристаллы, используемые в электрооптических приборах, обычно имеют достаточно низкую проводимость, позволяющую рассматривать их как идеальные диэлектрики. Из-за вытянутой формы и упорядоченной ориентации образующих их молекул жидкие кристаллы имеют анизотропные оптические свойства с одноосной симметрией (см. подразд. 6.3.1). Диэлектрическая проницаемость равна $\varepsilon_\parallel$ для электрических полей, направленных вдоль молекул, и $\varepsilon_\perp$ для полей в перпендикулярном направлении. Жидкие кристаллы, у которых $\varepsilon_\parallel > \varepsilon_\perp$ (положительные одноосные), обычно выбирают для электрооптических приложений.

При наложении постоянного (или низкочастотного) электрического поля индуцируются электрические диполи и появляющиеся электрические силы

создают вращающий момент, действующий на молекулы. Молекулы поворачиваются в таком направлении, чтобы свободная электростатическая энергия

$$-\frac{1}{2}\boldsymbol{E}\quad \boldsymbol{D} = -\frac{1}{2}\left(\varepsilon_{\perp}E_1^2 + \varepsilon_{\perp}E_2^2 + \varepsilon_{\parallel}E_3^2\right)$$

была минимальна (здесь E_1, E_2 и E_3 — компоненты вектора $\boldsymbol{E}$ вдоль главных осей). Поскольку $\varepsilon_{\parallel} > \varepsilon_{\perp}$, для данного направления электрического поля минимум энергии достигается, когда молекулы выстроены вдоль поля, так что $E_1 = E_2 = 0$, $\boldsymbol{E} = (0, 0, E)$, а энергия при этом равна $-(1/2)\varepsilon_{\parallel}E^2$. При полном выстраивании молекулярные оси направлены вдоль электрического поля (рис. 20.22). Очевидно, обращение электрического поля вызывает такой же поворот молекул. Такое же действие оказывает переменное поле, если приложить переменное напряжение.

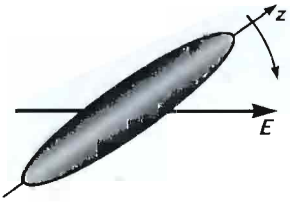


Рис. 20.22. Молекулы положительного одноосного жидкого кристалла поворачиваются и выстраиваются вдоль приложенного электрического поля

Фазовые пластинки и модуляторы на нематических жидких кристаллах

Нематическая жидкокристаллическая ячейка представляет собой тонкий слой нематического жидкого кристалла, помещенный между двумя параллельными стеклянными пластинками и натертый так, чтобы все молекулы были параллельны друг другу. Материал тогда ведет себя как одноосный кристалл с оптической осью, параллельной оси ориентации молекул. Для волн, бегущих в направлении z (перпендикулярно стеклянным пластинкам), нормальные моды линейно поляризованы в направлениях x и y (параллельно и перпендикулярно молекулярным осям соответственно), как показано на рис. 20.23, *а*. Показателями преломления являются необыкновенный и обыкновенный показатели преломления $n_н$ и $n_о$. Ячейка толщиной d создает запаздывание по фазе $\Gamma = 2\pi(n_н - n_о)d/\lambda_0$.

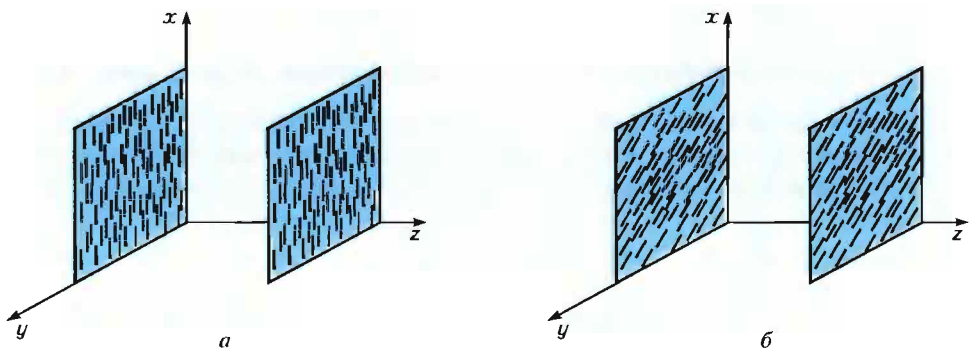


Рис. 20.23. Ориентация молекул в жидкокристаллической ячейке в отсутствие постоянного электрического поля (*а*) и при наложении постоянного электрического поля (*б*). Оптическая ось направлена вдоль осей молекул:

а — состояние без наклона; *б* — состояние с наклоном

Пусть теперь *в направлении оси z* приложено электрическое поле (посредством подачи напряжения V на прозрачные проводящие электроды, нанесенные на внутреннюю поверхность стеклянных пластин). Тогда возникающие электрические силы стремятся наклонить молекулы в сторону выстраивания вдоль поля, однако силы упругости на поверхностях стеклянных пластин сопротивляются такому движению. Если приложенное электрическое поле достаточно велико, большинство молекул наклоняются к оси z, за исключением непосредственно прилегающих к стеклянным поверхностям. Равновесный угол наклона θ для большинства молекул является монотонно возрастающей функцией V, которую можно описать выражением[1]

$$\theta = \begin{cases} 0, & V \le V_{кр}; \\ \dfrac{\pi}{2} - 2\,\mathrm{arctg}\left[\exp\left(-\dfrac{V - V_{кр}}{V_0}\right)\right] & V > V_{кр}, \end{cases} \tag{20.54}$$

где V — приложенное действующее напряжение; $V_{кр}$ — критическое напряжение, при котором процесс наклона начинается; V_0 — постоянная. Когда $V - V_{кр} = V_0$, $\theta \approx 50°$; при дальнейшем росте $V - V_{кр}$ угол θ приближается к 90°, как отмечено на рис. 20.24, *а*.

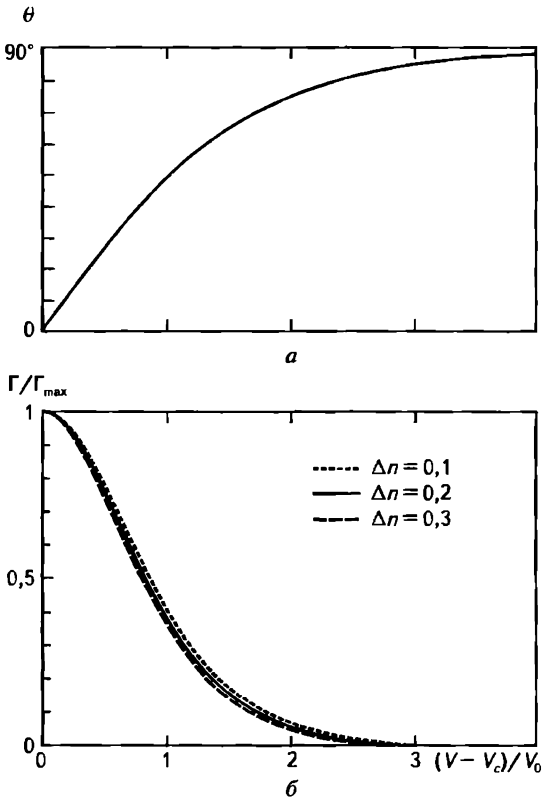


Рис. 20.24. Зависимость угла наклона θ молекул в сторону направления поля (ось z) от нормированного действующего напряжения (*а*). Зависимость нормированной фазовой задержки

$$\frac{\Gamma}{\Gamma_{max}} = \frac{n(\theta) - n_o}{n_н - n_o}$$

от нормированного действующего напряжения при $n_0 = 1{,}5$ для нескольких значений $\Delta n = n_н - n_o$, отмеченных на рисунке (*б*). Графики получены из (20.54) и (20.55)

[1] См., например, *de Gennes P.-G.* The Physics of Liquid Crystals. Oxford University Press, 2nd ed. 1995.

Когда электрическое поле убирают, ориентация молекул вблизи стеклянных пластин остается неизменной, а остальные молекулы поворачиваются обратно в плоскостях, параллельных стеклянным пластинам, и возвращаются в исходное положение. В этом смысле жидкокристаллический материал можно рассматривать как *жидкость с памятью*.

Для угла наклона θ нормальные моды оптической волны, распространяющейся в направлении z, поляризованы вдоль осей x и y и имеют показатели преломления $n(\theta)$ и n_o, где

$$\frac{1}{n^2(\theta)} = \frac{\cos^2\theta}{n_н^2} + \frac{\sin^2\theta}{n_o^2}, \tag{20.55}$$

так что фазовая задержка составляет $\Gamma = 2\pi(n(\theta) - n_o)d/\lambda_0$ (см. подразд. 6.3.3). Она достигает максимального значения $\Gamma_{max} = 2\pi(n_н - n_o)d/\lambda_0$, когда молекулы не наклонены ($\theta = 0$) и монотонно убывает к нулю по мере приближения угла наклона к 90°, как показано на рис. 20.24, *б*. Заметим, что для угла наклона θ угол между оптической осью и направлением распространения составляет $90° - \theta$, поэтому (20.55) отличается от (6.58).

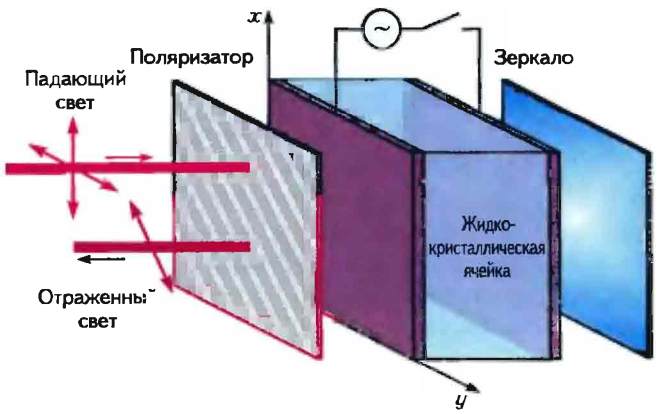


Рис. 20.25. Жидкокристаллическая ячейка вносит разность фаз $\Gamma = \pi/2$ в отсутствие поля (состояние «выключено») и $\Gamma = 0$ в присутствии поля (состояние «включено»). В состоянии «выключено» после отражения от зеркала и прохождения через кристалл туда и обратно плоскость поляризации поворачивается на 90°, так что свет блокируется. В состоянии «включено» поворот не происходит и отраженный свет не блокируется

Ячейку легко использовать как управляемый напряжением *фазовый модулятор*. Для оптической волны, распространяющейся в направлении z и поляризованной линейно вдоль оси x (параллельной ненаклоненным молекулярным осям), фазовый сдвиг равен $\varphi = 2\pi n(\theta)d/\lambda_0$. Для волн, поляризованных под углом 45° к оси x в плоскости x—y, ячейка действует как управляемая напряжением *фазовая пластинка*. При помещении между двумя скрещенными поляризаторами (под углом ±45°) полуволновая пластинка ($\Gamma = \pi$) превращается в управляемый напряжением *модулятор интенсивности*. Аналогично, четверть-

волновая пластинка ($\Gamma = \pi/2$), помещенная между зеркалом и поляризатором под 45° к оси x, служит модулятором интенсивности, как показано на рис. 20.25.

Жидкокристаллическая ячейка герметизируется между оптически плоскими стеклянными окнами с антиотражающими покрытиями. Типичная толщина жидкокристаллического слоя составляет $d = 10$ мкм, а типичные значения $\Delta n = n_н - n_o$ лежат в интервале от 0, 1 до 0,3. Фазовая задержка обычно выражается через разность хода $r = (n_н - n_o)d$, так что $\Gamma = 2\pi r/\lambda_0$. Типичные разности хода составляют несколько сотен нанометров (например, разность хода 300 нм соответствует разности фаз π при $\lambda_0 = 600$ нм).

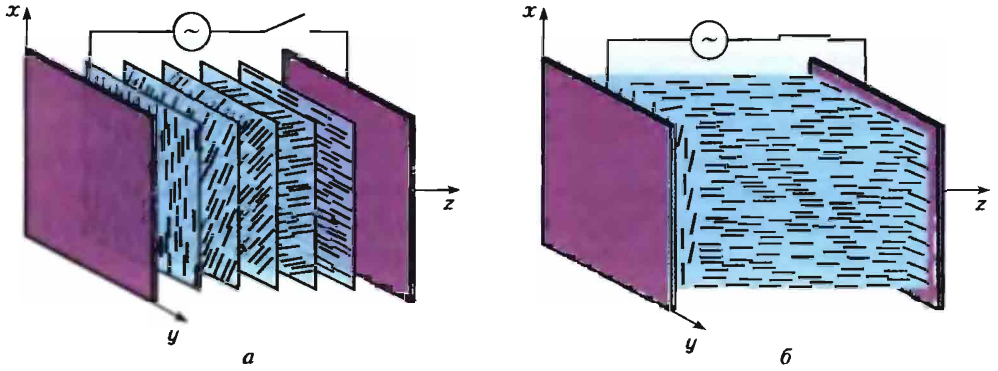
График приложенного напряжения обычно имеет форму прямоугольных импульсов с частотой в диапазоне от десятков герц до нескольких килогерц. Работа на более низких частотах может вызывать электромеханические эффекты, которые нарушают выстраивание молекул и снижают срок службы прибора. Частоты выше 100 кГц приводят к большему потреблению мощности из-за возрастания проводимости. Критическое действующее напряжение $V_{кр}$ обычно составляет несколько вольт.

Жидкие кристаллы инерционны. Время их отклика зависит от толщины жидкокристаллического слоя, вязкости материала, температуры и природы приложенного управляющего напряжения. Время нарастания — порядка десятков миллисекунд, если рабочее напряжение находится вблизи критического напряжения $V_{кр}$, но уменьшается до нескольких миллисекунд при более высоких напряжениях. Время затухания нечувствительно к рабочему напряжению, однако его можно уменьшить, используя ячейки меньшей толщины.

Модуляторы на твист-нематических жидких кристаллах

Твист-нематическая жидкокристаллическая ячейка — это тонкий слой нематического жидкого кристалла между двумя параллельными стеклянными пластинами, натертый таким образом, что ориентация молекул вращается по спирали вокруг оси, нормальной по отношению к пластинам (оси кручения). Например, если угол кручения равен 90°, то молекулы выстроены вдоль оси x у одной пластины и вдоль оси y — у другой (рис. 20.26, *а*). Поперечные слои материала действуют как одноосные кристаллы, оптическая ось которых поворачивается по спирали вокруг оси кручения. В разд. 6.5 было показано, что плоскость поляризации линейно поляризованного света, распространяющегося в направлении оси кручения, поворачивается вместе с молекулами, так что ячейка действует как вращатель плоскости поляризации.

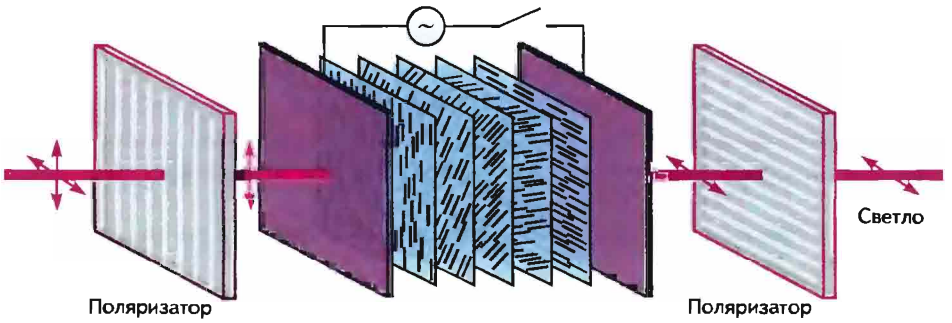
Если электрическое поле приложить вдоль направления оси кручения (оси z), молекулы наклоняются в сторону поля. Когда наклон составляет 90°, ориентация молекул утрачивает свой закрученный характер (кроме непосредственно прилегающих к поверхностям стеклянных пластин), так что способность поворачивать плоскость поляризации также утрачивается. Если электрическое поле убрать, ориентация слоев вблизи поверхностей доминирует, вызывая возврат молекул в первоначальное закрученное состояние. Способность поворачивать плоскость поляризации восстанавливается.



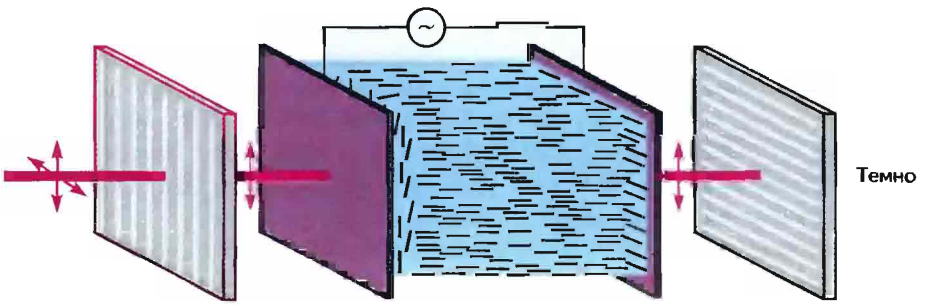
а б

Рис. 20.26. В присутствии достаточно сильного электрического поля молекулы твист-нематического жидкого кристалла наклоняются и их ориентация теряет закрученный характер:

а — закрученное состояние; *б* — наклоненное (незакрученное) состояние



а



б

Рис. 20.27. Твист-нематический жидкокристаллический переключатель:

а — в отсутствие электрического поля жидкокристаллическая ячейка действует как вращатель плоскости поляризации; свет проходит; *б* — при включении поля вращательная способность ячейки временно исчезает и свет блокируется

Поскольку вращательную способность можно выключать и включать путем включения и выключения электрического поля, можно создать оптический затвор, помещая ячейку с кручением на 90° между двумя скрещенными поляризаторами. Система пропускает свет в отсутствие электрического поля и блокирует, когда поле приложено, как показано на рис. 20.27.

Возможна также работа в режиме отражения, как показано на рис. 20.28. Здесь угол кручения составляет 45°, с одной стороны ячейки размещено зеркало, а с другой — поляризатор. В отсутствие электрического поля плоскость поляризации испытывает полный поворот на 90° при распространении туда и обратно через ячейку; следовательно, отраженный свет блокируется поляризатором. В присутствии электрического поля вращательная способность временно исчезает, и отраженный свет проходит через поляризатор. Возможны и другие режимы работы на пропускание и отражение с различными углами кручения.

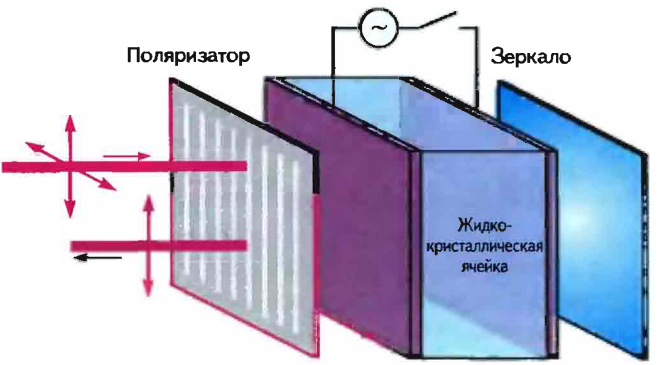


Рис. 20.28. Твист-нематическая жидкокристаллическая ячейка с углом закручивания 45° обеспечивает поворот плоскости поляризации на 90° при полном обходе в отсутствие электрического поля (состояние блокировки) и не поворачивает плоскость поляризации, когда поле наложено (состояние без блокировки). Прибор работает как выключатель

Жидкокристаллическая ячейка с кручением, помещенная между двумя скрещенными поляризаторами, может действовать и как аналоговый модулятор. При промежуточных углах кручения получается комбинация вращателя плоскости поляризации и фазовой пластинки. Анализ пропускания поляризованного света через среду из закрученных и наклоненных молекул весьма сложен, но полный эффект есть частичное пропускание интенсивности. Существует приблизительно линейный участок перехода от полного пропускания в полностью закрученном (ненаклоненном) состоянии и нулевым пропусканием в полностью наклоненном (незакрученном) состоянии. Однако этот динамический диапазон весьма ограничен.

Сегнетоэлектрические жидкие кристаллы

Смектические жидкие кристаллы организованы в слои, как показано на рис. 6.34, *б*. В смектической С-фазе молекулярная ориентация имеет наклон на угол θ по отношению к нормали к слоям (ось x), как показано на рис. 20.29. Материал обладает сегнетоэлектрическими свойствами. При помещении между двумя близко расположенными стеклянными пластинами повер-

хностные взаимодействия оставляют возможными только два устойчивых состояния молекулярной ориентации под углами $\pm\theta$, как показано на рис. 20.29. Когда электрическое поле $+E$ прикладывается в направлении z, возникает вращающий момент, переключающий молекулярную ориентацию в устойчивое состояние $+\theta$ (рис. 20.29, *а*). Молекулы можно переключить в состояние $-\theta$ с помощью электрического поля противоположной полярности $-E$ (рис. 20.29, *б*). Таким образом, ячейка действует как одноосный кристалл, оптическая ось которого может переключаться между двумя ориентациями.

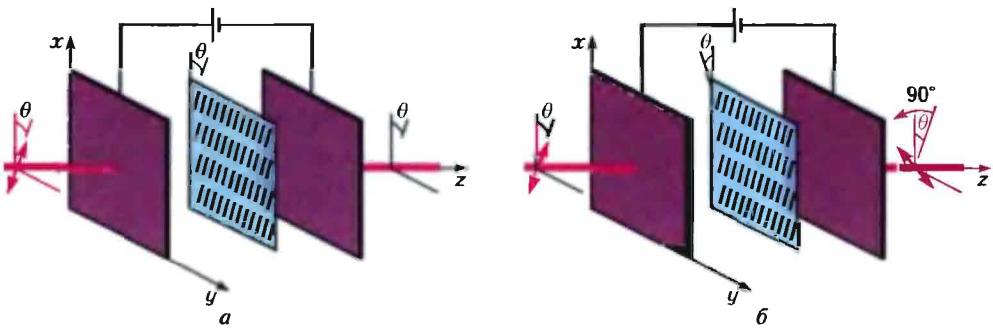


Рис. 20.29. Два состояния сегнетоэлектрической жидкокристаллической ячейки

В геометрии, показанной на рис. 20.29, падающий свет линейно поляризован под углом θ к оси x в плоскости x—y. В состоянии $+\theta$ поляризация параллельна оптической оси и волна распространяется с необыкновенным показателем преломления $n_н$ без приобретения разности фаз. В состоянии $-\theta$ плоскость поляризации образует угол 2θ с оптической осью. Если $2\theta = 45°$, волна приобретает разность фаз между компонентами $\Gamma = 2\pi(n_н - n_о)d/\lambda_0$, где d — толщина ячейки, а $n_о$ — обыкновенный показатель преломления. Если d подобрать таким, чтобы $\Gamma = \pi$, то плоскость поляризации повернется на 90° Таким образом, обращение внешнего электрического поля имеет своим результатом поворот плоскости поляризации на 90°.

Можно создать модулятор интенсивности путем помещения ячейки между двумя скрещенными поляризаторами. Время отклика сегнетоэлектрических жидкокристаллических переключателей в типичном случае <20 мкс при комнатной температуре, что намного быстрее, чем для нематических жидких кристаллов. Типичное напряжение переключения составляет ±10 В.

20.3.2. Пространственные модуляторы света

Жидкокристаллические дисплеи

Жидкокристаллический дисплей (ЖКД) изготавливается путем помещения прозрачных электродов различного рисунка на стеклянные пластины отражательной жидкокристаллической ячейки (нематической, твист-не-

матической или сегнетоэлектрической). Подавая напряжение на избранные электроды, можно получить картину, состоящую из отражаюших и неотражающих областей. На рис. 20.30 показан шаблон из семи полос для изображения цифр от 0 до 9. Большими числами электродов можно управлять последовательно. Сейчас для адресации жидкокристаллических дисплеев используются приборы с зарядовой связью (ПЗС). Разрешение прибора зависит от числа сегментов на единицу площади. ЖКД используются в потребительских изделиях, таких как цифровые часы, карманные калькуляторы, компьютерные мониторы, сотовые телефоны и телевизоры.

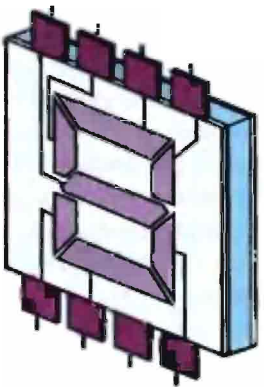


Рис. 20.30. Электроды ЖКД с семью полосами-сегментами

По сравнению с дисплеями на светодиодах (СИД) принципиальным преимуществом ЖКД является низкая потребляемая мощность. Однако ЖКД обладают рядом недостатков.

- Они являются пассивными устройствами, модулирующими свет, который уже существует, а не излучаюшими свой собственный свет. Поэтому они бесполезны в темноте.
- Нематические жидкие кристаллы относительно медленны.
- Оптическая эффективность ограничена в результате использования поляризаторов, которые поглощают, по крайней мере, 50 % неполяризованного падающего света.
- Угол обзора ограничен; контраст модулируемого света уменьшается по мере увеличения угла падения/отражения.

Оптически управляемые пространственные модуляторы света

Большинство ЖКД управляется электрически. Однако оптически управляемые пространственные модуляторы света (ПМС) привлекательны для приложений, включающих обработку изображений и оптических данных. Свет с распределением интенсивности $I_W(x, y)$ — «пишущим» изображением — преобразуется оптоэлектронным сенсором в распределение электрического поля $E(x, y)$, которое управляет коэффициентом отражения $\mathcal{R}(x, y)$ жидкокристаллической ячейки в отражательном режиме. Другая оптическая волна с однородным распределением интенсивности отражается прибором и создает «читаемое» изображение $I(x, y) \propto \mathcal{R}(x, y)$. Таким образом, пишущее изображение управляет читаемым изображением (см. рис. 20.14).

Если пишущее изображение переносится некогерентным светом, а читаемое формируется когерентным светом, то прибор служит в качестве простран-

ственного преобразователя некогерентного света в когерентный, напоминающий ПСОМ из подразд. 20.1.5. Более того, длины волн пишущего и читаемого пучков не обязаны быть одинаковыми. Читающий свет также может быть более интенсивным, чем пишущий, так что прибор можно использовать как усилитель изображения.

Для преобразования пишущего изображения $I_W(x, y)$ в распределение электрического поля $E(x, y)$ в жидкокристаллической ячейке существует несколько средств. Можно использовать слой фотопроводящего материала, например, сульфида кадмия (CdS), помещенный между пластинами конденсатора. При освещении с интенсивностью $I_W(x, y)$ проводимость $G(x, y)$ меняется пропорционально. Конденсатор разряжается в каждой точке в соответствии с локальной проводимостью, так что результирующее распределение электрического поля $E(x, y) \propto 1/I_W(x, y)$ есть негатив исходного изображения. Альтернативный способ — использовать пленочный фотодиод [например, *p*—*i*—*n*-фотодиод из гидрированного аморфного кремния (α-Si:H)]. Обратносмещенный фотодиод проводит в присутствии света, тем самым создавая разность потенциалов, пропорциональную локальной интенсивности света.

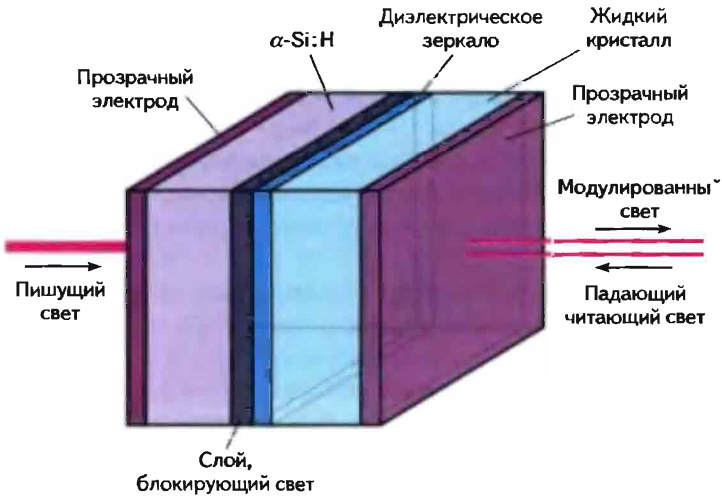


Рис. 20.31. Схема пространственного модулятора с параллельным выстраиванием молекул фирмы Хамамацу (PAL-SLM). Этот оптически управляемый ПМС имеет два главных слоя — слой аморфного кремния, который чувствует интенсивность пишущего света, и жидкокристаллический (ЖК) слой, который служит отражающим фазовым модулятором читающего света. Эти слои разделены диэлектрическим материалом, блокирующим свет. Прибор заключен в стеклянные обкладки (не показаны)

Примером промышленного жидкокристаллического (ЖК) пространственного модулятора света (ПМС) является пространственный модулятор с параллельным выстраиванием молекул фирмы Хамамацу (Hamamatsu Parallel Aligned

Spatial Light Modulator (PAL-SLM)), изображенный на рис. 20.31. В этом приборе в качестве записывающей среды используется α-Si:H, а в качестве фазового модулятора — нематический ЖК с молекулами, выстроенными параллельно. В каждой точке импеданс слоя аморфного кремния меняется пишущим светом, в результате чего в каждой соответственной точке слоя ЖК прикладывается напряжение, пропорциональное оптической интенсивности. Это приводит к повороту анизотропных молекул ЖК, стремящихся выстроиться вдоль приложенного поля. Вследствие этого читающий пучок претерпевает пропорциональный фазовый сдвиг при распространении через слой ЖК. PAL-SLM — непрерывный модулятор (т. е. не разделен на пиксели). Он имеет высокое пространственное разрешение, соответствующее 480 × 480 точек на активную область 2 × 2 см^2, а его время нарастания (спада) составляет 30 (40) мс.

*20.4. ФОТОРЕФРАКТИВНОСТЬ

Фоторефрактивные материалы проявляют фотопроводящие и электрооптические свойства и способны детектировать и сохранять пространственные распределения оптической интенсивности в виде распределения измененного показателя преломления. Фотоиндуцированные заряды создают пространственное распределение заряда, которое, в свою очередь, порождает внутреннее электрическое поле, меняющее показатель преломления за счет электрооптического эффекта.

Обычные фотопроводящие материалы часто являются хорошими изоляторами в темноте. При освещении фотоны поглощаются, генерируются свободные носители заряда (электронно-дырочные пары) и проводимость увеличивается. При выключении света процесс фотогенерации носителей заряда прекращается и проводимость возвращается к своему темновому значению по мере рекомбинации избыточных электронов и дырок. Фотопроводники используются как *детекторы* фотонов (см. разд. 18.2).

При освещении *фоторефрактивного* материала свободные носителя заряда (электроны или дырки) генерируются путем возбуждения с примесных уровней энергии в соответствующую энергетическую зону со скоростью, пропорциональной оптической мощности. Этот процесс сильно напоминает фотопроводимость примесного полупроводника (см. подразд. 18.2.2). Носители заряда диффундируют из точек высокой интенсивности, где они были сгенерированы, оставляя за собой неподвижные заряды противоположного знака (связанные с примесными ионами). Свободные носители могут быть захвачены ионизированными примесями в местах их расположения, вкладывая туда свой заряд при рекомбинации. Результатом является создание неоднородного распределения заряда, которое сохраняется в течение некоторого времени после того, как свет убрали. Это распределение заряда создает *внутреннее* электрическое поле, которое модулирует локальный показатель преломления материала за счет электрооптического эффекта (Поккельса). К изображению возможен оптический доступ

путем мониторирования пространственного распределения показателя преломления пробной оптической волной. Материал можно вернуть в исходное состояние (стереть изображение) путем освещения однородным светом или нагрева. Таким образом, материал можно использовать для записи и хранения изображений, как это делает фотографическая эмульсия. Процесс иллюстрируется рис. 20.32 для легированного ниобата лития ($LiNbO_3$).

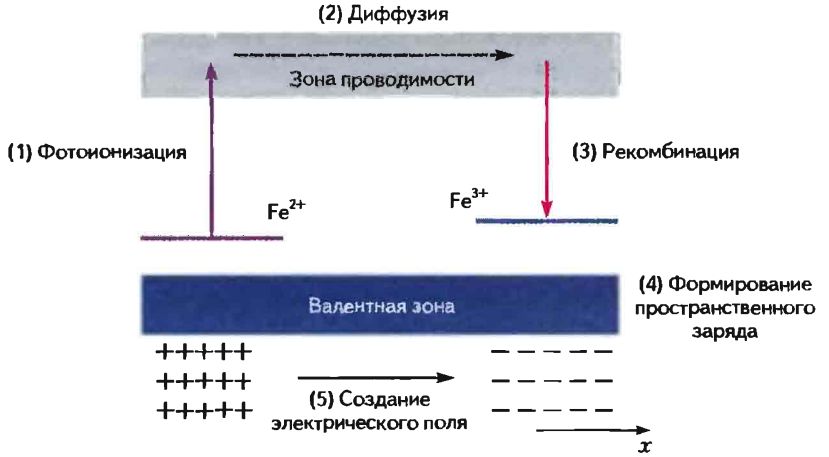


Рис. 20.32. Схема энергетических уровней $LiNbO_3$, иллюстрирующая процессы фотоионизации, диффузии, рекомбинации, формирования пространственного заряда и генерации электрического поля. В качестве доноров используются примесные центры Fe^{2+}. При ионизации они превращаются в Fe^{3+}, которые действуют как ловушки и становятся Fe^{2+} после рекомбинации

Важными фоторефрактивными материалами являются титанат бария ($BaTiO_3$), силикат висмута ($Bi_{12}SiO_{20}$), ниобат лития ($LiNbO_3$), ниобат калия ($KNbO_3$), арсенид галлия (GaAs) и ниобат бария-стронция (SBN).

Упрощенная теория фоторефрактивности

Когда фоторефрактивный материал освещается светом с интенсивностью $I(x)$, меняющейся в направлении x, показатель преломления меняется на $\Delta n(x)$. Ниже приводится описание последовательности промежуточных процессов этого эффекта (см. рис. 20.32) и упрощенная система уравнений, которым они подчиняются.

• *Фотогенерация носителей.* Поглощение фотона в точке x переводит электрон с донорного уровня в зону проводимости. Скорость фотоионизации $G(x)$ пропорциональна как оптической интенсивности, так и плотности числа неионизированных донорных центров. Следовательно,

$$G(x) = s\left(N_D - N_D^+\right) I(x), \tag{20.56}$$

где N_D — плотность числа доноров; N_D^+ — плотность числа ионизированных доноров; s —постоянная, называемая сечением фотоионизации.

• *Диффузия.* Поскольку $I(x)$ неоднородна, плотность числа возбужденных электронов $n(x)$ также неоднородна. В результате электроны диффундируют из областей высокой концентрации в области низкой.

• *Рекомбинация.* Электроны рекомбинируют со скоростью $R(x)$, пропорциональной плотности их числа $n(x)$ и плотности числа ионизированных доноров (ловушек) N_D^+, так что

$$R(x) = \gamma_R n(x) N_D^+, \tag{20.57}$$

где γ_R — постоянная. В состоянии равновесия скорость рекомбинации равна скорости фотоионизации $R(x) = G(x)$, так что

$$sI(x)\left(N_D - N_D^+\right) = \gamma_R n(x) N_D^+, \tag{20.58}$$

откуда

$$n(x) = \frac{s}{\gamma_R} \frac{N_D - N_D^+}{N_D^+} I(x). \tag{20.59}$$

• *Пространственный заряд.* Каждый электрон после фотовозбуждения оставляет после себя положительный ионный заряд. При захвате электрона (рекомбинации) его отрицательный заряд переносится в другую точку. В результате формируется неоднородное пространственное распределение заряда.

• *Электрическое поле.* Этот неоднородный пространственный заряд создает зависящее от местоположения электрическое поле $E(x)$, которое можно определить, замечая, что в стационарном состоянии плотности дрейфового и диффузионного электрического тока должны иметь одинаковую величину и противоположные знаки, так что полная плотность тока обращается в нуль, т. е.

$$J = e\mu_e n(x) E(x) - k_B T \mu_e \frac{dn}{dx} = 0, \tag{20.60}$$

где μ_e — подвижность электронов; k_B — постоянная Больцмана; T — температура. Таким образом,

$$E(x) = \frac{k_B T}{e} \frac{1}{n(x)} \frac{dn}{dx}. \tag{20.61}$$

• *Показатель преломления.* Поскольку материал электрооптический, внутреннее электрическое поле $E(x)$ локально модифицирует показатель преломления в соответствии с формулой

$$\Delta n(x) = -\frac{1}{2} n^3 r E(x), \tag{20.62}$$

где n и r — соответствующие значения показателя преломления и электрооптического коэффициента данного материала [см. (20.4)].

Связь между интенсивностью падающего света $I(x)$ и результирующим изменением показателя преломления $\Delta n(x)$ можно легко получить, если предположить, что отношение $N_D/(N_D^+ - 1)$ в (20.59) приблизительно постоянно и не зависит от x. В этом случае $n(x)$ пропорционально $I(x)$, так что (20.61) дает

$$E(x) = \frac{k_B T}{e} \frac{1}{I(x)} \frac{dI}{dx}. \tag{20.63}$$

Окончательно, подставляя этот результат в (20.62), получаем выражение для зависящего от координат изменения показателя преломления как функции интенсивности

$$\boxed{\Delta n(x) = -\frac{1}{2} n^3 r \frac{k_B T}{e} \frac{1}{I(x)} \frac{dI}{dx}.} \tag{20.64}$$

Изменение показателя преломления

Это уравнение легко обобщается на случай двух измерений, когда оно описывает действие фоторефрактивного материала в качестве прибора для записи изображений.

Для того чтобы сделать теорию простой, было сделано множество допущений. При выводе (20.63) из (20.61) предполагалось, что отношение плотностей числа неионизированных и ионизированных доноров приблизительно постоянно, хотя процесс фотоионизации испытывает пространственные вариации. Это предположение применимо, когда ионизация, кроме $I(x)$, вызвана другими более эффективными процессами, которые не имеют координатной зависимости. Мы пренебрегли темновой проводимостью и объемным фотогальваническим эффектом. Дырки игнорировались. Предполагалось, что внешние электрические поля не налагались, хотя на самом деле это используется в некоторых приложениях. Теория справедлива только для стационарного состояния, хотя ясно, что временная динамика фоторефрактивного процесса очень важна, поскольку именно она определяет скорость отклика фоторефрактивного материала на включение света. Однако, несмотря на все эти допущения, упрощенная теория отражает сущность поведения фоторефрактивных материалов.

Пример 20.7

Запись синусоидального пространственного распределения интенсивности

Рассмотрим распределение интенсивности в виде синусоидальной функции с периодом Λ, контрастом m и средней интенсивностью I_0:

$$I(x) = I_0\left(1 + m\cos\frac{2\pi x}{\Lambda}\right), \tag{20.65}$$

как показано на рис. 20.33. Подставляя это в (20.63) и (20.64), получаем распределения внутреннего электрического поля и показателя преломления

$$E(x) = E_{max}\frac{-\sin(2\pi x/\Lambda)}{1+m\cos(2\pi x/\Lambda)}; \quad \Delta n(x) = \Delta n_{max}\frac{\sin(2\pi x/\Lambda)}{1+m\cos(2\pi x/\Lambda)}, \qquad (20.66)$$

где E_{max} и Δn_{max} — максимальные значения $E(x)$ и $\Delta n(x)$, соответственно,

$$E_{max} = 2\pi\frac{k_B T}{e\Lambda}; \quad \Delta n_{max} = \frac{1}{2}n^3 r E_{max}.$$

Если, например, $\Lambda = 1$ мкм, $m = 1$ и $T = 300$ °K, то $E_{max} = 1{,}6 \cdot 10^5$ В/м. Это внутреннее поле эквивалентно напряжению 1,6 кВ, приложенному к кристаллу толщиной 1 см. Максимальное изменение показателя преломления Δn_{max} прямо пропорционально контрасту m и электрооптическому коэффициенту r и обратно пропорционально пространственному периоду Λ. Решетка $\Delta n(x)$ совершенно нечувствительна к уровню однородного освещения I_0.

Когда контраст изображения мал, вторым членом в знаменателях (20.66) можно пренебречь. Тогда внутреннее электрическое поле и изменение показателя преломления образуют синусоидальные пространственные распределения, смещенные на 90° по отношению к картине интенсивности падающего света:

$$\Delta n(x) \approx \Delta n_{max}\sin\frac{2\pi x}{\Lambda}. \qquad (20.67)$$

Эти распределения показаны на рис. 20.33.

Применение фоторефрактивного эффекта

Изображение $I(x, y)$ можно записать в фоторефрактивном кристалле в виде распределения показателя преломления $\Delta n(x, y)$. Это изображение можно считывать, используя кристалл в качестве пространственного модулятора фазы для пробной однородной плоской оптической волны. Фазовая модуляция может быть преобразована в амплитудную модуляцию, например, путем помещения ячейки в интерферометр.

Ввиду способности записывать изображения, фоторефрактивные материалы интересны для использования в голографии, осуществляемой в реальном масштабе времени (о голографии см. в разд. 4.5). Голографическая запись предметной волны осуществляется посредством смешения ее с опорной волной, как показано для двух плоских волн на рис. 20.34. Интенсивность суммы двух волн образует синусоидальную интерференционную картину, которая записывается в фоторефрактивном кристалле в виде распределения показателя преломления. После этого кристалл может служить объемной фазовой голограммой (см. разд. 4.5, рис. 4.51). Для восстановления записанной предметной волны кристалл освещается опорной волной. Действуя как объемная дифракционная решетка, кристалл отражает опорную волну и воспроизводит предметную волну.

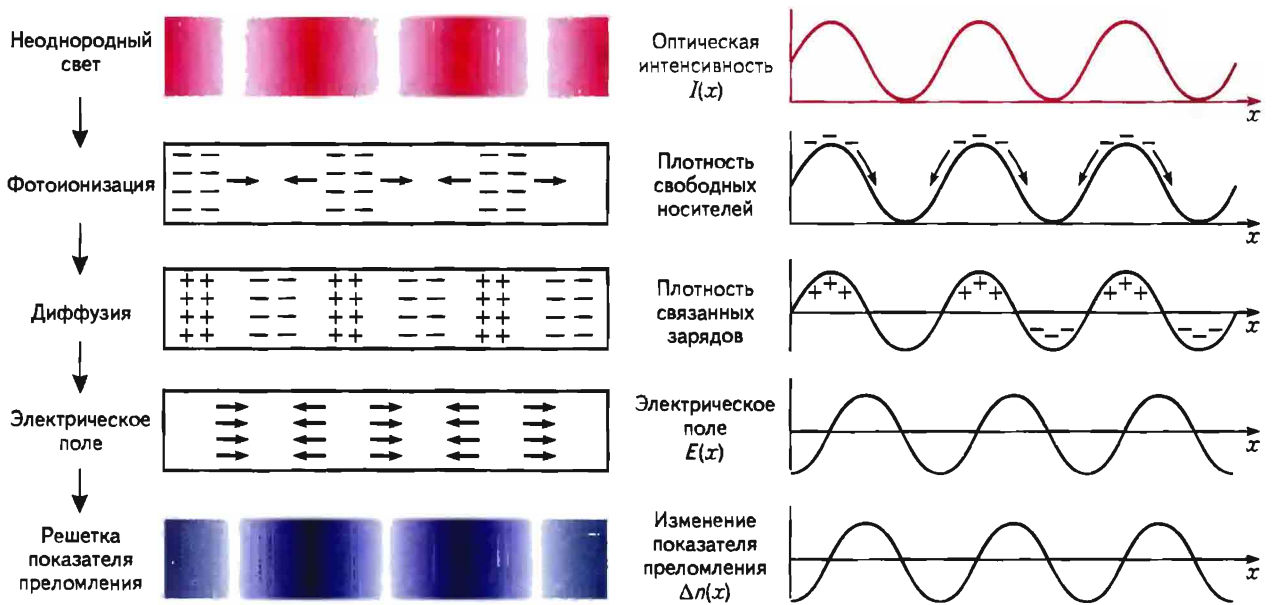


Рис. 20.33. Отклик фоторефрактивного материала на синусоидальное пространственное распределение интенсивности света

Поскольку процесс записи сравнительно быстр, процесс воспроизведения может быть осуществлен одновременно. Предметная и опорная волны вместе распространяются в среде и обмениваются энергией посредством отражения на создаваемой решетке. Такой процесс называется **двухволновым смешением**. Как показано на рис. 20.34 (см. также рис. 4.49), волны 1 и 2 интерферируют и образуют объемную решетку. Волна 1 отражается от решетки и добавляется к волне 2, волна 2 также отражается от решетки и добавляется к волне 1. Таким образом, две волны взаимодействуют друг с другом через посредство решетки, которую они сами создают в среде. Следовательно, распространение волны 1 через среду управляется волной 2 и наоборот. Например, волна 1 может усиливаться за счет волны 2.

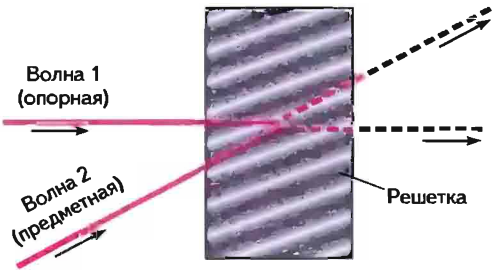


Рис. 20.34. Смешение двух волн как способ динамической голографии

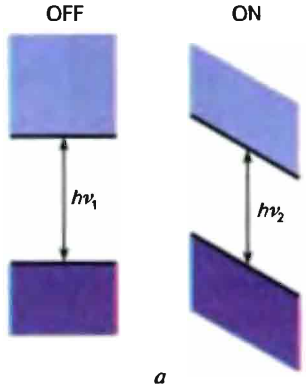
Смешение двух (и более) волн происходит также в других нелинейных оптических материалах, свойства которых зависят от света, что будет обсуждаться в гл. 21. Смешение волн имеет многочисленные приложения в оптической обработке информации (см. гл. 21 и 23), включая усиление изображений, устранение аберраций, определение взаимной корреляции изображений и оптронную связь.

20.5. ЭЛЕКТРОПОГЛОЩЕНИЕ

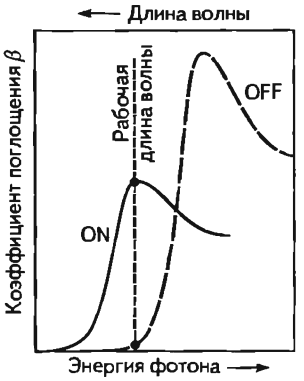
Электропоглощение — это изменение поглощения света в среде под действием приложенного извне электрического поля. В объемном полупроводнике наложение внешнего электрического поля приводит к туннелированию электронов, благодаря которому край поглощения сдвигается в запрещенную зону. В результате энергетическая ширина запрещенной зоны материала уменьшается до значений ниже, чем эффективная ширина с учетом хвоста зоны или хвоста Урбаха, так что $h\nu_2 < h\nu_1$ при включеном поле, как показано на рис. 20.35, *а*. Это явление, известное как эффект Франца—Келдыша, смещает спектр поглощения в сторону более длинных волн (рис. 20.35, *б*). Приложенное электрическое поле приводит также к уширению, а в конечном счете — к исчезновению экситонных пиков поглощения (см. подразд. 16.2.3).

Указанный эффект можно использовать в **электропоглощающих модуляторах** и **переключателях**. В отсутствие электрического поля (OFF) падающий пучок на рабочей длине волны, длина которой больше, чем соответствующая нормальной ширине запрещенной зоны, проходит через среду без поглощения

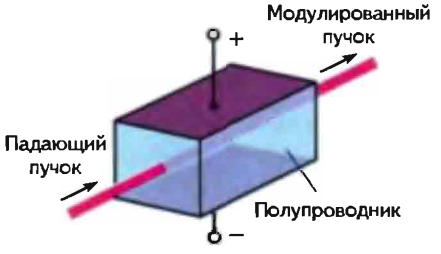
(см. рис. 20.35, *б*). При наложении электрического поля (ON) свет поглощается. Такие модуляторы часто конструктивно выполняются в виде волновода, на который электрическое поле накладывается в перпендикулярном направлении, как показано на рис. 20.35, *в*. По сравнению с электрооптическими модуляторами, которые работают на основе изменения показателя преломления в ответ на приложенное внешнее электрическое поле (см. подразд. 20.1.2 и 20.2.2), электропоглощающие модуляторы, как правило, работают с большими скоростями и при меньших напряжениях. Поскольку они могут быть интегрированы в единый чип с полупроводниковыми источниками света, они удобны для использования в оптоволоконных системах связи. Они также обладают меньшим чирпом, чем непосредственно модулируемые лазерные диоды (см. разд. 22.1).



а



б



в

Рис. 20.35. Эффект Франца—Келдыша:

а — запрещенная зона в отсутствие электрического поля (OFF) при его включении (ON) становится уже; *б* — изменение спектра поглощения, вызванное присутствием электрического поля. Пик поглощения сдвигается в сторону более длинных волн; *в* — модулятор на основе электропоглощения в волноводной конфигурации

Эффект электропоглощения более выражен в полупроводниковых структурах с множественными квантовыми ямами (МКЯ) (см. подразд. 16.1.7 и 17.2.4). Электрическое поле, приложенное в плоскости квантовой ямы, вызывает последствия, аналогичные эффекту Франца—Келдыша, включая сдвиг края поглощения в сторону более длинных волн и диссоциацию экситонов. Однако электрическое поле в направлении ограничения свободы электронов вызывает дополнительные явления, в совокупности называемые **квантоворазмерным эффектом Штарка** (КРЭШ), которые иллюстрируются рис. 20.36.

• Разность энергий между уровнями энергии зоны проводимости и валентной зоны уменьшается с ростом электрического поля ($h\nu_2 < h\nu_1$).

• Наклон зон вызывает сдвиг локализации волновых функций в сторону краев ямы.

• Ионизация экситонов подавляется, а экситонные уровни энергии остаются неуширенными даже при высоких значениях поля, поскольку электрон и дырка продолжают находиться в тесной близости друг к другу благодаря ограничивающему потенциалу.

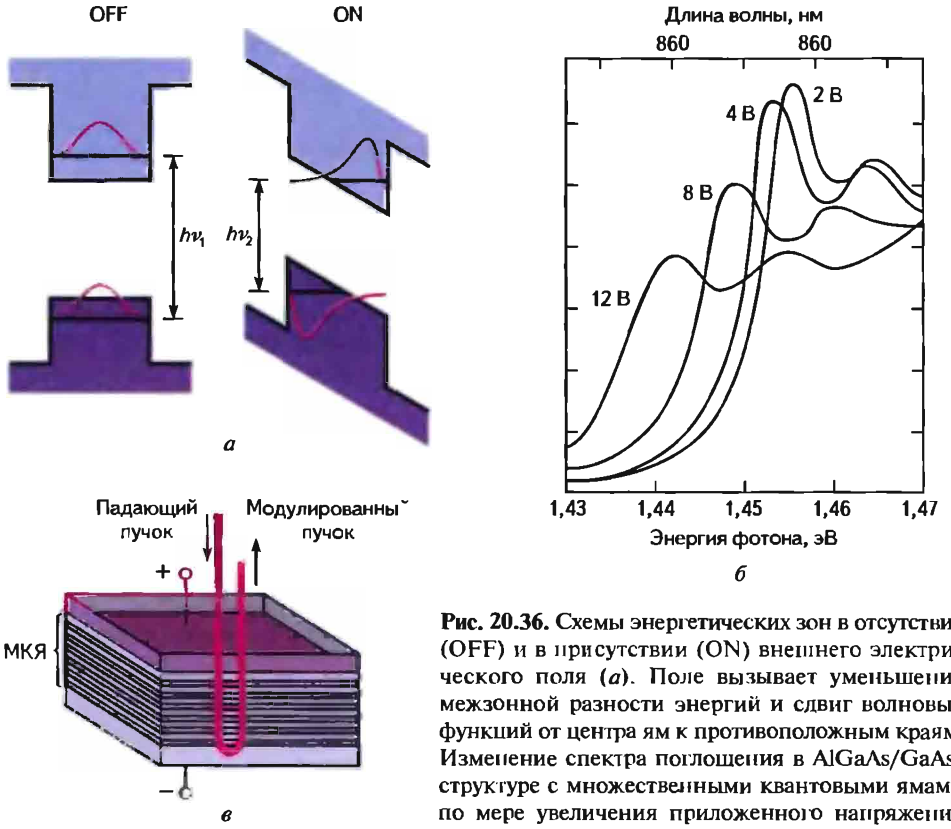


Рис. 20.36. Схемы энергетических зон в отсутствие (OFF) и в присутствии (ON) внешнего электрического поля (*а*). Поле вызывает уменьшение межзонной разности энергий и сдвиг волновых функций от центра ям к противоположным краям. Изменение спектра поглощения в AlGaAs/GaAs-структуре с множественными квантовыми ямами по мере увеличения приложенного напряжения (поля) (*б*). Экситонный пик поглощения смещается в сторону более длинных волн. (Заимствовано из *Miller D.A.B., Chemla D.S., Damen T.C., Wood T.H., Burrus, Jr. C.A., Gossard A.C., Wiegmann W.* The Quantum Well Self-Electrooptic Effect Device: Optoelectronic Bistability and Oscillation, and Self-Linearized Modulation. IEEE Journal of Quantum Electronics. Vol. 21, У1985 IEEE. P. 1462—1476, Fig. 1.) Схема модулятора с электропоглощением на МКЯ поверхностно-нормальной конструкции (*в*)

В результате всего сказанного структуры с МКЯ имеют больший сдвиг пика поглощения и более крутой край поглощения, чем объемные полупроводники. Электропоглощающие модуляторы на основе КРЭШ имеют превосходные характеристики, включающие:

- высокие скорости;
- большие коэффициенты контрастности;
- низкие управляющие напряжения;
- малый чирп.

В простейшей конструкции, работающей на проход, свет направляется через собственную МКЯ-структуру, заключенную между *p*- и *n*-областями, к которым приложено напряжение. Переключение осуществляется просто включением и выключением этого напряжения. Прибор такого рода можно также изготовить в волноводной конфигурации и интегрировать в один чип с РОС-лазером. КРЭШ-модуляторы и переключатели можно также изготавливать в виде матриц, работающих в режиме двойного прохода с поверхностно-нормальной архитектурой, как показано на рис. 20.36, *в*.

Рекомендуемая литература

ОБЩАЯ

См. также списки литературы в гл. 6 и 21, а также книги по оптоэлектронике из гл. 17.

Lines M.E., *Glass A.M.* Principles and Applications of Ferroelectrics and Related Materials. Clarendon, 1977; Oxford University Press, paperback 2nd ed. 2001.

Chigrinov V.G. Liquid Crystal Devices: Physics and Applications. Artech, 1999.

Boreman G.D. Basic Electro-Opticsfor Electrical Engineers. SPIE Optical Engineering Press, 1998.

Efron U., ed. Spatial Light Modulator Technology: Materials, Devices, and Applications. Marcel Dekker, 1995.

Agulló-López F., *Cabrera J.M.*, *Agulló-Rueda F.* Electrooptics: Phenomena, Materials and Applications. Academic Press, 1994.

Karim M.A. Electro-Optical Displays. Marcel Dekker, 1992.

Tamir T., ed. Guided-Wave Optoelectronics. Springer-Verlag, 1988, 2nd ed. 1990.

Hutcheson L.D., ed. Integrated Optical Circuits and Components: Design and Applications. Marcel Dekker, 1987.

Gottlieb M., *Ireland C.L.M.*, *Ley J.M.* Electro-Optic and Acousto-Optic Scanning and Deflection. Marcel Dekker, 1983.

Narasimhamurty T.S. Photoelastic and Electro-Optic Properties of Crystals. Plenum, 1981.

Pankove J.I., ed. Display Devices. Vol. 40: Topics in Applied Physics. Springer-Verlag, 1980.

Elion G.R., *Elion H.A.* Electro-Optics Handbook. Marcel Dekker, 1979.

Nelson D.F. Electric, Optic, and Acoustic Interactions in Dielectrics. Wiley, 1979.

Kaminow I.P. An Introduction to Electrooptic Devices. Academic Press, 1974.

ФОТОРЕФРАКТИВНЫЕ МАТЕРИАЛЫ

Frejlich J. Photo refractive Materials: Fundamental Concepts, Holographic Recording and Materials Characterization. Wiley, 2006.

Günter P., *Huignard J.-P.*, eds. Photo refractive Materials and Their Applications. 3: Applications. Springer-Verlag, 2006.

Günter P., *Huignard J.-P.*, eds. Photo refractive Materials and Their Applications. 2: Materials. Springer-Verlag, 2006.

Günter P., Huignard J.-P., eds. Photorefractive Materials and Their Applications. 1: Basic Effects. Springer-Verlag, 2005.

Yu F.T.S., Yin S. Photorefractive Optics: Materials, Properties, and Applications. Academic Press, 2000.

Davidson F.M., ed. Selected Papers on Photorefractive Materials. SPIE Optical Engineering Press (Milestone Series. Vol. 86), 1994.

Yeh P. Introduction to Photorefractive Nonlinear Optics. Wiley, 1993.

Special issue on photorefractive materials, effects, and devices. Journal of the Optical Society of America B. Vol. 7, № 12, 1990.

Pepper D.M., Feinberg J., Kukhtarev N.K. The Photorefractive Effect. Scientific American. Vol. 263, № 4, 1990. P. 62—74.

Feinberg J. Photorefractive Nonlinear Optics. Physics Today. Vol. 41, № 10, 1988. P 46—52.

Zel'dovich B.Ya., Pilipetsky N.F., Shkunov V.V. Principles of Phase Conjugation. Springer-Verlag, 1985.

Fisher R.A., ed. Optical Phase Conjugation. Academic Press, 1983.

СТАТЬИ

Special issue on electrooptic materials and devices. IEEE Journal of Quantum Electronics. Vol. QE-23, № 12, 1987.

Miller D.A.B., Chemla D.S., Damen T.C., Wood T.H., Burrus, Jr. C.A., Gossard A.C., Wiegmann W. The Quantum Well Self-Electrooptic Effect Device: Optoelectronic Bistability and Oscillation and Self-Linearized Modulation. IEEE Journal of Quantum Electronics. Vol. 21, 1985. P 1462—476.

Wemple S.H., DiDomenico, Jr. M. Electro-Optical and Nonlinear Optical Properties of Crystals. In Applied Solid State Science: Advances in Materials and Device Research. R. Wolfe, ed. Academic Press. Vol. 3, 1972. P. 263—383.

Задачи

К РАЗДЕЛУ 20.1

1. **Время отклика фазового модулятора**. Кристалл GaAs с показателем преломления $n = 3{,}6$ и электрооптическим коэффициентом $r = 1{,}6$ пм/В используется в качестве электрооптического модулятора фазы, работающего на $\lambda_0 = 1{,}3$ мкм в продольной конфигурации. Кристалл имеет длину 3 см и площадь поперечного сечения 1 см^2. Определите полуволновое напряжение V_π, время прохода света через кристалл и электрическую емкость прибора (низкочастотная диэлектрическая постоянная GaAs равна $\varepsilon/\varepsilon_0 = 13{,}5$). Напряжение подается от источника с сопротивлением 50 Ом. Какой фактор ограничивает скорость работы прибора: время прохода света через прибор или время отклика электрической цепи?

2. **Чувствительность интерферометрического электрооптического модулятора интенсивности**. Интегрально-оптический модулятор интенсивности с интерферометром Маха—Цендера, показанный на рис. 20.6, используется как линейный аналоговый модулятор. Если полуволновое напряжение равно $V_\pi = 10$ В, то какова чувствительность прибора (отношение приращения коэффициента пропускания по интенсивности к приращению приложенного напряжения)?

3. **Пьезооптический тензодатчик деформаций.** Пьезооптический материал меняет свой показатель преломления пропорционально деформации. Предложите конструкцию тензодатчика, основанную на этом эффекте. Рассмотрите его интегрально-оптическую версию. Если материал при этом является и электрооптическим, для измерений можно использовать компенсацию изменений показателя преломления за счет деформации и приложенного электрического поля. Измеряемой величиной является электрическое поле, которое обращает в нуль показания фотоприемника в интерферометре Маха—Цендера.

4. **Магнитооптические модуляторы.** Опишите, каким образом фарадеевский вращатель (см. подразд. 6.4.2) можно использовать в качестве модулятора интенсивности.

К РАЗДЕЛУ 20.2

1. **Интегрально-оптический кварцевый модулятор фазы.** Поскольку объемный плавленый кварц обладает центром симметрии, в нем невозможен линейный электрооптический эффект Поккельса. Однако термически ориентированный кварц имеет коэффициенты Поккельса, достаточные для использования в оптических модуляторах. Определите фазовый сдвиг, вносимый интегрально-оптическим фазовым модулятором из ориентированного кварца, конфигурация которого показана на рис. 20.4. Пусть длина электродов $L = 25$ мм, расстояние между ними $d = 30$ мкм, а длина волны $\lambda = 1{,}55$ мкм. Оптическая волна поляризована в направлении y, электрическое поле создается приложенным напряжением $V = 400$ В и направлено вдоль оси y, а волна бежит вдоль электродов в направлении z. Материал ориентирован так, что его главные оси (x_1, x_2, x_3) направлены вдоль осей z, x и y соответственно. Показатель преломления ориентированного материала $m = 1{,}445$, а коэффициенты Поккельса описываются матрицей

$$\begin{bmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{13} & 0 \\ r_{13} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

где $r_{13} = 0{,}15$ пм/В.

2. **Каскадные фазовые модуляторы.**

а. Кристалл KDP ($r_{41} = 8$ пм/В, $r_{63} = 11$ пм/В; $n_o = 1{,}507$, $n_н = 1{,}467$ при $\lambda_0 = 633$ нм) используется в качестве продольного фазового модулятора. Ориентация кристаллических осей и приложенного электрического поля показана в примерах 20.2 и 20.6. Определите полуволновое напряжение V_π при $\lambda_0 = 633$ нм.

6. Электрооптический модулятор фазы состоит из девяти кристаллов KDP, разделенных электродами, на которые подается смещение, как показано на рис. 20.37. Как должны быть ориентированы пластины по отношению друг к другу, чтобы полная фазовая модуляция была максимальной? Вычислите V_π для составного модулятора.

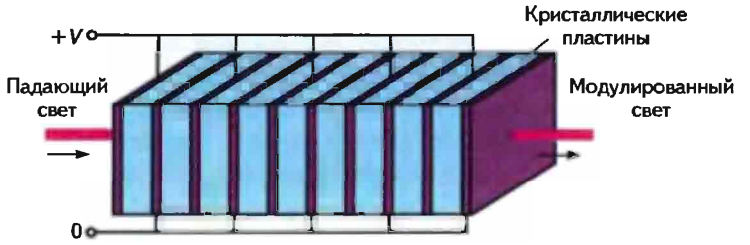


Рис. 20.37

3. **Двухтактный модулятор интенсивности.** В оптическом модуляторе интенсивности использованы два интегрированных электрооптических модулятора фазы и 3-дБ-направленный ответвитель, как показано на рис. 20.38. Входящая волна разделяется на две волны равной амплитуды, каждая из которых подвергается фазовой модуляции, отражается от зеркала, снова модулируется по фазе, после чего обе возвратившиеся волны складываются с помощью направленного ответвителя и образуют выходную волну. Выразите коэффициент пропускания прибора по интенсивности через приложенное напряжение, длину волны, размеры и физические параметры фазового модулятора.

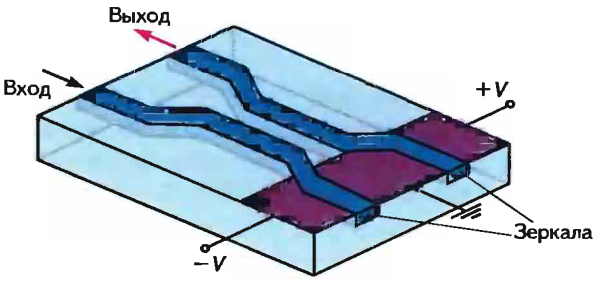


Рис. 20.38

4. **Интегрально-оптический модулятор интенсивности из $LiNbO_3$.** Спроектируйте интегрально-оптический модулятор интенсивности из $LiNbO_3$, используя интерферометр Маха—Цендера, показанный на рис. 20.5. Выберите ориентацию кристалла и поляризацию направляемой волны для наименьшего полуволнового напряжения V_π. Считайте, что активная область имеет длину $L = 1$ мм и ширину $d = 5$ мкм. Длина волны $\lambda_0 = 0{,}85$ мкм, показатели преломления

равны $n_o = 2{,}29$ и $n_н = 2{,}17$, а электрооптические коэффициенты $r_{33} = 30{,}9$, $r_{13} = 8{,}6$, $r_{22} = 3{,}4$ и $r_{42} = 28$ пм/В.

5. **Двойное преломление в электрооптическом кристалле.**

а. Неполяризованный пучок He—Ne-лазера ($\lambda_0 = 633$ нм) проходит через пластину $LiNbO_3$ толщиной 1 см ($n_н = 2{,}17$, $n_o = 2{,}29$, $r_{33} = 30{,}9$ пм/В, $r_{13} = 8{,}6$ пм/В). Пучок перпендикулярен пластине, а оптическая ось лежит в плоскости падения света под углом 45° к пучку. Пучок дважды преломляется (см. подразд. 6.3.5). Определите поперечное смещение и запаздывание между обыкновенным и необыкновенным пучками.

б. Если в направлении, параллельном оптической оси, приложено электрическое поле $E = 30$ В/м, то каково будет его действие на проходящие пучки? Каковы возможные применения такого устройства?

ГЛАВА 21 НЕЛИНЕЙНАЯ ОПТИКА



Николас Бломберген (род. в 1920 г.) в начале 1960 гг. выполнил пионерские исследования по нелинейной оптике. В 1981 г. разделил Нобелевскую премию с Артуром Шавловом.

В течение долгой истории оптики и до сравнительно недавнего времени считалось, что все оптические среды линейны. Из этого допущения вытекают весьма далеко идущие последствия.

- Оптические свойства материалов, такие как показатель преломления и коэффициент поглощения, не зависят от интенсивности света.
- Можно применять принцип суперпозиции — фундаментальный догмат классической оптики.
- Частота света никогда не меняется при прохождении через среду.
- Два пучка света в одной и той же области среды не оказывают воздействия друг на друга, так что свет не может использоваться для управления светом.

Появление первого лазера в 1960 г. дало возможность исследовать поведение света в оптических материалах при значительно больших интенсивностях, чем доступные ранее. Эксперименты, выполненные после появления лазеров, ясно показали, что оптические среды на самом деле проявляют нелинейные свойства. Приведем некоторые примеры.

- Показатель преломления и, следовательно, скорость света в нелинейной оптической среде зависят от интенсивности света.
- Принцип суперпозиции нарушается в нелинейной оптической среде.
- Частота света меняется при прохождении через нелинейную оптическую среду. Например, свет может из красного превратиться в синий.
- Фотоны взаимодействуют друг с другом в нелинейной оптической среде, так что на самом деле свет можно использовать для управления светом.

В области нелинейной оптики проявляется множество удивительных явлений, многие из которых в высшей степени полезны.

Нелинейно-оптическое поведение не наблюдается при распространении лазерного излучения в свободном пространстве. «Нелинейность» — свойство среды, в которой распространяется свет, а не света как такового. Взаимодействие света со светом, следовательно, требует наличия вещества, играющего роль посредника: присутствие оптического поля меняет свойства среды, что, в свою очередь, влияет на другое оптическое поле или на само исходное поле и модифицирует его.

Как обсуждалось в гл. 5, свойства диэлектрической среды, через которую распространяется оптическая электромагнитная волна, описываются соотношением между вектором плотности поляризации $\mathcal{P}(\boldsymbol{r}, t)$ и вектором электрического поля $\mathcal{E}(\boldsymbol{r}, t)$. Полезно рассмотреть $\mathcal{P}(\boldsymbol{r}, t)$ как выходной сигнал системы, на вход которой подается $\mathcal{E}(\boldsymbol{r}, t)$. Математическая связь между векторными функциями $\mathcal{P}(\boldsymbol{r}, t)$ и $\mathcal{E}(\boldsymbol{r}, t)$, определяемая свойствами среды, полностью характеризует систему. Говорят, что среда нелинейна, если нелинейной является эта связь (см. разд. 5.2).

О данной главе

В гл. 5 диэлектрические среды классифицировались по дисперсии, однородности и изотропии (см. разд. 5.2). Чтобы сосредоточиться на основном предмете данной главы — нелинейности, — мы вначале ограничимся средами без дисперсии, однородными и изотропными. Следовательно, векторы $\mathcal{P}$ и $\mathcal{E}$ параллельны в каждой точке и в каждый момент времени и могут изучаться покомпонентно.

Теория нелинейной оптики и ее приложения представлены на двух уровнях. Упрощенный подход дается в разд. 21.1—21.3. Затем следует более подробный анализ тех же явлений (разд. 21.4 и 21.5).

Распространение света в средах, характеризуемых квадратичным (второго порядка) соотношением между $\mathcal{P}$ и $\mathcal{E}$, описано в разд. 21.2 и 21.4. Приложения включают удвоение частот монохроматической волны (*генерацию второй гармоники*), смешение двух монохроматических волн с образованием третьей волны на суммарной или разностной частоте (*преобразование частоты*), использование двух волн для усиления третьей волны (*параметрическое усиление*) и включение обратной связи в параметрический усилитель для получения генератора (*параметрическая генерация*). Распространение волн в среде с кубической (третьего порядка) связью между $\mathcal{P}$ и $\mathcal{E}$ обсуждается в разд. 21.3 и 21.5. Приложения включают *генерацию третьей гармоники, самомодуляцию фазы, самофокусировку, четырехволновое смешение* и *обращение волнового фронта*.

Нелинейная оптика в других главах

Принципиальное допущение, на котором основано рассмотрение явлений в данной главе, состоит в том, что среда пассивна, т. е. она не обменивается энергией со световыми волнами. Волны различных частот могут обме-

ниваться энергией друг с другом благодаря нелинейным свойствам среды, однако их полная энергия не меняется. Этот класс нелинейных явлений называется **параметрическими взаимодействиями**. Некоторые нелинейные оптические непараметрические взаимодействия рассмотрены в других главах настоящей книги.

• *Лазерные взаимодействия.* Взаимодействие света со средой вблизи резонанса с атомными или молекулярными переходами включает такие явления, как поглощение, вынужденное и спонтанное испускание, как описано в разд. 13.3. Эти взаимодействия становятся нелинейными, когда свет достаточно интенсивен, чтобы заметно изменить заселенности различных уровней энергии. Нелинейные оптические эффекты проявляются в насыщении лазерных усилителей и поглотителей излучения (см. разд. 14.4).

• *Многофотонное поглощение.* При высокой интенсивности света может происходить одновременное поглощение нескольких фотонов, суммарная энергия которых равна разности энергий уровней атомного перехода. Для k-фотонного поглощения скорость пропорциональна I^k, где I — оптическая интенсивность. Это нелинейно-оптическое явление кратко описано в подразд. 13.5.2.

• *Нелинейное рассеяние.* Нелинейное неупругое рассеяние включает взаимодействие света с колебательными или акустическими модами среды. Примерами являются вынужденное комбинационное рассеяние и вынужденное рассеяние Мандельштама—Бриллюэна, описанные в подразд. 13.5.3 и 14.3.4.

Во всей этой главе предполагается также, что свет описывается стационарными волнами. К нестационарным нелинейно-оптическим явлениям относятся:

• *Нелинейная оптика импульсов света.* Параметрическое взаимодействие оптических импульсов с нелинейной средой описано в разд. 22.5.

• *Оптические солитоны.* Это импульсы света, распространяющиеся на исключительно большие расстояния в нелинейной среде с дисперсией без изменения ширины или формы. Это нелинейное явление — результат баланса между дисперсией и нелинейной самомодуляцией фазы. Оно описано в подразд. 22.5.2. Использование солитонов в волоконных системах связи описано в подразд. 24.2.5.

Еще один нелинейный оптический эффект — *оптическая бистабильность.* Она возникает в нелинейных оптических системах с обратной связью. Ее применение в фотонных переключателях описано в разд. 23.4.

21.1. НЕЛИНЕЙНЫЕ ОПТИЧЕСКИЕ СРЕДЫ

Линейная диэлектрическая среда характеризуется линейным соотношением между плотностью поляризации и электрическим полем $\mathcal{P} = \varepsilon_0 \chi \mathcal{E}$, где ε_0 — диэлектрическая проницаемость вакуума; χ — диэлектрическая восприимчивость среды (см. подразд. 5.2.1). Нелинейная диэлектрическая среда, напротив, характеризуется нелинейным соотношением между $\mathcal{P}$ и $\mathcal{E}$ (см. подразд. 5.2.2), что иллюстрируется рис. 21.1.

Нелинейность может иметь микроскопическую или макроскопическую природу. Плотность поляризации $\mathcal{P} = N\mathbf{p}$ есть произведение дипольного момента $\mathbf{p}$ отдельной частицы, индуцированного приложенным электрическим полем $\mathcal{E}$,

и плотности числа таких частиц N. Нелинейное поведение может быть связано как с $\mathfrak{p}$, так и с N.

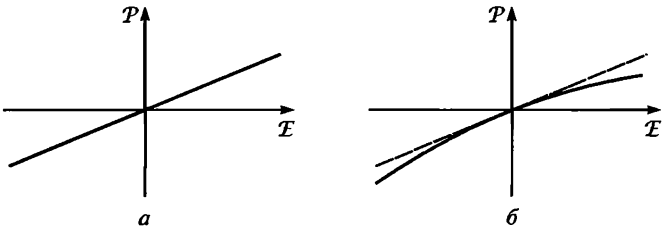


Рис. 21.1. Связь между $\mathcal{P}$ и $\mathcal{E}$ для линейной (*а*) и нелинейной (*б*) диэлектрических сред

Связь между $\mathfrak{p}$ и $\mathcal{E}$ линейна, когда $\mathcal{E}$ мало, но становится нелинейной, когда $\mathcal{E}$ достигает значений, сравнимых с внутриатомными электрическими полями, типичная величина которых составляет $\sim 10^5$—10^8 В/м. Это можно понять в рамках простой модели Лоренца, в которой дипольный момент равен $\mathfrak{p} = -ex$, где x — смещение заряда $-e$ под действием приложенной к нему силы $-e\mathcal{E}$ (см. подразд. 5.5.3). Если возвращающая упругая сила пропорциональна смещению (т. е. справедлив закон Гука), то равновесное смещение x пропорционально $\mathcal{E}$. В этом случае $\mathcal{P}$ пропорционально $\mathcal{E}$ и среда линейна. Однако если возвращающая сила является нелинейной функцией смещения, то равновесное смещение x и плотность поляризации $\mathcal{P}$ будут нелинейными функциями $\mathcal{E}$ и среда нелинейна. Динамика модели ангармонического осциллятора, описывающей диэлектрическую среду с такими свойствами, обсуждается в разд. 21.7.

Другим возможным источником нелинейного отклика оптического материала на действие света является зависимость плотности числа частиц N от оптического поля. Примером является активная среда лазера, где число атомов на энергетических уровнях, участвующих в поглощении и испускании света, зависит от интенсивности самого света (см. разд. 14.4).

Поскольку внешние электрические поля, как правило, малы по сравнению с характерными внутриатомными или кристаллическими полями даже при использовании сфокусированного излучения лазера, нелинейность обычно мала. Зависимость $\mathcal{P}$ от $\mathcal{E}$ примерно линейна для малых $\mathcal{E}$ и лишь слегка отличается от линейной с ростом $\mathcal{E}$ (см. рис. 21.1). При таких условиях функцию, связывающую $\mathcal{P}$ и $\mathcal{E}$, можно разложить в ряд Тейлора в точке $\mathcal{E} = 0$

$$\mathcal{P} = a_1\mathcal{E} + \frac{1}{2}a_2\mathcal{E}^2 + \frac{1}{6}a_3\mathcal{E}^3 + \ldots, \tag{21.1}$$

и рассматривать лишь несколько первых членов разложения. Коэффициенты a_1, a_2 и a_3 представляют собой первую, вторую и третью производные от $\mathcal{P}$ по $\mathcal{E}$, взятые в точке $\mathcal{E} = 0$. Эти коэффициенты являются характеристическими параметрами среды. Первый член, являющийся линейным, преобладает при малых $\mathcal{E}$. Ясно, что $a_1 = \varepsilon_0\chi$, где χ — линейная восприимчивость, связанная с диэлектрической постоянной и показателем преломления материала соотно-

шением $n^2 = \varepsilon/\varepsilon_0 = 1 + \chi$ [см. (5.26)]. Второй член представляет нелинейность второго порядка (квадратичную), третий — нелинейность третьего порядка (кубичную) и т. д.

Удобно записывать (21.1) в виде[1]

$$\boxed{\mathcal{P} = \varepsilon_0 \chi \mathcal{E} + 2d\mathcal{E}^2 + 4\chi^{(3)}\mathcal{E}^3 + \ldots,} \qquad (21.2)$$

где $d = (1/4)a_2$ и $\chi^{(3)} = (1/24)a_3$ — коэффициенты, характеризующие величину нелинейных эффектов второго и третьего порядка соответственно.

Выражение (21.2) математически полностью определяет свойства нелинейной оптической среды. Материальная дисперсия, неоднородность и анизотропия не учитываются с целью упрощения формул и концентрации внимания на основном предмете — проявлениях нелинейности. Разделы 21.6 и 21.7 посвящены анизотропным и диспергирующим нелинейным средам соответственно.

Среды с центром симметрии обладают симметрией относительно инверсии, т. е. свойства среды не меняются при преобразовании $\boldsymbol{r} \to -\boldsymbol{r}$. В таких средах функция $\mathcal{P}(\mathcal{E})$ должна быть нечетной, так чтобы обращение $\mathcal{E}$ приводило к обращению $\mathcal{P}$ без каких-либо других изменений. Тогда нелинейный коэффициент второго порядка должен исчезнуть, и самый низкий порядок возможной нелинейности — третий.

Типичные значения коэффициента нелинейности второго порядка d для диэлектрических кристаллов, полупроводников и органических материалов, используемых в фотонике, лежат в интервале $d = 10^{-24}$—10^{-21} (Кл/В^2 в единицах МКС). Типичные значения коэффициента нелинейности третьего порядка $\chi^{(3)}$ для стекол, кристаллов, полупроводников, стекол, легированных полупроводниками, и органических материалов, используемых в фотонике, находятся в области $\chi^{(3)} = 10^{-34}$—10^{-29} [(Кл м)/В^3 в единицах МКС] Смещенные или асимметричные квантовые ямы позволяют получить большие нелинейности в среднем и дальнем инфракрасном диапазоне.

Упражнение 21.1

Интенсивность света, требуемая для проявления нелинейных эффектов

а. Определите интенсивность света (Вт/см^2), при которой отношение второго члена к первому в правой части (21.2) составляет 1 % в кристалле APD ($NH_4H_2PO_4$) с $n = 1{,}5$ и $d = 6{,}8 \cdot 10^{-24}$ Кл/В^2 при $\lambda_0 = 1{,}06$ мкм.

б. Определите интенсивность света, при которой третий член в (21.2) составляет 1 % от первого члена в дисульфиде углерода (CS_2), для которого $n = 1{,}6$, $d = 0$ и $\chi^{(3)} = 4{,}4 \cdot 10^{-32}$ (Кл м)/В^3 при $\lambda_0 = 694$ нм.

[1] Такие обозначения используются в ряде книг, например *Yariv A.* Quantum Electronics. Wiley, 3rd ed. 1989. В других книгах используются альтернативные обозначения $\mathcal{P} = \varepsilon_0(\chi\mathcal{E} + \chi^{(2)}\mathcal{E}^2 + \chi^{(3)}\mathcal{E}^3)$, см., например, *Shen Y.R.* The Principles of Nonlinear Optics. Wiley, 1984; paperback ed. 2002.

Замечание. В соответствии с (5.70) интенсивность света выражается как

$$I = \frac{|\mathcal{E}|^2}{2\eta} = \frac{\langle \mathcal{E}^2 \rangle}{\eta},$$

где $\eta = \eta_0/n$ — импеданс среды; $\eta_0 = (\mu_0/\varepsilon_0)^{1/2} \approx 377$ Ом — импеданс свободного пространства (см. разд. 5.4).

Нелинейное волновое уравнение

Распространение света в нелинейной среде подчиняется волновому уравнению (5.40), которое было получено из уравнений Максвелла для произвольной однородной изотропной диэлектрической среды. Изотропия среды гарантирует, что векторы $\mathcal{P}$ и $\mathcal{E}$ всегда параллельны, так что их можно рассматривать покомпонентно, что дает:

$$\nabla^2 \mathcal{E} - \frac{1}{c_0^2}\frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathcal{P}}{\partial t^2}. \tag{21.3}$$

Удобно записать плотность поляризации в (21.2) как сумму линейной ($\varepsilon_0 \chi \mathcal{E}$) и нелинейной ($\mathcal{P}_{\text{нл}}$) частей

$$\mathcal{P} = \varepsilon_0 \chi \mathcal{E} + \mathcal{P}_{\text{нл}}; \tag{21.4}$$

$$\mathcal{P}_{\text{нл}} = 2d\mathcal{E}^2 + 4\chi^{(3)}\mathcal{E}^3 + \ldots. \tag{21.5}$$

Использование (21.4) вместе с соотношениями

$$c = \frac{c_0}{n}; \quad n^2 = 1 + \chi; \quad c_0 = \frac{1}{(\varepsilon_0 \mu_0)^{1/2}}$$

[см. формулы (5.26), (5.27)] позволяет записать (21.3) как

$$\boxed{\nabla^2 \mathcal{E} - \frac{1}{c^2}\frac{\partial^2 \mathcal{E}}{\partial t^2} = -S;} \tag{21.6}$$

$$\boxed{S = -\mu_0 \frac{\partial^2 \mathcal{P}_{\text{нл}}}{\partial t^2}.} \tag{21.7}$$

Волновое уравнение в нелинейной среде

Удобно рассматривать (21.6) как волновое уравнение, в котором член S является источником, излучающим в линейной среде с показателем преломления n. Поскольку $\mathcal{P}_{\text{нл}}$ и, следовательно, S являются нелинейной функцией $\mathcal{E}$, (21.6) является нелинейным дифференциальным уравнением относительно $\mathcal{E}$. Это основное уравнение нелинейной оптики.

Для решения этого нелинейного волнового уравнения можно привлечь два приближенных подхода. Первый — это итерационный подход, известный как

борновское приближение. Это приближение лежит в основе упрощенного введения в нелинейную оптику, представленного в разд. 21.2 и 21.3. Второй подход — теория связанных волн, в которой нелинейное волновое уравнение используется для вывода связанных линейных дифференциальных уравнений в частных производных, описывающих взаимодействующие волны. Этот подход — основа более строгой теории волновых взаимодействий в нелинейных средах, представленной в разд. 21.4 и 21.5.

Теория рассеяния в применении к нелинейной оптике: борновское приближение

Радиационный источник S в (21.6) — функция поля $\mathcal{E}$, которое он сам излучает. Чтобы подчеркнуть это обстоятельство, будем писать $S = S(\mathcal{E})$ и проиллюстрируем процесс с помощью простой блок-схемы, показанной на рис. 21.2. Предположим, что оптическое поле $\mathcal{E}_0$ падает на нелинейную среду, ограниченную некоторым объемом, как показано на рисунке. Поле создает источник излучения $S(\mathcal{E}_0)$, который испускает оптическое поле $\mathcal{E}_1$. Соответствующий источник излучения $S(\mathcal{E}_1)$ испускает поле $\mathcal{E}_2$ и так далее. Этот процесс дает итеративное решение, первый шаг которого называется **первым борновским приближением**. Второе борновское приближение дает следующий шаг процесса и так далее. Первое борновское приближение справедливо, если интенсивность света достаточно мала, так что нелинейность является слабой. В этом приближении распространение света сквозь нелинейную среду трактуется как процесс рассеяния, в котором поле падающего света рассеивается средой. Рассеянный свет определяется через падающий в два этапа.

1. Падающее поле $\mathcal{E}_0$ используется для определения нелинейной части плотности поляризации $\mathcal{P}_{\text{нл}}$, из которой определяется источник излучения $S(\mathcal{E}_0)$.

2. Излученное (рассеянное) поле $\mathcal{E}_1$ определяется по найденному источнику посредством сложения сферических волн, связанных с различными его точками (как в теории дифракции, обсуждавшейся в разд. 4.3).

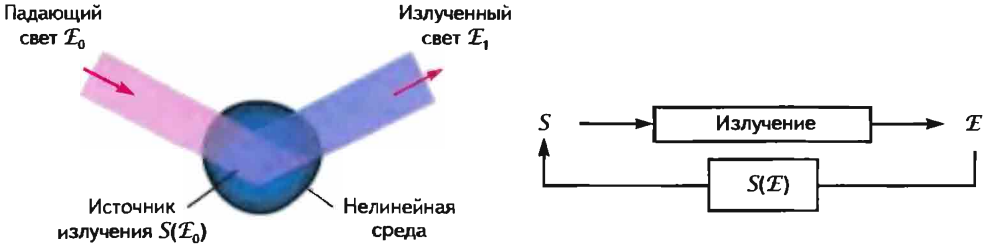


Рис. 21.2. Первое борновское приближение. Падающее оптическое поле $\mathcal{E}_0$ создает источник $S(\mathcal{E}_0)$, который излучает оптическое поле $\mathcal{E}_1$

Теория, представленная в разд. 21.2 и 21.3, основана на первом борновском приближении. Исходное поле $\mathcal{E}_0$ предполагается состоящим из одной или не-

скольких монохроматических волн различных частот. Соответствующая нелинейная поляризация $\mathcal{P}_{нл}$ определяется с помощью (21.5), а функция источника $S(\mathcal{E}_0)$ рассчитывается с помощью (21.7). Поскольку $S(\mathcal{E}_0)$ — нелинейная функция, создаются новые частоты. Следовательно, источник излучает оптическое поле $\mathcal{E}_1$ с новыми частотами, которых не было в исходной волне $\mathcal{E}_0$. Отсюда возникает множество интересных явлений, которые используются для создания нелинейно-оптических приборов.

21.2. НЕЛИНЕЙНАЯ ОПТИКА ВТОРОГО ПОРЯДКА

В этом разделе мы изучим оптические свойства нелинейной среды, у которой нелинейности выше второго порядка пренебрежимо малы, так что

$$\boxed{\mathcal{P}_{нл} = 2d\mathcal{E}^2.} \tag{21.8}$$

Рассмотрим электрическое поле $\mathcal{E}$, состоящее из одной или двух гармонических компонент, и определим спектральные компоненты $\mathcal{P}_{нл}$. В первом борновском приближении источник излучения S содержит те же спектральные компоненты, что и $\mathcal{P}_{нл}$, следовательно, те же компоненты содержит и излученное (рассеянное) поле.

21.2.1. Генерация второй гармоники (ГВГ) и оптическое выпрямление

Рассмотрим отклик такой нелинейной среды на гармоническое электрическое поле с угловой частотой ω (длина волны $\lambda_0 = 2\pi c_0/\omega$) и комплексной амплитудой $E(\omega)$

$$\mathcal{E}(t) = \mathrm{Re}\{E(\omega)\exp(j\omega t)\} = \frac{1}{2}\left[E(\omega)\exp(j\omega t) + E^*(\omega)\exp(-j\omega t)\right]. \tag{21.9}$$

Соответствующая плотность нелинейной поляризации $\mathcal{P}_{нл}$ получается подстановкой (21.9) в (21.8)

$$\mathcal{P}_{нл}(t) = P_{нл}(0) + \mathrm{Re}\{P_{нл}(2\omega)\exp(j2\omega t)\}, \tag{21.10}$$

где

$$P_{нл}(0) = dE(\omega)E^*(\omega); \tag{21.11}$$

$$P_{нл}(2\omega) = dE^2(\omega). \tag{21.12}$$

Графическая схема этого процесса приведена на рис. 21.3.

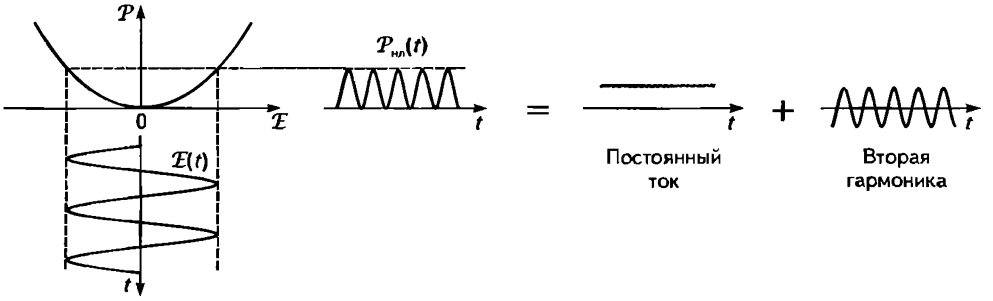


Рис. 21.3. Синусоидальное электрическое поле с угловой частотой ω в средс с нелинейностью второго порядка создает поляризацию с компонентой на 2ω (вторая гармоника) и постоянной компонентой

Генерация второй гармоники (ГВГ)

Источник

$$S(t) = -\mu_0 \frac{\partial^2 \mathcal{P}_{нл}}{\partial t^2},$$

соответствующий (21.10), имеет компоненту на частоте 2ω с комплексной амплитудой $S(2\omega) = 4\mu_0\omega^2 d E(\omega)E(\omega)$, которая излучает оптическое поле на частоте 2ω (длина волны $\lambda_0/2$). Следовательно, рассеянное поле имеет компоненту на второй гармонике падающего оптического поля. Поскольку амплитуда излучаемого света второй гармоники пропорциональна $S(2\omega)$, ее интенсивность пропорциональна $|S(2\omega)|^2$, что, в свою очередь, пропорционально квадрату интенсивности падающей волны

$$I(\omega) = \frac{|E(\omega)|^2}{2\eta}$$

и квадрату коэффициента нелинейности d.

Поскольку испускаемые поля складываются когерентно, интенсивность волны второй гармоники пропорциональна квадрату длины взаимодействия L.

Эффективность генерации второй гармоники $\eta_{ГВГ} = I(2\omega)/I(\omega)$ получается пропорциональной $L^2 I(\omega)$. Так как $I(\omega) = P/A$, где P — мощность падающего света, а A — площадь поперечного сечения области взаимодействия, эффективность ГВГ часто выражают в виде

$$\boxed{\eta_{ГВГ} = C^2 \frac{L^2}{A} P,} \qquad (21.13)$$

Эффективность ГВГ

где C^2 — постоянная (в единицах $Вт^{-1}$), пропорциональная d^2 и ω^2. Выражение для C^2 будет приведено далее [формула (21.106)].

В соответствии с (21.29) для получения максимальной эффективности ГВГ важно, чтобы падающая волна имела наибольшую возможную мощность P. Это достигается путем использования импульсных лазеров, у которых энергия сжата во времени для получения больших пиковых мощностей. Кроме того, для получения максимального отношения L^2/A волна должна быть сфокусирована на минимально возможную площадь A с обеспечением максимально возможной длины взаимодействия L. Если размеры нелинейного кристалла не являются ограничивающими факторами, то максимальное L при заданной площади A ограничивается дифракцией пучка. Например, гауссов пучок, сфокусированный в пятно шириной W_0, сохраняет площадь сечения $A = \pi W_0^2$ на длине перетяжки

$$L = 2z_0 = \frac{2\pi W_0^2}{\lambda}$$

[см. (3.22)], так что получается отношение

$$\frac{L^2}{A} = \frac{2L}{\lambda} = \frac{4A}{\lambda^2}.$$

Пучок нужно фокусировать в пятно максимального размера для достижения наибольшей длины перетяжки. В этом случае эффективность пропорциональна L. Для тонкого кристалла L определяется его толщиной, и пучок нужно фокусировать в пятно минимальной площади A (рис. 21.4, *а*). Для толстого кристалла пучок нужно фокусировать в пятно максимального размера, укладывающегося на поперечном сечении кристалла (рис. 21.4, *б*).

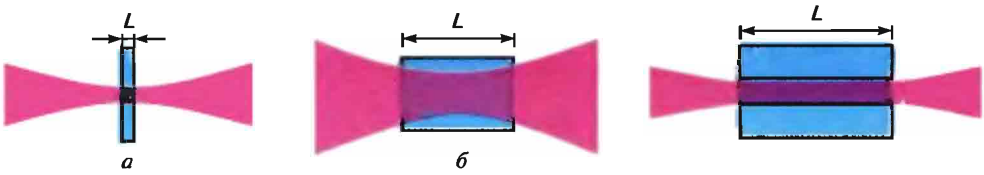


Рис. 21.4. Область взаимодействия в тонком (*а*) и толстом (*б*) кристаллах и в волноводе (*в*)

Волноводные структуры имеют преимущество в том, что они ограничивают свет в поперечном сечении на больших длинах (рис. 21.4, *в*). Поскольку A определяется размером направляемой моды, эффективность пропорциональна L^2. Оптические волноводы могут быть планарными, канальными (см. гл. 8) или волоконными (см. гл. 9). Хотя первоначально волокна из кварцевого стекла вычеркивались из списка кандидатов на генерацию второй гармоники, поскольку плавленый кварц обладает центром симметрии (и поэтому в нем должно быть $d = 0$), в реальных кварцевых волокнах имеет место генерация второй гармоники, которая приписывается электрическим квадрупольным и магнитным дипольным взаимодействиям, а также дефектам и центрам окраски, присутствующим в сердцевине волокна.

На рис. 21.5 (см. цв. вклейку) показаны некоторые схемы генерации второй гармоники в объемных материала и волноводах, в которых инфракрасный свет преобразуется в видимый, а видимый — в ультрафиолетовый.

Оптическое выпрямление

Компонента $\mathcal{P}_{\text{нл}}(0)$ в (21.5) соответствует постоянной (не зависящей от времени) плотности поляризации, которая создает постоянную разность потенциалов на пластинах конденсатора, внутри которого помещен нелинейный материал (рис. 21.6). Генерация постоянного напряжения под действием интенсивного оптического поля представляет собой оптическое выпрямление (аналогичное преобразованию синусоидального переменного напряжения в постоянное напряжение в обычном электронном выпрямителе). Например, оптический импульс с пиковой мощностью в несколько мегаватт может создать напряжение в несколько сотен микровольт.

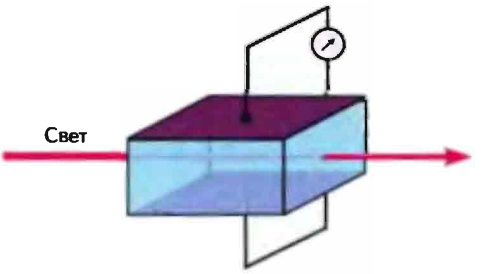


Рис. 21.6. Прохождение интенсивного пучка света через нелинейный кристалл генерирует на нем постоянное напряжение

21.2.2. Электрооптический эффект

Рассмотрим теперь электрическое поле $\mathcal{E}(t)$, состоящее из гармоники на оптической частоте ω и постоянной компоненты (на частоте $\omega = 0$):

$$\mathcal{E}(t) = E(0) + \text{Re}\left\{E(\omega)\exp(j\omega t)\right\}. \tag{21.14}$$

Чтобы различить эти две компоненты, мы обозначили постоянное поле через $E(0)$, а оптическое — через $E(\omega)$. Конечно, обе компоненты являются электрическими полями.

Подставляя (21.14) в (21.8), получим

$$\mathcal{P}_{\text{нл}}(t) = P_{\text{нл}}(0) + \text{Re}\left\{P_{\text{нл}}(\omega)\exp(j\omega t)\right\} + \text{Re}\left\{P_{\text{нл}}(2\omega)\exp(j2\omega t)\right\}, \tag{21.15}$$

где

$$P_{\text{нл}}(0) = d\left[2E^2(0) + |E(\omega)|^2\right]; \tag{21.16а}$$

$$P_{\text{нл}}(\omega) = 4dE(0)E(\omega); \tag{21.16б}$$

$$P_{\text{нл}}(2\omega) = dE^2(\omega), \tag{21.16в}$$

так что плотность поляризации содержит компоненты на угловых частотах 0, ω и 2ω.

Если оптическое поле существенно меньше по величине, чем постоянное электрическое поле, т. е. $|E(\omega)|^2 \ll |E(0)|^2$, то второй гармоникой поляризации $P_{\text{нл}}(2\omega)$ можно пренебречь по сравнению с компонентами $P_{\text{нл}}(0)$ и $P_{\text{нл}}(\omega)$. Это эквивалентно линеаризации $\mathcal{P}_{\text{нл}}$ как функции $\mathcal{E}$, т. е. приближению этой зависимости прямой линией, наклон которой равен производной при $\mathcal{E} = E(0)$, как показано на рис. 21.7.

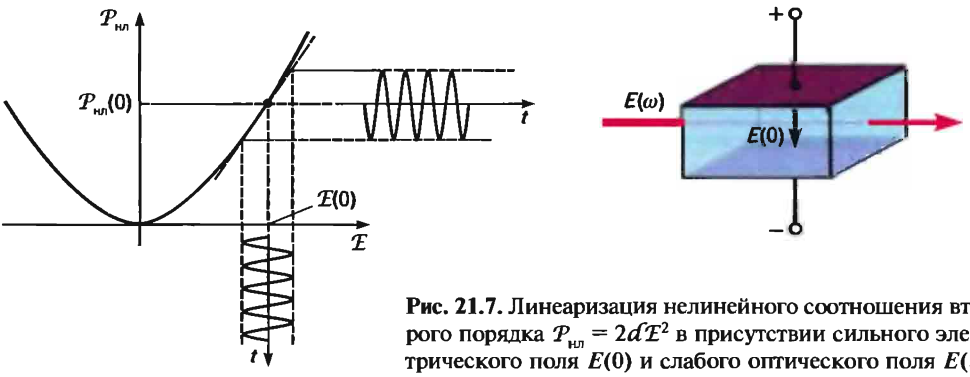


Рис. 21.7. Линеаризация нелинейного соотношения второго порядка $\mathcal{P}_{\text{нл}} = 2d\mathcal{E}^2$ в присутствии сильного электрического поля $E(0)$ и слабого оптического поля $E(\omega)$

Выражение (21.16б) устанавливает линейную связь между $P_{\text{нл}}(\omega)$ и $E(\omega)$, которую мы запишем в виде

$$P_{\text{нл}}(\omega) = \varepsilon_0 \Delta\chi E(\omega),$$

где $\Delta\chi$ — увеличение восприимчивости, пропорциональное электрическому полю $E(0)$,

$$\Delta\chi = \frac{4d}{\varepsilon_0} E(0).$$

Соответствующее приращение показателя преломления получается дифференцированием соотношения $n^2 = 1 + \chi$ и дает $2n\Delta n = \Delta\chi$, откуда

$$\Delta n = \frac{2d}{n\varepsilon_0} E(0). \qquad (21.17)$$

Среда оказывается эффективно линейной с показателем преломления $n + \Delta n$, который линейно управляется электрическим полем $E(0)$.

Нелинейная природа среды создает связь между электрическим полем $E(0)$ и оптическим полем $E(\omega)$, обеспечивая возможность управлять одним полем с помощью другого, так что нелинейная среда проявляет линейные электрооптические свойства (эффект Поккельса), обсуждавшиеся в гл. 20. Этот эффект характеризуется соотношением

$$\Delta n = -\frac{1}{2} n^3 r E(0),$$

где τ — коэффициент Поккельса. Сравнивая эту формулу с (21.17), мы приходим к выводу, что коэффициент Поккельса τ связан с коэффициентом оптической нелинейности второго порядка d соотношением

$$\tau \approx -\frac{4}{\varepsilon_0 n^4} d. \tag{21.18}$$

Хотя это выражение выявляет общую основу эффекта Поккельса и нелинейности среды, оно не согласуется с экспериментально наблюдаемыми величинами τ и d. Это происходит потому, что мы использовали неявное допущение об отсутствии дисперсии в среде (т. е. о том, что отклик среды не зависит от частоты). Это предположение, очевидно, не выполняется, когда одна из компонент поля имеет оптическую частоту ω, а другая постоянна, т. е. имеет нулевую частоту. Роль дисперсии обсуждается в разд. 21.7.

21.2.3. Трехволновое смешение

Рассмотрим теперь случай поля $\mathcal{E}(t)$, содержащего две гармонические компоненты на оптических частотах ω_1 и ω_2:

$$\mathcal{E}(t) = \mathrm{Re}\left\{E(\omega_1)\exp(j\omega_1 t) + E(\omega_2)\exp(j\omega_2 t)\right\}. \tag{21.19}$$

Нелинейная часть поляризации $\mathcal{P}_{\text{нл}} = 2d\mathcal{E}^2$ тогда содержит компоненты на пяти частотах: 0, $2\omega_1$, $2\omega_2$, $\omega_+ = \omega_1 + \omega_2$ и $\omega_- = \omega_1 - \omega_2$ с амплитудами

$$P_{\text{нл}}(0) = d\left[\left|E(\omega_1)\right|^2 + \left|E(\omega_2)\right|^2\right]; \tag{21.20а}$$

$$P_{\text{нл}}(2\omega_1) = dE(\omega_1)E(\omega_1); \tag{21.20б}$$

$$P_{\text{нл}}(2\omega_2) = dE(\omega_2)E(\omega_2); \tag{21.20в}$$

$$P_{\text{нл}}(\omega_+) = 2dE(\omega_1)E(\omega_2); \tag{21.20г}$$

$$P_{\text{нл}}(\omega_-) = 2dE(\omega_1)E^*(\omega_2). \tag{21.20д}$$

Таким образом, среда с нелинейностью второго порядка может использоваться для смешения двух оптических волн различной частоты и генерации третьей волны на разностной или суммарной частоте. Первый из этих процессов называется **преобразованием частоты с понижением**, или **генерацией разностной частоты**, а второй — **преобразованием частоты с повышением (ап-конверсией)**, или **генерацией суммарной частоты**. Пример преобразования частоты вверх показан на рис. 21.8: свет от двух лазеров с вакуумными длинами волн $\lambda_{01} = 1{,}06$ мкм и $\lambda_{02} = 10{,}6$ мкм входит в кристалл прустита и генерирует третью волну с длиной волны $\lambda_{03} = 0{,}96$ мкм (при этом $\lambda_{03}^{-1} = \lambda_{01}^{-1} + \lambda_{02}^{-1}$).

Хотя пара падающих волн на частотах ω_1 и ω_2 создает плотность поляризации на частотах 0, $2\omega_1$, $2\omega_2$, $\omega_+ = \omega_1 + \omega_2$ и $\omega_- = \omega_1 - \omega_2$, соответствующие

гармоники поля не обязательно генерируются, поскольку для этого должны удовлетворяться некоторые дополнительные условия (условия синхронизма), к объяснению которых мы сейчас переходим.

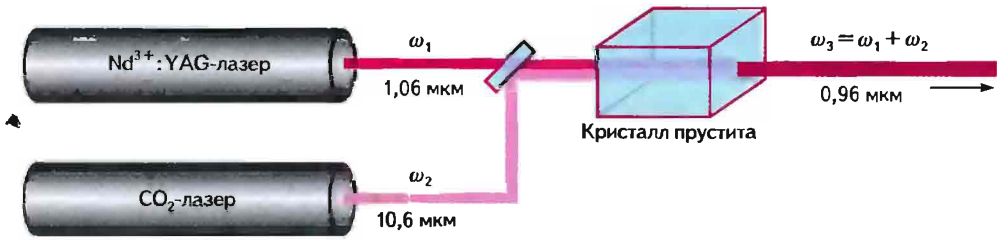


Рис. 21.8. Пример генерации суммарной частоты (ГСЧ), называемой также ап-конверсией, в нелинейном кристалле

Согласование частот и фазовый синхронизм

Если волны 1 и 2 плоские с волновыми векторами $\boldsymbol{k}_1$ и $\boldsymbol{k}_2$, так что

$$E(\omega_1) = A_1 \exp(-j\boldsymbol{k}_1 \cdot \boldsymbol{r}); \quad E(\omega_2) = A_2 \exp(-j\boldsymbol{k}_2 \cdot \boldsymbol{r}),$$

то в соответствии с (21.20г)

$$P_{\text{нл}}(\omega_3) = 2dE(\omega_1)E(\omega_2) = dA_1A_2 \exp(-j\boldsymbol{k}_3 \cdot \boldsymbol{r}),$$

где

$$\boxed{\omega_1 + \omega_2 = \omega_3;} \qquad (21.21)$$

Условие согласования частот

$$\boxed{\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_3.} \qquad (21.22)$$

Условие фазового синхронизма



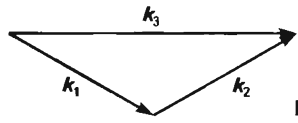


Рис. 21.9. Условие фазового синхронизма

Следовательно, среда действует как источник света частоты $\omega_3 = \omega_1 + \omega_2$ с комплексной амплитудой, пропорциональной $\exp(-j\boldsymbol{k}_3 \cdot \boldsymbol{r})$, т. е. излучает волну с волновым вектором $\boldsymbol{k}_3 = \boldsymbol{k}_1 + \boldsymbol{k}_2$, как показано на рис. 21.9. Равенство (21.22) можно трактовать как условие фазового синхронизма между волновыми фронтами трех волн, которое аналогично условию согласования частот $\omega_1 + \omega_2 = \omega_3$.

Поскольку аргументом комплексной волновой функции является $\omega t - \boldsymbol{k} \cdot \boldsymbol{r}$, оба эти условия вместе гарантируют фазовое согласование трех волн, как во времени, так и в пространстве, что необходимо для поддержания их взаимодействия в течение продолжительного времени и в протяженной области пространства.

Способы трехволнового смешения

При распространении двух оптических волн с угловыми частотами ω_1 и ω_2 через среду с нелинейностью второго порядка они смешиваются и создают плотность поляризации на ряде частот. Предположим, что условия фазового синхронизма выполняются только для гармоники на суммарной частоте $\omega_3 = \omega_1 + \omega_2$. Другие частоты не поддерживаются, поскольку предполагается, что условия синхронизма для них не выполняются.

Раз волна 3 генерируется, она взаимодействует с волной 1 и генерирует волну на разностной частоте $\omega_2 = \omega_3 - \omega_1$. Очевидно, что для такого процесса условие фазового синхронизма также выполняется. Волны 3 и 2 аналогично комбинируют и излучают на частоте ω_1. Следовательно, все три волны связаны, и каждая пара волн взаимодействует, внося вклад в третью волну. Такой процесс называется **трехволновым смешением**.

В общем случае **двухволновое смешение** невозможно. Две волны с произвольными частотами ω_1 и ω_2 не могут связываться средой без помощи третьей волны. Смешение двух волн может происходить только в вырожденном случае, когда $\omega_2 = 2\omega_1$ и вторая гармоника волны 1 дает вклад в волну 2, а субгармоника $\omega_2/2$ волны 2, попадающая на разностную частоту $\omega_2 - \omega_1$, — в волну 1.

Трехволновое смешение относится к **параметрическим процессам**. Оно принимает множество форм в зависимости от того, какая из трех волн подается на вход, а какая выделяется на выходе. Рассмотрим некоторые примеры (рис. 21.10, см. цв. вклейку).

• **Оптическое преобразование частоты (ОПЧ)**. Волны 1 и 2 смешиваются в **ап-конверторе**, генерируя волну на суммарной частоте $\omega_3 = \omega_1 + \omega_2$. Этот процесс называется также **генерацией суммарной частоты (ГСЧ)**. Он уже был проиллюстрирован на рис. 21.8. Генерация второй гармоники (ГВГ) является вырожденным частным случаем ГСЧ. Противоположный процесс **преобразования частоты вниз**, или **генерации разностной частоты**, осуществляется при взаимодействии волн 3 и 1 с образованием волны 2 на разностной частоте $\omega_2 = \omega_3 - \omega_1$. Преобразователи частоты вверх и вниз используются для получения когерентного света на частотах, где нет подходящих лазеров, а также как оптические смесители в оптических системах связи.

• **Оптический параметрический усилитель (ОПУ)**. Волны 1 и 3 взаимодействуют так, что волна 1 растет, при этом создается вспомогательная волна 2. Прибор работает как когерентный усилитель на частоте ω_1 и называется ОПУ. Волна 3, называемая **накачкой**, обеспечивает необходимую энергию, а волна 2 называется **холостой**. Усиливаемая волна называется **сигналом**. Очевидно, усиление зависит от мощности накачки. ОПУ используются для детектирования слабого света на таких длинах волн, где нет чувствительных приемников.

• **Параметрический генератор света (ПГС).** При наличии соответствующей обратной связи параметрический усилитель превращается в параметрический генератор, на который подается только волна накачки. ПГС используются для генерации когерентного света и цугов импульсов с синхронизацией мод в непрерывном диапазоне частот, обычно в таких частотных полосах, где имеется дефицит перестраиваемых лазерных источников.

• **Спонтанное параметрическое рассеяние света (СПРС) с понижением частоты.** В этом случае на вход нелинейного кристалла подается только волна накачки 3, а преобразование в волны низкой частоты 1 и 2 происходит спонтанно. Условия согласования частот и фазового синхронизма (21.21) и (21.22) приводят к множеству решений, каждое из которых дает пару волн 1 и 2 со своими частотами и направлениями. Преобразованный свет пониженной частоты принимает вид многочастотного конуса, как показано на рис. 21.10.

Дальнейшие подробности, касающиеся этих параметрических приборов, приведены в разд. 21.4.

Смешение волн как процесс взаимодействия фотонов.

Процесс трехволнового смешения можно рассматривать с фотонной точки зрения как процесс взаимодействия трех фотонов, при котором два фотона более низкой частоты ω_1 и ω_2 уничтожаются, а фотон более высокой частоты ω_3 рождается, как показано на рис. 21.11, *а*. В другом случае уничтожение фотона высокой частоты ω_3 сопровождается рождением двух фотонов на более низких частотах ω_1 и ω_2, как показано на рис. 21.11, *б*. Поскольку $\hbar\omega$ и $\hbar\boldsymbol{k}$ представляют собой энергию и импульс фотона с частотой ω и волновым вектором $\boldsymbol{k}$ (см. разд. 12.1), законы сохранения энергии и импульса требуют выполнения равенств

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3; \tag{21.23}$$

$$\hbar\boldsymbol{k}_1 + \hbar\boldsymbol{k}_2 = \hbar\boldsymbol{k}_3, \tag{21.24}$$

где $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ и $\boldsymbol{k}_3$ — волновые векторы фотонов. Эти соотношения, очевидно, совпадают с условиями согласования частот и фазового синхронизма (21.21) и (21.22).

Энергетическая схема процесса трехволнового смешения, показанная на рис. 21.11, *б*, напоминает аналогичную схему для генерации трехуровневого лазера с оптической накачкой, показанную на рис. 21.11, *в* (см. подразд. 14.2.2). Однако между этими двумя процессами имеются существенные различия.

• Один из трех переходов, участвующих в процессе лазерной генерации, безызлучательный.

• При лазерной генерации происходит обмен энергией между полем и средой.

• Энергетические уровни, участвующие в процессе лазерной генерации, сравнительно узкие и определены строением атомной или молекулярной системы, тогда как энергетические уровни параметрического процесса диктуются энергией фотонов и условиями синхронизма и могут перестраиваться в широкой области спектра.

Процесс смешения волн включает обмен энергией между взаимодействующими волнами. Очевидно, энергия должна сохраняться, что гарантируется условием согласования частот $\omega_1 + \omega_2 = \omega_3$. Числа фотонов также должны сохраняться соответственно их взаимодействию. Рассмотрим процесс расщепления фотона, представленный на рис. 21.11, *б*. Если $\Delta\phi_1$, $\Delta\phi_2$ и $\Delta\phi_3$ — полные изменения потоков фотонов (число фотонов в секунду) в процессе взаимодействия (поток выходящих фотонов минус поток входящих фотонов) на частотах ω_1, ω_2 и ω_3, то $\Delta\phi_1 = \Delta\phi_2 = -\Delta\phi_3$, так что на каждый потерянный фотон ω_3 приходится по одному возникшему фотону ω_1 и ω_2.

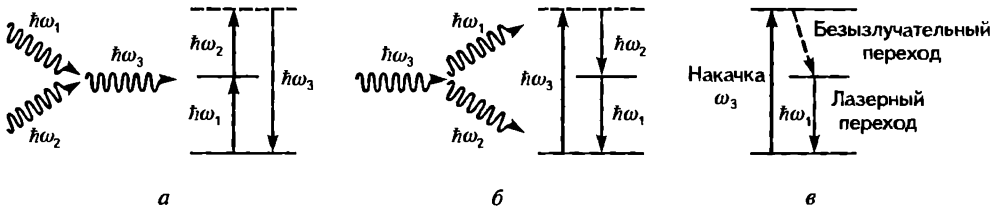


Рис. 21.11. Сравнение параметрических процессов в среде с нелинейностью второго порядка и лазерной генерации:

а — уничтожение двух низкочастотных фотонов и рождение одного высокочастотного. Штриховая линия для верхнего уровня означает, что он виртуальный; *б* — уничтожение одного высокочастотного фотона и рождение двух низкочастотных; *в* — трехуровневый лазер с оптической накачкой — непараметрический процесс, при котором среда участвует в энергетическом обмене

Если три волны распространяются в одном направлении, например вдоль оси z, то взяв в качестве области взаимодействия цилиндр с единичной площадью основания и бесконечно малой высотой $\Delta z \to 0$, мы приходим к заключению, что плотности потоков ϕ_1, ϕ_2 и ϕ_3 [число фотонов/(с · м^2)] для трех волн должны удовлетворять соотношению

$$\boxed{\frac{d\phi_1}{dz} = \frac{d\phi_2}{dz} = -\frac{d\phi_3}{dz}.} \qquad (21.25)$$

Сохранение числа фотонов

Поскольку интенсивности волн (Вт/м^2) равны $I_1 = \hbar\omega_1\phi_1$, $I_2 = \hbar\omega_2\phi_2$ и $I_3 = \hbar\omega_3\phi_3$, формула (21.25) дает

$$\boxed{\frac{d}{dz}\left(\frac{I_1}{\omega_1}\right) = \frac{d}{dz}\left(\frac{I_2}{\omega_2}\right) = -\frac{d}{dz}\left(\frac{I_3}{\omega_3}\right).} \qquad (21.26)$$

Соотношения Мэнли—Роу

Равенства (21.26) известны как соотношения Мэнли—Роу. Они были выведены при рассмотрении взаимодействия волн в нелинейных электронных системах. Соотношения Мэнли—Роу можно вывести в рамках волновой оптики, не привлекая понятий о фотонах (см. упражнение 21.7).

21.2.4. Фазовый синхронизм и кривые настройки

Фазовый синхронизм при коллинеарном трехволновом смешении

Если три смешиваемые волны коллинеарны, т. е. распространяются в одном направлении, а среда лишена дисперсии, то условие фазового синхронизма (21.22) приводит к скалярному уравнению

$$\frac{n\omega_1}{c_0} + \frac{n\omega_2}{c_0} = \frac{n\omega_3}{c_0},$$

которое автоматически следует из условия согласования частот $\omega_1 + \omega_2 = \omega_3$. Однако, поскольку все материалы в действительности обладают дисперсией, три волны распространяются с различными скоростями, соответствующими различным показателям преломления n_1, n_2 и n_3. В этом случае условия согласования частот и фазового синхронизма независимы:

$$\boxed{\omega_1 + \omega_2 = \omega_3; \quad \omega_1 n_1 + \omega_2 n_2 = \omega_3 n_3,} \qquad (21.27)$$

Условия согласования

и должны удовлетворяться одновременно. Поскольку обычно это невозможно, для компенсации дисперсии часто используется двулучепреломление, присутствующее в анизотропных средах.

В анизотропных средах три показателя преломления n_1, n_2 и n_3 в общем случае зависят от поляризации волн и направления их распространения относительно главных осей (см. подразд. 6.3.3). Это создает дополнительные степени свободы для удовлетворения условий согласования. Точный контроль показателей преломления на трех частотах часто достигается соответствующим выбором поляризации, ориентации кристалла и, в некоторых случаях, температуры.

На практике среда часто является одноосным кристаллом с обыкновенным и необыкновенным показателями преломления $n_o(\omega)$ и $n_н(\omega)$, зависящими от частоты. Каждая из трех волн может быть обыкновенной (о) или необыкновенной (н), и процесс обозначается соответственно этому. Например, н-о-о означает, что волны 1, 2 и 3 являются н-, о- и о-волнами соответственно. Для о-волны $n(\omega) = n_o(\omega)$, для н-волны $n(\omega) = n_н(\theta, \omega)$ зависит от угла θ между направлением распространения волны и оптической осью кристалла согласно формуле

$$\frac{1}{n^2(\theta, \omega)} = \frac{\cos^2\theta}{n_o^2(\omega)} + \frac{\sin^2\theta}{n_н^2(\omega)}, \qquad (21.28)$$

которая графически представляется эллипсом [см. (6.58) и рис. 6.50]. Если поляризации сигнала и холостой волны одинаковы, то смешение волн относят к **типу I**, если ортогональны — к **типу II**

Пример 21.1

Коллинеарная генерация второй гармоники (ГВГ) типа I

Для ГВГ волны 1 и 2 имеют одинаковые частоты ($\omega_1 = \omega_2 = \omega$) и $\omega_3 = 2\omega$. Для смешения первого типа волны 1 и 2 имеют одинаковые поляризации, так что $n_1 = n_2$. Тогда из (21.27) следует условие фазового синхронизма $n_3 = n_1$, т. е. основная волна должна иметь такой же показатель преломления, что и вторая гармоника. Из-за дисперсии это условие обычно не выполняется, кроме случая, когда поляризации указанных волн различны. Для одноосных кристаллов соответствующий процесс либо о-о-н, либо н-н-о. В любом случае направление входа волны в кристалл выбирается так, чтобы $n_3 = n_1$, т. е. чтобы двулучепреломление точно компенсировало дисперсию.

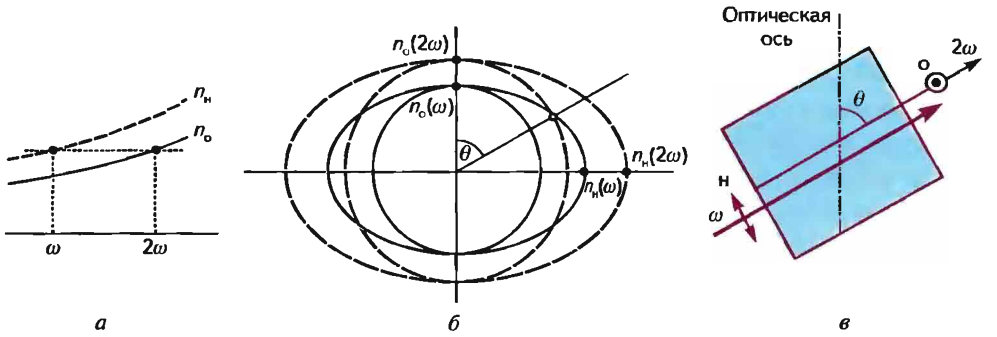


Рис. 21.12. Фазовый синхронизм при н-н-о ГВГ:

а — совпадение показателей преломления н-волны при ω и о-волны при 2ω; *б* — поверхности показателя преломления при ω (сплошные линии) и при 2ω (штриховые линии); *в* — волна направляется под таким углом θ к оптической оси, чтобы необыкновенный показатель преломления $n_н(\theta, \omega)$ волны с частотой ω равнялся обыкновенному показателю преломления $n_о(2\omega)$ волны с частотой 2ω

Для н-н-о-процесса, такого как на рис. 21.12, основная волна необыкновенная, а вторая гармоника обыкновенная, $n_1 = n(\theta, \omega)$ и $n_3 = n_о(2\omega)$, так что условие синхронизма есть $n(\theta, \omega) = n_о(2\omega)$. Его выполнение достигается подбором угла θ, для которого

$$\boxed{n(\theta, \omega) = n_о(2\omega),} \qquad (21.29)$$

ГВГ н-н-о типа I

где $n(\theta, \omega)$ дается выражением (21.28). Графическое пояснение приведено на рис. 21.12, где показаны обыкновенный и необыкновенный показатели преломления (окружность и эллипс) при ω (сплошные линии) и при 2ω (штриховые линии). Угол, при котором выполняется условие синхронизма, соответствует пересечению круга при 2ω с эллипсом при ω.

Например, для KDP при основной длине волны $\lambda = 694$ нм $n_о(\omega) = 1{,}506$, $n_н(\omega) = 1{,}466$, а при $\lambda/2 = 347$ нм $n_о(2\omega) = 1{,}534$, $n_н(2\omega) = 1{,}490$. В этом случае (21.29) и (21.28) дают $\theta = 52°$. Этот угол называется углом среза кристалла.

Такие же уравнения можно написать для ГВГ в конфигурации о-о-н. В этом случае для KDP на фундаментальной длине волны $\lambda = 1{,}06$ мкм $\theta = 41°$

Пример 21.2

Коллинеарный параметрический генератор света (ПГС)

Частоты генерации ПГС определяются условиями согласования частот и фазового синхронизма. Для о-о-н конфигурации смешения типа I

$$\boxed{\omega_1 + \omega_2 = \omega_3; \quad \omega_1 n_o(\omega_1) + \omega_2 n_o(\omega_2) = \omega_3 n(\theta, \omega_3).} \qquad (21.30)$$

ПГС о-о-н типа I

Для н-о-н-смешения типа II

$$\boxed{\omega_1 + \omega_2 = \omega_3; \quad \omega_1 n(\theta, \omega_1) + \omega_2 n_o(\omega_2) = \omega_3 n(\theta, \omega_3).} \qquad (21.31)$$

ПГС н-о-н типа II

Функции $n_o(\omega)$ и $n_н(\omega)$ определяются из уравнения Зельмейера (5.111), а необыкновенный показатель преломления $n(\theta, \omega)$ определяется как функция угла θ между оптической осью кристалла и направлением распространения волн с помощью (21.28). Для заданной частоты накачки ω_3 найденные из уравнений (21.30) и (21.31) частоты ω_1 и ω_2 часто изображают в зависимости от угла θ. Такие графики называют настроечными кривыми. Примеры таких кривых приведены на рис. 21.13 (см. цв. вклейку).

Фазовый синхронизм при неколлинеарном трехволновом смешении

В неколлинеарном случае условие фазового синхронизма $\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_3$ эквивалентно

$$\omega_1 n_1 \hat{u}_1 + \omega_2 n_2 \hat{u}_2 = \omega_3 n_3 \hat{u}_3,$$

где $\hat{u}_1$, $\hat{u}_2$, $\hat{u}_3$ — единичные векторы направления распространения волн. Показатели преломления n_1, n_2, n_3 зависят от направления распространения волн по отношению к осям кристалла, а также от поляризации и частоты. Это векторное равенство эквивалентно двум скалярным, так что условия синхронизма приобретают вид

$$\boxed{\begin{array}{l} \omega_1 + \omega_2 = \omega_3; \quad \omega_1 n_1 \sin\theta_1 = \omega_2 n_2 \sin\theta_2; \\ \omega_1 n_1 \cos\theta_1 + \omega_2 n_2 \cos\theta_2 = \omega_3 n_3, \end{array}} \qquad (21.32)$$

где θ_1 и θ_2 — углы, образованные направлениями волн 1 и 2 с направлением волны 3. Центральной задачей проектирования приборов для трехволнового смешения является выбор направлений и поляризаций, удовлетворяющих приведенным уравнениям. Покажем это на ряде примеров и упражнений.

Упражнение 21.2

Неколлинеарная генерация второй гармоники (ГВГ) типа II

На рис. 21.14 показана неколлинеарная о-н-н ГВГ II типа. Обыкновенная и необыкновенная волны, обе на основной частоте ω, создают необыкновенную волну второй гармоники на частоте 2ω. Предполагается, что направления распространения трех волн и оптическая ось лежат в одной плоскости, две фундаментальные волны образуют с оптической осью углы θ_1 и θ_2, а волна второй гармоники — угол θ. Показатели преломления, входящие в условия фазового синхронизма, равны $n_1 = n_o(\omega)$, $n_2 = n(\theta + \theta_2, \omega)$ и $n_3 = n(\theta, 2\omega)$, т. е.

$$n_o(\omega)\sin\theta_1 = n(\theta+\theta_2,\ \omega)\sin\theta_2;$$
$$n_o\cos\theta_1 + n(\theta+\theta_2,\ \omega)\cos\theta_2 = 2n(\theta,\ 2\omega). \qquad (21.33)$$

о-н-н ГВГ II типа

Для кристалла KDP и основной волны с длиной волны 1,06 мкм (Nd:YAG-лазер) определите ориентацию кристалла и углы θ_1 и θ_2, обеспечивающие эффективную генерацию второй гармоники.

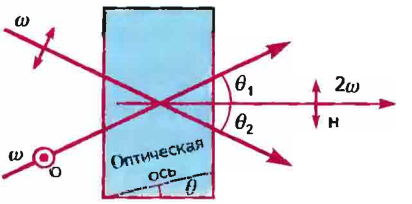


Рис. 21.14. Неколлинеарная генерация второй гармоники типа II

Пример 21.3

Спонтанное параметрическое рассеяние света (СПРС) с преобразованием частоты вниз

При СПРС волна накачки с частотой ω_3 рождает пару волн с частотами ω_1 и ω_2 под углами θ_1 и θ_2, удовлетворяющими условиям согласования частот и фазового синхронизма (21.32). Например, в случае о-о-н типа I $n_1 = n_o(\omega_1)$, $n_2 = n_o(\omega_2)$ и $n_3 = n(\theta, \omega_3)$. Эти соотношения вместе с уравнениями Зельмейера для $n_o(\omega)$ и $n_н(\omega)$ порождают множество решений (ω_1, θ_1), (ω_2, θ_2) для сигнальной и холостой волн, как видно из примера на рис. 21.15 (см. цв. вклейку).

Допустимая фазовая расстройка и длина когерентности

Небольшая фазовая расстройка $\Delta\boldsymbol{k} = \boldsymbol{k}_3 - \boldsymbol{k}_1 - \boldsymbol{k}_2 \neq 0$ может приводить к существенному снижению эффективности смешения волн. Если волны 1 и 2 — плоские с волновыми векторами $\boldsymbol{k}_1$ и $\boldsymbol{k}_2$, так что $E(\omega_1) = A_1 \exp(-j\boldsymbol{k}_1 \cdot \boldsymbol{r})$ и $E(\omega_2) = A_2 \exp(-j\boldsymbol{k}_2 \cdot \boldsymbol{r})$, то, в соответствии с (21.20г),

$$P_{нл}(\omega_3) = 2dE(\omega_1)E(\omega_2) = 2dA_1A_2\exp\left[-j(\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{r}\right] =$$
$$= 2dA_1A_2\exp(j\Delta\boldsymbol{k}\cdot\boldsymbol{r})\exp(-j\boldsymbol{k}_3\cdot\boldsymbol{r}).$$

Эта поляризация создает источник света (21.7) с угловой частотой ω_3, волновым вектором k_3 и комплексной амплитудой $2\omega_3^2\mu_0 dA_1A_2 \exp(j\Delta k \cdot r)$. Можно показать (см. задачу 5 к разд. 21.2), что интенсивность генерируемой волны пропорциональна квадрату интеграла от амплитуды источника по объему взаимодействия V

$$I_3 \propto \left| \int_V dA_1 A_2 \exp(j\Delta k \cdot r)\, dr \right|^2 \tag{21.34}$$

Поскольку вклады различных точек области взаимодействия складываются как комплексные числа, зависящая от координат фаза $\Delta k \cdot r$ при наличии расстройки снижает полную интенсивность по сравнению со случаем точного синхронизма. Рассмотрим частный случай одномерного взаимодействия. Пусть длина области взаимодействия вдоль оси z равна L:

$$I_3 \propto \left| \int_0^L \exp(j\Delta k z)\, dz \right|^2 = L^2 \operatorname{sinc}^2\left(\frac{\Delta k L}{2\pi}\right),$$

где Δk — z-компонента вектора Δk; $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$. Отсюда следует, что при наличии расстройки волнового вектора Δk величина I_3 уменьшается на множитель $\operatorname{sinc}^2(\Delta kL/2\pi)$, который равен единице при $\Delta k = 0$ и падает с ростом Δk, достигая значения $(2/\pi)^2 \approx 0{,}4$ при $|\Delta k| = \pi/L$. При $|\Delta k| = 2\pi/L$ (рис. 21.16) эта величина обращается в нуль. Для данного L расстройка Δk, соответствующая заданному уменьшению эффективности преобразования, обратно пропорциональна L, так что требование фазового синхронизма становится более строгим по мере увеличения L. Для данной расстройки Δk длина

$$\boxed{L_к = \frac{2\pi}{|\Delta k|}} \tag{21.35}$$

Длина когерентности

является мерой максимального расстояния, на котором параметрический процесс эффективен; величину $L_к$ часто называют **длиной когерентности** смешения волн.



Рис. 21.16. Фактор снижения эффективности трехволнового смешения в результате наличия фазовой расстройки ΔkL между волнами, взаимодействующими в пределах расстояния L

Например, для генерации второй гармоники

$$|\Delta k| = 2\frac{2\pi}{\lambda_0}|n_3 - n_1|,$$

где λ_0 — длина волны основной гармоники в свободном пространстве; n_1 и n_3 — показатели преломления основной волны и второй гармоники. В этом случае длина

$$L_к = \frac{\lambda_0}{2}|n_3 - n_1|$$

обратно пропорциональна величине $|n_3 - n_1|$, определяемой материальной дисперсией. Например, для $|n_3 - n_1| = 10^2$ получаем $L_к = 50\lambda$.

То, что процесс взаимодействия возможен при наличии фазовой расстройки, можно рассматривать как следствие неопределенности волнового вектора $\Delta k \propto 1/L$, связанной с ограничением волн в пределах расстояния L [см. (А.13) в приложении А]. Соответствующая неопределенность импульса $\Delta p = \hbar\Delta k \propto 1/L$ объясняет очевидное нарушение закона сохранения импульса в процессе смешения волн.

Ширина полосы фазового синхронизма

Как отмечалось выше, для конечной длины взаимодействия L допустима фазовая расстройка $|\Delta k| \le 2\pi/L$. Если точный фазовый синхронизм достигается на множестве номинальных частот смешиваемых волн, то малые уходы частоты от номинала допустимы, поскольку выполняется условие $\omega_1 + \omega_2 = \omega_3$. Ширина спектральных полос, связанных с такой допустимостью, определяется условием $|\Delta k| \le 2\pi/L$.

Например, при ГВГ имеется две волны с частотами $\omega_1 = \omega$ и $\omega_3 = 2\omega$. Расстройка Δk является функцией $\Delta k(\omega)$ основной частоты ω. Прибор проектируется в расчете на номинальную основную частоту ω_0, т. е. $\Delta k(\omega_0) = 0$. Тогда ширина полосы $\Delta\omega$ устанавливается условием $|\Delta k(\omega_0 + \Delta\omega)| = 2\pi/L$. Если $\Delta\omega$ достаточно мало, можно написать $\Delta k(\omega_0 + \Delta\omega) = \Delta k'\Delta\omega$, где $\Delta k' = (\mathrm{d}/\mathrm{d}\omega)\Delta k$ при ω_0. Следовательно, $\Delta\omega = 2\pi/|\Delta k'|L$, откуда спектральная ширина (Гц):

$$\boxed{\Delta\nu = \frac{1}{|\Delta k'|L}.} \qquad (21.36)$$

Ширина полосы фазового синхронизма

Поскольку $\Delta k(\omega) = k_3(2\omega) - 2k_1(\omega)$, производная

$$\Delta k' = \frac{\mathrm{d}k_3(2\omega)}{\mathrm{d}\omega} - \frac{\mathrm{d}k_1(\omega)}{\mathrm{d}\omega} = 2\left[\frac{\mathrm{d}k_3(2\omega)}{\mathrm{d}(2\omega)} - \frac{\mathrm{d}k_1(\omega)}{\mathrm{d}\omega}\right] = 2\left(\frac{1}{v_3} - \frac{1}{v_1}\right),$$

где v_1 и v_3 — групповые скорости волн 1 и 3 на частотах ω и 2ω соответственно (см. разд. 5.6). Спектральная ширина, следовательно, связана с длиной и расстройкой групповых скоростей соотношением

$$\boxed{\Delta\nu = \frac{1}{2}\left|\frac{L}{v_3} - \frac{L}{v_1}\right|^{-1} = \frac{c_0}{2L}\frac{1}{|N_3 - N_1|},} \qquad (21.37)$$

Ширина полосы фазового синхронизма

где N_1 и N_3 — групповые показатели преломления материала на основной частоте и частоте второй гармоники.

Очевидно, что генерация второй гармоники широкополосной волны или ультракороткого импульса (см. разд. 23.5) может быть осуществлена с помощью тонкого кристалла (за счет более низкой эффективности преобразования) при дополнительном конструктивном ограничении — совпадении групповых скоростей $v_3 \approx v_1$ или $N_3 \approx N_1$. Допустимые отклонения от фазового синхронизма при СПРС показаны на рис 21.15 толщиной линий.

21.2.5. Квазисинхронизм

При наличии расстройки волновых векторов $\Delta\boldsymbol{k}$ различные точки области взаимодействия излучают с фазами $\Delta\boldsymbol{k}\cdot\boldsymbol{r}$, зависящими от координат, поэтому амплитуда параметрически генерируемой волны существенно снижается. Поскольку фазового синхронизма добиться бывает трудно, либо это связано с сильными ограничениями в выборе нелинейного коэффициента или формы кристалла, которая обеспечивает максимальную эффективность преобразования волны, разработан подход, при котором допустимая фазовая расстройка компенсируется за счет использования среды с периодической зависимостью коэффициента нелинейности от координат. Такая периодичность вносит противоположный вклад в фазу и улучшает фазовое согласование излучающих элементов среды. Метод носит название **квазисинхронизма**.

Если среда имеет коэффициент нелинейности $d(\boldsymbol{r})$, зависящий от координат, то (21.34) приобретает вид

$$I_3 \propto \left|\int_V d(\boldsymbol{r})\exp(j\Delta\boldsymbol{k}\cdot\boldsymbol{r})\,d\boldsymbol{r}\right|^2 \tag{21.38}$$

Если $d(\boldsymbol{r})$ — гармоническая функция $d(\boldsymbol{r}) = d_0\exp(-j\boldsymbol{G}\cdot\boldsymbol{r})$, то при $\boldsymbol{G} = \Delta\boldsymbol{k}$ фазовая расстройка полностью исключается. Соответственно этому условие фазового синхронизма заменяется на следующее

$$\boxed{\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{G} = \boldsymbol{k}_3.} \tag{21.39}$$

Фактически нелинейная среда служит фазовой решеткой (или продольной решеткой Брэгга) с волновым вектором $\boldsymbol{G}$.

Вообще говоря, трудно изготовить среду с непрерывно меняющимся по гармоническому закону коэффициентом нелинейности $d(\boldsymbol{r}) = d_0\exp(-j\boldsymbol{G}\cdot\boldsymbol{r})$, но можно изготовить более простые периодические структуры, например, среды, у которых коэффициент нелинейности постоянен по величине, но периодически меняет свой знак. Поскольку каждая периодическая функция может быть разложена в ряд Фурье, состоящий из гармоник, одна из них может служить для коррекции фазовой расстройки, а другие не будут играть никакой роли в процессе смешения волн, поскольку они вносят фазовые расстройки большей величины.

Квазисинхронизм при коллинеарном смешении волн

Для коллинеарных волн, распространяющихся вдоль оси z и имеющих фазовую расстройку Δk, требуемая фазовая решетка имеет вид $\exp(-jGz)$, где $G = \Delta k$. Такая решетка может быть получена с помощью периодического коэффициента нелинейности $d(z)$, описываемого рядом Фурье

$$d(z) = \sum_{m=-\infty}^{\infty} d_m \exp\left(-\frac{j2\pi mz}{\Lambda}\right),$$

где Λ — период; $\{d_m\}$ — коэффициенты Фурье. Любая из этих компонент может быть использована для согласования фаз. Например, для m-й гармоники $G = m2\pi/\Lambda = \Delta k$, так что

$$\boxed{\Lambda = \frac{m2\pi}{\Delta k} = mL_к,} \qquad (21.40)$$

Условие квазисинхронизма

т. е. период решетки Λ равен целому числу длин когерентности $L_к = 2\pi/\Delta k$.

Равенство (21.39) вместе с условием согласования частот дает

$$\boxed{\omega_1 + \omega_2 = \omega_3; \quad \omega_1 n_1 + \omega_2 n_2 + \frac{m2\pi c}{\Lambda} = \omega_3 n_3.} \qquad (21.41)$$

Настроечные кривые при квазисинхронизме

Эти уравнения используются вместо (21.27) для определения настроечных кривых и углов среза кристалла при проектировании параметрических приборов. Очевидно, что квазисинхронизм обеспечивает некоторую гибкость в получении настроечных кривых желаемой формы.

Квазисинхронизм в среде с периодическим изменением знака коэффициента нелинейности

Простейший вид периодической зависимости коэффициента нелинейности — это периодическая смена его знака, т. е. переключение между двумя постоянными значениями $+d_0$ и $-d_0$ через каждые $\Lambda/2$, как показано на рис. 21.17.

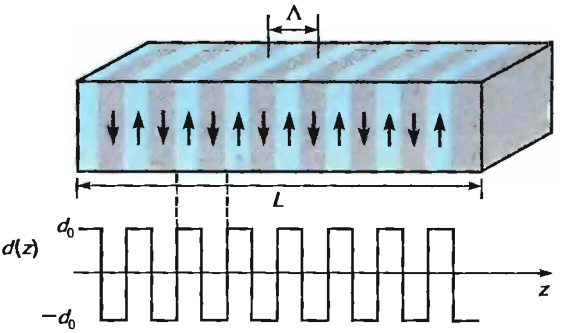


Рис. 21.17. Нелинейный кристалл с периодически меняющимся коэффициентом нелинейности $d(z)$ с периодом Λ

Физический механизм, благодаря которому обращение знака нелинейности служит для компенсации координатной зависимости фазы излучения, иллюстрируется рисунком 21.18 для случая $m = 1$, т. е. когда период решетки Λ равняется длине когерентности $L_к = 2\pi/\Delta k$.

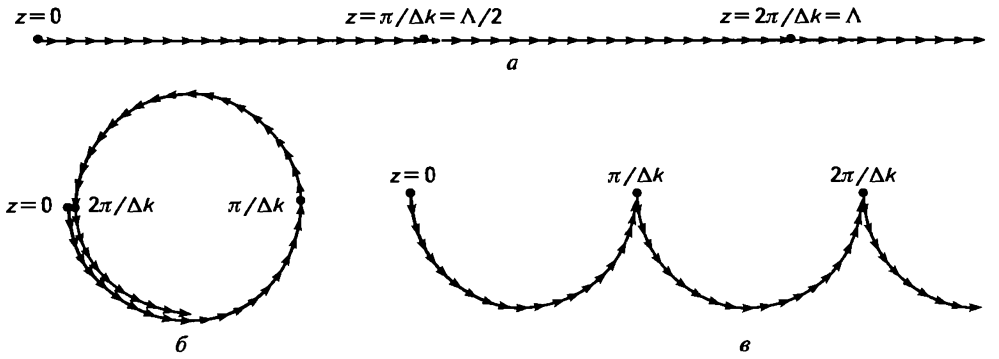


Рис. 21.18. Комплексные амплитуды волн, излучаемых малыми элементами в различных точках z нелинейной среды:

а — при фазовом синхронизме ($\Delta k = 0$) все векторы на комплексной плоскости выстроены в линию и достигается максимальная эффективность преобразования; *б* — при наличии фазового рассогласования Δk векторы на комплексной плоскости имеют разную ориентацию, и эффективность преобразования значительно уменьшается; *в* — при квазисинхронизме направление векторов на комплексной плоскости периодически меняется на обратное за счет изменения знака коэффициента нелинейности с интервалом $\Lambda/2$. Эффективность преобразования частично восстанавливается

Повышение эффективности преобразования за счет квазисинхронизма можно количественно определить следующим образом. Согласно теории рядов Фурье $d_m = (2/m\pi)d_0$ для нечетных m, в противном случае $d_m = 0$. Если согласование фаз обеспечивается за счет m-й гармоники, т. е. $\Lambda = mL_к$, то эффективность параметрического преобразования пропорциональна

$$d_m^2 = \left(\frac{2}{\pi m}\right)^2 d_0^2.$$

В отличие от этого, для однородной среды с коэффициентом нелинейности d_0 при той же длине L, но с расстройкой волновых векторов Δk, эффективность преобразования равна

$$d_0^2 \operatorname{sinc}^2\left(\frac{\Delta k L}{2\pi}\right) = d_0^2 \operatorname{sinc}^2\left(\frac{L}{L_к}\right)$$

и спадает как $(d_0^2/\pi^2)(L/L_к)^2$ при $L \gg L_к$. Так как $L_к = \Lambda/m$, увеличение эффективности преобразования происходит в $4(L/\Lambda)^2$ раз, т. е. пропорционально квадрату числа периодов структуры. Очевидно, что использование периодической среды может дать существенное улучшение эффективности преобразования.

Наиболее сложной задачей при осуществлении квазисинхронизма является изготовление периодической нелинейной структуры. Однородный нелинейный кристалл можно сделать периодическим путем обращения направления главной оси в чередующихся слоях, что обеспечивает периодическую смену знака коэффициента d. Этого можно добиться, если литографическим образом подвергнуть кристалл действию периодического электрического поля, которое обращает направление постоянной электрической поляризации кристалла. Такой метод называется **поляризацией (poling)** кристалла. Этот подход применялся к сегнетоэлектрическим кристаллам, таким как $LiTaO_3$, KTP и $LiNbO_3$; последний породил технологию, известную как технология **периодически поляризованного ниобата лития (ППНЛ)** (periodically poled lithium niobate, PPLN). Полупроводниковые кристаллы, такие как GaAs, также использовались для этой цели.

21.3. НЕЛИНЕЙНАЯ ОПТИКА ТРЕТЬЕГО ПОРЯДКА

В средах, облалающих центральной симметрией, нелинейный член второго порядка отсутствует, поскольку поляризация должна менять знак при изменении знака электрического поля. Тогда преобладающей нелинейностью является нелинейность третьего порядка

$$\boxed{\mathcal{P}_{\text{нл}} = 4\chi^{(3)}\mathcal{E}^3} \qquad (21.42)$$

(рис. 21.19), и материал называется **керровской средой**. Реакцией керровских сред на оптические поля является генерация третьей гармоники, а также сумм и разностей трех частот.

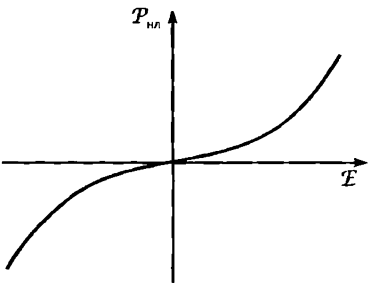


Рис. 21.19. Нелинейность третьего порядка

Упражнение 21.3

Электрооптический эффект Керра в средах с оптической нелинейностью третьего порядка

Монохроматическое оптическое поле $E(\omega)$ падает на среду с нелинейностью третьего порядка в присутствии постоянного электрического поля $E(0)$. Оптическое поле намного слабее постоянного электрического, так что $|E(\omega)|^2 \ll |E(0)|^2$. С помощью (21.42) покажите, что компонента нелинейной поляризации $\mathcal{P}_{\text{нл}}$ на частоте ω приближенно выражается как $P_{\text{нл}}(\omega) \approx 12\chi^{(3)}E^2(0)E(\omega)$. Покажите, что эта компонента поляризации эквивалентна изменению показателя преломления $\Delta n = -(1/2)sn^3E^2(0)$, где

$$s = -\frac{12}{\varepsilon_0 n^4}\chi^{(3)}. \qquad (21.43)$$

Пропорциональная зависимость между изменением показателя преломления и квадратом электрического поля — это квадратичный электрооптический эффект, или эффект Керра, описанный в подразд. 20.1.1, а s называется коэффициентом Керра.

21.3.1. Генерация третьей гармоники (ГТГ) и оптический эффект Керра

Генерация третьей гармоники (ГТГ)

В соответствии с (21.42) отклик среды с нелинейностью третьего порядка на монохроматическое оптическое поле $\mathcal{E}(t) = \mathrm{Re}\{E(\omega)\exp(j\omega t)\}$ есть нелинейная поляризация $\mathcal{P}_{\text{нл}}(t)$, содержащая компоненту на частоте ω и еще одну компоненту на частоте 3ω:

$$P_{\text{нл}}(\omega) = 3\chi^{(3)}|E(\omega)|^2 E(\omega); \tag{21.44а}$$

$$P_{\text{нл}}(3\omega) = \chi^{(3)}E^3(\omega). \tag{21.44б}$$

Наличие компоненты поляризации на частоте 3ω отражает факт генерации третьей гармоники света. Однако в большинстве случаев энергетическая эффективность такого преобразования мала. Реальное получение третьей гармоники часто достигается посредством генерации второй гармоники с последующей генерацией суммарной частоты первой и второй гармоник.

Оптический эффект Керра

Компонента поляризации на частоте ω в (21.44а) соответствует малому изменению восприимчивости $\Delta\chi$ на частоте ω, которое определяется выражением

$$\varepsilon_0\Delta\chi = \frac{P_{\text{нл}}(\omega)}{E(\omega)} = 3\chi^{(3)}|E(\omega)|^2 = 6\chi^{(3)}\eta I, \tag{21.45}$$

где $I = |E(\omega)|^2/2\eta$ — оптическая интенсивность исходной волны. Так как $n^2 = 1 + \chi$, имеем $2n\Delta n = \Delta\chi$, так что эквивалентное изменение показателя преломления составляет $\Delta n = \Delta\chi/2n$:

$$\Delta n = \frac{3\eta}{\varepsilon_0 n}\chi^{(3)}I \equiv n_2 I, \tag{21.46}$$

где

$$\boxed{n_2 = \frac{3\eta_0}{n^2\varepsilon_0}\chi^{(3)}.} \tag{21.47}$$

Оптический коэффициент Керра

Таким образом, изменение показателя преломления пропорционально оптической интенсивности. Полный показатель преломления тогда является линейной функцией[2] интенсивности

$$\boxed{n(I) = n + n_2 I.} \qquad (21.48)$$

Оптический эффект Керра

Этот эффект называют оптическим эффектом Керра по причине его сходства с электрооптическим эффектом Керра, который обсуждался в подразд. 20.1.1 и при котором Δn пропорционально квадрату постоянного электрического поля. Оптический эффект Керра является самоиндуцированным эффектом, при котором фазовая скорость волны зависит от собственной интенсивности этой волны. Это пример **нелинейной рефракции**.

Порядок величины коэффициента n_2 (в единицах см^2/Вт) составляет от 10^{-16} до 10^{-14} в стеклах, от 10^{-14} до 10^{-7} в примесных стеклах, от 10^{-10} до 10^{-8} в органических материалах и от 10^{-10} до 10^{-2} в полупроводниках. Он чувствителен к рабочей длине волны (см. разд. 21.7) и зависит от поляризации.

21.3.2. Самомодуляция фазы (СМФ), самофокусировка и пространственные солитоны

Самомодуляция фазы (СМФ)

В результате оптического эффекта Керра оптическая волна, распространяющаяся в среде с нелинейностью третьего порядка, испытывает **самомодуляцию фазы (СМФ)**. Сдвиг фазы оптического пучка мощностью P с поперечным сечением A при прохождении длины L в среде составляет

$$\varphi = -n(I)k_0 L = \frac{2\pi n(I)L}{\lambda_0} = -2\pi\left(n + n_2\frac{P}{A}\right)\frac{L}{\lambda_0},$$

так что его изменение равно

$$\Delta\varphi = -2\pi n_2 \frac{L}{\lambda_0 A} P, \qquad (21.49)$$

что пропорционально оптической мощности P. Самомодуляция фазы находит приложение в задачах управления света светом.

Для максимального эффекта L должно быть велико, а A мало. Эти требования хорошо выполняются в оптических волноводах. Оптическая мощность, при которой $\Delta\varphi = -\pi$, равна $P_\pi = \lambda_0 A/2Ln_2$. Например, волокно из примесного стекла длиной $L = 1$ м с площадью сечения $A = 10^{-2}$ мм^2 и $n_2 = 10^{-10}$ см^2/Вт, работающее на длине волны $\lambda_0 = 1$ мкм, переключает фазу на π при оптичес-

[2] Уравнение (21.48) часто пишут в альтернативной форме $n(I) = n + n_2|E|^2/2$, где n_2 отличается от (21.47) множителем η.

кой мощности $P_\pi = 0{,}5$ Вт. Материалы с большим значением n_2 могут использоваться в волноводах сантиметровой длины для достижения сдвига фазы на π при мощностях в несколько милливатт.

Модуляция фазы может быть преобразована в модуляцию интенсивности с использованием таких же схем, как и в электрооптических модуляторах (см. подразд. 20.1.2):

1) с помощью интерферометра (например, Маха—Цендера);

2) с использованием разности модулированных фаз двух поляризационных компонент, дающих эффект фазовой пластинки между скрещенными поляризаторами;

3) с использованием интегрально-оптического направленного ответвителя (см. подразд. 8.5.2). В результате получается полностью оптический модулятор, в котором слабый оптический пучок управляется сильным оптическим пучком.

Полностью оптические переключатели обсуждаются в подразд. 23.3.3.

Самофокусировка

Еще один интересный эффект, связанный с самомодуляцией фазы, — **самофокусировка**. Если интенсивный оптический пучок проходит через тонкий слой материала с керровской нелинейностью, как показано на рис. 21.20, то изменение показателя преломления копирует распределение интенсивности в поперечной плоскости. Если, например, пучок имеет наибольшую интенсивность в центре, то максимальное изменение показателя преломления также происходит в центре. Слой тогда действует как среда с градиентом показателя преломления, сообщающая волне неоднородный фазовый сдвиг, что приводит к искривлению волнового фронта. При определенных условиях среда может действовать как линза с зависящим от мощности фокусным расстоянием, как показано в упражнении 21.4. Фокусировка керровской линзой используется при синхронизации мод (см. подразд. 15.4.4).

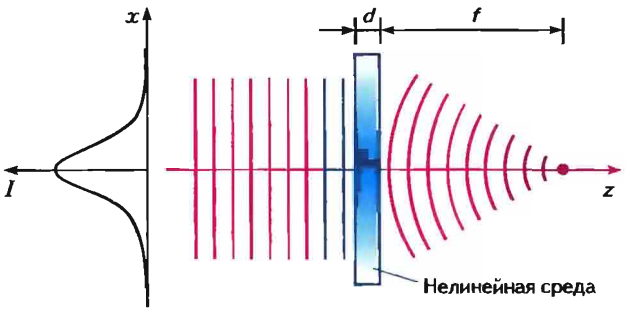


Рис. 21.20. Среда с нелинейностью третьего порядка действует как линза, оптическая сила которой зависит от интенсивности падающего пучка

Упражнение 21.4

Оптическая керровская линза

Пучок света, распространяющийся вдоль оси z, проходит через тонкий слой нелинейного материала, в котором наблюдается оптический эффект Керра, $n(I) = n + n_2 I$. Слой лежит в плоскости x—y и имеет малую толщину d, так что

его комплексный амплитудный коэффициент пропускания равен $\exp(-jnk_0 d)$. Пучок имеет приблизительно плоский волновой фронт и распределение интенсивности в точках вблизи оси пучка ($x, y \ll W$)

$$I \approx I_0\left(1 - \frac{x^2 + y^2}{W^2}\right),$$

где I_0 — максимальная интенсивность; W — ширина пучка. Покажите, что среда действует как тонкая линза с фокусным расстоянием, обратно пропорциональным I_0.

Замечание. Линза с фокусным расстоянием f имеет комплексный амплитудный коэффициент пропускания, пропорциональный $\exp[jk_0(x^2 + y^2)/2f]$; см. формулу (2.40) и упражнение 2.8.

Пространственные солитоны

Когда интенсивный оптический пучок распространяется через нелинейную однородную среду значительной толщины, а не через тонкий слой, показатель преломления изменяется неоднородно, так что среда может действовать как градиентный волновод. Таким образом, пучок может сам создавать свой собственный волновод. Если интенсивность пучка имеет такое же пространственное распределение в поперечной плоскости, как одна из мод волновода, который создается самим пучком, то пучок распространяется самосогласованно, без изменения своего пространственного распределения. При этих условиях *дифракция* компенсируется *самомодуляцией фазы*, и пучок ограничен волноводом, который создается им самим. Такие самонаправляющиеся пучки называются **пространственными солитонами**. Аналогичное поведение импульсов имеет место во времени, когда дисперсия групповой скорости компенсируется самомодуляцией фазы. Как обсуждается в подразд. 22.5.2, это ведет к формированию временных солитонов, которые распространяются без изменения формы.

Самонаправление света в оптической керровской среде математически описывается уравнением Гельмгольца

$$\nabla^2 E + n^2(I)k_0^2 E = 0,$$

где

$$n(I) = n + n_2 I; \quad k_0 = \frac{\omega}{c_0}; \quad I = \frac{|E|^2}{2\eta}.$$

Это нелинейное дифференциальное уравнение относительно E, которое упрощается заменой $E = A\exp(-jkz)$, где $k = nk_0$, в предположении, что огибающая $A(x, z)$ в направлении z меняется медленно (по сравнению с длиной волны $\lambda = 2\pi/k$) и вообще не меняется в направлении y (см. подразд. 2.2.3). В приближении

$$\frac{\partial^2}{\partial z^2}[A\exp(-ikz)] \approx \left(-2ik\frac{\partial A}{\partial z} - k^2 A\right)\exp(-ikz)$$

уравнение Гельмгольца принимает вид

$$\frac{\partial^2 A}{\partial x^2} - 2jk\frac{\partial A}{\partial z} + k_0^2\left[n^2(I) - n^2\right]A = 0. \tag{21.50}$$

В силу малости нелинейного эффекта ($n_2 I \ll n$) имеем

$$\left[n^2(I) - n^2\right] = \left[n(I) - n\right]\left[n(I) + n\right] \approx \left[n_2 I\right][2n] = \frac{2n_2 n|A|^2}{2\eta} = \frac{n^2 n_2}{\eta_0}|A|^2, \tag{21.51}$$

так что (21.50) принимает вид

$$\frac{\partial^2 A}{\partial x^2} + \frac{n_2}{\eta_0}k^2|A|^2 A = 2jk\frac{\partial A}{\partial z}. \tag{21.52}$$

Уравнение (21.52) называется нелинейным уравнением Шредингера. Одно из его решений

$$\boxed{A(x,\ z) = A_0\,\mathrm{sech}\left(\frac{x}{W_0}\right)\exp\left(-j\frac{z}{4z_0}\right).} \tag{21.53}$$

Пространственный солитон

где W_0 — постоянная; sech — гиперболический секанс, A_0 удовлетворяет соотношению

$$n_2\frac{A_0^2}{2\eta_0} = \frac{1}{k^2 W_0^2};$$

z_0 — релеевская длина [см (3.22)],

$$z_0 = \frac{1}{2}kW_0^2 = \frac{\pi W_0^2}{\lambda}.$$

Распределение интенсивности

$$I(x,\ z) = \frac{|A(x,\ z)|^2}{2\eta} = \frac{A_0^2}{2\eta}\mathrm{sech}^2\left(\frac{x}{W_0}\right) \tag{21.54}$$

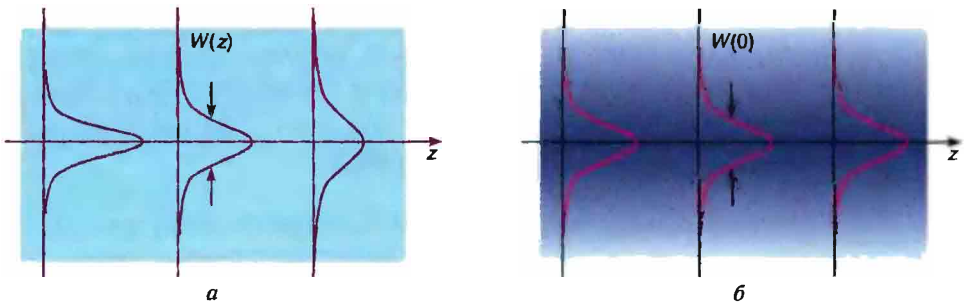
не зависит от z и имеет ширину W_0, как показано на рис. 21.21. Распределение (21.53) — это мода градиентного волновода с показателем преломления

$$n + n_2 I = n\left(1 + \frac{1/k^2}{W_0^2}\right)\mathrm{sech}^2\left(\frac{x}{W_0}\right),$$

так что самосогласованность гарантируется. Поскольку $E = A\exp(-ikz)$ волна имеет постоянную распространения

$$k + \frac{1}{4}z_0 = k\left(1 + \frac{\lambda^2}{8\pi^2 W_0^2}\right)$$

и скорость $c/(1 + \lambda^2/8\pi^2 W_0^2)$.



а *б*

Рис. 21.21. Сравнение гауссова пучка, распространяющегося в линейной среде (*а*), и пространственного солитона (самонаправляющегося оптического пучка) в нелинейной среде (*б*)

Скорость меньше, чем c для локализованных пучков (малые W_0), но приближается к c для больших W_0.

Усиление при вынужденном комбинационном рассеянии

Коэффициент нелинейности $\chi^{(3)}$ в общем случае комплексный,

$$\chi^{(3)} = \chi_R^{(3)} + j\chi_I^{(3)}.$$

Поэтому приращение фазы в (21.49):

$$\Delta\varphi = 2\pi n_2 \frac{L}{\lambda_0 A} P = \frac{6\pi\eta_0}{\varepsilon_0}\frac{\chi^{(3)}}{n^2}\frac{L}{\lambda_0 A} P, \qquad (21.55)$$

также получается комплексным. Таким образом, фазовый множитель $\exp(-j\varphi)$, связанный с распространением, представляет собой комбинацию фазового сдвига

$$\Delta\varphi = \frac{6\pi\eta_0}{\varepsilon_0}\frac{\chi_R^{(3)}}{n^2}\frac{L}{\lambda_0 A} P$$

и усиления $\exp(\gamma_R L/2)$, где коэффициент усиления

$$\boxed{\gamma_R = \frac{12\pi\eta_0}{\varepsilon_0}\frac{\chi_I^{(3)}}{n^2}\frac{1}{\lambda_0}\frac{P}{A}} \qquad (21.56)$$

Коэффициент усиления вынужденного комбинационного рассеяния

пропорционален оптической мощности P. Этот эффект называется **усилением за счет вынужденного комбинационного рассеяния (ВКР)** (в англоязычной литературе — **рамановским** усилением). Его природа связана с взаимодействием света с колебательными модами среды, которые могут служить источником энергии. Когда указанное усиление превосходит потери, среда может действовать как оптический усилитель (см. подразд. 14.3.4). При наличии соответствующей обратной связи ВКР-усилитель превращается в ВКР-лазер (см. подразд. 15.3.1). Феноменологическая конструкция комплексного коэффициента нелинейности $\chi^{(3)}$ не отличается от комплексной линейной восприимчивости, построенной для описания потерь и усиления в линейных средах (см. разд. 5.5).

21.3.3. Фазовая кросс-модуляция (ФКМ)

Теперь рассмотрим отклик среды с нелинейностью третьего порядка на оптическое поле, состоящее из двух монохроматических волн с угловыми частотами ω_1 и ω_2,

$$\mathcal{E}(t) = \mathrm{Re}\{E(\omega_1)\exp(j\omega_1 t)\} + \mathrm{Re}\{E(\omega_2)\exp(j\omega_2 t)\}.$$

При подстановке в (21.42) компонента $P_{\text{нл}}(\omega_1)$ нелинейной поляризации на частоте ω_1 оказывается равной

$$P_{\text{нл}}(\omega_1) = \chi^{(3)}\left[3|E(\omega_1)|^2 + 6|E(\omega_2)|^2\right]E(\omega_1). \qquad (21.57)$$

Если считать, что обе волны имеют одинаковый показатель преломления n, то это соотношение можно представить в виде $P_{\text{нл}}(\omega_1) = 2\varepsilon_0 n\Delta n E(\omega_1)$, где

$$\boxed{\Delta n = n_2(I_1 + 2I_2);} \qquad (21.58)$$

КМФ

$$n_2 = \frac{3\eta_0\chi^{(3)}}{\varepsilon_0 n^2}.$$

Величины I_1 и I_2 — интенсивности волн 1 и 2, соответственно,

$$I_1 = \frac{|E(\omega_1)|^2}{2\eta}; \quad I_2 = \frac{|E(\omega_2)|^2}{2\eta}.$$

Следовательно, волна 1 распространяется с показателем преломления $n + \Delta n$, управляемым как ее собственной интенсивностью, так и интенсивностью волны 2. Волна 2 испытывает такое же влияние, так что обе волны связаны.

Поскольку фазовый сдвиг, приобретаемый волной 1, модулируется интенсивностью волны 2, это явление называют **фазовой кросс-модуляцией** (ФКМ). Она может приводить к взаимным помехам между каналами оптической связи на близких частотах, например, в системах уплотнения с разделением сигналов по длине волны (см. подразд. 24.3.3).

Как мы видели в подразд. 21.2.3, смешение двух волн невозможно в среде с нелинейностью второго порядка (за исключением вырожденного случая). Заметим, однако, что двухволновое смешение может происходить в фоторефрактивных средах, как показано на рис. 20.33.

Упражнение 21.5

Оптический эффект Керра в присутствии трех волн

Три монохроматические волны с частотами ω_1, ω_2 и ω_3 распространяются в среде с нелинейностью третьего порядка. Определите комплексную амплитуду компоненты $\mathcal{P}_{\text{нл}}(t)$ в (21.42) на частоте ω_1. Покажите, что эта волна распространяется со скоростью $c_0/(n + \Delta n)$, где

$$\Delta n = n_2 (I_1 + 2I_2 + 2I_3); \tag{21.59}$$

$$n_2 = \frac{3\eta_0 \chi^{(3)}}{\varepsilon_0 n^2},$$

при

$$I_q = \frac{|E(\omega_q)|^2}{2\eta}, \quad q = 1,\ 2,\ 3.$$

21.3.4. Четырехволновое смешение (ЧВС)

Исследуем теперь **четырехволновое смешение (ЧВС)** в среде с нелинейностью третьего порядка. Начнем с определения отклика среды на суперпозицию трех волн с угловыми частотами ω_1, ω_2 и ω_3

$$\begin{aligned}\mathcal{E}(t) = \operatorname{Re}\{E(\omega_1)\exp(j\omega_1 t)\} + \operatorname{Re}\{E(\omega_2)\exp(j\omega_2 t)\} + \\ + \operatorname{Re}\{E(\omega_3)\exp(j\omega_3 t)\}.\end{aligned} \tag{21.60}$$

Удобно записать $\mathcal{E}(t)$ как сумму шести членов

$$\mathcal{E}(t) = \sum_{q=\pm1,\ \pm2,\ \pm3} \frac{1}{2} E(\omega_q)\exp(j\omega_q t), \tag{21.61}$$

где $\omega_{-q} = -\omega_q$ и $E(-\omega_q) = E^*(\omega_q)$. Подставляя (21.61) в (21.42), получим $\mathcal{P}_{\text{нл}}$ в виде суммы $6^3 = 216$ членов:

$$\mathcal{P}_{\text{нл}}(t) = \frac{1}{8}\chi^{(3)} \sum_{q,\ r,\ l=\pm1,\ \pm2,\ \pm3} E(\omega_q) E(\omega_r) E(\omega_l) \exp\left[j(\omega_q + \omega_r + \omega_l)t\right]. \tag{21.62}$$

Таким образом, $\mathcal{P}_{\text{нл}}$ — сумма гармонических компонент на частотах ω_1, ..., $3\omega_1$, ..., $2\omega_1 \pm \omega_2$, ..., $\pm\omega_1 \pm \omega_2 \pm \omega_3$. Амплитуду $P_{\text{нл}}(\omega_q + \omega_r + \omega_l)$ компоненты на

частоте $\omega_q + \omega_r + \omega_l$ можно определить складывая соотвстствующие перестановки q, r, l в (21.62). Например, $P_{\text{нл}}(\omega_1 + \omega_2 - \omega_3)$ включает шесть перестановок:

$$P_{\text{нл}}(\omega_1 + \omega_2 - \omega_3) = 6\chi^{(3)} E(\omega_1) E(\omega_2) E^*(\omega_3), \tag{21.63}$$

Уравнение (21.63) отражает тот факт, что четыре волны с частотами ω_1, ω_2, ω_3 и ω_4 смешиваются средой, если $\omega_4 = \omega_1 + \omega_2 - \omega_3$ или

$$\boxed{\omega_1 + \omega_2 = \omega_3 + \omega_4.} \tag{21.64}$$

Условие согласования частот

Это равенство выражает условие согласования частот для ЧВС.

Пусть волны плоские и имеют волновые векторы $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, $\boldsymbol{k}_3$, так что

$$E(\omega_q) \propto \exp(-j\boldsymbol{k}_q \boldsymbol{r}), \quad q = 1, 2, 3,$$

тогда (21.62) дает

$$\begin{aligned} P_{\text{нл}}(\omega_4) &\propto \exp(-j\boldsymbol{k}_1 \cdot \boldsymbol{r}) \exp(-j\boldsymbol{k}_2 \cdot \boldsymbol{r}) \exp(-j\boldsymbol{k}_3 \cdot \boldsymbol{r}) = \\ &= \exp\left[-j(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3) \cdot \boldsymbol{r}\right], \end{aligned} \tag{21.65}$$

так что волна 4 также является плоской и имеет волновой вектор $\boldsymbol{k}_4 = \boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3$, откуда

$$\boxed{\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_3 + \boldsymbol{k}_4.} \tag{21.66}$$

Условие синхронизма

Равенство (21.66) выражает условие фазового синхронизма для ЧВС.

Несколько процессов ЧВС происходят одновременно, удовлетворяя условиям согласования частот и фазового синхронизма. Как показано ранее, волны 1, 2 и 3 взаимодействуют и генерируют волну 4 в соответствии с (21.63). Аналогично, волны 3, 4 и 1 взаимодействуют с образованием волны 2 в соответствии с

$$P_{\text{нл}}(\omega_2) = 6\chi^{(3)} E(\omega_3) E(\omega_4) E^*(\omega_1), \tag{21.67}$$

и так далее.

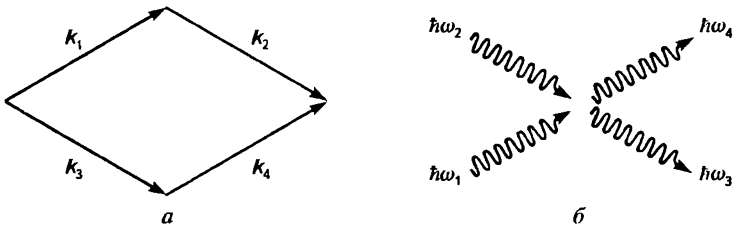


Рис. 21.22. Четырехволновое смешение (ЧВС):

а — условие синхронизма; *б* — взаимодействие четырех фотонов

Процесс ЧВС можно интерпретировать как взаимодействие четырех фотонов. Фотоны с частотами ω_3 и ω_4 уничтожаются с рождением пары фотонов ω_1

и ω_3, как показано на рис. 21.22. Равенства (21.64) и (21.66) выражают законы сохранения энергии и импульса соответственно.

Трехволновое смешение

В частично вырожденном случае, когда две из четырех волн имеют одинаковые частоты, $\omega_3 = \omega_4 \equiv \omega_0$, имеем три волны с частотами, связанными соотношением

$$\omega_1 + \omega_2 = 2\omega_0, \tag{21.68}$$

так что частоты ω_1 и ω_2 симметрично расположены относительно центральной частоты ω_0, подобно боковым частотам гармонической волны с амплитудной модуляцией или стоксовым и антистоксовым частотам при комбинационном рассеянии. Компоненты плотности нелинейной поляризации на ω_1, ω_2 и ω_3 включают члены вида

$$P_{\text{нл}}(\omega_1) = 3\chi^{(3)}E^2(\omega_3)E^*(\omega_2); \tag{21.69а}$$

$$P_{\text{нл}}(\omega_2) = 3\chi^{(3)}E^2(\omega_3)E^*(\omega_1); \tag{21.69б}$$

$$P_{\text{нл}}(\omega_3) = 6\chi^{(3)}E(\omega_1)E(\omega_2)E^*(\omega_3). \tag{21.69в}$$

Эти члены отвечают за трехволновое смешение, т. е. излучение на частоте каждой волны генерируется за счет смешения остальных волн. Такие процессы можно использовать для преобразования оптической частоты, параметрического усиления и параметрической генерации, а также спонтанного параметрического рассеяния с преобразованием частоты вниз, как и трехволновое смешение в средах с нелинейностью второго порядка. Волны на ω_1, ω_2 и ω_3 можно рассматривать как сигнальную, холостую волну и волну накачки параметрического процесса. Заметим, однако, что такой процесс *трехволнового* смешения происходит с участием *четырех* фотонов. Например, происходит уничтожение двух фотонов на частоте ω_0 и рождение двух фотонов на частотах ω_1 и ω_2. Пример параметрического усиления в среде с $\chi^{(3)}$, такой как волокно из кварцевого стекла, показан на рис. 21.23.



Рис. 21.23. Трехволновый четырехфотонный оптоволоконный параметрический усилитель света

21.3.5. Обращение волнового фронта (ОВФ)

Условие согласования частот (21.64) выполняется и в том случае, когда все четыре волны имеют одинаковые частоты

$$\omega_1 = \omega_2 = \omega_3 = \omega_4 = \omega. \tag{21.70}$$

В этом случае процесс называется **вырожденными четырехволновым смешением.**

Предположим, что две волны (3 и 4) — однородные плоские волны, распространяющиеся в противоположных направлениях,

$$E_3(\boldsymbol{r}) = A_3 \exp(-j\boldsymbol{k}_3 \ \boldsymbol{r}); \quad E_4(\boldsymbol{r}) = A_4 \exp(-j\boldsymbol{k}_4 \cdot \boldsymbol{r}), \tag{21.71}$$

где

$$\boldsymbol{k}_4 = -\boldsymbol{k}_3. \tag{21.72}$$

Подставим (21.71) и (21.72) в (21.67). Видно, что плотность поляризации волны 2: $6\chi^{(3)}A_3A_4E_1^*(\boldsymbol{r})$. Этот член соответствует источнику, излучающему оптическую волну (волну 2) с комплексной амплитудой

$$\boxed{E_2(\boldsymbol{r}) \propto A_3A_4E_1^*(\boldsymbol{r}).} \tag{21.73}$$

Обращение волнового фронта

Поскольку A_3 и A_4 — постоянные величины, волна 2 пропорциональна комплексно сопряженной волне 1. Прибор служит для **обращения волнового фронта.** Волны 3 и 4 называются волнами **накачки**, а волны 1 и 2 — **пробной** и **сопряженной** волнами соответственно. Как вскоре будет показано, сопряженная волна идентична пробной волне за исключением того, что она распространяется в противоположном направлении. Комплексное сопряжение фазы действует как специальное зеркало, которое отражает волну назад точно на себя без изменения формы ее волновых фронтов.

Для лучшего понимания процесса обращения волнового фронта рассмотрим два простых примера:

Пример 21.4

Обращение плоской волны

Если волна 1 — однородная плоская волна, $E_1(\boldsymbol{r}) = A_1 \exp(-j\boldsymbol{k}_1 \cdot \boldsymbol{r})$, распространяющаяся в направлении $\boldsymbol{k}_1$, то $E_2(\boldsymbol{r})$ — однородная плоская волна, распространяющаяся в противоположном направлении $\boldsymbol{k}_2 = -\boldsymbol{k}_1$, как показано на рис. 21.24, *б*,

$$E_2(\boldsymbol{r}) = A_1^* \exp(j\boldsymbol{k}_1 \cdot \boldsymbol{r}).$$

Таким образом, условие синхронизма (21.66) выполнено. Среда действует как специальное «зеркало», которое отражает падающую плоскую волну обратно на себя, независимо от угла падения.

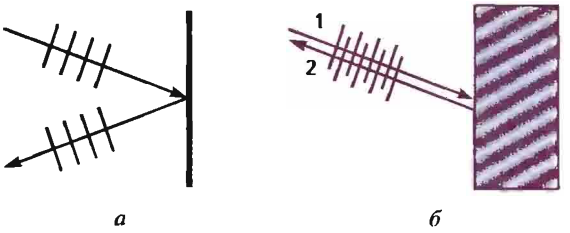


Рис. 21.24. Отражение плоской волны от обычного зеркала (*а*), от зеркала, обращающего волновой фронт (*б*)

Пример 21.5

Обращение сферической волны

Если волна 1 сферическая с центром в начале координат $\boldsymbol{r} = 0$,

$$E_1(\boldsymbol{r}) \propto \frac{1}{r}\exp(-jkr),$$

то волна 2 имеет комплексную амплитуду

$$E_2(\boldsymbol{r}) \propto \frac{1}{r}\exp(+jkr).$$

Это сферическая волна, бегущая обратно и сходящаяся в центр, как показано на рис. 21.25, *б*.

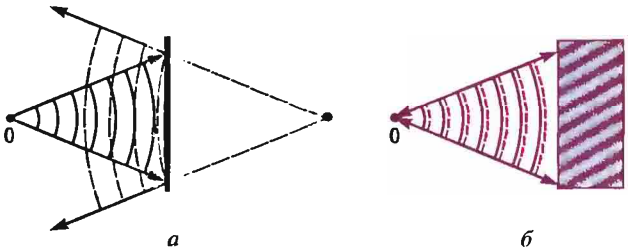


Рис. 21.25. Отражение сферической волны от обычного зеркала (*а*) и от зеркала, обращающего волновой фронт (*б*)

Поскольку произвольная пробная волна может рассматриваться как суперпозиция плоских волн (см. гл. 4), каждая из которых отражается точно на себя при обращении волнового фронта, сопряженная волна идентична падающей волне во всем пространстве, но имеет противоположное направление распространения. Сопряженная волна повторяет падающую волну, распространяясь назад и сохраняя ту же форму волновых фронтов.

Фазовое сопряжение аналогично обращению времени. Это можно понять, рассматривая поле сопряженной волны

$$\mathcal{E}_2(\boldsymbol{r}, t) = \mathrm{Re}\{E_2(\boldsymbol{r})\exp(j\omega t)\} \propto \mathrm{Re}\{E_1^*(\boldsymbol{r})\exp(j\omega t)\}.$$

Поскольку действительная часть комплексного числа равна действительной части сопряженного ему числа, $\mathcal{E}_2(\boldsymbol{r}, t) \propto \mathrm{Re}\{E_1(\boldsymbol{r})\exp(-j\omega t)\}$. Сравнивая

это с полем пробной волны $\mathcal{E}_1(\boldsymbol{r}, t) = \mathrm{Re}\{E_1(\boldsymbol{r}) \exp(j\omega t)\}$, сразу видим, что одна получается из другой заменой $t \to -t$, так что сопряженная волна получается из пробной волны обращением времени.

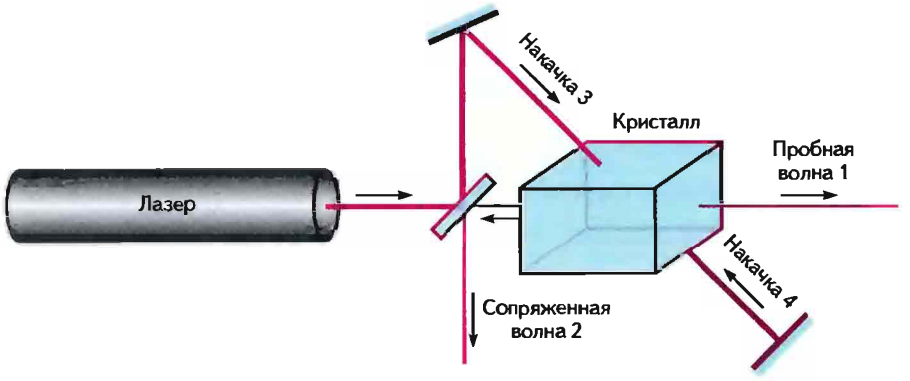


Рис. 21.26. Оптическая система для вырожденного четырехволнового смешения с использованием нелинейного кристалла. Волны накачки 3 и 4 и пробная волна 1 получаются от лазера с помощью светоделителя и двух зеркал. Сопряженная волна 2 создается внутри кристалла

Сопряженная волна может нести больше энергии, чем пробная волна. В этом можно убедиться, если заметить, что интенсивность сопряженной волны (волны 2) пропорциональна произведению интенсивностей волн 3 и 4 [см. (21.73)]. Когда мощности волн накачки возрастают так, что сопряженная волна 2 несет большую мощность, чем пробная волна 1, среда действует как «усиливающее зеркало». Пример оптической установки для демонстрации обращения волнового фронта показан на рис. 21.26.

Вырожденное четырехволновое смешение как разновидность голографии в реальном времени

Процесс вырожденного четырехволнового смешения аналогичен объемной голографии (см. разд. 4.5). Голография является двухступенчатым процессом, при котором интерференционная картина, образованная суперпозицией предметной волны E_1 и опорной волны E_3, записывается на фотоэмульсии. Затем другая опорная волна E_4 пропускается через эмульсию или отражается от нее, создавая предметную волну $E_2 \propto E_4 E_3 E_1^*$ или ее реплику $E_2 \propto E_4 E_1 E_3^*$, в зависимости от геометрии (см. рис. 4.51, *а* и *б*). Нелинейная среда позволяет одновременно производить голографическую запись и воспроизведение в реальном времени. Это происходит и в керровской среде, и в фоторефрактивных средах (см. разд. 20.4).

Когда четыре волны смешиваются в нелинейном кристалле, каждая пара волн интерферирует и создает решетку, от которой отражается третья волна, создавая четвертую волну. Роли предметной и опорной волн распределены

между четырьмя волнами, так что существуют два типа решеток, как показано на рис. 21.27. Рассмотрим сначала процесс, изображенный на рис. 21.27, *а* (см. также рис. 4.51, *а*). Предположим, что две опорные волны (обозначенные 3 и 4) являются плоскими и распространяются навстречу друг другу. Процесс голографии состоит из двух ступеней.

1. Предметная волна 1 складывается с опорной волной 3 и интенсивность их суммы записывается в среде в виде объемной решетки (голограммы).

2. Реконструирующая опорная волна 4 претерпевает брэгговское отражение от решетки и создает сопряженную волну (волну 2).

Такая решетка называется пропускающей.

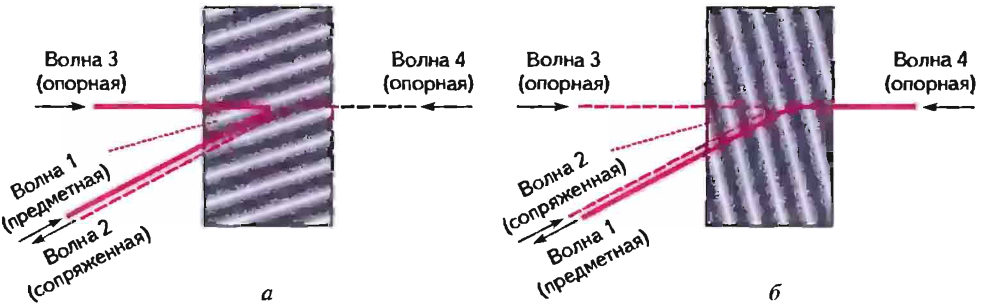


Рис. 21.27. Четырехволновое смешение в нелинейной среде. Опорная и предметная волны интерферируют и создают решетку, от которой вторая опорная волна отражается и создает сопряженную волну. Имеется две возможности, соответствующие пропускающей (*а*) и отражающей решетке (*б*)

Вторая возможность, показанная на рис 21.27, *б*, состоит в интерференции опорной волны 4 с предметной волной 1, в результате чего образуется решетка, называемая отражающей. Вторая опорная волна 3 отражается от этой решетки и создает сопряженную волну 2. Обе решетки могут существовать одновременно, но, как правило, они имеют различные эффективности.

Итак, четырехволновое смешение позволяет осуществить голографию в реальном времени и обращение волнового фронта, которые имеют множество приложений в обработке оптических сигналов.

Использование обращения волнового фронта для восстановления волн

Способность нелинейных сред отражать волну на саму себя с полным воспроизведением пройденного пути в обратном направлении предполагает множество полезных применений, включающих устранение аберраций волнового фронта. Идея основана на принципе обратимости, который иллюстрируется рис. 21.28. Лучи, проходящие через линейную оптическую среду слева направо, при обращении проходят теми же путями в обратном направлении. Тот же принцип применим и к волнам.

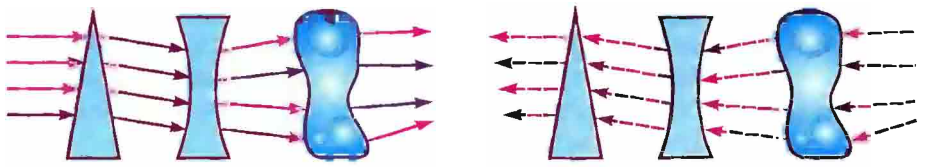


Рис. 21.28. Оптическая взаимность

Если волновой фронт оптического пучка искажен средой, вносящей аберрации, то исходная волна может быть восстановлена с помощью устройства, обращающего волновой фронт и направляющего пучок назад через ту же среду, как показано на рис. 21.29.

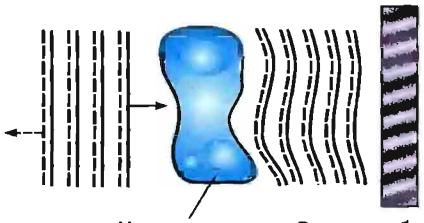


Рис. 21.29. Обращающее волновой фронт зеркало отражает искаженную волну точно на себя, так что при обратном прохождении искажения компенсируются

Важное приложение возможно в оптических резонаторах (см. гл. 10). Если резонатор содержит среду, вносящую аберрации, то замена одного из обычных зеркал на зеркало, обращающее волновой фронт, гарантирует устранение искажения на каждом обходе резонатора, так что моды резонатора будут иметь неискаженные волновые фронты на выходе из обычного зеркала, как показано на рис. 21.30.

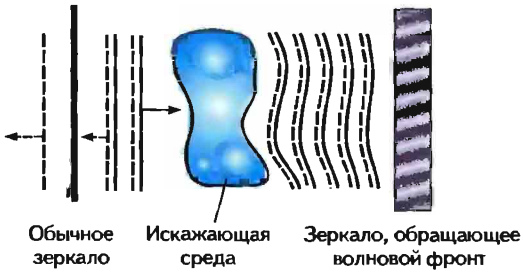


Рис. 21.30. Оптический резонатор с одним обычным и одним обращающим волновой фронт зеркалом

*21.4. НЕЛИНЕЙНАЯ ОПТИКА ВТОРОГО ПОРЯДКА: ТЕОРИЯ СВЯЗАННЫХ ВОЛН

Количественный анализ процесса трехволнового смешения в среде с нелинейностью второго порядка проведен в этом разделе на основе теории связанных волн. В отличие от рассмотрения в разд. 21.2, все три волны рассматриваются на равноправной основе. Для упрощения анализа рассмотрение эффектов анизотропии и дисперсии отложено до разд. 21.6 и 21.7 соответственно.

Уравнения для связанных волн

В соответствии с (21.6) и (21.7) распространение волн в среде с нелинейностью второго порядка определяется основным волновым уравнением

$$\nabla^2 \mathcal{E} - \frac{1}{c^2}\frac{\partial^2 \mathcal{E}}{\partial t^2} = -S, \tag{21.74}$$

где

$$S = -\mu_0 \frac{\partial^2 \mathcal{P}_{\text{нл}}}{\partial t^2} \tag{21.75}$$

играет роль источника излучения, а $\mathcal{P}_{\text{нл}}$ — нелинейная часть плотности поляризации,

$$\mathcal{P}_{\text{нл}} = 2d\mathcal{E}^2 \tag{21.76}$$

Поле $\mathcal{E}(t)$ — суперпозиция трех волн с угловыми частотами ω_1, ω_2 и ω_3 и комплексными амплитудами E_1, E_2 и E_3 соответственно:

$$\mathcal{E}(t) = \sum_{q=1,\ 2,\ 3} \text{Re}\left\{E_q \exp\left(j\omega_q t\right)\right\} = \sum_{q=1,\ 2,\ 3} \frac{1}{2}\left[E_q \exp\left(j\omega_q t\right) + E_q^* \exp\left(-j\omega_q t\right)\right]. \tag{21.77}$$

Удобно переписать (21.77) в компактной форме

$$\mathcal{E}(t) = \sum_{q=\pm1,\ \pm2,\ \pm3} \frac{1}{2} E_q \exp\left(j\omega_q t\right), \tag{21.78}$$

где $\omega_{-q} = -\omega_q$ и $E_{-q} = E_q^*$. Соответствующая плотность поляризации, получающаяся подстановкой в (21.76), представляет собой сумму $6 \times 6 = 36$ членов:

$$\mathcal{P}_{\text{нл}}(t) = 2d\frac{1}{4} \sum_{q,\ r=\pm1,\ \pm2,\ \pm3} E_q E_r \exp\left[j\left(\omega_q + \omega_r\right)t\right]. \tag{21.79}$$

Соответствующая функция источника

$$S = \frac{1}{2}\mu_0 d \sum_{q,\ r=\pm1,\ \pm2,\ \pm3} \left(\omega_q + \omega_r\right)^2 E_q E_r \exp\left[j\left(\omega_q + \omega_r\right)t\right] \tag{21.80}$$

генерирует сумму гармонических компонент, частоты которых представляют собой суммы и разности исходных частот ω_1, ω_2 и ω_3.

Подставляя (21.78) и (21.80) в волновое уравнение (21.74), получим одно дифференциальное уравнение с многими членами, каждый из которых — гармоническая функция на некоторой частоте. Если частоты ω_1, ω_2 и ω_3 различны, данное уравнение можно разделить на три не зависящих от времени дифференциальных уравнения путем приравнивания членов в обеих частях (21.74) на

каждой из частот ω_1, ω_2 и ω_3 отдельно. Результат представляется в виде трех уравнений Гельмгольца со своим источником каждое:

$$\left(\nabla^2 + k_1^2\right) E_1 = -S_1; \tag{21.81а}$$

$$\left(\nabla^2 + k_2^2\right) E_2 = -S_2; \tag{21.81б}$$

$$\left(\nabla^2 + k_3^2\right) E_3 = -S_3, \tag{21.81в}$$

где S_q — амплитуды компонент S на частотах ω_q и $k_q = n\omega_q/c_0$, $q = 1, 2, 3$. Каждая из комплексных амплитуд трех волн удовлетворяет уравнению Гельмгольца с функцией источника, равной амплитуде компоненты S на соответствующей частоте. При некоторых условиях источник для одной волны зависит от электрических полей двух других волн, так что все три волны связаны друг с другом.

В отсутствие нелинейности $d = 0$, так что функция источника S исчезает, и каждая из трех волн удовлетворяет уравнению Гельмгольца, не зависящему от двух других, как и ожидается в линейной оптике.

Если частоты ω_1, ω_2 и ω_3 не соизмеримы (одна частота не является суммой или разностью двух других или удвоенной другой частотой), то функция источника S не содержит компонент на частотах ω_1, ω_2 или ω_3. Тогда компоненты S_1, S_2 и S_3 исчезают и три волны не взаимодействуют друг с другом.

Чтобы три волны были связаны средой, их частоты должны быть соизмеримы. Например, предположим, что одна из частот является суммой двух других

$$\omega_1 + \omega_2 = \omega_3. \tag{21.82}$$

Тогда источник S содержит компоненты на частотах ω_1, ω_2 и ω_3. Исследование 36 членов в (21.80) дает

$$S_1 = 2\mu_0\omega_1^2 d E_3 E_2^*; \tag{21.83}$$

$$S_2 = 2\mu_0\omega_2^2 d E_3 E_1^*; \tag{21.84}$$

$$S_3 = 2\mu_0\omega_3^2 d E_1 E_2. \tag{21.85}$$

Источник волны 1 пропорционален $E_3E_2^*$ (поскольку $\omega_1 = \omega_3 - \omega_2$), так что волны 2 и 3 вместе вносят вклад в рост волны 1. Аналогично, источник для волны 3 пропорционален E_1E_2 (поскольку $\omega_3 = \omega_1 + \omega_2$), так что волны 1 и 2 в сочетании друг с другом усиливают волну 3, и так далее. Все три волны оказываются связанными или «смешанными» средой в процессе, описываемом тремя связанными дифференциальными уравнениями относительно E_1, E_2 и E_3:

$$\boxed{\left(\nabla^2 + k_1^2\right) E_1 = -2\mu_0\omega_1^2 d E_3 E_2^*;} \tag{21.86а}$$

$$\boxed{\left(\nabla^2 + k_2^2\right) E_2 = -2\mu_0\omega_2^2 d E_3 E_1^*;} \tag{21.86б}$$

$$\boxed{\left(\nabla^2 + k_3^2\right) E_3 = -2\mu_0\omega_3^2 d E_1 E_2.} \tag{21.86в}$$

Трехволновое смешение
Связанные уравнения

Упражнение 21.6

ГВГ как вырожденное трехволновое смешение

Уравнения (21.86) верны, только когда частоты ω_1, ω_2 и ω_3 различны. Рассмотрим теперь вырожденный случай, для которого $\omega_1 = \omega_2 = \omega$ и $\omega_3 = 2\omega$, так что вместо трех имеется две волны с амплитудами E_1 и E_3. Это соответствует генерации второй гармоники (ГВГ). Покажите, что такие волны удовлетворяют уравнению Гельмгольца с источниками

$$S_1 = 2\mu_0\omega_1^2 d E_3 E_1^*; \tag{21.87}$$

$$S_3 = \mu_0\omega_3^2 d E_1 E_1, \tag{21.88}$$

так что связанные волновые уравнения имеют вид

$$\left(\nabla^2 + k_1^2\right)E_1 = -2\mu_0\omega_1^2 d E_3 E_1^*; \tag{21.89а}$$

$$\left(\nabla^2 + k_3^2\right)E_3 = -\mu_0\omega_3^2 d E_1 E_1. \tag{21.89б}$$

Связанные уравнения ГВГ

Заметим, что эти уравнения не получаются из уравнений трехволнового смешения (21.86) путем подстановки $E_1 = E_2$ [в (21.89б) отсутствует множитель 2].

Смешение трех коллинеарных однородных плоских волн

Допустим, что три волны плоские и распространяются вдоль оси z с комплексными амплитудами $E_q = A_q \exp(-jk_q z)$, комплексными огибающими A_q и волновыми числами $k_q = \omega_q/c$, $q = 1, 2, 3$. Удобно нормировать комплексные огибающие, вводя переменные

$$a_q = \frac{A_q}{\left(2\eta\hbar\omega_q\right)^{1/2}},$$

где $\eta = \eta_0/n$ — импеданс среды; $\eta_0 = (\mu_0/\varepsilon_0)^{1/2}$ — импеданс свободного пространства; $\hbar\omega_q$ — энергия фотона с угловой частотой ω_q. Тогда

$$E_q = \sqrt{2\eta\hbar\omega_q}\, a_q \exp\left(-jk_q z\right); \quad q = 1,\ 2,\ 3, \tag{21.90}$$

и интенсивности трех волн

$$I_q = \frac{|E_q|^2}{2\eta} = \hbar\omega_q |a_q|^2$$

Плотности потока фотонов, связанные с этими волнами [число фотонов/(с · м^2)]:

$$\phi_q = \frac{I_q}{\hbar\omega_q} = |a_q|^2 \tag{21.91}$$

Переменная a_q, таким образом, представляет комплексную огибающую волны q, нормированную таким образом, что $|a_q|^2$ — плотность потока фотонов. Такая нормировка удобна, поскольку в процессе смешения волн должно выполняться условие сохранения числа фотонов (см. подразд. 21.2.3).

В результате взаимодействия между тремя волнами комплексные огибающие a_q меняются вдоль z, так что $a_q = a_q(z)$. Если взаимодействие слабое, $a_q(z)$ меняются медленно, так что их можно приближенно считать постоянными на расстояниях порядка длины волн. Это делает возможным применение приближения медленно меняющихся огибающих, в котором величинами d^2a_q/dz^2 пренебрегают по сравнению с

$$k_q \frac{da_q}{dz} = \frac{2\pi}{\lambda_q}\frac{da_q}{dz};$$

$$\left(\nabla^2 + k_q^2\right) = \left[a_q \exp\left(-jk_q z\right)\right] \approx -j2k_q \frac{da_q}{dz}\exp\left(-jk_q z\right) \tag{21.92}$$

(см. подразд. 2.2.3). В этом приближении уравнения (21.86) сводятся к более простым уравнениям, схожим с параксиальными уравнениями Гельмгольца, в которых учтена фазовая расстройка:

$$\frac{da_1}{dz} = -jga_3a_2^* \exp\left(-j\Delta kz\right); \tag{21.93а}$$

$$\frac{da_2}{dz} = -jga_3a_1^* \exp\left(-j\Delta kz\right); \tag{21.93б}$$

$$\frac{da_3}{dz} = -jga_1a_2 \exp\left(j\Delta kz\right), \tag{21.93в}$$

Уравнения для связанных волн
Трехволновое смешение

где

$$g^2 = 2\hbar\omega_1\omega_2\omega_3\eta^3 d^2; \tag{21.94}$$

$$\Delta k = k_3 - k_2 - k_1 \tag{21.95}$$

представляют собой расстройку синхронизма. Таким образом, изменения a_1, a_2 и a_3 вдоль оси z определяются тремя связанными обыкновенными дифференциальными уравнениями первого порядка (21.93), к решению которых при различных граничных условиях, отвечающих различным приложениям, мы и переходим. Полезно, однако, сначала найти некоторые инварианты процесса смешения волн. Они будут функциями a_1, a_2 и a_3, не зависящими от z. Инварианты полезны тем, что позволяют сократить число независимых переменных. В упражнениях 21.8 и 21.7 предлагается получить инварианты на основе законов сохранения энергии и числа фотонов.

Упражнение 21.7

Сохранение числа фотонов: соотношения Мэнли—Роу

С помощью (21.93) покажите, что

$$\frac{\mathrm{d}}{\mathrm{d}z}|a_1|^2 = \frac{\mathrm{d}}{\mathrm{d}z}|a_2|^2 = -\frac{\mathrm{d}}{\mathrm{d}z}|a_3|^2, \tag{21.96}$$

откуда следует соотношение Мэнли—Роу (21.26), выведенное ранее из закона сохранения числа фотонов. Уравнение (21.95) подразумевает, что $|a_1|^2 + |a_3|^2$ и $|a_2|^2 + |a_3|^2$ также являются инвариантами в процессе смешения волн.

Упражнение 21.8

Сохранение энергии

Покажите, что сумма интенсивностей $I_q = \hbar\omega_q|a_q|^2$, $q = 1, 2, 3$, трех волн, удовлетворяющих (21.93), не зависит от z, так что

$$\frac{\mathrm{d}}{\mathrm{d}z}(I_1 + I_2 + I_3) = 0. \tag{21.97}$$

21.4.1. Генерация второй гармоники (ГВГ)

Генерация второй гармоники (ГВГ) — вырожденный случай трехволнового смешения, при котором

$$\omega_1 = \omega_2 = \omega; \quad \omega_3 = 2\omega. \tag{21.98}$$

Происходит два типа взаимодействий: два фотона частоты ω рождают один фотон частоты 2ω (вторая гармоника), либо один фотон частоты 2ω распадается на два фотона с частотой ω каждый (вырожденное параметрическое преобразование частоты вниз).

Взаимодействие двух волн описывается параксиальными уравнениями Гельмгольца с источниками. Закон сохранения импульса требует, чтобы

$$2k_1 = k_3. \tag{21.99}$$

Упражнение 21.9

Уравнения связанных волн для ГВГ

Примените приближение медленной огибающей (21.92) к уравнению Гельмгольца (21.89), описывающему две коллинеарные волны в вырожденном случае, и покажите, что

$$\frac{\mathrm{d}a_1}{\mathrm{d}z} = -jga_3a_1^* \exp(-j\Delta kz); \tag{21.100а}$$

$$\frac{\mathrm{d}a_3}{\mathrm{d}z} = -j\frac{g}{2}a_1a_1 \exp(j\Delta kz), \tag{21.100б}$$

где

$$\Delta k = k_3 - 2k_1;$$

$$g^2 = 4\hbar\omega^3\eta^3 d^2. \qquad (21.101)$$

Если две коллинеарные волны точно удовлетворяют условию фазового синхронизма ($\Delta k = 0$), то уравнения (21.100) сводятся к виду

$$\boxed{\frac{da_1}{dz} = -jga_3a_1^*;} \qquad (21.102а)$$

$$\boxed{\frac{da_3}{dz} = -j\frac{g}{2}a_1a_1.} \qquad (21.102б)$$

Связанные уравнения ГВГ

На входе в прибор ($z = 0$) амплитуда второй гармоники предполагается равной нулю, $a_3(0) = 0$, а амплитуда основной волны $a_1(0)$ — действительной. Ищем решение, для которого $a_1(z)$ является действительной для любого z. Закон сохранения энергии

$$a_1^2(z) + 2|a_3(z)|^2 = a_1^2(0),$$

формула (21.102б) даст дифференциальное уравнение относительно $a_3(z)$:

$$\frac{da_3}{dz} = -j\frac{g}{2}\left[a_1^2(0) - 2|a_3(z)|^2\right], \qquad (21.103)$$

решение которого можно подставить в (21.102а) и получить общее решение

$$a_1(z) = a_1(0)\operatorname{sech}\left(\frac{1}{\sqrt{2}}ga_1(0)z\right); \qquad (21.104а)$$

$$a_3(z) = -\frac{j}{\sqrt{2}}a_1(0)\tanh\left(\frac{1}{\sqrt{2}}ga_1(0)z\right). \qquad (21.104б)$$

Следовательно, плотности потока фотонов $\phi_1(z) = |a_1(z)|^2$ и $\phi_3(z) = |a_3(z)|^2$ меняются в соответствии с выражениями

$$\boxed{\phi_1(z) = \phi_1(0)\operatorname{sech}^2\frac{\gamma z}{2};} \qquad (21.105а)$$

$$\boxed{\phi_3(z) = \frac{1}{2}\phi_1(0)\tanh^2\frac{\gamma z}{2},} \qquad (21.105б)$$

где

$$\frac{\gamma}{2} = \frac{ga_1(0)}{\sqrt{2}},$$

т. е.

$$\gamma^2 = 2g^2 a_1^2(0) = 2g^2\phi_1(0) = 8\mathcal{d}^2\eta^3\hbar\omega^3\phi_1(0) = 8\mathcal{d}^2\eta^3\omega^2 I_1(0). \qquad (21.106)$$

Поскольку $\operatorname{sech}^2 x + \tanh^2 x = 1$, $\phi_1(z) + 2\phi_3(z) = \phi_1(0)$ остается постоянным, отражая тот факт, что в каждой точке z фотоны волны 1 преобразуются в половинное число фотонов волны 3. Спад $\phi_1(z)$ и нарастание $\phi_3(z)$ с ростом z показаны на рис. 21.31, *б*.

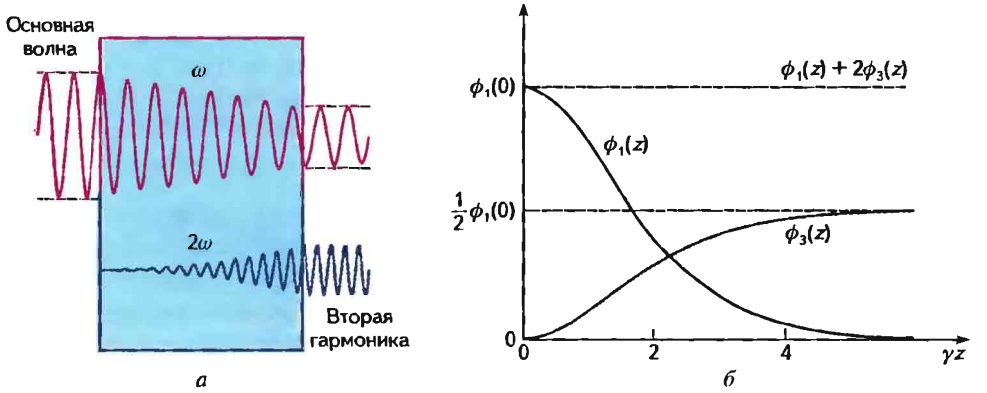


Рис. 21.31. Генерация второй гармоники:

а — волна с частотой ω, падающая на нелинейный кристалл, генерирует волну на частоте 2ω; *б* — по мере того, как плотность потока фотонов $\phi_1(z)$ основной волны убывает, плотность потока фотонов второй гармоники $\phi_3(z)$ растет. Поскольку число фотонов сохраняется, сумма $\phi_1(z) + 2\phi_3(z) = \phi_1(0)$ постоянна; *в* — два фотона частоты ω превращаются в один фотон частоты 2ω

Эффективность ГВГ

Эффективность генерации второй гармоники для области взаимодействия длиной L определяется коэффициентом преобразования во вторую гармонику

$$\eta_{\text{ГВГ}} = \frac{I_3(L)}{I_1(0)} = \frac{\hbar\omega_3\phi_3(L)}{\hbar\omega_1\phi_1(0)} = \frac{2\phi_3(L)}{\phi_1(0)} = \tanh^2\frac{\gamma L}{2}. \qquad (21.107)$$

Для больших γL (длинная ячейка, большая входная интенсивность или высокая нелинейность) эффективность приближается к единице. Это значит, что вся входная мощность (на частоте ω) полностью преобразуется во вторую гармонику; все фотоны с частотой ω преобразуются в половинное число фотонов частоты 2ω.

Для малых γL [малая длина кристалла L, малая нелинейность $\mathcal{d}$ или низкая плотность потока фотонов на входе $\phi_1(0)$] аргумент функции tanh мал и, следо-

вательно, можно использовать приближение $\tanh(x) \approx x$. Коэффициент преобразования во вторую гармонику в этом случае равен

$$\eta_{\text{ГВГ}} = \frac{I_3(L)}{I_1(0)} \approx \frac{1}{4}\gamma^2 L^2 = \frac{1}{2}g^2 L^2 \phi_1(0) = 2d^2\eta^3\hbar\omega^3 L^2 \phi_1(0) = 2d^2\eta^3\omega^2 L^2 I_1(0), \quad (21.108)$$

так что

$$\boxed{\eta_{\text{ГВГ}} = C^2 \frac{L^2}{A} P; \quad C^2 = 2\omega^2\eta_0^3 \frac{d^2}{n^3},} \quad (21.109)$$

Эффективность ГВГ

где $P = I_1(0)A$ — мощность падающего света на основной частоте; A — площадь поперечного сечения. Этот результат воспроизводит выражение (21.13) и показывает, что постоянная C^2 пропорциональна материальному параметру d^2/n^3, который является основным критерием сравнения различных нелинейных материалов друг с другом.

Пример 21.6

Коэффициент преобразования ГВГ

Для материала с $d^2/n^3 = 10^{-46}$ Кл/В^2 (см. типичные значения d в табл. 21.3) и длине волны первой гармоники 1 мкм $C^2 = 38 \cdot 10^{-9}$ Вт$^{-1}$ = 0,038 МВт$^{-1}$. В этом случае коэффициент преобразования ГВГ равен 10 %, если $PL^2/A = 2{,}63$ МВт. отношение длины рабочей области к ширине равно 1000, т. е. $L^2/A = 10^6$, то требуемая мощность равна 2,63 Вт. Это можно реализовать, например, при $L = 1$ см и $A = 100$ мкм^2, что соответствует плотности мощности $P/A = 2{,}63 \cdot 10^6$ Вт/см^2. Эффективность ГВГ можно повысить, повышая плотность мощности, увеличивая длину взаимодействия или выбирая материал с большим коэффициентом d^2/n^3.

Фазовая расстройка при ГВГ

Для изучения влияния расстройки фазы (импульса) нужно использовать общие уравнения (21.100) при $\Delta k \neq 0$. Для простоты ограничимся случаем слабой связи, для которого $\gamma L \ll 1$. В этом случае амплитуда основной волны $a_1(z)$ меняется незначительно (см. рис. 21.6, *а*) и может считаться приблизительно постоянной. Подставляя $a_1(z) \approx a_1(0)$ в (21.100б) и выполняя интегрирование, получаем

$$a_3(L) = -j\frac{g}{2}a_1^2(0)\int_0^L \exp(j\Delta kz)\,dz = -\left(\frac{g}{2\Delta k}\right)a_1^2(0)\left[\exp(j\Delta kL) - 1\right], \quad (21.110)$$

откуда

$$\phi_3(L) = |a_3(L)|^2 = \left(\frac{g}{\Delta k}\right)^2 \phi_1^2(0)\sin^2\left(\frac{\Delta kL}{2}\right),$$

где $a_1(0)$ считается действительным. Коэффициент преобразования во вторую гармонику тогда равен

$$\eta_{\text{ГВГ}} = \frac{I_3(L)}{I_1(0)} = \frac{2\phi_3(L)}{\phi_1(0)} = C^2 \frac{L^2}{A} P \operatorname{sinc}^2\left(\frac{\Delta kL}{2\pi}\right), \tag{21.111}$$

где $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$.

Таким образом, влияние фазовой расстройки состоит в уменьшении эффективности генерации второй гармоники на множитель $\operatorname{sinc}^2(\Delta kL/2\pi)$. Это подтверждает предыдущие результаты, показанные на рис. 21.16. Для заданной расстройки Δk процесс ГВГ эффективен на длинах, меньших длины когерентности $L_к = 2\pi/|\Delta k|$.

21.4.2. Преобразование оптической частоты (ПОЧ)

Преобразователь частоты вверх (рис. 21.32, см. цв. вклейку) превращает волну частоты ω_1 в волну более высокой частоты ω_3 с помощью вспомогательной волны с частотой ω_2, которая называется **накачкой**. Фотон накачки $\hbar\omega_2$ складывается с фотоном $\hbar\omega_1$ **сигнальной** волны и формирует фотон $\hbar\omega_3$ **преобразованного сигнала** на повышенной частоте $\omega_3 = \omega_1 + \omega_2$.

Процесс частотной конверсии подчиняется трем связанным уравнениям (21.93). Для простоты предположим, что три волны удовлетворяют условию синхронизма ($\Delta k = 0$), что накачка достаточно сильна, так что ее амплитуда не меняется существенно в пределах рассматриваемой длины взаимодействия, т. е. $a_2(z) \approx a_2(0)$ для всех z от 0 до L. Тогда три уравнения (21.93) сводятся к двум

$$\boxed{\frac{da_1}{dz} = -j\frac{\gamma}{2}a_3;} \tag{21.112а}$$

$$\boxed{\frac{da_3}{dz} = -j\frac{\gamma}{2}a_1,} \tag{21.112б}$$

Связанные уравнения ГВГ

где $\gamma = 2ga_2(0)$ и $a_2(0)$ предполагается действительным. Это простые дифференциальные уравнения с гармоническими решениями

$$a_1(z) = a_1(0)\cos\frac{\gamma z}{2}; \tag{21.113а}$$

$$a_3(z) = -ja_1(0)\sin\frac{\gamma z}{2}. \tag{21.113б}$$

Соответствующие плотности потока фотонов даются выражениями

$$\boxed{\phi_1(z) = \phi_1(0)\cos^2\frac{\gamma z}{2};} \tag{21.114а}$$

$$\boxed{\phi_3(z) = \phi_1(0)\sin^2\frac{\gamma z}{2}.} \tag{21.114б}$$

Зависимости плотностей потока фотонов ϕ_1 и ϕ_3 от z схематически изображены на рис. 21.32, *б*. Происходит периодический обмен фотонами между двумя волнами. В области между $z = 0$ и $z = \pi/\gamma$ входящие фотоны ω_1 комбинируют с фотонами накачки ω_2 и генерируют фотоны повышенной частоты ω_3. Следовательно, волна 1 ослабляется, а волна 3 усиливается. В области между $z = \pi/\gamma$ и $z = 2\pi/\gamma$ фотоны ω_3 распадаются на фотоны ω_1 и ω_2, так что волна 3 ослабляется, а волна 1 усиливается. Процесс повторяется периодически по мере распространения волн через среду.

Коэффициент преобразования частоты вверх для среды длиной L равен

$$\eta_{\text{поч}} = \frac{I_3(L)}{I_1(0)} = \frac{\omega_3}{\omega_1}\sin^2\frac{\gamma L}{2}. \tag{21.115}$$

При $\gamma L \ll 1$ с использованием (21.94) получаем приближенное выражение

$$\frac{I_3(L)}{I_1(0)} \approx \frac{\omega_3}{\omega_1}\left(\frac{\gamma L}{2}\right)^2 = \frac{\omega_3}{\omega_1} g^2 L^2 \phi_2(0) = 2\omega_3^2 L^2 d^2 \eta^3 I_2(0),$$

откуда

$$\boxed{\eta_{\text{поч}} = C^2 \frac{L^2}{A} P_2; \quad C^2 = 2\omega_3{}^2 \eta_0^3 \frac{d^2}{n^3},} \tag{21.116}$$

Коэффициент преобразования частоты вверх

где A — площадь поперечного сечения; $P_2 = I_2(0)A$ — мощность накачки. Это выражение аналогично (21.109) для коэффициента преобразования во вторую гармонику.

Упражнение 21.10

Преобразование частоты вверх для инфракрасного излучения

Преобразователь частоты изготовлен на основе кристалла прустита ($d = 1{,}5 \cdot 10^{-22}$ Кл/В^2, $n = 2{,}6$, $d^2/n^3 = 1{,}3 \cdot 10^{-45}$ Кл^2/В^4). Входная волна получается от CO_2-лазера с длиной волны 10,6 мкм, а накачка — от Nd^{3+}:YAG-лазера мощностью 1 Вт с длиной волны 1,06 мкм, пучок которого сфокусирован в пятно с площадью 10^{-2} мм^2 (см. рис. 21.8). Определите длину волны, преобразованной с повышением частоты, и коэффициент преобразования, если волны коллинеарны, а длина взаимодействия 1 см.

21.4.3. Параметрическое усиление (ПУ) и параметрическая генерация (ПГ) света

ПУ основано на трехволновом смешении в нелинейном кристалле, обеспечивающем оптическое усиление (рис. 21.33, *а*, см. цв. вклейку). Процесс описывается теми же тремя связанными уравнениями (21.93) для следующих

трех волн. Волна 1 представляет собой усиливаемый **сигнал**; она падает на кристалл с малой интенсивностью $I_1(0)$. Волна 3 — **накачка** высокой интенсивности, обеспечивающая необходимую для усиления энергию. Волна 2 называется **холостой** и возникает в процессе взаимодействия.

В предположении идеального фазового синхронизма ($\Delta k = 0$) и без учета ослабления накачки, $a_3(z) \approx a_3(0)$, уравнения для связанных волн (21.93) дают

$$\boxed{\frac{da_1}{dz} = -j\frac{\gamma}{2}a_2^*;} \tag{21.117а}$$

$$\boxed{\frac{da_2}{dz} = -j\frac{\gamma}{2}a_1^*,} \tag{21.117б}$$

где $\gamma = 2ga_3(0)$. Если $a_3(0)$ действительно, то действительно и γ, и дифференциальные уравнения имеют решение

$$a_1(z) = a_1(0)\operatorname{ch}\frac{\gamma z}{2} - ja_2^*(0)\operatorname{sh}\frac{\gamma z}{2}; \tag{21.118а}$$

$$a_2(z) = -ja_1^*(0)\operatorname{sh}\frac{\gamma z}{2} + a_2(0)\operatorname{ch}\frac{\gamma z}{2}. \tag{21.118б}$$

Если $a_2(0) = 0$, т. е. начальное поле равно нулю, то соответствующие плотности потока фотонов

$$\boxed{\phi_1(z) = \phi_1(0)\operatorname{ch}^2\frac{\gamma z}{2};} \tag{21.119а}$$

$$\boxed{\phi_2(z) = \phi_1(0)\operatorname{sh}^2\frac{\gamma z}{2}.} \tag{21.119б}$$

И $\phi_1(z)$, и $\phi_2(z)$ растут монотонно с увеличением z, как показывает рис. 21.33, *б*. Этот рост насыщается, когда от накачки забирается значительная энергия, так что предположение о неизменности ее амплитуды перестает быть верным.

Полное усиление в среде длиной L равно

$$G = \frac{\phi_1(L)}{\phi_1(0)} = \operatorname{sh}^2\left(\frac{\gamma L}{2}\right).$$

В пределе $\gamma L \gg 1$

$$G = \frac{\left(e^{\gamma L/2} + e^{-\gamma L/2}\right)^2}{4} \approx \frac{e^{\gamma L}}{4},$$

так что усиление экспоненциально растет с увеличением γL. Коэффициент усиления

$$\gamma = 2ga_3(0) = 2d\sqrt{2\hbar\omega_1\omega_2\omega_3\eta^3}a_3(0),$$

откуда

$$\boxed{\gamma = 2C\sqrt{I_3(0)} = 2C\sqrt{\frac{P_3}{A}}; \quad C^2 = 2\omega_1\omega_2\eta_0^3\frac{d^2}{n^3},} \qquad (21.120)$$

Коэффициент усиления ПУ

где $P_3 = I_3(0)A$ — мощность накачки; A — площадь поперечного сечения; C^2 — параметр, аналогичный описывающему ГВГ, и параметрическое преобразование частоты вверх.

Взаимодействие фотонов выглядит как расщепление фотона накачки $\hbar\omega_3$ на фотон $\hbar\omega_1$, усиливающий полезный сигнал, и фотон $\hbar\omega_2$ холостой волны (рис. 21.33, *в*).

Упражнение 21.11

Усиление параметрического усилителя

ПУ усиливает свет на 2,5 мкм с использованием кристалла KTP длиной 2 см с накачкой от Nd : YAG-лазера с длиной волны 1,064 мкм. Определите длину волны холостого излучения и коэффициент C в формуле (21.120). Определите лазерную мощность и площадь сечения пучка, обеспечивающие полное усиление 3 дБ. Считайте, что для KTP $n = 1{,}75$, $d = 2{,}3 \cdot 10^{-23}$ Кл/В^2.

Параметрический генератор света (ПГС)

Параметрический генератор получается, если в параметрическом усилителе обеспечить обратную связь для сигнальной и холостой волн или хотя бы одной из них, как показано на рис. 21.34 (см. цв. вклейку). В первом случае генератор называют **дважды резонансным**, а во втором — **однократно резонансным** ПГС.

Частоты генерации ω_1 и ω_2 определяются условиями согласования частот и условиями синхронизма, которые в коллинеарном случае имеют вид $\omega_1 + \omega_2 = \omega_3$ и $n_1\omega_1 + n_2\omega_2 = n_3\omega_3$. Решение этих двух уравнений дает ω_1 и ω_2, как показано в подразд. 21.2.4. Кроме того, эти частоты должны совпадать с резонансными частотами мод резонатора, аналогично тому, как это происходит в обычных лазерах (см. подразд. 15.1.2). Поэтому может оказаться, что система имеет избыточные ограничения, особенно в случае двойного резонанса, когда и сигнальная, и холостая частоты должны совпадать с частотами мод резонатора.

Еще одним условием генерации является превышение усиления над потерями, вносимыми зеркалами, за один полный обход резонатора. Приравнивая

усиления и потери, можно найти выражение для порогового усиления и соответствующую пороговую интенсивность накачки, как будет показано ниже для ПГС с одним и двумя резонансами.

Однократный резонанс

На пороге генерации амплитуда усиленного и дважды отраженного сигнала $a_1(L)r_1^2$ равна исходной амплитуде $a_1(0)$, где L — длина нелинейной среды; r_1 — модуль амплитудного коэффициента отражения зеркала (предполагается, что оба зеркала одинаковы, а набег фазы за полный обход не учитывается, поскольку он составляет целое число 2π). Используя (21.118а) вместе с граничным условием $a_2(0) = 0$, получаем

$$r_1^2 \operatorname{ch}\left(\frac{\gamma L}{2}\right) = 1,$$

откуда

$$\mathcal{R}_1^2 \operatorname{ch}^2\left(\frac{\gamma L}{2}\right) = 1. \tag{21.121}$$

Здесь $\mathcal{R}_1 = r_1^2$ — коэффициент отражения зеркала по интенсивности на частоте сигнала. Поскольку $\mathcal{R}_1$ в типичном случае немного меньше единицы, $\operatorname{ch}^2(\gamma L/2)$ немного больше единицы, т. е. $\gamma L/2 \ll 1$, и можно пользоваться приближением $\operatorname{ch}^2(x) \approx 1 + x^2$. С помощью (21.120) получаем пороговую интенсивность, откуда находим пороговую мощность накачки

$$\boxed{P_3|_{\text{пор}}(0) \approx \frac{1}{C^2}\frac{A}{L^2}\frac{1-\mathcal{R}_1^2}{\mathcal{R}_1^2},} \tag{21.122}$$

Пороговая мощность ПГС
Однократный резонанс

где

$$C^2 = \frac{2\omega_1\omega_2\eta_0^2 d^2}{n^3};$$

A — площадь поперечного сечения. Например, если $L^2/A = 10^6$, $C^2 = 10^{-7}$ Вт$^{-1}$, а $\mathcal{R}_1 = 0{,}9$, то $P_3|_{\text{пор}}(0) \approx 2{,}3$ Вт.

Двойной резонанс

На пороге генерации должны выполняться два условия: $a_1(L)r_1^2 = a_1(0)$ и $a_2(L)r_2^2 = a_2(0)$, где r_1 и r_2 — модули амплитудных коэффициентов зеркал на частоте сигнала и холостой волны, соответственно. Подставляя $a_1(L)$ из (21.118а) и $a_2(L)$ из (21.118б) и выполняя сопряжение, получаем

$$(1-\mathcal{R}_1)\operatorname{ch}\frac{\gamma z}{2}a_1(0) + j\mathcal{R}_1 \operatorname{sh}\frac{\gamma z}{2}a_2^*(0) = 0; \tag{21.123а}$$

$$-j\mathcal{R}_2 \operatorname{sh}\frac{\gamma z}{2}a_1(0) + (1-\mathcal{R}_2)\operatorname{ch}\frac{\gamma z}{2}a_2^*(0) = 0, \tag{21.123б}$$

где $\mathcal{R}_1 = r_1^2$ и $\mathcal{R}_2 = r_2^2$ — коэффициенты отражения зеркала по интенсивности на частоте сигнала и холостой волны соответственно. Приравнивая значения отношения $a_1(0)/a_2^*(0)$, полученные из (21.123а) и (21.123б), получаем

$$\text{th}^2\left(\frac{\gamma L}{2}\right) = \frac{(1-\mathcal{R}_1)(1-\mathcal{R}_2)}{\mathcal{R}_1\mathcal{R}_2}. \tag{21.124}$$

Поскольку правая часть (21.124) намного меньше единицы, можно использовать приближение $\text{th}(x) \approx x$ и записать

$$\left(\frac{\gamma L}{2}\right)^2 = \frac{(1-\mathcal{R}_1)(1-\mathcal{R}_2)}{\mathcal{R}_1\mathcal{R}_2},$$

откуда получаем пороговую мощность накачки

$$\boxed{P_3\big|_{\text{пор}}(0) \approx \frac{1}{C^2}\frac{A}{L^2}\frac{(1-\mathcal{R}_1)(1-\mathcal{R}_2)}{\mathcal{R}_1\mathcal{R}_2}.} \tag{21.125}$$

Пороговая мощность ПГС
Двойной резонанс

Отношение пороговых мощностей накачки для двойного и однократного резонанса, как показывает сравнение формул (21.122) и (21.125), равно $(\mathcal{R}_1/\mathcal{R}_2)(1-\mathcal{R}_2)/(1+\mathcal{R}_1)$. Так как $\mathcal{R}_1 \approx 1$ и $\mathcal{R}_2 \approx 1$, это отношение приближенно равно $(1-\mathcal{R}_2)/2$, что представляет собой малую величину. Таким образом, порог генерации для двойного резонанса значительно ниже, чем для однократного. К сожалению, параметрические генераторы с двойным резонансом более чувствительны к флуктуациям длины резонатора в силу требования, чтобы как частота сигнала, так и частота холостой волны одновременно совпадали с частотами мод резонатора. В силу этих причин ПСГ с двойным резонансом часто имеют плохую стабильность и изрезанные пиками спектры.

*21.5. НЕЛИНЕЙНАЯ ОПТИКА ТРЕТЬЕГО ПОРЯДКА: ТЕОРИЯ СВЯЗАННЫХ ВОЛН

21.5.1. Четырехволновое смешение (ЧВС)

Выведем теперь связанные уравнения, описывающие ЧВС в среде с нелинейностью третьего порядка, в рамках того же подхода, который был использован при рассмотрении трехволнового смешения в разд. 21.4.

Уравнения для связанных волн

Четыре волны, составляющие полное поле,

$$\mathcal{E}(t) = \sum_{q=1,\,2,\,3,\,4} \text{Re}\left[E(\omega_q)\exp(j\omega_q t)\right] = \sum_{q=\pm1,\,\pm2,\,\pm3,\,\pm4} \frac{1}{2}E(\omega_q)\exp(j\omega_q t) \tag{21.126}$$

распространяются в среде с плотностью нелинейной поляризации

$$\mathcal{P}_{\text{нл}} = 4\chi^{(3)}\mathcal{E}^3 \tag{21.127}$$

Соответствующий источник изучения

$$S = -\mu_0 \frac{\partial^2 \mathcal{P}_{\text{нл}}}{\partial t^2}$$

представляет собой сумму $8^3 = 512$ членов

$$S = \frac{1}{2}\mu_0\chi^{(3)} \sum_{q,\,p,\,r=\pm1,\,\pm2,\,\pm3,\,\pm4} \left(\omega_q + \omega_p + \omega_r\right)^2 E_q E_p E_r \exp\left[j\left(\omega_q + \omega_p + \omega_r\right)t\right]. \tag{21.128}$$

Подставляя (21.126) и (21.128) в волновое уравнение (21.74) и приравнивая члены на каждой из четырех частот ω_1, ω_2, ω_3, ω_4, получаем четыре уравнения Гельмгольца с соответствующими источниками

$$\left(\nabla^2 + k_q^2\right)E_q = -S_q, \quad q = 1,\ 2,\ 3,\ 4, \tag{21.129}$$

где S_q — амплитуда компоненты источника S на частоте ω_q.

Чтобы волны были связаны, их частоты должны быть соизмеримы. Рассмотрим, например, случай, когда сумма двух частот оказывается равной сумме двух других частот

$$\omega_1 + \omega_2 = \omega_3 + \omega_4, \tag{21.130}$$

и предположим, что частоты различны. Тогда три волны могут комбинировать и создавать источник на четвертой частоте. Используя (21.130), найдем амплитуды компонент в (21.128) на четырех частотах:

$$S_1 = \mu_0\omega_1^2\chi^{(3)}\left\{6E_3E_4E_2^* + 3E_1\left[|E_1|^2 + 2|E_2|^2 + 2|E_3|^2 + 2|E_4|^2\right]\right\}; \tag{21.131а}$$

$$S_2 = \mu_0\omega_2^2\chi^{(3)}\left\{6E_3E_4E_1^* + 3E_2\left[|E_2|^2 + 2|E_1|^2 + 2|E_3|^2 + 2|E_4|^2\right]\right\}; \tag{21.131б}$$

$$S_3 = \mu_0\omega_3^2\chi^{(3)}\left\{6E_1E_2E_4^* + 3E_3\left[|E_3|^2 + 2|E_2|^2 + 2|E_1|^2 + 2|E_4|^2\right]\right\}; \tag{21.131в}$$

$$S_4 = \mu_0\omega_4^2\chi^{(3)}\left\{6E_1E_2E_3^* + 3E_4\left[|E_4|^2 + 2|E_1|^2 + 2|E_2|^2 + 2|E_3|^2\right]\right\}. \tag{21.131г}$$

Следовательно, каждая волна возбуждается источником с двумя компонентами. Первая компонента — результат смешения трех других волн. Например, первый член в S_1 пропорционален $E_3E_4E_2^*$ и, следовательно, представляет смешение волн 2, 3 и 4, создающее источник для волны 1. Вторая компонента пропорциональна комплексной амплитуде самой волны. Например, второй член в S_1 пропорционален E_1, так что отражает модуляцию показателя преломления и, следовательно, описывает оптический эффект Керра (упражнение 21.5).

Поэтому удобно разделить два вклада в эти источники:

$$S_q = \bar{S}_q + \left(\frac{\omega_q}{c_0}\right)^2 \Delta\chi_q E_q, \quad q = 1,\ 2,\ 3,\ 4, \tag{21.132}$$

где

$$\bar{S}_1 = 6\mu_0\omega_1^2\chi^{(3)}E_3E_4E_2^*; \tag{21.133а}$$

$$\bar{S}_2 = 6\mu_0\omega_2^2\chi^{(3)}E_3E_4E_1^*; \tag{21.133б}$$

$$\bar{S}_3 = 6\mu_0\omega_3^2\chi^{(3)}E_1E_2E_4^*; \tag{21.133в}$$

$$\bar{S}_4 = 6\mu_0\omega_4^2\chi^{(3)}E_1E_2E_3^*; \tag{21.133г}$$

$$\Delta\chi_q = 6\frac{\eta}{\varepsilon_0}\chi^{(3)}\left(2I - I_q\right), \quad q = 1,\ 2,\ 3,\ 4. \tag{21.134}$$

Здесь $I_q = |E_q|^2/2\eta$ — интенсивности волн; $I = I_1 + I_2 + I_3 + I_4$ — полная интенсивность, которая постоянна ввиду сохранения энергии; η — импеданс среды. Это дает возможность переписать уравнения Гельмгольца (21.129) в виде

$$\left(\nabla^2 + \bar{k}_q^2\right)E_q = -\bar{S}_q, \quad q = 1,\ 2,\ 3,\ 4, \tag{21.135}$$

где

$$\bar{k}_q = \bar{n}_q\frac{\omega_q}{c_0}; \tag{21.136}$$

$$\bar{n}_q^2 = n^2 + 2nn_2\left(2I - I_q\right); \tag{21.137}$$

$$n_2 = \frac{3\eta_0}{\varepsilon_0 n^2}\chi^{(3)}, \tag{21.138}$$

что совпадает с (21.47). Если второй член в (21.137) много меньше первого, то

$$\boxed{\bar{n}_q \approx n + n_2\left(2I - I_q\right).} \tag{21.139}$$

Оптический эффект Керра

Таким образом, уравнение Гельмгольца для каждой волны претерпело изменения двух видов.

1. Появился источник, представляющий комбинированное действие трех других волн. Это может приводить к усилению существующей волны или генерации новой волны на данной частоте.

2. Показатель преломления для каждой волны изменен и является функцией интенсивностей четырех волн.

Эти уравнения используются для получения четырех связанных дифференциальных уравнений, которые можно решить относительно полей или их комплексных огибающих при соответствующих граничных условиях. Это тот же подход, который был использован для нелинейных процессов второго порядка. Теперь мы применим его к частным случаям нелинейных процессов третьего порядка.

21.5.2. Трехволновое смешение и генерация третьей гармоники (ГТГ)

Рассмотрим теперь вырожденные случаи, для которых две или три волны имеют одинаковые частоты.

Трехволновое смешение

В вырожденном случае, когда две из четырех волн имеют одинаковую частоту $\omega_3 = \omega_4 \equiv \omega_0$, мы имеем три волны, частоты которых связаны соотношением $\omega_1 + \omega_2 = 2\omega_0$. Теория связанных волн для такого процесса трехволнового смешения может быть сформулирована путем определения источников излучения на всех трех частотах:

$$S_1 = \mu_0\omega_1^2\chi^{(3)}\left\{6E_0^2E_2^* + 3E_1\left[|E_1|^2 + 2|E_2|^2 + 2|E_0|^2\right]\right\}; \qquad (21.140а)$$

$$S_2 = \mu_0\omega_2^2\chi^{(3)}\left\{6E_0^2E_1^* + 3E_2\left[|E_2|^2 + 2|E_1|^2 + 2|E_0|^2\right]\right\}; \qquad (21.140б)$$

$$S_0 = \mu_0\omega_0^2\chi^{(3)}\left\{6E_1E_2E_0^* + 3E_0\left[|E_0|^2 + 2|E_1|^2 + 2|E_2|^2\right]\right\}. \qquad (21.140в)$$

При подстановке в уравнения Гельмгольца

$$\left(\nabla^2 + k_q^2\right)E_q = -S, \quad q = 0,\ 1,\ 2,$$

получается система связанных уравнений, которые в принципе можно решить при соответствующих граничных условиях.

Для коллинеарных волн, распространяющихся вдоль оси z, $E_q(r) = A_q\exp(-jk_qz)$. Как и для нелинейных процессов второго порядка, используем приближение медленных огибающих

$$\left(\nabla^2 + k_q^2\right)\left[A_q\exp\left(-jk_qz\right)\right] \approx -j2k_q\frac{dA_q}{dz}\exp\left(-jk_qz\right)$$

и выразим комплексные амплитуды $A_q = \sqrt{2\eta\hbar\omega_q}a_q$ через переменные a_q, нормированные так, что $\phi_q = |a_q|^2$ — плотность потока фотонов. Анализ упрощается, если при вычислении коэффициентов связи считать, что $\omega_1 \approx \omega_2 \approx \omega_0$. В результате получаем следующую систему связанных уравнений:

$$\frac{da_1}{dz} = -jg\left[a_0^2a_2^*\exp(-j\Delta kz) + a_1\left(|a_1|^2 + 2|a_2|^2 + 2|a_0|^2\right)\right]; \qquad (21.141а)$$

$$\frac{da_2}{dz} = -jg\left[a_0^2a_1^*\exp(-j\Delta kz) + a_2\left(|a_2|^2 + 2|a_1|^2 + 2|a_0|^2\right)\right]; \qquad (21.141б)$$

$$\frac{da_0}{dz} = -jg\left[2a_1a_2a_0^*\exp(j\Delta kz) + a_0\left(|a_0|^2 + 2|a_1|^2 + 2|a_2|^2\right)\right], \qquad (21.141в)$$

где

$$g = \hbar\omega_0\left(\frac{\omega_0}{c_0}\right)n_2; \tag{21.142}$$

$$\Delta k = 2k_0 - k_1 - k_2 \tag{21.143}$$

представляет расстройку фазового синхронизма.

Эта система нелинейных уравнений может быть легко решена в приближении неистощаемой накачки ($|a_1|, |a_2| \ll |a_0|$), поскольку в этом случае $a_0(z)$ приблизительно постоянна. В случае полного фазового синхронизма ($\Delta k = 0$) (21.141) аппроксимируется двумя дифференциальными уравнениями

$$\frac{da_1}{dz} = -j\gamma\left(a_2^* + 2a_1\right); \tag{21.144а}$$

$$\frac{da_2}{dz} = -j\gamma\left(a_1^* + 2a_2\right), \tag{21.144б}$$

где $\gamma = ga_0^2$ — постоянная, пропорциональная неизменной интенсивности накачки. Решение этих уравнений выражается через начальные величины обеих волн:

$$a_1(z) = \left[(1 - j\gamma z)a_1(0) - j\gamma z a_2^*(0)\right]\exp(-j\gamma z); \tag{21.145а}$$

$$a_2(z) = \left[-j\gamma z a_1^*(0) + (1 - j\gamma z)a_2(0)\right]\exp(-j\gamma z). \tag{21.145б}$$

Если начальная амплитуда холостой волны $a_2(0) = 0$, то плотность потока фотонов $\phi_1(z) = |a_1(z)|^2$ сигнальной волны растет как $\phi_1(z) = (1 + \gamma^2 z^2)\phi_1(0)$. Скорость роста чувствительна к начальной величине и фазе холостой волны. Например, если $a_2(0) = re^{j\varphi}a_1(0)$, то

$$\phi_1(z) = \left[1 + (2r\sin\varphi)\gamma z + \left(1 + r^2 + 2r\cos\varphi\right)\gamma^2 z^2\right]\phi_1(0), \tag{21.146}$$

что является функцией разности фаз φ, которая достигает максимума при $\operatorname{tg}\varphi = 2/\gamma z$. При малых z наибольший рост происходит, когда $\varphi = \pi/2$. Очевидно, что прибор является **фазочувствительным усилителем**.

Для исследования влияния истощения накачки и фазовой расстройки нужно решать полную систему уравнений (21.141). Один из шагов в этом направлении — записать комплексные амплитуды $a_q = b_q\exp(j\varphi_q)$, выделив их модули b_q и фазы φ_q. Поставляя это в (21.141) и приравнивая действительные и мнимые части в каждом уравнении, приходим к следующей системе нелинейных уравнений для действительных переменных:

$$\frac{db_1}{dz} = gb_0^2 b_2\sin\varphi; \tag{21.147а}$$

$$\frac{db_2}{dz} = gb_0^2 b_1\sin\varphi; \tag{21.147б}$$

$$\frac{db_0}{dz} = -gb_0 b_1 b_2\sin\varphi; \tag{21.147в}$$

$$\frac{d\varphi}{dz} = \Delta k + g\left[2b_0^2 - b_1^2 - b_2^2\right] + g\left[\frac{b_0^2 b_1}{b_2} + \frac{b_0^2 b_2}{b_1} - 4b_1 b_2\right]\cos\varphi, \qquad (21.147г)$$

где $\phi = \Delta kz + \phi_1 + \phi_2 - 2\phi_0$. Легко определить два инварианта. В согласии с сохранением оптической интенсивности сумма $b_1^2 + b_2^2 + b_3^2$ должна быть постоянна. Кроме того, в соответствии с законом сохранения числа фотонов, разность $b_1^2 - b_2^2$ должна быть постоянна (вариант соотношения Мэнли—Роу). Можно найти и другие инварианты, содержащие фазу[3], и использовать их для изучения роли фазовой расстройки, разности начальных амплитуд и фаз между сигналом и холостой волной. Например, из (21.147а) видно, что начальная скорость роста сигнала есть, когда $\sin\varphi = 0$, т. е. $\varphi = \pi/2$.

Генерация третьей гармоники (ГТГ)

Другой случай вырождения при четырехволновом смешении — генерация третьей гармоники. Здесь три из четырех волн имеют одинаковые частоты $\omega_1 = \omega_2 = \omega_4 = \omega$, а четвертая имеет суммарную частоту $\omega_3 = \omega_1 + \omega_2 + \omega_4 = 3\omega$. В действительности имеется две волны 1 и 3, амплитуды которых связаны средой с нелинейностью третьего порядка. Теория связанных волн будет построена на основе того же подхода, который был использован при рассмотрении четырех- и трехволнового смешения. Он приводит к двум уравнениям Гельмгольца

$$\left(\nabla^2 + k_q^2\right)E_q = -S,$$

где

$$S_1 = \mu_0\omega_1^2\chi^{(3)}\left\{3E_3E_1^*E_1^* + 3E_1\left[|E_1|^2 + 2|E_3|^2\right]\right\}; \qquad (21.148а)$$

$$S_3 = \mu_0\omega_3^2\chi^{(3)}\left\{E_1^3 + 3E_3\left[|E_3|^2 + 2|E_1|^2\right]\right\}. \qquad (21.148б)$$

Эти уравнения можно использовать для вывода связанных уравнений относительно E_1 и E_3, как это было сделано в предыдущих случаях.

Упражнение 21.12

ГТГ в приближении неистощаемой накачки

Предположим, что основная волна и волна третьей гармоники — плоские волны, распространяющиеся в направлении z, с комплексными огибающими A_q, $q = 1, 3$. Используйте приближение медленной огибающей, чтобы записать связанные дифференциальные уравнения для A_1 и A_3. Покажите, что в приближении неистощаемой накачки [$|A_1 \ll A_3$ и $A_1(z) \approx A_1(0)$]

$$\frac{da_3}{dz} = -jga_1^3\exp(-j\Delta kz), \qquad (21.149)$$

[3] См. *Capellini G., Trillo S.* Third-Order Three-Wave Mixing in Single-Mode Fibers: Exact Solutions and Spatial Instability Effects. Journal of the Optical Society of America B. Vol. 8, 1991. P. 824—838.

где

$$A_q = \sqrt{2\eta\hbar\omega_q}\,a_q; \quad \Delta k = 3k_1 - k_3.$$

Выведите выражение для g.

21.5.3. Обращение волнового фронта (ОВФ)

Теперь получим и решим связанные уравнения в полностью вырожденном случае, когда частоты всех четырех волн одинаковы: $\omega_1 = \omega_2 = \omega_3 = \omega_4 = \omega$. Как было принято в подразд. 21.3.5, две волны (3 и 4), называемые волнами накачки, представляют собой плоские волны, распространяющиеся в противоположных направлениях, с комплексными амплитудами $E_3(\boldsymbol{r}) = A_3 \exp(-j\boldsymbol{k}_3 \cdot \boldsymbol{r})$ и $E_4(\boldsymbol{r}) = A_4 \exp(-j\boldsymbol{k}_4 \cdot \boldsymbol{r})$, где $\boldsymbol{k}_4 = -\boldsymbol{k}_3$. Предполагается, что их интенсивности много больше интенсивностей волн 1 и 2, так что они практически не истощаются в процессе взаимодействия, и их комплексные огибающие A_3 и A_4 можно считать постоянными. Полная интенсивность всех четырех волн тогда тоже приблизительно постоянна:

$$I \approx \frac{|A_3|^2 + |A_4|^2}{2\eta}.$$

Члены $2I - I_1$ и $2I - I_2$, определяющие эффективный показатель преломления $\bar{n}$ для волн 1 и 2 в (21.139), приближенно равны $2I$ и, следовательно, также приблизительно постоянны, так что оптический эффект Керра сводится к постоянному изменению показателя преломления. Поэтому его влияние учитываться не будет.

С этими допущениями задача сводится к взаимодействию двух связанных волн 1 и 2. Уравнения (21.135) и (21.133) дают

$$\boxed{\left(\nabla^2 + k^2\right) E_1 = -\xi E_2^*;} \tag{21.150а}$$

$$\boxed{\left(\nabla^2 + k^2\right) E_2 = -\xi E_1^*,} \tag{21.150б}$$

где

$$\xi = 6\mu\omega^2 \chi^{(3)} E_3 E_4 = 6\mu\omega^2 \chi^{(3)} A_3 A_4; \tag{21.151}$$

$$k = \frac{\bar{n}\omega}{c_0}, \quad \bar{n} \approx n + 2n_2 I.$$

Здесь $\bar{n}$ — постоянная.

Четыре нелинейных связанных дифференциальных уравнения сведены, таким образом, к двум *линейным* связанным уравнениям, каждое из которых имеет вид уравнения Гельмгольца с источником. Источник для волны 1 пропорционален сопряженной комплексной амплитуде волны 2 и наоборот.

Обращение волнового фронта

Пусть волны 1 и 2 также плоские и распространяются в противоположных направлениях вдоль оси z, как показано на рис. 21.35 (см. цв. вклейку),

$$E_1 = A_1 \exp(-jkz); \quad E_2 = A_2 \exp(jkz). \tag{21.152}$$

Это предположение согласуется с условием фазового синхронизма, поскольку $k_1 + k_2 = k_3 + k_4$.

Подставляя (21.152) в (21.150) и используя приближение медленно меняющихся огибающих (21.92), сведем уравнения (21.150) к двум дифференциальным уравнениям первого порядка

$$\frac{dA_1}{dz} = -j\gamma A_2^*; \tag{21.153а}$$

$$\frac{dA_2}{dz} = j\gamma A_1^*, \tag{21.153б}$$

где γ — коэффициент связи,

$$\gamma = \frac{\xi}{2k} = 3\omega\eta_0 \frac{\chi^{(3)}}{\bar{n}} A_3 A_4, \tag{21.154}$$

абсолютная величина которого может быть записана в виде

$$\boxed{|\gamma| = 2C\sqrt{I_3 I_4}, \quad C = 3\omega\eta_0^2 \frac{\chi^{(3)}}{\bar{n}^2}.} \tag{21.155}$$

Коэффициент связи

Здесь I_3 и I_4 — интенсивности двух волн,

$$I_3 = \frac{|A_3|^2}{2\eta}; \quad I_4 = \frac{|A_4|^2}{2\eta};$$

$$\eta = \frac{\eta_0}{\bar{n}}.$$

Для простоты предположим, что $A_3 A_4$ действительно, так что действительно и γ. Решение (21.153) тогда представляет собой две гармонические функции $A_1(z)$ и $A_2(z)$ с разностью фаз 90°. Если нелинейная среда заполняет пространство между плоскостями $z = -L$ и $z = 0$, как показано на рис. 21.35, то волна 1 имеет амплитуду $A_1(-L) = A_i$ в плоскости входа, а волна 2 имеет нулевую амплитуду в плоскости выхода $A_2(0) = 0$. При таких граничных условиях решение (21.153):

$$A_1(z) = \frac{A_i}{\cos\gamma L}\cos\gamma z; \tag{21.156}$$

$$A_2(z) = j\frac{A_i^*}{\cos\gamma L}\sin\gamma z. \tag{21.157}$$

Амплитуда отраженной волны в плоскости входа $A_r = A_2(-L)$:

$$\boxed{A_r = -jA_i^* \operatorname{tg} \gamma L,} \qquad (21.158)$$

Амплитуда отраженной волны

тогда как амплитуда прошедшей волны $A_t = A_1(0)$:

$$\boxed{A_t = \frac{A_i}{\cos \gamma L}.} \qquad (21.159)$$

Амплитуда прошедшей волны

Уравнения (21.158) и (21.159) указывают на ряд приложений.

• Отраженная волна представляет собой сопряженную падающую волну. Прибор работает как **отражатель, обращающий волновой фронт** (см. подразд. 21.3.6).

• Коэффициент отражения по интенсивности

$$\frac{|A_r|^2}{|A_i|^2} = \operatorname{tg}^2 \gamma L$$

может быть больше или меньше 1, что соответствует усилению или ослаблению. Следовательно, среда может служить **отражательным усилителем** («усиливающим зеркалом»).

• Амплитудный коэффициент пропускания

$$\frac{|A_t|^2}{|A_i|^2} = \frac{1}{\cos^2 \gamma L}$$

всегда больше 1, так что среда всегда действует как **усилитель проходящего сигнала**.

• Когда $\gamma L = \pi/2$ или нечетному числу $\pi/2$, коэффициенты отражения и пропускания бесконечны, что говорит о неустойчивости. Устройство тогда становится **генератором**.

*21.6. АНИЗОТРОПНЫЕ НЕЛИНЕЙНЫЕ СРЕДЫ

В анизотропной среде каждая из трех компонент вектора поляризации $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3)$ в общем случае зависит от трех компонент вектора электрического поля $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3)$. Эти функции линейны для малых $\mathcal{E}$ (см. разд. 6.3), но начинают слегка отклоняться от линейности по мере роста $\mathcal{E}$. Следовательно, их можно разложить в ряды Тейлора по степеням трех компонент $\mathcal{E}$ так же, как это делалось при скалярном анализе, представленном в разд. 21.1:

$$\boxed{\mathcal{P}_i = \varepsilon_0 \sum_j \chi_{ij} \mathcal{E}_i + 2 \sum_{jk} d_{ijk} \mathcal{E}_i \mathcal{E}_k + 4 \sum_{jkl} \chi^{(3)}_{ijkl} \mathcal{E}_j \mathcal{E}_k \mathcal{E}_l}, \quad i, j, k, l = 1, 2, 3. \qquad (21.160)$$

Коэффициенты χ_{ij}, d_{ijk} и $\chi^{(3)}_{ijkl}$ — компоненты тензоров, соответствующих скалярным коэффициентам χ, d и $\chi^{(3)}$, а формула (21.160) является векторным обобщением соотношения (21.2). Поскольку d_{ijk} пропорционален $\partial^2 \mathcal{P}_i / \partial \mathcal{E}_j \partial \mathcal{E}_k$, он инвариантен относительно перестановки j и k. Аналогично, величина $\chi^{(3)}_{ijkl}$ инвариантна относительно перестановок j, k и l. Для среды без потерь и дисперсии имеются дополнительные внутренние симметрии: $\chi_{ij} = \chi_{ji}$, как показано в подразд. 6.3.1. Аналогично, d_{ijk} и $\chi^{(3)}_{ijkl}$ также инвариантны относительно перестановки своих индексов. Эта полная симметрия относительно перестановок, вообще говоря, не имеет места в нелинейных средах с дисперсией.

С использованием условия симметрии $d_{ijk} = d_{ikj}$ компоненты тензора d_{ijk} обычно записываются в виде матрицы d_{iJ} размером 3×6, где шесть независимых комбинаций $(j, k) = 11, 22, 33, 23, 31, 12$ нумеруются одним индексом $J = 1, 2, 3, 4, 5, 6$ в порядке следования (см. табл. 20.1). Например, d_{25} обозначает коэффициенты $d_{231} = d_{213}$.

Коэффициенты третьего порядка $\chi^{(3)}_{ijkl}$ аналогично записываются в виде матрицы $\chi^{(3)}_{IK}$, где пара (i, j) нумеруется одним индексом $I = 1, 2, ..., 6$, а пара (k, l) — индексом $K = 1, 2, ..., 6$.

Структурная симметрия кристаллов накладывает дополнительные ограничения на компоненты тензоров d_{ijk} и $\chi^{(3)}_{ijkl}$. При совпадении осей координат (1, 2, 3) с главными осями кристалла, которые определяются тензором χ_{ij}, некоторые элементы массивов d_{iJ} и $\chi^{(3)}_{IK}$ обращаются в нуль, а другие совпадают или связаны простыми правилами. Примеры приведены в табл. 21.1 и 21.2. Величины коэффициентов d_{iJ} для ряда распространенных нелинейных кристаллов приведены в табл. 21.3. Хотя кубические кристаллы имеют изотропные линейные оптические свойства, их хорошо определенные кристаллические оси (задаваемые их структурной симметрией) наделяют анизотропными нелинейными оптическими свойствами.

Таблица 21.1. **Нелинейные коэффициенты второго порядка для некоторых кристаллических групп**

$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{14} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & -d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$
Кубический $\bar{4}3m$ (GaAs, CdTe, InAs)	Тетрагональный $\bar{4}2m$ (KDP, ADP)	Тригональный $3m$ (BBO, $LiNbO_3$, $LiTaO_3$)

Тензоры d_{ijk} и $\chi^{(3)}_{ijkl}$ тесно связаны с тензорами Поккельса и Керра $\mathfrak{r}_{ijk}$ и $\mathfrak{s}_{ijkl}$ соответственно, как демонстрируется в задаче 2 к разд. 21.6, и имеют ту же симметрию, как можно видеть при сравнении табл. 21.1 и 21.2, где приведены d_{iJ} и $\chi^{(3)}_{IK}$, с табл. 20.2 и 20.3, где приведены $\mathfrak{r}_{Ik}$ и $\mathfrak{s}_{IK}$ для ряда кристаллических групп. Заметим, что матрица d_{iJ} аналогична транспонированной матрице $\mathfrak{r}_{Ik}$.

Таблица 21.2. **Нелинейные коэффициенты третьего порядка $\chi_{ik}^{(3)}$ для изотропной среды**

$$\begin{bmatrix} \chi_{11}^{(3)} & \chi_{12}^{(3)} & \chi_{12}^{(3)} & 0 & 0 & 0 \\ \chi_{12}^{(3)} & \chi_{11}^{(3)} & \chi_{12}^{(3)} & 0 & 0 & 0 \\ \chi_{12}^{(3)} & \chi_{12}^{(3)} & \chi_{11}^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \chi_{44}^{(3)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \chi_{44}^{(3)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \chi_{44}^{(3)} \end{bmatrix}, \qquad \chi_{44}^{(3)} = \frac{1}{2}\left(\chi_{11}^{(3)} - \chi_{12}^{(3)}\right).$$

Таблица 21.3. **Примеры величин нелинейных оптических коэффициентов второго порядка для некоторых материалов***

Кристалл	d_{ij}, Кл/В^2	d_{ij}/ε_0, пм/В**
β-BaB_2O_4 (BBO)	$d_{22} = 2{,}0 \cdot 10^{-23}$ $d_{31} = 3{,}0 \cdot 10^{-25}$	2,2 0,04
LiB_3O_5 (LBO)	$d_{31} = 5{,}9 \cdot 10^{-24}$ $d_{32} = 7{,}5 \cdot 10^{-24}$ $d_{33} = 3{,}5 \cdot 10^{-25}$	0,67 0,85 0,04
$LiIO_3$	$d_{31} = 3{,}9 \cdot 10^{-23}$ $d_{33} = 4{,}1\ \ 10^{-23}$	4,4 4,6
$LiNbO_3$	$d_{22} = 1{,}9 \cdot 10^{-23}$ $d_{31} = 4{,}1 \cdot 10^{-23}$ $d_{33} = 2{,}2 \cdot 10^{-22}$	2,1 4,6 25,2
$KNbO_3$	$d_{31} = 1{,}1 \cdot 10^{-22}$ $d_{32} = 1{,}2 \cdot 10^{-22}$	11,9 13,7
$KTiOPO_4$ (KTP)	$d_{31} = 2{,}0 \cdot 10^{-23}$ $d_{32} = 3{,}3 \cdot 10^{-23}$ $d_{33} = 1{,}3 \cdot 10^{-22}$	2,2 3,7 14,6
KH_2PO_4 (KDP)	$d_{36} = 3{,}1 \cdot 10^{-24}$	0,38
$NH_4H_2PO_4$ (ADP)	$d_{36} = 4{,}2 \cdot 10^{-24}$	0,47
α-SiO_2 (кварц)	$d_{11} = 2{,}7 \cdot 10^{-24}$	0,30
$KBe_2BO_3F_2$ (KBBF)	$d_{11} = 4{,}3 \cdot 10^{-24}$	0,49
GaAs	$d_{14} = 1{,}5 \cdot 10^{-21}$	170,0
Te	$d_{11} = 5{,}8 \cdot 10^{-21}$	650,0

* Большая часть коэффициентов взята из книги *Nikogosyan D.N.* Nonlinear Optical Crystals: A Complete Survey. Springer-Verlag, 2005. Значения приведены на длине волны $\lambda_0 = 1{,}06$ мкм за исключением Te, для которого $\lambda_0 = 10{,}6$ мкм.

** Коэффициенты d/ε_0 в единицах пм/В часто используются на практике. Нелинейные оптические коэффициенты в единицах Кл/В^2 (СИ) легко преобразуются в пм/В путем деления d на $10^{-12}\varepsilon_0 \approx 8{,}85 \cdot 10^{-24}$.

Трехволновое смешение в анизотропных средах с нелинейностью второго порядка

Когда оптическое поле, состоящее из двух монохроматических линейно поляризованных волн с угловыми частотами ω_1 и ω_2 и комплексными амплитудами $\boldsymbol{E}(\omega_1)$ и $\boldsymbol{E}(\omega_2)$ распространяется через кристалл с нелинейностью второго порядка, вектор наведенной нелинейной плотности поляризации $\boldsymbol{P}(\omega_3)$ на частоте $\omega_3 = \omega_1 + \omega_2$ имеет компоненты

$$P_i(\omega_3) = 2\sum_{jk} d_{ijk} E_j(\omega_1) E_k(\omega_2), \quad i, j, k = 1, 2, 3, \tag{21.161}$$

где $E_j(\omega_1)$, $E_k(\omega_2)$ и $P_i(\omega_3)$ — компоненты векторов в системе главных осей кристалла. Это соотношение является обобщением формулы (21.20г).

В сокращенных обозначениях $(j, k) = J$ (21.161) удобно записать в матричном виде

$$\begin{bmatrix} P_1(\omega_3) \\ P_2(\omega_3) \\ P_3(\omega_3) \end{bmatrix} = 2 \begin{bmatrix} d_{11} & d_{12} & d_{16} \\ d_{21} & d_{22} & d_{26} \\ d_{31} & d_{32} & d_{36} \end{bmatrix} \begin{bmatrix} E_1(\omega_1) E_1(\omega_2) \\ E_2(\omega_1) E_2(\omega_2) \\ E_3(\omega_1) E_3(\omega_2) \\ E_2(\omega_1) E_3(\omega_2) + E_3(\omega_1) E_2(\omega_2) \\ E_3(\omega_1) E_1(\omega_2) + E_1(\omega_1) E_3(\omega_2) \\ E_1(\omega_1) E_2(\omega_2) + E_2(\omega_1) E_1(\omega_2) \end{bmatrix}. \tag{21.162}$$

Эффективная величина d

Если $E_j(\omega_1) = E(\omega_1)\cos\theta_{1j}$ и $E_k(\omega_2) = E(\omega_2)\cos\theta_{2k}$, где θ_{1j} и θ_{2k} — углы, образованные векторами $\boldsymbol{E}(\omega_1)$ и $\boldsymbol{E}(\omega_2)$ и главными осями, то (21.161) можно переписать в виде

$$P_i(\omega_3) = 2\left[\sum_{jk} d_{ijk} \cos\theta_{1j} \cos\theta_{2k}\right] E(\omega_1) E(\omega_2). \tag{21.163}$$

Поскольку вектор плотности поляризации $\boldsymbol{P}(\omega_3)$ является источником волны 3, вклад в излучение дает только компонента $\boldsymbol{P}_\perp(\omega_3)$, лежащая в плоскости, ортогональной волновому вектору $\boldsymbol{k}_3$; компонента, параллельная $\boldsymbol{k}_3$, не может излучать ТЕМ-волну. Если $\boldsymbol{P}_\perp(\omega_3)$ образует углы $\bar{\theta}_{3i}$ с главными осями, то ее модуль равен

$$P_\perp(\omega_3) = \sum_i P_i(\omega_3) \cos\bar{\theta}_{3i}. \tag{21.164}$$

Из (21.163) и (21.164) следует, что

$$P_\perp(\omega_3) = 2d_{\text{эфф}} E(\omega_1) E(\omega_2), \tag{21.165}$$

с эффективным коэффициентом нелинейности второго порядка

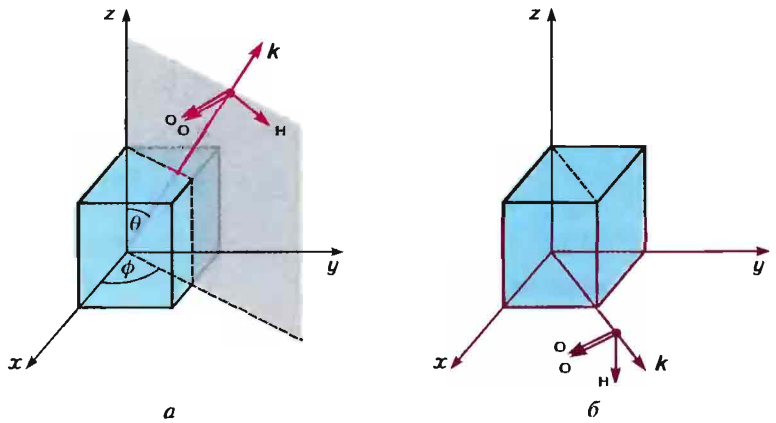
$$d_{\text{эфф}} = \sum_{ijk} d_{ijk} \cos\bar{\theta}_{3i} \cos\theta_{1j} \cos\theta_{2k}. \qquad (21.166)$$

Уравнение (21.165) принимает такой же вид, как и в скалярной формулировке, представленной в подразд. 21.2.3 и разд. 21.4; $d_{\text{эфф}}$ играет роль коэффициента d. Пример 21.7 иллюстрирует прямое вычисление $d_{\text{эфф}}$ для трехволнового смешения в анизотропном кристалле.

Пример 21.7

Коллинеарное трехволновое смешение типа I в кристалле KDP

В этом примере мы определим эффективный нелинейный коэффициент $d_{\text{эфф}}$ для трех коллинеарных волн, распространяющихся в кристалле KDP в произвольном направлении, определенном углами (θ, ϕ) в сферической системе координат с осью z, направленной вдоль оси кристалла, как показано на рис. 21.36. Волны 1 и 2 — обыкновенные волны на частотах ω_1 и ω_2, а волна 3 — необыкновенная волна с частотой $\omega_3 = \omega_1 + \omega_2$.



а *б*

Рис. 21.36. Геометрия коллинеарного о-о-н трехволнового смешения I типа в одноосном кристалле, оптическая ось которого направлена вдоль z (*а*). Направление распространения, при котором достигается максимум $d_{\text{эфф}}$ (*б*)

С помощью (21.161) и табл. 21.1 для кристаллов симметрии $\bar{4}2m$, таких как KDP, нелинейные компоненты вектора плотности поляризации выражаются как

$$\begin{aligned} P_1(\omega_3) &= 2d_{14}\left[E_2(\omega_1)E_3(\omega_2) + E_3(\omega_1)E_2(\omega_2)\right]; \\ P_2(\omega_3) &= 2d_{14}\left[E_3(\omega_1)E_1(\omega_2) + E_1(\omega_1)E_3(\omega_2)\right]; \\ P_3(\omega_3) &= 2d_{36}\left[E_1(\omega_1)E_2(\omega_2) + E_2(\omega_1)E_1(\omega_2)\right]. \end{aligned} \qquad (21.167)$$

В такой геометрии компоненты электрического поля волн 1 и 2

$$E_1(\omega_1) = E(\omega_1)\sin\phi, \quad E_2(\omega_1) = -E(\omega_1)\cos\phi, \quad E_3(\omega_1) = 0;$$

$$E_1(\omega_2) = E(\omega_2)\sin\phi, \quad E_2(\omega_2) = -E(\omega_2)\cos\phi, \quad E_3(\omega_2) = 0.$$

Следовательно, на основе (21.167) компоненты вектора плотности поляризации для волны 3 получаются следующими:

$$P_1(\omega_3) = 0; \quad P_2(\omega_3) = 0; \quad P_3(\omega_3) = -4d_{36}\sin\phi\cos\phi E(\omega_1)E(\omega_2). \quad (21.168)$$

В этом случае компонента $P_\perp(\omega_3) = -P_3(\omega_3)\sin\theta$, так что

$$\boxed{d_{\text{эфф}} = -d_{36}\cos\theta\cos 2\phi.} \quad (21.169)$$

Этот результат можно получить путем непосредственного использования (21.166) с соответствующими углами и коэффициентами.

Эффективный коэффициент нелинейности (21.169) имеет максимальное значение d_{36} при углах $\theta = 90°$ и $\phi = 45°$, как показано на рис. 21.36.

---

*21.7. НЕЛИНЕЙНЫЕ СРЕДЫ С ДИСПЕРСИЕЙ

В этом разделе кратко обсуждается природа дисперсии и ее влияние на нелинейные оптические процессы. Для простоты влияние анизотропии не рассматривается. Диспергирующая среда — это среда с памятью (см. разд. 5.2); плотность поляризации $\mathcal{P}(t)$, вызванная приложенным электрическим полем $\mathcal{E}(t)$, не возникает мгновенно. Напротив, отклик $\mathcal{P}(t)$ в момент времени t зависит от приложенного электрического поля во все предыдущие моменты $t' \le t$. Когда среда еще и нелинейная, функциональная связь между $\mathcal{P}(t)$ и $\{\mathcal{E}(t), t' \le t\}$ также становится нелинейной. Имеется два средства описания таких нелинейных динамических систем.

1. Феноменологическое интегральное соотношение между $\mathcal{P}(t)$ и $\mathcal{E}(t)$, основанное на разложении в ряд Вольтерра, аналогичное разложению в ряд Тейлора. Коэффициенты разложения характеризуют среду феноменологически.

2. Нелинейное дифференциальное уравнение для $\mathcal{P}(t)$ с $\mathcal{E}(t)$ в качестве вынуждающей силы, полученное из физической модели процесса поляризации, подобной модели Лоренца для линейной среды.

Описание нелинейной среды с дисперсией при помощи интегральных преобразований

Если отклонение от линейности мало, то для описания связи между $\mathcal{P}(t)$ и $\mathcal{E}(t)$ можно использовать разложение в ряд Вольтерра. Первый член этого разложения — линейная комбинация $\mathcal{E}(t')$ для всех $t' \le t$

$$\mathcal{P}(t) = \varepsilon_0 \int_{-\infty}^{\infty} \chi(t - t')\mathcal{E}(t')\,dt' \quad (21.170)$$

Это соотношение описывает линейную систему с функцией отклика на импульсное воздействие $\varepsilon_0 x(t)$ [см. разд. 5.2, особенно (5.38), и приложение Б].

Второй член в разложении — суперпозиция произведений $\mathcal{E}(t')\mathcal{E}(t'')$ для *пар* моментов времени $t' \le t$ и $t'' \le t$:

$$\mathcal{P}(t) = \varepsilon_0 \int_{-\infty}^{\infty}\int x^{(2)}(t-t',\ t-t'')\mathcal{E}(t')\mathcal{E}(t'')\,dt'\,dt'', \tag{21.171}$$

где $x^{(2)}(t', t'')$ — функция двух переменных, характеризующая дисперсию нелинейности второго порядка. Третий член представляет нелинейность третьего порядка, характеризуемую функцией $x^{(3)}(t', t'', t''')$ и аналогичным выражением, содержащим тройной интеграл.

Линейный вклад (21.170) в поляризацию с учетом дисперсии можно также полностью охарактеризовать откликом на монохроматическое поле. Если $\mathcal{E}(t) = \mathrm{Re}\{E(\omega)\exp(j\omega t)\}$, то $\mathcal{P}(t) = \mathrm{Re}\{P(\omega)\exp(j\omega t)\}$, где $P(\omega) = \varepsilon_0\chi(\omega)E(\omega)$; $\chi(\omega)$ — Фурье-образ функции $x(t)$ при $n = \omega/2\pi$. Тогда среда полностью характеризуется зависящей от частоты восприимчивостью $\chi(\omega)$.

Вклад *нелинейности второго порядка*, описываемый формулой (21.171), характеризуется откликом на суперпозицию двух монохроматических волн с угловыми частотами ω_1 и ω_2. Подставляя

$$\mathcal{E}(t) = \mathrm{Re}\left\{E(\omega_1)\exp(j\omega_1 t) + E(\omega_2)\exp(j\omega_2 t)\right\} \tag{21.172}$$

в (21.171), можно показать, что компонента плотности поляризации на частоте $\omega_3 = \omega_1 + \omega_2$ имеет амплитуду

$$\boxed{P(\omega_3) = 2d(\omega_3;\ \omega_1,\ \omega_2)E(\omega_1)E(\omega_2).} \tag{21.173}$$

Коэффициент $d(\omega_3; \omega_1, \omega_2)$ представляет собой зависящий от частоты коэффициент d в (21.20г). Связь между этим коэффициентом и функцией отклика $x^{(2)}(t', t'')$ устанавливается определением

$$X^{(2)}(\omega_1,\ \omega_2) = \int_{-\infty}^{\infty}\int \chi^{(2)}(t',\ t'')\exp\left[-j(\omega_1 t' + \omega_2 t'')\right]dt'\,dt'', \tag{21.174}$$

которое представляет собой двумерное преобразование Фурье функции $x^{(2)}(t', t'')$ при $n_1 = -\omega_1/2\pi$ и $n_2 = -\omega_2/2\pi$ [см. (А.27) в приложении А]. Подставляя (21.172) в (21.171) и используя (21.174), получаем

$$d(\omega_3;\ \omega_1,\ \omega_2) = \varepsilon_0 X^{(2)}(\omega_1,\ \omega_2). \tag{21.175а}$$

Таким образом, среда с дисперсией нелинейности второго порядка полностью характеризуется любой из двух частотно-зависимых функций: $X^{(2)}(\omega_1, \omega_2)$ или $d(\omega_3; \omega_1, \omega_2)$.

Вырожденный случай — генерация второй гармоники в среде с нелинейностью второго порядка — также легко описывается подстановкой $\mathcal{E}(t) = \mathrm{Re}\{E(\omega)\exp(j\omega t)\}$

в (21.171) с использованием (21.174). Результирующая поляризация имеет компоненту на частоте 2ω с амплитудой $P(2\omega) = d(2\omega;\ \omega,\ \omega)E(\omega)E(\omega)$, где

$$d(2\omega;\ \omega,\ \omega) = \frac{1}{2}\varepsilon_0\chi^{(2)}(\omega,\ \omega). \qquad (21.175б)$$

Другие коэффициенты d, представляющие различные процессы смешения волн, можно аналогичным образом связать с двумерной функцией $\chi^{(2)}(\omega_1, \omega_2)$. Например, электрооптический эффект получается в результате взаимодействия между постоянным электрическим полем ($\omega_1 = 0$) и оптическим полем ($\omega_2 = \omega$). Коэффициент, соответствующий этому взаимодействию, есть $d(\omega; 0, \omega) = 2\varepsilon_0\chi^{(2)}(\omega, 0)$; он определяет коэффициент Поккельса r в соответствии с (21.18).

В среде с *нелинейностью третьего порядка* электрическое поле, состоящее из трех гармонических функций с угловыми частотами ω_1, ω_2 и ω_3, создает плотность поляризации с суммарной частотой $\omega_4 = \omega_1 + \omega_2 + \omega_3$ и амплитудой

$$\boxed{P(\omega_4) = 6\chi^{(3)}(\omega_4;\ \omega_1,\ \omega_2,\ \omega_3)E(\omega_1)E(\omega_2)E(\omega_3),} \qquad (21.176)$$

где функция $\chi^{(3)}(\omega_4; \omega_1, \omega_2, \omega_3)$ заменяет коэффициент $\chi^{(3)}$, описывающий случай без дисперсии. Функцию $\chi^{(3)}(\omega_4; \omega_1, \omega_2, \omega_3)$ можно определить из $x^{(3)}(t', t'', t''')$ с помощью соотношений, аналогичных (21.175а).

Короче говоря, вследствие дисперсии коэффициенты нелинейности второго и третьего порядка d и $\chi^{(3)}$ приобретают зависимость от частоты волн, участвующих в процессе смешения.

Описание нелинейных сред с дисперсией при помощи дифференциальных уравнений

Примером нелинейной динамической связи между $\mathcal{P}(t)$ и $\mathcal{E}(t)$ является дифференциальное уравнение

$$\frac{d^2\mathcal{P}}{dt^2} + \sigma\frac{d\mathcal{P}}{dt} + \omega_0^2\mathcal{P} + \omega_0^2\varepsilon_0\chi_0 b\mathcal{P}^2 = \omega_0^2\varepsilon_0\chi_0\mathcal{E}, \qquad (21.177)$$

где σ, ω_0, χ_0 и b — постоянные. В отсутствие нелинейного члена $\omega_0^2\varepsilon_0\chi_0 b\mathcal{P}^2$ уравнение (21.177) сводится к уравнению (5.99), которое справедливо для линейной резонансной диэлектрической среды, описываемой моделью осциллятора Лоренца (см. подразд. 5.5.3). В этом случае каждый атом характеризуется гармоническим осциллятором, в котором электрон массы m подвергается действию силы внешнего электрического поля $-e\mathcal{E}$, упругой возвращающей силы $-\kappa x$ и силы трения $m\sigma dx/dt$, где x — смещение электрона относительно положения равновесия; $\omega_0 = \sqrt{\kappa/m}$ — резонансная угловая частота. Среда линейна и обладает дисперсией, которая описывается формулой [см. (5.101)]

$$\boxed{\chi(\omega) = \chi_0\frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega\sigma}.} \qquad (21.178)$$

Линейная восприимчивость (модель гармонического осциллятора)

Когда возвращающая сила нелинейно зависит от смещения, например, как $-\kappa x - \kappa_2 x^2$, где κ и κ_2 — постоянные, получается ангармонический осциллятор, описываемый уравнением (21.177), где величина b пропорциональна κ_2. Тогда среда нелинейна.

Упражнение 21.13

Плотность поляризации для среды из ангармонических осцилляторов

Покажите, что для среды, содержащей в единице объема N атомов, каждый из которых моделируется ангармоническим (нелинейным) осциллятором с возвращающей силой $-\kappa x - \kappa_2 x^2$, связь между $\mathcal{P}(t)$ и $\mathcal{E}(t)$ дается нелинейным дифференциальным уравнением (21.177), где

$$\chi_0 = \frac{Ne^2}{\varepsilon_0 m\omega_0^2}; \quad b = \frac{\kappa_2}{e^3 N^2}.$$

Уравнение (21.177) невозможно решить точно. Однако, если нелинейный член мал, приближенное решение можно получить с помощью итерационного процесса. Запишем (21.177) в виде

$$\mathcal{L}\{\mathcal{P}\} = \mathcal{E} - b\mathcal{P}^2, \tag{21.179}$$

где $\mathcal{L}$ — линейный дифференциальный оператор,

$$\mathcal{L} = \frac{1}{\omega_0^2 \varepsilon_0 \chi_0}\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \sigma\frac{\mathrm{d}}{\mathrm{d}t} + \omega_0^2\right).$$

Итерационное решение (21.179) включает следующие шаги:

1. Пренебрегая нелинейным членом $b\mathcal{P}^2$ в (21.179), находят приближение первого порядка $\mathcal{P}_1$ как решение *линейного* уравнения

$$\mathcal{L}\{\mathcal{P}_1\} \approx \mathcal{E}. \tag{21.180}$$

2. Это приближенное решение используется для определения малого нелинейного члена $b\mathcal{P}_1^2$.

3. Приближение второго порядка получается путем решения (21.179), где член $b\mathcal{P}^2$ заменен на $b\mathcal{P}_1^2$. Решение получающегося *линейного* уравнения обозначается $\mathcal{P}_2$

$$\mathcal{L}\{\mathcal{P}_2\} = \mathcal{E} - b\mathcal{P}_1^2. \tag{21.181}$$

4. Процесс повторяется и получается приближение третьего порядка, как показано на блок-схеме рис. 21.37.

Вначале исследуем частный случай монохроматического света $\mathcal{E} = \mathrm{Re}\{E(\omega)\exp(j\omega t)\}$. В первом приближении $\mathcal{P}_1(t) = \mathrm{Re}\{P_1(\omega)\exp(j\omega t)\}$, где

$P_1(\omega) = \varepsilon_0\chi(\omega)E(\omega)$; $\chi(\omega)$ определяется (21.178). Во втором приближении линейная система находится под действием вынуждающей силы

$$\mathcal{E} - bP_1^2 = \mathrm{Re}\left\{E(\omega)e^{j\omega t}\right\} - b\left[\mathrm{Re}\left\{\varepsilon_0\chi(\omega)e^{j\omega t}\right\}\right]^2 =$$

$$= \mathrm{Re}\left\{E(\omega)e^{j\omega t}\right\} - \frac{1}{2}b\,\mathrm{Re}\left\{\left[\varepsilon_0\chi(\omega)E(\omega)\right]^2 e^{j2\omega t}\right\} - \frac{1}{2}b\left|\varepsilon_0\chi(\omega)E(\omega)\right|^2$$

Поскольку эти три члена имеют частоты ω, 2ω и 0, линейная система откликается с восприимчивостями $\chi(\omega)$, $\chi(2\omega)$ и $\chi(0)$ соответственно. Компонента $\mathcal{P}_2$ на частоте 2ω имеет амплитуду

$$P_2(2\omega) = \varepsilon_0\chi(2\omega)\left\{-\frac{1}{2}b\left[\varepsilon_0\chi(\omega)E(\omega)\right]^2\right\}.$$

Поскольку $P(2\omega) = d(2\omega;\ \omega,\ \omega)E(\omega)E(\omega)$, мы приходим к выводу, что

$$\boxed{d(2\omega;\ \omega,\ \omega) = -\frac{1}{2}b\varepsilon_0^3\left[\chi(\omega)\right]^2\chi(2\omega).} \qquad (21.182)$$

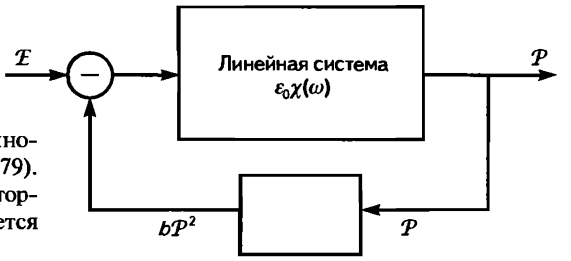


Рис. 21.37. Блок-схема решения нелинейного дифференциального уравнения (21.179). Линейная система, представленная операторным уравнением $\mathcal{L}\{\mathcal{P}\} = \mathcal{E}$, характеризуется передаточной функцией $\varepsilon_0\chi(\omega)$

Упражнение 21.14

Правило Миллера

Для нелинейной резонансной среды, описываемой уравнением (21.177) в случае, когда свет состоит из двух монохроматических волн с угловыми частотами ω_1 и ω_2, покажите, что приближение второго порядка, описываемое (21.180) и (21.181), дает компоненту плотности поляризации на частоте $\omega_3 = \omega_1 + \omega_2$ с амплитудой $P_2(\omega_3) = 2d(\omega_3;\ \omega_1,\ \omega_2)E(\omega_1)E(\omega_2)$, где

$$\boxed{d(\omega_3;\ \omega_1,\omega_2) = -\frac{1}{2}b\varepsilon_0^3\chi(\omega_1)\chi(\omega_2)\chi(\omega_3).} \qquad (21.183)$$

Правило Миллера

Соотношение (21.183) известно как *правило Миллера.*

Правило Миллера утверждает, что коэффициент нелинейности второго порядка для генерации волны с частотой $\omega_3 = \omega_1 + \omega_2$ из двух волн с частотами ω_1

и ω_2 пропорционален произведению линейных восприимчивостей на трех частотах $\chi(\omega_1)\chi(\omega_2)\chi(\omega_3)$. Все три частоты, следовательно, должны лежать в оптическом окне прозрачности среды (вне резонанса). Если эти частоты намного меньше, чем резонансная частота ω_0, то (21.78) дает $\chi(\omega) = \chi_0$ и тогда (21.183) дает

$$d(\omega_3;\ \omega_1, \omega_2) = -\frac{1}{2} b\varepsilon_0^2 \chi_0^3,$$

что не зависит от частоты. Тогда среда является приближенно недиспергирующей, и можно применять результаты предыдущих разделов, в которых дисперсия не учитывалась. Правило Миллера также показывает, что материалы с большими показателями преломления (большими χ_0) стремятся иметь большие d.

Анизотропные диспергирующие среды

Если одновременно рассматривать и анизотропные, и диспергирующие свойства среды, то трехволновое смешение в среде с нелинейностью второго порядка будет описываться более общим соотношением

$$\boxed{P_i(\omega_3) = 2\sum_{jk} d_{ijk}(\omega_3;\ \omega_1,\ \omega_2) E_j(\omega_1) E_k(\omega_2),} \qquad (21.184)$$

где $\omega_3 = \omega_1 + \omega_2$. Коэффициенты d_{ijk} теперь зависят от частот смешиваемых волн. Это соотношение аналогично соотношению

$$P_i(\omega) = \sum_j \chi_{ij}(\omega) E_j(\omega),$$

которое описывает линейные среды. Аналогично, четырехволновое смешение в среде с нелинейностью третьего порядка теперь описывается соотношением

$$\boxed{P_i(\omega_4) = 6\sum_{jkl} \chi^{(3)}_{ijkl}(\omega_4;\ \omega_1,\ \omega_2,\ \omega_3) E_j(\omega_1) E_k(\omega_2) E_l(\omega_3),} \qquad (21.185)$$

где $\omega_4 = \omega_1 + \omega_2 + \omega_3$.

Зависящие от частоты компоненты тензоров d_{ijk} и $\chi^{(3)}_{ijkl}$ подчиняются ряду соотношений внутренней симметрии, аналогичных соотношению

$$\chi^*_{ij}(\omega) = \chi_{ij}(-\omega)$$

в линейной оптике:

$$d^*_{ijk}(\omega_3;\ \omega_1,\ \omega_2) = d_{jki}(\omega_1;\ -\omega_2,\ \omega_3) = d_{kij}(\omega_2;\ \omega_3,\ -\omega_1); \qquad (21.186)$$

$$\chi^{(3)*}_{ijkl}(\omega_4;\ \omega_1,\ \omega_2,\ \omega_3) = \chi^{(3)}_{jkli}(\omega_1;\ -\omega_2,\ -\omega_3,\ \omega_4) = \ldots = \\ = \chi^{(3)}_{jikl}(\omega_3;\ \omega_4,\ -\omega_1,\ -\omega_2). \qquad (21.187)$$

В этих формулах коэффициент $d_{jki}(\omega_1; -\omega_2, \omega_3)$, например, представляет процесс преобразования частоты вниз, при котором волна частоты ω_2 с поляризацией k смешивается с волной на частоте ω_3 с поляризацией i и генерирует волну частоты $\omega_1 = \omega_3 - \omega_2$ с поляризацией j. Другие коэффициенты интерпретируются аналогично. Данный тип внутренней симметрии, конечно, дополняется другими соотношениями структурной симметрии, которым подчиняются различные классы кристаллов.

Рекомендуемая литература

книги

См. также списки литературы в гл. 5, 6, 15 и 20, а также книги по оптоэлектронике в гл. 17.

Agrawal G.P. Nonlinear Fiber Optics. Academic Press, 1991; 4th ed. 2006.

Menzel R. Photonics: Linear and Nonlinear Interactions of Laser Light and Matter. Springer-Verlag, 2001; 2nd ed. 2006.

Wegener M. Extreme Nonlinear Optics: An Introduction. Springer-Verlag, 2005.

Nikogosyan D.N. Nonlinear Optical Crystals: A Complete Survey. Springer-Verlag, 2005.

Banerjee P.P. Nonlinear Optics: Theory, Numerical Modeling, and Applications. Marcel Dekker, 2004.

Brignon A., Huignard J.-P., eds. Phase Conjugate Laser Optics. Wiley, 2004.

Boyd R.W. Nonlinear Optics. Academic Press, 1992; 2nd ed. 2003.

Suhara T., Fujimura M. Waveguide Nonlinear-Optic Devices. Springer-Verlag, 2003.

Sutherland R.L. Handbook of Nonlinear Optics. Marcel Dekker, 2nd ed. 2003.

Risk W.P., Gosnell T.R., Nurmikko A.V. Compact Blue-Green Lasers. Cambridge University Press, 2003.

Shen Y.R. The Principles of Nonlinear Optics. Wiley, 1984; paperback ed. 2002.

Miyata S., Sasabe H., eds. Light Wave Manipulation Using Organic Nonlinear Optical Materials. CRC Press, 2000.

Robieux J. High Power Laser Interactions. Lavoisier, 2000.

He G.S., Liu S.H. Physics of Nonlinear Optics. World Scientific, 1999.

Maimistov A.I., Basharov A.M. Nonlinear Optical Waves. Springer-Verlag, 1999.

Dmitriev V.G., Gurzadyan G.G., Nikogosyan D.N. Handbook of Nonlinear Optical Crystals. Springer-Verlag, 1991; 3rd ed. 1999.

Mills D.L. Nonlinear Optics: Basic Concepts. Springer-Verlag, 1991; 2nd ed. 1998.

Kajzar F., Reinisch R., eds. Beam Shaping and Control with Nonlinear Optics. Plenum Press, 1998.

Nalwa H.S., Miyata S., eds. Nonlinear Optics of Organic Molecules and Polymers. CRC Press, 1997.

Bloembergen N. Nonlinear Optics. World Scientific, 1965; 4th ed., 1996.

Sauter E.G. Nonlinear Optics. Wiley, 1996.

Tang C.L., Cheng L.K. Fundamentals of Optical Parametric Processes and Oscillators. Harwood, 1995.

Zhang J.-Y., Huang J.Y., Shen Y.R. Optical Parametric Generation and Amplification. Harwood, 1995.

Zyss J. Molecular Nonlinear Optics: Materials, Physics, and Devices. Academic Press, 1994.

Hopf F.A., Stegeman G.I. Applied Classical Electrodynamics. Vol. 2: Nonlinear Optics. Wiley, 1986; reprinted 1992.

Prasad P.N., Williams D.J. Introduction to Nonlinear Optical Effects in Molecules and Polymers. Wiley, 1991.

Butcher P.N., Cotter D. The Elements of Nonlinear Optics. Cambridge University Press, 1990; paperback ed. 1991.

Butylkin V.S., Kaplan A.E., Khronopulo Yu.G., Yakubovich E.I. Resonant Nonlinear Interactions of Light with Matter. Springer-Verlag, 1989.

Schubert M., Wilhelmi B. Nonlinear Optics and Quantum Electronics. Wiley, 1986.

Zel'dovich B.Y., Pilipetsky N.F., Shkunov V.V. Principles of Phase Conjugation. Springer-Verlag, 1985.

Fisher R.A., ed. Optical Phase Conjugation. Academic Press, 1983.

Rabin H., Tang C.L. Quantum Electronics. Academic Press, 1975.

Kaminow I.P. An Introduction to Electrooptic Devices. Academic Press, 1974.

СТАТЬИ

Issue on nonlinear optics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 3, 2006.

Gmachl C., Malis O., Belyanin A. Optical Nonlinearities in Intersubband Transitions and Quantum Cascade Lasers. In Intersubband Transitions in Quantum Structures. R. Paiella, ed. McGraw-Hill. 2006. P. 181—235.

Saltiel S.M., Sukhorukov A.A., Kivshar Y.S. Multistep Parametric Processes in Nonlinear Optics. In Progress in Optics. E. Wolf, ed. Elsevier. Vol. 47, 2005. P. 1—73.

Issue on nonlinear optics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 10, № 5, 2004.

Issue on nonlinear optics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, № 3, 2002.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Gosnell T.R., ed. Selected Papers on Upconversion Lasers. SPIE Optical Engineering Press (Milestone Series. Vol. 161), 2000.

Byer R.L. Quasi-Phasematched Nonlinear Interactions and Devices. Journal of Nonlinear Optical Physics and Materials. Vol. 6, 1997. P. 549—592.

Hunt J.H., ed. Selected Papers on Optical Parametric Oscillators and Amplifiers and Their Applications. SPIE Optical Engineering Press (Milestone Series. Vol. 140), 1997.

Roberts D.A. Simplified Characterization of Uniaxial and Biaxial Nonlinear Optical Crystals: A Plea for Standardization of Nomenclature and Conventions. IEEE Journal of Quantum Electronics. Vol. 8, 1992. P. 2057—2074.

Brandt H.E., ed. Selected Papers on Nonlinear Optics. SPIE Optical Engineering Press (Milestone Series. Vol. 32), 1991.

Khoo I.C. Nonlinear Optics of Liquid Crystals. In Progress in Optics. E. Wolf, ed. North-Holland. Vol. 26. 1988.

Pepper D.M. Applications of Optical Phase Conjugation. Scientific American. Vol. 254, № 1, 1986. P. 74—83.

Shkunov V.V., Zel'dovich B.Y. Optical Phase Conjugation. Scientific American. Vol. 253, № 6, 1985. P. 54—59.

Bloembergen N. Nonlinear Optics and Spectroscopy (Nobel lecture). Reviews of Modern Physics. Vol. 54, 1982. P. 685—695.

Mikaelian A.L. Self-Focusing Media with Variable Index of Refraction. In Progress in Optics. E. Wolf, ed. North-Holland. Vol. 17, 1980.

Brunner W., Paul H. Theory of Optical Parametric Amplification and Oscillation. In Progress in Optics. E. Wolf, ed. North-Holland. Vol. 15, 1977.

Hellwarth R.W. Third-Order Optical Susceptibilities of Liquids and Solids. Progress in Quantum Electronics. Vol. 5, 1977. P. 1—68.

Задачи

К РАЗДЕЛУ 21.2

1. **Энергообмен при преобразовании частоты вверх.** Кристалл $LiNbO_3$ с показателем преломления $n = 2{,}2$ используется для преобразования света с вакуумной длиной волны 1,3 мкм в свет с вакуумной длиной волны 0,5 мкм с помощью процесса трехволнового смешения. Все три волны представляют собой коллинеарные плоские волны, распространяющиеся вдоль оси z. Определите длину волны третьей компоненты (накачки). Если известно, что на длине волны 1,3 мкм при прохождении малого расстояния Δz мощность волны падает на 1 мВт, то каково усиление мощности для волны, частота которой преобразуется верх, и каковы потери или усиление накачки на том же расстоянии?

2. **Условия согласования для коллинеарной ГВГ типа II.** Определите угол θ для кристалла KDP, используемого для генерации второй гармоники типа II излучения с $\lambda = 1{,}06$ мкм в конфигурациях о-н-о и о-н-н. Используйте уравнения Зельмейера из табл. 5.1 для определения зависимости показателей преломления от длины волны.

3. **Фазовый синхронизм при вырожденном параметрическом преобразовании частоты вниз.** Для вырожденного параметрического преобразования частоты вниз используется кристалл KDP. Длина волны повышается с 0,6 до 1,2 мкм. Если обе волны коллинеарны, каковы должны быть направления их распространения относительно оптической оси кристалла и их поляризации, чтобы выполнялось условие фазового синхронизма? KDP — одноосный кристалл со следующими показателями преломления: при $\lambda_0 = 0{,}6$ мкм $n_o = 1{,}509$ и $n_н = 1{,}468$; при $\lambda_0 = 1$, 2 мкм $n_o = 1{,}490$ и $n_н = 1{,}459$.

4. **Условия согласования для трехволнового смешения в среде с дисперсией.** Показатель преломления нелинейной среды является функцией длины волны, аппроксимируемой выражением $n(\lambda_0) \approx n_0 - \xi\lambda_0$, где λ_0 — длина волны в свободном пространстве; n и ξ — постоянные. Покажите, что три волны с длинами волн λ_{01}, λ_{02} и λ_{03}, распространяющиеся в одном направлении, не могут эффективно взаимодействовать в среде с нелинейностью второго порядка. Будет ли эффективное взаимодействие возможно, если одна из волн распространяется в противоположном направлении?

5. **Допустимая фазовая расстройка.**

а. Уравнение Гельмгольца с источником $\nabla^2 E + k^2 E = -S$ имеет решение

$$E(\boldsymbol{r}) = \int_V S(\boldsymbol{r}') \frac{\exp\left(-jk_0 |\boldsymbol{r} - \boldsymbol{r}'|\right)}{4\pi |\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}\boldsymbol{r}',$$

где V — объем источника; $k_0 = 2\pi/\lambda_0$. Это уравнение можно использовать для определения поля, излученного в точке $\boldsymbol{r}$, если функция источника задана во всех точках $\boldsymbol{r}'$ в пределах объема источника. Если этот объем

занимает небольшую область с центром в начале координат $\boldsymbol{r} = \mathbf{0}$, а точка $\boldsymbol{r}$ находится достаточно далеко от источника, так что $r' \ll r$ для всех r' внутри источника, то

$$|\boldsymbol{r} - \boldsymbol{r}'| = \left(r^2 + r'^2 - 2\boldsymbol{r}\cdot\boldsymbol{r}'\right)^{1/2} \approx r\left(1 - \frac{\boldsymbol{r}\cdot\boldsymbol{r}'}{r^2}\right);$$

$$E(\boldsymbol{r}) \approx \frac{\exp(-ik_0 r)}{4\pi r}\int_V S(\boldsymbol{r}')\exp\left(jk_0\hat{\boldsymbol{r}}\cdot\boldsymbol{r}'\right)d\boldsymbol{r}',$$

где $\hat{\boldsymbol{r}}$ — единичный вектор в направлении $\boldsymbol{r}$. Предполагая, что объем V представляет собой куб со стороной L и источник является гармонической функцией $S(\boldsymbol{r}) = \exp(-j\boldsymbol{k}_s\ \boldsymbol{r})$, покажите, что при $L \gg \lambda_0$ интенсивность излученного света имеет максимум при $\boldsymbol{k}_0\hat{\boldsymbol{r}} = \boldsymbol{k}_s$ и резко падает при отклонении от этого условия. Таким образом, гармонический источник с размерами много больше длины волны излучает плоскую волну приблизительно с тем же волновым вектором.

б. Используйте соотношение из пункта (а) и первое борновское приближение для определения рассеянного поля, когда поле, падающее на среду с нелинейностью второго порядка, представляет собой сумму двух волн с волновыми векторами $\boldsymbol{k}_1$ и $\boldsymbol{k}_2$. Получите условие синхронизма $\boldsymbol{k}_3 = \boldsymbol{k}_1 + \boldsymbol{k}_2$ и определите наименьшую величину расстройки $\Delta\boldsymbol{k} = \boldsymbol{k}_3 - \boldsymbol{k}_1 - \boldsymbol{k}_2$, при которой рассеянное поле E исчезает.

6. **ГВГ назад при квазисинхронизме.** Покажите, что периодически поляризованный кристалл можно использовать для генерации волны второй гармоники, бегущей назад по отношению к основной волне. Запишите уравнение, определяющее согласование фаз в этой системе с квазисинхронизмом. Если это уравнение удовлетворяется для седьмой гармоники периодической структуры среды, определите отношение периода поляризации среды к длине волны основной гармоники света в среде.

К РАЗДЕЛУ 21.3

1. **Инварианты четырехволнового смешения.** Выведите уравнения, описывающие сохранение энергии и числа фотонов (соотношения Мэнли—Роу) для четырехволнового смешения.

2. **Мощность пространственного солитона.** Найдите выражение для проинтегрированной интенсивности пространственного солитона, описываемого (21.53), и покажите, что она обратно пропорциональна ширине пучка W_0.

3. **Оптооптический фазовый модулятор.** Спроектируйте систему для модуляции фазы оптического пучка с длиной волны 546 нм и шириной $W = 0{,}1$ мм с использованием ячейки Керра с CS_2 длиной 10 см. Модулятор управляется светом импульсного лазера с длиной волны 694 нм. CS_2 имеет показатель преломления $n = 1{,}6$ и коэффициент нелинейности третьего порядка $\chi^{(3)} = 4{,}4 \cdot 10^{-32}$ (Кл м)/В^3. Оцените оптическую мощность P_π управляющего света, необходимую для изменения фазы управляемого света на π.

4. **ГВГ в среде с нелинейностью третьего порядка с участием статического электрического поля.** Покажите, что ГВГ может происходить в среде с нелинейностью третьего порядка при наложении постоянного электрического поля. Какие физические параметры определяют эффективность такого процесса?

К РАЗДЕЛУ 21.4

1. **Усиление параметрического усилителя.** Параметрический усилитель на основе кристалла KDP длиной 4 см ($n \approx 1{,}49$, $d = 8{,}3 \cdot 10^{-24}$ Кл/В^2) предназначен для усиления света с длиной волны 550 нм. Длина волны накачки составляет 335 нм, а ее интенсивность 10^6 Вт/см^2. Предполагая, что сигнальная волна, волна накачки и холостая волна коллинеарны, определите коэффициент усиления среды и полное усиление прибора.

2. **Вырожденное параметрическое преобразование частоты вниз.** Напишите и решите связанные уравнения, описывающие взаимодействие волн при вырожденном параметрическом преобразовании частоты вниз при накачке на частоте $\omega_3 = 2\omega$ и сигналах на частоте $\omega_1 = \omega_2 = \omega$. Все волны распространяются вдоль оси z. Выведите выражения для плотностей потока фотонов на ω и 2ω и для коэффициента преобразования при длине взаимодействия L. Убедитесь в сохранении энергии и числа фотонов.

3. **Пороговая интенсивность накачки для параметрической генерации.** В параметрическом генераторе используется кристалл $LiNbO_3$ длиной 5 см с коэффициентом нелинейности второго порядка $d = 4 \cdot 10^{-23}$ Кл/В^2 и показателем преломления $n = 2{,}2$ (который предполагается почти одинаковым на всех рассматриваемых частотах). Накачка получается от Nd : YAG-лазера с длиной волны 1,06 мкм, частота которого удваивается с помощью генератора второй гармоники. Кристалл помещен в резонатор с одинаковыми зеркалами, имеющими коэффициент отражения 0,98. Фазовый синхронизм достигается при равенстве частот сигнальной и холостой волн. Определите минимальную интенсивность накачки, необходимую для параметрической генерации.

К РАЗДЕЛУ 21.5

1. **Комбинация генерации второй гармоники и суммарной частоты.** Две волны с угловыми частотами ω_1 и ω_2, их вторые гармоники $2\omega_1$ и $2\omega_2$ и волна суммарной частоты $\omega_1 + \omega_2$ одновременно взаимодействуют в среде с нелинейностью второго порядка. Предполагая, что условия фазового синхронизма выполняются для обоих процессов ГВГ и процесса генерации суммарной частоты, напишите связанные уравнения для данного процесса смешения пяти волн. Решите эти уравнения численно и покажите, что присутствие второй волны может подавить процесс ГВГ для первой.

2. **Уравнения для связанных волн при вырожденном четырехволновом смешении.** Рассмотрите коллинеарное четырехволновое взаимодействие в среде с нелинейностью третьего порядка в случае вырождения $\omega_4 = \omega_3$ и $\omega_1 + \omega_2 = 2\omega_3$. Выведите связанные волновые уравнения для амплитуд A_1, A_2 и A_3, считая условие фазового синхронизма точно выполненным.

К РАЗДЕЛУ 21.6

1. **Коллинеарное трехволновое смешение типа II в кристалле BBO.** Повторите анализ, выполненный в примере 21.6, чтобы показать, что эффективный коэффициент нелинейности $d_{эфф}$ для трехволнового о-н-н-смешения типа II в кристалле группы $3m$, таком как BBO, определяется формулой $d_{эфф} = d_{22}\cos^2\theta\cos 3\phi$.

2. **Связь между коэффициентами оптической нелинейности и электрооптическими коэффициентами.** Покажите, что электрооптические коэффициенты связаны с коэффициентами оптической нелинейности формулами

$$r_{ijk} = -\frac{4\varepsilon_0 d_{ijk}}{\varepsilon_{ii}\varepsilon_{jj}}; \quad s_{ijkl} = -\frac{12\varepsilon_0 \chi^{(3)}_{ijkl}}{\varepsilon_{ii}\varepsilon_{jj}}.$$

Эти формулы являются обобщениями (21.18) и (21.43) соответственно.

Указание. Если две матрицы **A** и **B** связаны соотношением $\mathbf{B} = \mathbf{A}^{-1}$, то приращения матриц $\Delta\mathbf{A}$ и $\Delta\mathbf{B}$ связаны как $\Delta\mathbf{B} = -\mathbf{A}^{-1}\Delta\mathbf{A}\mathbf{A}^{-1}$.

ГЛАВА
22 ОПТИКА СВЕРХБЫСТРЫХ ПРОЦЕССОВ





В 1980 г. **Л. Молленауэр**, **Р. Столен** и **Дж. Гордон** продемонстрировали распространение оптического солитона в стеклянном волокне.

Интерес к ультракоротким оптическим импульсам появился с изобретением лазера. С тех пор происходит продвижение в область все более и более коротких импульсов. Первые твердотельные и полупроводниковые лазеры были естественно импульсными, и разработка лазеров непрерывного действия потребовала значительных дополнительных усилий. За развитием техники наносекундных импульсов последовали пикосекундные импульсы, что затем привело к получению фемтосекундных и, уже в самое последнее время, аттосекундных импульсов. Однако последовательное продвижение в область более коротких импульсов становится все более сложной задачей. Этот прогресс сопровождается появлением множества важных приложений, включающих связь со сверхвысокими скоростями передачи данных и зондирование сверхбыстрых физических, химических и биологических явлений. В этих приложениях используются либо предельно малые длительности, либо сверхвысокие интенсивности (или напряженности поля) импульсов.

В применении к оптике термины *сверхбыстрый* и *сверхкороткий* обычно описывают импульсы от наносекундной до пикосекундной длительности или более короткие. В электронике, однако, эти термины относятся к импульсам длительностью от наносекунд до десятков пикосекунд, поскольку максимальный предел скорости в электронике намного ниже, чем в оптике. Наносекундный электри-

ческий импульс имеет спектральную ширину порядка гигагерца, и для его передачи требуется широкополосная микроволновая цепь. Пикосекундный импульс имеет спектральную ширину порядка терагерц, не поддерживаемую существующими электрическими или микроволновыми цепями. Если бы понадобилось создать фемтосекундный электрический импульс, он покрыл бы спектральную полосу в сотни терагерц, что равно всему частотному диапазону от нуля до края видимой области (≈0,3 мкм). В дополнение к этому, из-за соотношения неопределенностей $\Delta E \Delta t \geq h/2$ такой импульс имел бы неопределенность в энергии, превышающую 1,5 эВ, т. е. величину порядка ширины запрещенной зоны типичных полупроводников, что сделало бы обычную электронику ненадежной.

О данной главе

Ультракороткие оптические импульсы можно генерировать с помощью комбинации специально разработанных лазеров, использующих различные методы переключения или синхронизации мод (см. разд. 15.4), однако эти методы недостаточны для генерации фемтосекундных импульсов. Импульсы, генерируемые такими лазерами, требуют дальнейшего сжатия и формирования с помощью специальных методов, основанных на использовании линейных или нелинейных диспергирующих оптических элементов и систем, что и будет обсуждаться в данной главе.

Глава начинается с описания основных временны́х и пространственных характеристик оптических импульсов (см. разд. 22.1) и их фильтрации при помощи:

1) линейных диспергирующих элементов, таких как призмы и решетки (см. разд. 22.2);

2) пропускания через линейные диспергирующие среды, такие как оптические волокна (см. разд. 22.3).

Затем мы обратимся к пространственным эффектам и изучим оптику импульсных волн со сверхшироким спектром (см. разд. 22.4). Далее мы перейдем к нелинейной оптике импульсов (см. разд. 22.5). Некоторые из нелинейных оптических явлений, рассмотренных в гл. 19 в стационарном режиме, будут обобщены на случай импульсных волн. Сюда относятся параметрическое смешение волн, самомодуляция фазы и оптические солитоны (см. подразд. 22.5.2). Наконец, ряд методов детектирования ультракоротких оптических импульсов с помощью «медленных» приемников освещен в разд. 22.6.

22.1. ХАРАКТЕРИСТИКИ ИМПУЛЬСОВ

22.1.1. Временны́е и спектральные характеристики

Импульс света описывается оптическим полем конечной временно́й протяженности. В данной главе мы используем скалярную теория и будем представлять компоненты поля с помощью комплексной волновой функции $U(\boldsymbol{r}, t)$, нормированной таким образом, чтобы оптическая интенсивность была

равна $I(\boldsymbol{r}, t) = |U(\boldsymbol{r}, t)|^2$ (Вт/м^2). Когда мы интересуемся только временны́ми или спектральными свойствами импульса в фиксированной точке пространства $\boldsymbol{r}$, мы будем использовать просто функции $U(t)$ и $I(t)$.

Временны́е и спектральные представления

Комплексная волновая функция, описывающая оптический импульс с центральной частотой n_0, записывается в виде $U(t) = \mathcal{A}(t)\exp(j\omega_0 t)$, где $\mathcal{A}(t)$ — комплексная огибающая, а $\omega_0 = 2\pi\nu_0$ — центральная угловая частота. Комплексная огибающая характеризуется модулем $|\mathcal{A}(t)|$ и фазой $\varphi(t) = \arg\{\mathcal{A}(t)\}$, так что $U(t) = |\mathcal{A}(t)|\exp(j[\omega_0 t + \varphi(t)])$. Оптическая интенсивность равна $I(t) = |U(t)|^2 = |\mathcal{A}(t)|^2$ (Вт/м^2), а площадь под графиком интенсивности как функции времени $\int I(t)\mathrm{d}t$ представляет собой плотность энергии (Дж/м^2).

Профили интенсивности типичных импульсов включают гауссову функцию

$$I(t) \propto \exp\left(-\frac{2t^2}{\tau^2}\right)$$

(которая детально изучена в подразд. 22.1.2), функцию Лоренца

$$I(t) \propto \frac{1}{1 + t^2/\tau^2}$$

и гиперболический секанс

$$I(t) \propto \operatorname{sech}^2\left(\frac{t}{\tau}\right)$$

(который появляется в подразд. 22.5.2 в связи с оптическими солитонами). Ширина каждого из этих импульсов пропорциональная постоянной времени τ.

В спектральной области импульс описывается Фурье-образом

$$V(\nu) = \int U(t)\exp(-j2\pi\nu t)\mathrm{d}t,$$

который является комплексной функцией

$$V(\nu) = |V(\nu)|\exp[j\psi(\nu)].$$

Квадрат амплитуды $S(\nu) = |V(\nu)|^2$ называется **спектральной интенсивностью**, а $\psi(\nu)$ — **спектральной фазой**. Функция $V(\nu)$ имеет центральную частоту ν_0 и обращается в нуль при отрицательных ν, поскольку $U(t)$ — комплексный аналитический сигнал (см. подразд. 2.6.1). Фурье-образ комплексной огибающей

$$A(\nu) = \int \mathcal{A}(t)\exp(-j2\pi\nu t)\mathrm{d}t = V(\nu - \nu_0)$$

центрирован на частоте $\nu = 0$. Если импульс имеет малую спектральную ширину, то комплексная огибающая является медленно меняющейся функцией времени (т. е. меняется мало на протяжении оптического периода $1/\nu_0$), однако это не так для сверхузких импульсов со сверхшироким спектром. На рис. 22.1 показаны различные временны́е и спектральные функции, характеризующие оптический импульс.

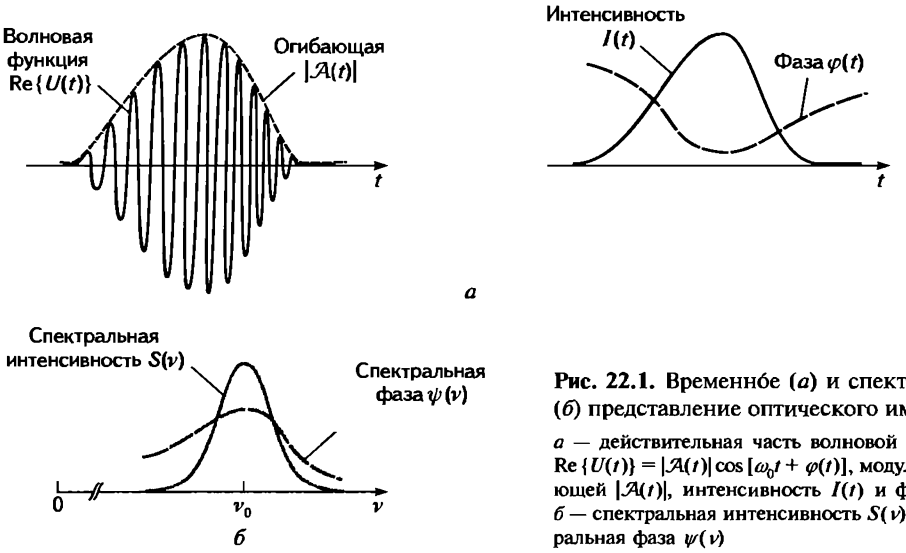


Рис. 22.1. Временно́е (*а*) и спектральное (*б*) представление оптического импульса:

а — действительная часть волновой функции $\mathrm{Re}\{U(t)\} = |\mathcal{A}(t)|\cos[\omega_0 t + \varphi(t)]$, модуль огибающей $|\mathcal{A}(t)|$, интенсивность $I(t)$ и фаза $\varphi(t)$; *б* — спектральная интенсивность $S(\nu)$ и спектральная фаза $\psi(\nu)$

Временна́я и спектральная ширина

Временна́я и пространственная ширина импульса — это ширина интенсивности $I(t) = |U(t)|^2$ и спектральной интенсивности $S(\nu) = |V(\nu)|^2$ соответственно, определяема любой из мер, приведенных в приложении А.2. По умолчанию будем использовать в качестве такой меры полную ширину на уровне половины максимума (ПШПМ) и обозначать временну́ю ширину $\tau_{\text{ПШПМ}}$, а спектральную $\Delta\nu$.

В силу того, что $U(t)$ и $V(\nu)$ связаны преобразованием Фурье, спектральная ширина обратно пропорциональна временно́й ширине. Коэффициент пропорциональности зависит от формы импульса и определения ширины. Это соотношение обратной пропорциональности иллюстрируется на рис. 22.2, *а* для импульса гауссовой формы, у которого $\tau_{\text{ПШПМ}}\Delta\nu = 0{,}44$.

Спектральная интенсивность $S(\nu)$ часто изображается графически как функция длины волны $S_\lambda(\lambda)$. Преобразование осуществляется с помощью соотношения

$$S_\lambda(\lambda) = S(\nu)\left|\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right| = \frac{c}{\lambda^2}S\left(\frac{c}{\lambda}\right).$$

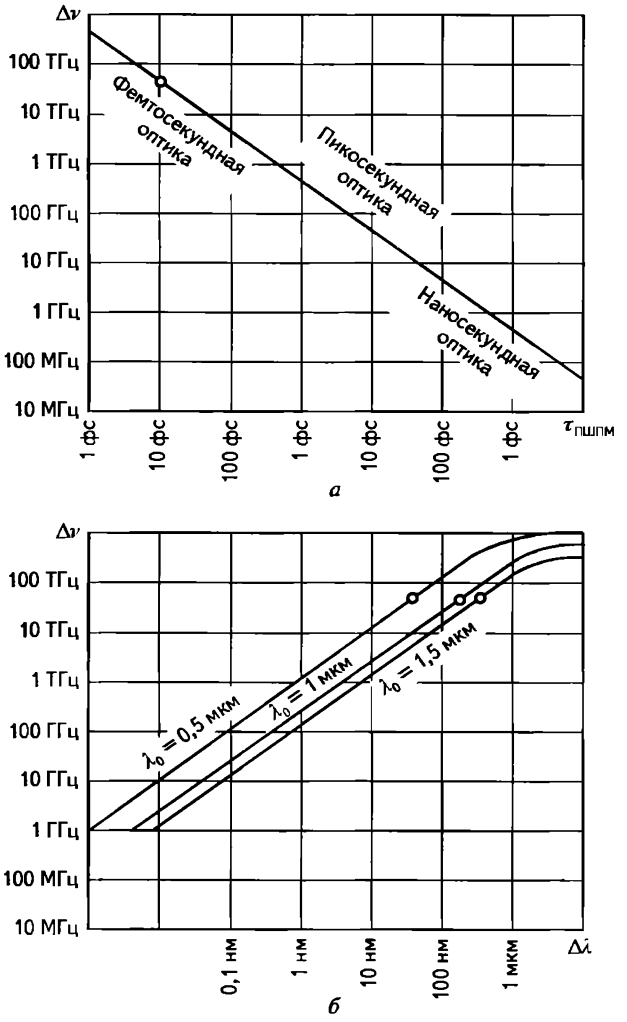


Рис. 22.2. Соотношение $\Delta\nu = 0{,}44/\tau_{\text{пшпм}}$ между спектральной шириной $\Delta\nu$ и временной шириной $\Delta\tau_{\text{пшпм}}$ для гауссова импульса (*а*). Соответствующая ширина $\Delta\lambda$ для импульса с центральной частотой ν_0, которой соответствует центральная длина волны $\lambda_0 = c/\nu_0 = 0{,}5$; 1 и 1,5 мкм (*б*). Например, 10-фемтосекундный импульс имеет спектральную ширину $\Delta\nu = 44$ ТГц, что соответствует $\Delta\lambda = 37$, 147 и 331 нм, если центральная частота равна $\lambda_0 = 0{,}5$; 1 и 1,5 мкм соответственно (светлые кружки на графике). Это соотношение линейно, если $\Delta\nu \ll \nu_0$ [см. (22.1)]

Спектральная ширина также может быть преобразована в шкалу длин волн. Если $\Delta\nu \ll \nu_0$, то спектральная ширина в единицах длины волны приближенно равна

$$\Delta\lambda \approx \left|\frac{d\lambda}{d\nu}\right| \Delta\nu$$

или

$$\boxed{\Delta\lambda \approx \frac{\lambda_0^2}{c}\Delta\nu,} \qquad (22.1)$$

Спектральная ширина

где $\lambda_0 = c/\nu_0$ — длина волны, соответствующая центральной частоте. Если $\Delta\nu$ выражено в терагерцах, λ_0 — в микрометрах, а $\Delta\lambda$ в нанометрах, то

$$\Delta\lambda \approx 3{,}3\lambda_0^2\Delta\nu \qquad \Delta\lambda\ [\text{нм}];\quad \lambda_0\ [\text{мкм}];\quad \Delta\nu\ [\text{ТГц}]. \qquad (22.2)$$

Например, спектральная ширина $\Delta\nu = 1$ ТГц соответствует $\Delta\lambda = 1$ нм при $\lambda_0 = 0{,}55$ мкм и 4 нм при $\lambda_0 = 1{,}1$ мкм. Это соотношение иллюстрируется на рис. 22.2, *б*.

Для сверхузких импульсов с большими $\Delta\nu$ точное выражение для $\Delta\lambda$ имеет вид

$$\Delta\lambda = \frac{c}{\nu_0 - \Delta\nu/2} - \frac{c}{\nu_0 + \Delta\nu/2} = \frac{\lambda_0^2}{c}\frac{\Delta\nu}{1 - (\Delta\nu/2\nu_0)^2}. \qquad (22.3)$$

Однако при этих условиях понятие спектральной ширины утрачивает свое значение. Импульс длительностью 2 фс имеет спектральную ширину $\Delta\nu = 220$ ТГц, что соответствует $\Delta\lambda = 847$ нм при $\lambda_0 = 1$ мкм, т. е. спектр весьма широк и охватывает видимый и инфракрасный диапазоны.

Мгновенная частота

Другим параметром для описания оптического импульса является временна́я зависимость его мгновенной частоты. Мгновенная угловая частота ω_i — это производная от фазы функции $U(t)$ по времени. Ей соответствует линейная частота $\nu_i = \omega_i/2\pi$, так что

$$\boxed{\omega_i = \omega_0 + \frac{d\varphi}{dt};\quad \nu_i = \nu_0 + \frac{1}{2\pi}\frac{d\varphi}{dt}.} \qquad (22.4)$$

Мгновенная частота

Если фаза является линейной функцией времени, $\varphi(t) = 2\pi ft$, то мгновенная частота $\nu_i = \nu_0 + f$, т. е. линейное изменение фазы соответствует постоянному сдвигу частоты. Нелинейная зависимость фазы от времени соответствует зависящей от времени мгновенной частоте.

Импульсы с чирпом

Говорят, что импульс имеет **чирп**, или частотную модуляцию (ЧМ), если его мгновенная частота меняется во времени. Если ν_i — возрастающая функция времени в центре импульса ($t = 0$), т. е.

$$\varphi'' = \frac{d^2\varphi}{dt^2} > 0,$$

то чирп называется **положительным**. Если ν_i — убывающая функция времени в центре импульса, т. е. $\varphi'' < 0$, то чирп называется **отрицательным**.

В частности, если фаза оптического импульса шириной τ является квадратичной функцией времени $\varphi(t) = at^2/\tau^2$, где a — постоянная, то $\varphi'' = 2a/\tau^2$, и мгновенная частота $\nu_i = \nu_0 + (a/\pi\tau^2)t$ является линейной функцией времени. В этом случае говорят, что импульс обладает **линейным чирпом**, а параметр

$$a = \frac{1}{2}\varphi''\tau^2 \qquad (22.5)$$

Параметр чирпа

называется **параметром чирпа**. Чирп импульса положительный, если $a > 0$, и отрицательный, если $a < 0$. При $t = \tau/2$ мгновенная частота возрастает на величину $a/2\pi\tau$ порядка $a\Delta\nu$. Таким образом, параметр чирпа отражает отношение между изменением мгновенной частоты в точке, соответствующей половине ширины импульса, и спектральной шириной импульса $\Delta\nu$. Примеры импульсов с линейным чирпом и их мгновенные частоты показаны на рис. 22.3.

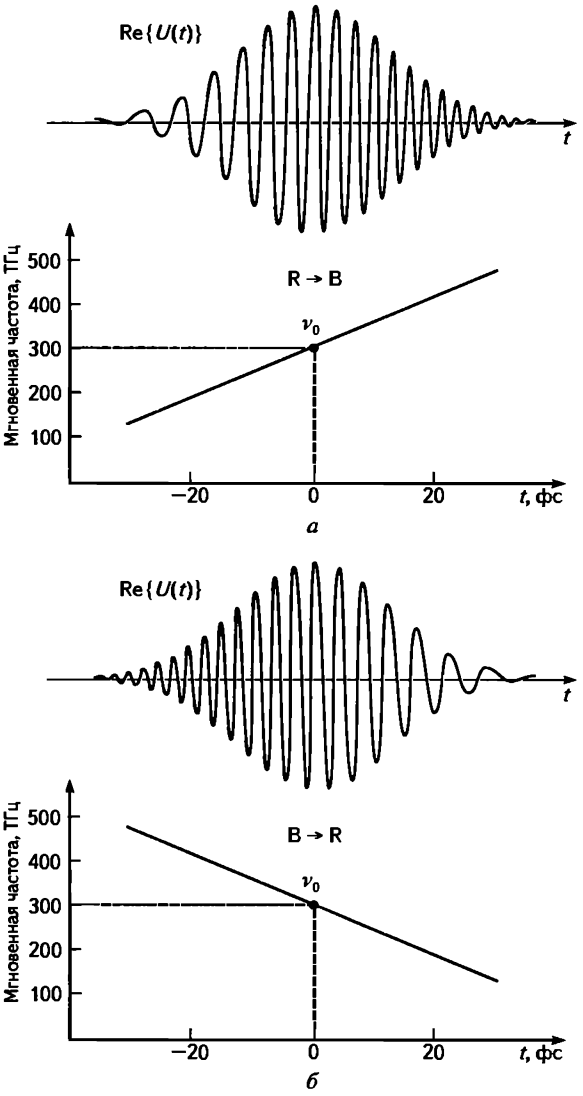


Рис. 22.3. Оптические импульсы с линейным положительным и отрицательным чирпом:

а — импульс с положительным чирпом имеет возрастающую мгновенную частоту; *б* — импульс с отрицательным чирпом имеет убывающую мгновенную частоту. На этом рисунке ширина импульса составляет 20 фс, а центральная частота ν_0 = 300 ТГц. Буквы R и B (red — красный и blue — синий) указывают области длинных и коротких волн соответственно

Если зависимость фазы φ от времени — произвольная нелинейная функция, как на рис. 22.1, то ее можно аппроксимировать разложением в ряд Тейлора в окрестности центра импульса, и тогда коэффициент чирпа a, определенный формулой (22.5), представляет чирп-эффект наинизшего порядка, обусловленный квадратичным членом разложения.

Спектр, зависящий от времени

Часто полезно проследить за изменениями спектра переменного импульса во времени. Такие изменения не отражаются преобразованием Фурье, которое обеспечивает усредненное спектральное представление всего сигнала в целом и не позволяет судить о том, какие частоты появляются в какое время. Это особенно очевидно, если сигнал составлен из последовательности сегментов, каждый из которых имеет свой особый спектральный состав.

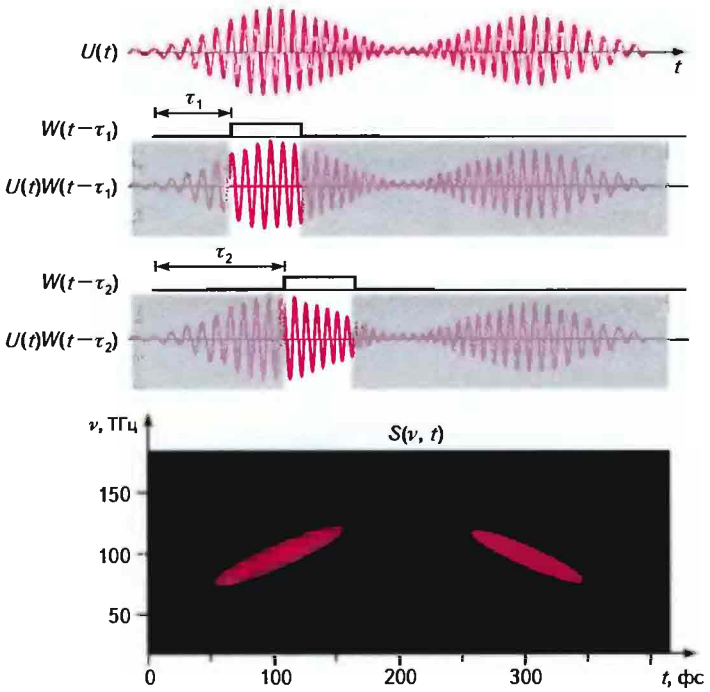


Рис. 22.4. Кратковременное преобразование Фурье функции $U(t)$ строится в виде последовательности преобразований Фурье этой функции, умноженной на функцию подвижного окна $W(t-\tau)$. Спектрограмма $S(\nu, t)$ — квадрат модуля этих Фурье-образов. В данном примере $U(t)$ состоит из двух гауссовых импульсов, каждый из которых имеет постоянную времени $\tau = 60$ фс и центральную частоту 100 ТГц. Первый импульс имеет положительный чирп ($a = 5$), а второй — отрицательный чирп ($a = -5$) и меньшую амплитуду. Функция окна $W(t)$ — гауссова с постоянной времени $\tau = 20$ фс

Хорошим примером является музыкальный сигнал, в котором спектральные изменения отражают процесс исполнения партитуры во времени (т. е. и представляют собой музыку, как таковую — *примеч. пер.*).

Хотя мгновенная частота может быть мерой переменной природы спектра, она не всегда адекватно ее отражает, поскольку основывается только на фазе и игнорирует амплитуду. Часто используется мера, основанная на использовании скользящего окна, которое выделяет короткий отрезок времени в каждый момент и в котором производится преобразование Фурье. Эта процедура повторяется при различных положения скользящего окна, как показано на рис. 22.4, а результат изображается графически как функция двух переменных: частоты и временно́й задержки. Такое преобразование называется **кратковременным преобразованием Фурье**, а получающаяся функция двух переменных — **кратковременным Фурье-образом**. Квадрат ее модуля называется **спектрограммой** и часто строится в виде картины с горизонтальными и вертикальными осями, представляющими время и частоту соответственно, как показано на рис. 22.4.

Если $W(t)$ — оконная функция с короткой длительностью T и началом при $t = 0$, а $U(t)$ — волновая функция импульса, то произведение $U(t)W(t-\tau)$ — сегмент импульса с продолжительностью T и началом в момент τ. Фурье-образ этого сегмента

$$\Phi(\nu, t) = \int U(t)W(t-\tau)\exp(-j2\pi\nu t)\,\mathrm{d}t. \qquad (22.6)$$

Кратковременное преобразование Фурье

Функция $\Phi(\nu, t)$ — кратковременный Фурье-образ, а ее квадрат модуля $S(\nu, t) = |\Phi(\nu, t)|^2$ — спектрограмма.

22.1.2. Гауссовы импульсы и гауссовы импульсы с чирпом

Спектрально ограниченные гауссовы импульсы

Спектрально ограниченный гауссов импульс имеет комплексную огибающую с постоянной фазой и гауссовым модулем

$$\mathcal{A}(t) = A_0 \exp\left(-\frac{t^2}{\tau^2}\right), \qquad (22.7)$$

где τ — действительная постоянная. Интенсивность

$$I(t) = I_0 \exp\left(-\frac{2t^2}{\tau^2}\right)$$

— также гауссова функция с максимальным значением $I_0 = |A_0|^2$, полной шириной на уровне $1/e\sqrt{2}\,\tau$ и длительностью на уровне половины максимума:

$$\tau_{\text{ПШПМ}} = \sqrt{2\ln 2}\tau = 1{,}18\tau. \qquad (22.8)$$

Преобразование Фурье комплексной огибающей $A(\nu) \propto \exp(-\pi^2\tau^2\nu^2)$ — также гауссова функция; таковой же является и спектральная интенсивность

$$S(\nu) \propto \exp\left[-2\pi^2\tau^2(\nu-\nu_0)^2\right]. \qquad (22.9)$$

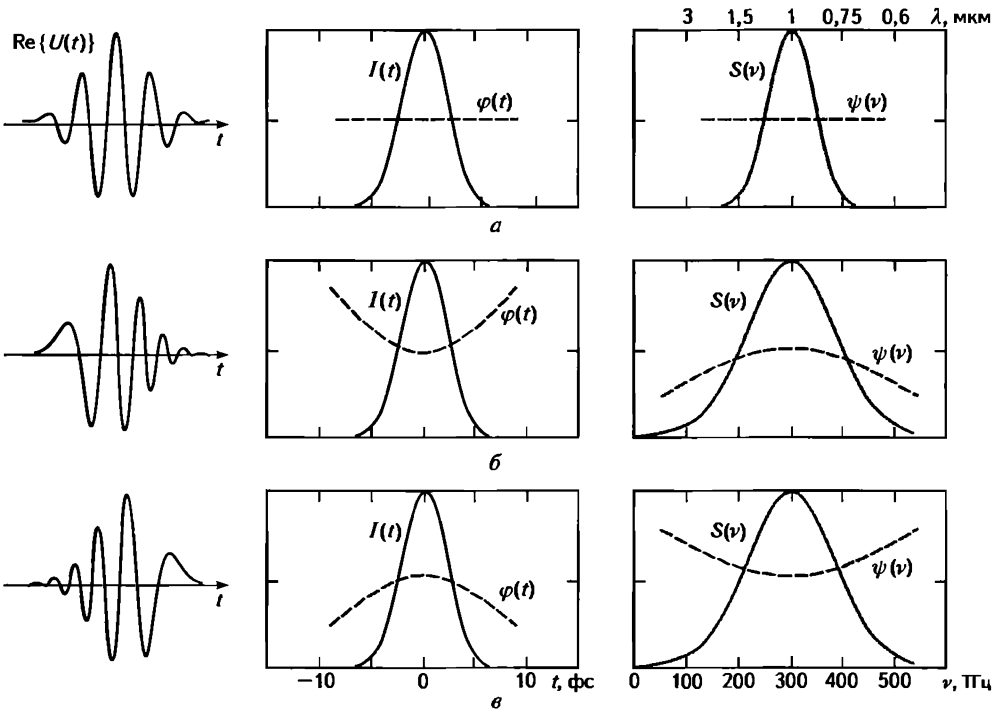


Рис. 22.5. Временные и спектральные профили трех гауссовых импульсов с центральной частотой $\nu_0 = 300$ ТГц (что соответствует длине волны 1 мкм и оптическому периоду 3,3 фс) и шириной $\tau_{\text{ПШПМ}} = 5$ фс ($\tau = 4$, 23 фс):

а — спектрально ограниченный импульс; спектральная ширина $\Delta\nu = 88$ ТГц ($\Delta\lambda = 73$ нм); *б* — импульс с положительным чирпом, $a = 2$; спектральная ширина в $\sqrt{1+a^2} = \sqrt{5}$ больше, чем в случае (*а*), так что $\Delta\nu = 197$ ТГц. Мгновенная частота — линейно возрастающая функция времени, принимающая значение $\nu_0 = 300$ ТГц при $t = 0$ (центр импульса) и значения $\nu_t = \nu_0(1 \pm at/\pi\nu_0\tau) = 300(1 \pm 0{,}497)$ ТГц при $t = \pm\tau$. Частота изменяется в интервале от 151 до 449 ТГц по мере изменения времени от $-\tau$ до $+\tau$. Это соответствует изменению длины волны от 0,67 до 1,99 мкм; *в* — то же, что в случае (*б*), но для импульса с отрицательным параметром чирпа $a = -2$

Полная ширина на уровне половины максимума для спектральной интенсивности составляет

$$\Delta\nu = \frac{0{,}375}{\tau} = \frac{0{,}44}{\tau_{\text{ПШПМ}}}, \qquad (22.10)$$

так что произведение временно́й и спектральной ширин на уровне половины максимума равно $\tau_{\text{пшпм}}\Delta\nu = 0{,}44$. На рис. 22.5, *а* показаны временны́е и спектральные характеристики спектрально ограниченного гауссова импульса.

Как обсуждается в приложении А.2, спектрально ограниченный гауссов импульс имеет минимальное произведение временно́й и спектральной ширин, откуда и происходит термин «спектрально ограниченный» (его также называют **ограниченным по преобразованию Фурье**).

Хотя гауссов импульс имеет идеальную форму, которая в точном виде не встречается на практике, он является полезным приближением, удобным для аналитического исследования.

Гауссов импульс с чирпом

Гауссов импульс более общего вида имеет комплексную огибающую $\mathcal{A}(t) = A_0 \exp(-\alpha t^2)$, где $\alpha = (1 - ja)/\tau^2$ — комплексный параметр; τ и a — действительные параметры, так что

$$\mathcal{A}(t) = A_0 \exp\left(-\frac{t^2}{\tau^2}\right)\exp\left(\frac{jat^2}{\tau^2}\right). \qquad (22.11)$$

Модуль комплексной огибающей является гауссовой функцией $|A_0|\exp(-t^2/\tau^2)$ и интенсивность тоже гауссова. Фаза является квадратичной функцией $\varphi = at^2/\tau^2$, так что мгновенная частота $\nu_t = \nu_0 + at/\pi\tau^2$ — линейная функция времени, т. е. импульс имеет линейный чирп с параметром чирпа a. Этот чирп положительный (отрицательный) при положительном (отрицательном) a. При $a = 0$ импульс спектрально ограниченный (не имеет чирпа). Фурье-образ комплексной огибающей $\mathcal{A}(t) = A_0 \exp(-\alpha t^2)$ пропорционален $\exp(-\pi^2\tau^2\nu^2/\alpha)$, что тоже является гауссовой функцией частоты. Спектральная интенсивность $S(\nu)$ пропорциональна $\exp[-2\pi^2\tau^2(\nu - \nu_0)^2/(1 + a^2)]$, что представляет собой гауссову функцию с шириной на уровне половины максимума

$$\Delta\nu = \frac{0{,}375}{\tau}\sqrt{1 + a^2} = \frac{0{,}44}{\tau_{\text{пшпм}}}\sqrt{1 + a^2}$$

Это в $\sqrt{1 + a^2}$ раз больше, чем для импульса без чирпа ($a = 0$) с такой же постоянной времени τ. Произведение временно́й и спектральной ширин на уровне половины максимума равно

$$\tau_{\text{пшпм}}\Delta\nu = 0{,}44\sqrt{1 + a^2},$$

так что гауссов импульс без чирпа ($a = 0$) имеет наименьшее произведение спектральной и временно́й ширин. Спектральная фаза $\psi(\nu) \propto a\nu^2$ — квадратичная функция частоты.

Основные формулы, характеризующие гауссов импульс с чирпом, сведены в табл. 22.1. На рис. 22.5 показаны временны́е и спектральные характеристики гауссовых импульсов с чирпом и без него.

Таблица 22.1. **Временные и спектральные свойства гауссова импульса с чирпом, имеющего максимальную амплитуду A_0, максимальную интенсивность $I_0 = |A_0|^2$, центральную частоту ν_0, постоянную времени τ и параметр чирпа a**

$\mathcal{A}(t) = A_0 \exp\left[-(1-ja)\frac{t^2}{\tau^2}\right]$	Комплексная огибающая	(22.12)
$I(t) = I_0 \exp\left(-\frac{2t^2}{\tau^2}\right)$	Интенсивность	(22.13)
$\int I(t)\mathrm{d}t = \sqrt{\frac{\pi}{2}} I_0 \tau$	Плотность энергии	(22.14)
$\tau_{1/e} = \sqrt{2}\tau$	Полуширина на уровне $1/e$	(22.15)
$\tau_{\text{пшпм}} = 1{,}18\tau$	Полуширина на уровне половины максимума	(22.16)
$\varphi(t) = \frac{at^2}{\tau^2}$	Фаза	(22.17)
$A(\nu) = \frac{A_0\tau}{2\sqrt{\pi(1-ja)}} \exp\left[-\frac{\pi^2\tau^2\nu^2}{1-ja}\right]$	Фурье-образ	(22.18)
$S(\nu) = \frac{I_0\tau^2}{4\pi\sqrt{1+a^2}} \exp\left[-\frac{2\pi^2\tau^2(\nu-\nu_0)^2}{1+a^2}\right]$	Спектральная интенсивность	(22.19)
$\Delta\nu_{1/e} = \frac{2}{\tau}\sqrt{1+a^2}$	Полуширина на уровне $1/e$	(22.20)
$\Delta\nu = \frac{0{,}375}{\tau}\sqrt{1+a^2} = \frac{0{,}44}{\tau_{\text{пшпм}}}\sqrt{1+a^2}$	Спектральная ширина на уровне половины максимума	(22.21)
$\psi(\nu) = -2\pi^2\tau^2\left[\frac{a}{1+a^2}\right]\nu^2$	Спектральная фаза	(22.22)
$\nu_i = \nu_0 + \left(\frac{a}{\pi\tau^2}\right)t$	Мгновенная частота	(22.23)

22.1.3. Пространственные характеристики

В этом разделе мы изучим несколько простых примеров импульсных оптических волн, распространяющихся в свободном пространстве или линейной, однородной среде без дисперсии. В таких средах волновая функция удовлетворяет волновому уравнению

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

Простейшими точными решениями этого уравнения являются импульсная плоская и импульсная сферическая волны. Мы обсудим эти решения, а также введем импульсный гауссов пучок. Более детальное исследование пространственных свойств импульсного света отложим до разд. 22.4.

Импульсная плоская волна

Импульсная плоская волна, распространяющаяся в направлении z, имеет комплексную волновую функцию вида

$$U(r,\ t) = \mathcal{A}\left(t - \frac{z}{c}\right)\exp\left[j\omega_0\left(t - \frac{z}{c}\right)\right],$$

где $\mathcal{A}(t)$ — произвольная функция. Соответствующая интенсивность есть $I(t - z/c)$, где $I(t) = |\mathcal{A}(t)|^2$. Если ширина $I(t)$ равна τ, то бегущий импульс занимает расстояние $\Delta z = c\tau$ в любой момент времени и распространяется без изменения скорости c, как показано на рис. 22.6. Численные значения временно́й и пространственной ширин в свободном пространстве таковы:

Временна́я ширина τ	1 нс	1 пс	1 фс	1 ас
Пространственная ширина $c\tau$	30 см	0,3 мм	0,3 мкм	0,3 нм

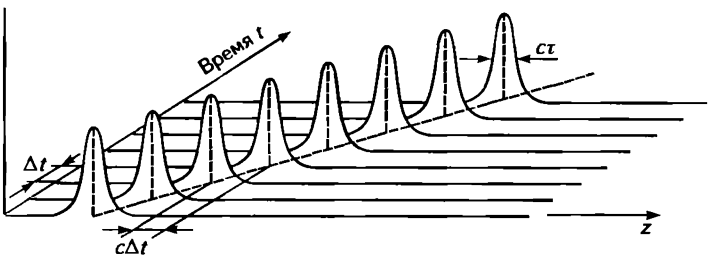


Рис. 22.6. Огибающая импульса плоской волны длительностью τ, распространяющегося в направлении z со скоростью c. Импульс занимает область длиной $c\tau$ в любой момент времени

Импульсная плоская волна, распространяющаяся под углом θ к оси z, имеет комплексную волновую функцию

$$U\ (r,\ t) = \mathcal{A}\left(t - \frac{x\sin\theta + z\cos\theta}{c}\right)\exp\left[-jk_0\ (x\sin\theta + z\cos\theta)\right]\exp\left(j\omega_0 t\right)$$

и интенсивность $I[t - (x\sin\theta + z\cos\theta)/c]$, где $I(t) = |\mathcal{A}(t)|^2$. Если эта интенсивность записывается как функция x и z в виде последовательности мгновенных снимков (каждый в фиксированный момент времени), то результат получается таким, как показано на рис. 22.7, *а*. Светлая полоса на каждом снимке представляет движущийся импульс в данный момент времени. Например, импульс

длительностью 100 фс в свободном пространстве виден как полоса шириной 30 мкм. Заметим, что одна вертикальная линия (z = const), пересекающая полосу на одном снимке, обеспечивает полную запись временно́го профиля импульса, поскольку на ней записывается функция $I[-x \sin \theta/c + \text{const}]$. Таким образом, временно́й профиль можно измерить по пространственному профилю на снимке импульса. Это можно использовать для детектирования импульсов, как будет обсуждаться в подразд. 22.5.2.

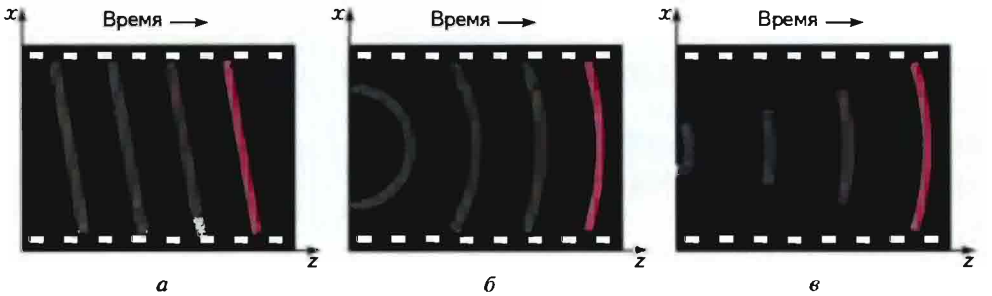


Рис. 22.7. Четыре мгновенных снимка (через равные промежутки времени) импульсной плоской волны, распространяющейся под углом (*а*). Каждый снимок представляет собой одну полосу толщиной $c\tau$ в направлении z, где τ — длительность импульса. Полоса движется слева направо по мере распространения импульса. То же, что в случае (*а*), но для сферической волны (*б*). То же, что в случае (*а*), но для гауссова пучка (*в*)

Импульсная сферическая волна

Другим простым решением волнового уравнения является импульсная сферическая волна

$$U(r,\ t) = \frac{1}{r} g\left(t - \frac{r}{c}\right) \exp\left[j\omega_0\left(t - \frac{r}{c}\right)\right],$$

где $g(t)$ — произвольная функция. Импульс распространяется в радиальных направлениях, а его волновые фронты — концентрические сферы, как показано на рис. 22.7, *б*. В любой момент времени он занимает сферический слой с радиальной толщиной $c\tau$, где τ — ширина функции $g(t)$.

**Параксиальная волна, модулированная медленно меняющимся импульсом*

Когда огибающая импульсной волны медленно меняется со временем, так что она почти постоянна на протяжении оптического периода, говорят, что волна имеет **медленную огибающую**. Из-за малой спектральной ширины $\Delta\nu = \nu_0$ пространственное поведение примерно такое же, как у монохроматической (стационарной) волны на центральной частоте ν_0 или длине

волны $\lambda_0 = c/\nu_0$. Следовательно, волну можно рассматривать как **квазистационарную** импульсную волну.

Если волна является также параксиальной (см. подразд. 2.2.3), то ее волновую функцию можно выразить через огибающую общей формулой

$$U(r, t) = \mathcal{A}(r, t) \exp(jk_0 z) \exp(j\omega_0 t),$$

где огибающая медленно меняется в зависимости от z так, что она почти постоянна на расстояниях порядка длины волны $\lambda_0 = 2\pi/k_0$, т. е. выполняется условие

$$\frac{\partial^2 \mathcal{A}}{\partial z^2} \ll k_0^2 \mathcal{A}.$$

Поскольку огибающая меняется также во времени, справедливо приближение

$$\frac{\partial^2 \mathcal{A}}{\partial t^2} \ll \omega_0^2 \mathcal{A}.$$

При таких условиях волновое уравнение

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

приводит к приближенному уравнению для огибающей

$$\boxed{\nabla_T^2 \mathcal{A} - j\frac{4\pi}{\lambda_0}\left(\frac{\partial \mathcal{A}}{\partial z} + \frac{1}{c}\frac{\partial \mathcal{A}}{\partial t}\right) = 0,} \qquad (22.24)$$

Параксиальное уравнение

где ∇_T^2 — поперечный оператор Лапласа,

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

Уравнение (22.24) называется параксиальным уравнением для медленной огибающей (МО). Для стационарной волны $\partial\mathcal{A}/\partial t = 0$ и (22.24) переходит в параксиальное уравнение Гельмгольца (2.27).

Как видно из прямой подстановки, (22.24) удовлетворяется функцией

$$\mathcal{A}(\rho, z, t) = g\left(t - \frac{z}{c}\right)\mathcal{A}_0(r),$$

где g — произвольная функция запаздывающего времени $t - z/c$; $\mathcal{A}_0(r)$ удовлетворяет параксиальному уравнению Гельмгольца

$$\nabla_T^2 \mathcal{A}_0 - j\frac{4\pi}{\lambda_0}\frac{\partial \mathcal{A}_0}{\partial z} = 0,$$

которое справедливо в стационарном случае. Следовательно, в данном приближении параксиальная волна с длиной волны λ_0 может модулироваться медленно меняющимся импульсом произвольной формы без изменения своего пространственного поведения.

Импульсный гауссов пучок

Одним из решений параксиального уравнения Гельмгольца является гауссов пучок, описываемый формулой (3.5). В случае квазистационарного импульса гауссов пучок задается выражением

$$\mathcal{A}(\rho, z, t) = g\left(t - \frac{z}{c}\right)\frac{jz_0}{z + jz_0}\exp\left(-j\frac{\pi}{\lambda_0}\frac{\rho^2}{z + jz_0}\right), \qquad (22.25)$$

где $g(t)$ — произвольная медленно меняющаяся функция запаздывающего времени $t - z/c$; z_0 — рэлеевская длина (также называемая дифракционной длиной). В этом приближении, за исключением запаздывания, отсутствует связь между координатами и временем, т. е. пучок сохраняет свой гауссов пространственный профиль все время и во всех точках пространства. Мгновенные снимки такого пучка показаны на рис. 22.7, *в*.

В разд. 22.4 будет показано, что для ультракоротких импульсов, когда приближение МО неприменимо, пространственно-временна́я связь может быть существенной, и волна, являющаяся гауссовой во времени и пространстве в данной поперечной плоскости, становится негауссовой как во времени, так и в пространстве, по мере ее свободного распространения.

22.2. ФОРМИРОВАНИЕ И КОМПРЕССИЯ ИМПУЛЬСОВ

Временно́й профиль короткого оптического импульса неизбежно меняется при распространении в оптической системе с дисперсией. Это происходит потому, что спектральные компоненты, составляющие импульс, ослабляются и/или сдвигаются по фазе по-разному. Влияние дисперсии более выражено для ультракоротких импульсов, поскольку они имеют большую спектральную ширину. Диспергирующие оптические элементы могут предназначаться также для целенаправленного изменения формы импульса, например его сжатия или растяжения.

В данном разделе мы рассмотрим только временны́е эффекты, т. е. только импульсные плоские волны. В разд. 22.4 будут рассмотрены пространственные эффекты в линейных оптических средах, включая дифракцию и распространение пучков в диспергирующих средах. Данный раздел также ограничен линейными диспергирующими системами; дисперсия в нелинейных системах изучается в разд. 22.5.

22.2.1. Фильтры с чирпом

Линейная фильтрация оптического импульса

Передача оптического импульса через произвольную линейную оптическую систему в общем случае описывается теорией линейных систем (см. приложение Б). Линейная система, не зависящая от времени, характеризуется передаточной функцией $H(\nu)$, на которую умножается Фурье-компонента входного импульса на частоте n, чтобы получить компоненту той же частоты на выходе. Если $U_1(t)$ и $U_2(t)$ — комплексные волновые функции исходного и фильтрованного импульса соответственно, то их Фурье-образы $V_1(\nu)$ и $V_2(\nu)$ связаны как

$$V_2(\nu) = H(\nu)V_1(\nu). \tag{22.26}$$

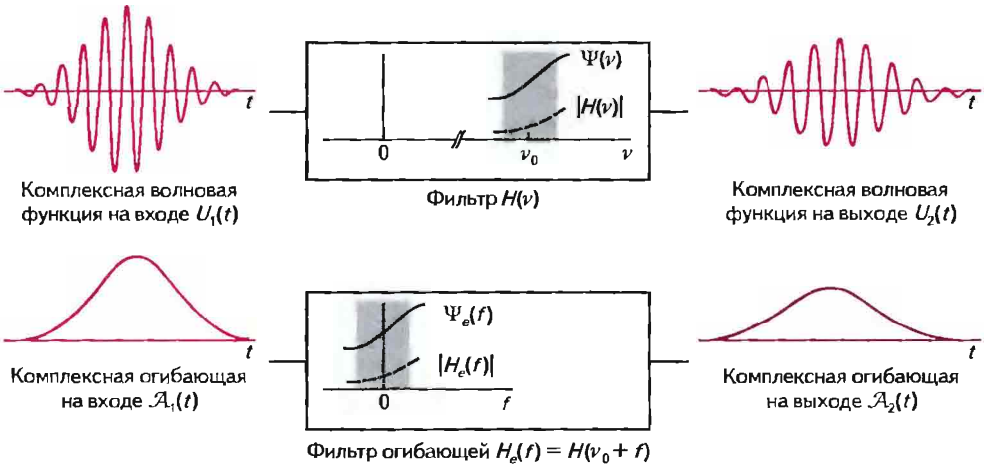


Рис. 22.8. Фильтрация волновой функции фильтром $H(\nu)$ (*верхний рисунок*) эквивалентна фильтрации огибающей фильтром $H_e(f) = H(\nu_0 + f)$ (*нижний рисунок*). Серым цветом показана рабочая область частот

При использовании (22.26) нужно знать $H(\nu)$ только на частотах, лежащих в пределах спектральной полосы импульса, т. е. области шириной $\Delta\nu$ с центральной частотой ν_0, как показано на рис. 22.8. Когда $\Delta\nu = \nu_0$, удобно работать с комплексной огибающей вместо волновой функции. С помощью соотношения $U(t) = \mathcal{A}(t)\exp(j2\pi\nu_0 t)$ и свойства сдвига преобразования Фурье $V(\nu) = A(\nu - \nu_0)$, где $A(\nu)$ — Фурье-образ $\mathcal{A}(t)$, из (22.26) получается $A_2(\nu - \nu_0) = H(\nu)A_1(\nu - \nu_0)$, где индексы 1 и 2 относятся к входному и выходному импульсам соответственно. Обозначая разность частот как $f = \nu - \nu_0$, получаем $A_2(f) = H(\nu_0 + f)A_1(f)$ или

$$A_2(f) = H_e(f)A_2(f), \tag{22.27}$$

где

$$\boxed{H_e(f) = H(\nu_0 + f)} \qquad (22.28)$$

Передаточная функция для огибающей

называется **передаточной функцией для огибающей**. Работать с (22.9) удобнее, чем с (22.8), поскольку частота f обычно намного меньше, чем ν. Иллюстрацией приведенных соотношений служит рис. 22.8.

Передаточные функции $H(\nu)$ и $H_e(f)$ являются комплексными функциями, причем

$$H(\nu) = |H(\nu)| \exp[-j\Psi(\nu)]; \quad H_e(f) = |H_e(f)| \exp[-j\Psi_e(f)],$$

где $\Psi_e(f) = \Psi(\nu_0 + f)$ — действительная функция, представляющая передачу фазы. Фаза, вносимая фильтром, зачастую играет более важную роль в изменении формы импульса, чем модуль. В этой главе будут обсуждаться **фазовые фильтры**, т. е. фильтры, для которых модуль $|H(\nu)|$ примерно постоянен в интересующем нас частотном интервале.

При переходе к временному представлению (22.9) переходит в свертку

$$\mathcal{A}_2(t) = \int_{-\infty}^{\infty} h_e(t - t')\mathcal{A}_1(t')\,dt', \qquad (22.29)$$

где $h_e(t)$ — результат обратного Фурье-преобразования функции $H_e(f)$.

Идеальный фильтр

Идеальный фильтр сохраняет форму огибающей входного импульса; он просто умножает ее на постоянный множитель (с модулем меньше 1 для ослабителя и больше 1 для усилителя), а также, возможно, задерживает импульс на заданное время. Передаточная функция имеет вид

$$H_e(f) = H_0 \exp(-j2\pi f\tau_з), \qquad (22.30)$$

где H_0 — постоянная; $G = |H_0|^2$ — коэффициент ослабления или усиления интенсивности; $\tau_з$ — время задержки. Фаза является линейной функцией частоты

$$\Psi_e(f) = \Psi_0 + 2\pi\tau_з f,$$

где $\Psi_0 = \arg\{H_0\}$ — постоянная начальная фаза (рис. 22.9, *а*). По основному свойству преобразования Фурье (см. приложение А) фаза $2\pi\tau_з f$ эквивалентна временно́й задержке $\tau_з$. Огибающие на входе и выходе связаны соотношением $\mathcal{A}_2(t) = H_0\mathcal{A}_1(t - \tau_з)$, а интенсивности — соотношением $I_2(t) = GI_1(t - \tau_з)$. Для распределенного аттенюатора/усилителя с коэффициентом ослабления/усиления α, скоростью света c и длиной d передаточная функция равна

$$H_e(f) = \exp\left(-\frac{\alpha d}{2}\right)\exp\left(-\frac{j2\pi fd}{c}\right),$$

так что $G = \exp(-\alpha d)$ и $\tau_з = d/c$. Слой идеального недиспергирующего материала с коэффициентом ослабления α и показателем преломления n является примером подобного фильтра, где $c = c_0/n$. В этом случает передаточная функция

$$H(\nu) = \exp\left(-\frac{\alpha d}{2}\right)\exp(-j\beta d),$$

где $\beta = 2\pi\nu/c$ — постоянная распространения (см. подразд. 5.5.1). Если α и n зависят от частоты, т. е. среда обладает дисперсией, то фильтр становится неидеальным, и форма импульса может существенно измениться, как будет показано в разд. 22.3.

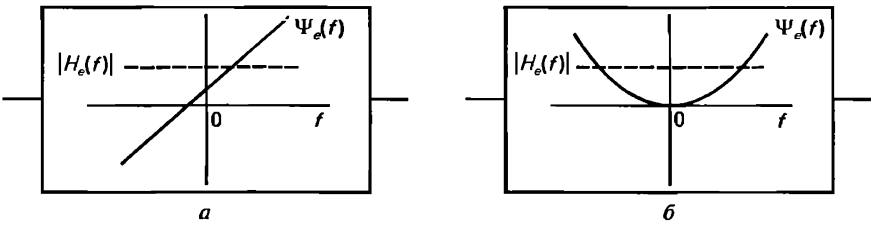


Рис. 22.9. Модуль и фаза передаточной функции для огибающей в случае идеального фильтра (*а*) и фильтра с чирпом (с $b > 0$) (*б*)

Фильтр с чирпом

Возможно, самым важным фильтром в оптике сверхбыстрых процессов является гауссов фильтр с чирпом, который часто называется просто **фильтром с чирпом**. Это фазовый фильтр, фаза которого является квадратичной функцией частоты, $\Psi_e(f) = b\pi^2 f^2$ (рис. 22.9, *б*), так что передаточная функция также гауссова:

$$\boxed{H_e(f) = \exp\left(-jb\pi^2 f^2\right),} \qquad (22.31)$$

Передаточная функция фильтра с чирпом

где b — действительный параметр (в единицах с^2), называемый **коэффициентом чирпа** фильтра. При $b > 0$ чирп, вносимый фильтром, является **положительным**, а $b < 0$ — **отрицательным**.

Соответствующая функция отклика на импульсное воздействие получается обратным преобразованием Фурье формулы (22.31) (см. табл. А1) и также представляет собой гауссову функцию

$$\boxed{h_e(t) = \frac{1}{\sqrt{j\pi b}}\exp\left(\frac{jt^2}{b}\right).} \qquad (22.32)$$

Функция отклика на импульсное воздействие для фильтра с чирпом

Она также имеет фазу, квадратично зависящую от времени, т. е. представляет собой функцию с линейным чирпом, положительным при $b > 0$ и отрицательным при $b < 0$.

Каскад из двух фильтров с чирпом с коэффициентами b_1 и b_2 эквивалентен одному фильтру с коэффициентом чирпа $b = b_1 + b_2$, поскольку передаточные функции перемножаются. Таким образом, фильтр с отрицательным чирпом может компенсировать действие фильтра с положительным чирпом, так что действие фильтра с чирпом является обратимым.

Как легко показать путем подстановки (22.32) в (22.29), огибающие импульса на входе и выходе фильтра с чирпом связаны соотношением

$$\mathcal{A}_2(t) = \frac{1}{\sqrt{j\pi b}} \int_{-\infty}^{\infty} \mathcal{A}_1(t') \exp\left[j \frac{(t-t')^2}{b} \right] \mathrm{d}t' \qquad (22.33)$$

Это преобразование математически эквивалентно соотношению, описывающему дифракцию Френеля [см (4.45) в подразд. 4.3.2], и при достаточно большом параметре чирпа b превращается в соотношение для дифракции Фраунгофера, т. е. становится эквивалентным преобразованию Фурье (см. подразд. 4.3.1). Аналогия между дифракцией в пространстве и дисперсией во времени, которую описывает фильтр с чипом, формально устанавливается в подразд. 22.3.4.

Аппроксимация произвольного фазового фильтра при помощи фильтра с чирпом

Когда модуль и фаза фильтра слабо меняются в пределах малой спектральной ширины импульса, можно предположить, что модуль приблизительно постоянен и равен своему значению на центральной частоте,

$$|H(\nu_0 + f)| \approx |H(\nu_0)| \equiv |H_0|,$$

а фазовую функцию $\Psi(\nu)$ разложить в ряд Тейлора в точке ν_0. Оставляя только первые три члена,

$$\Psi(\nu_0 + f) \approx \Psi_0 + \Psi' f + \frac{1}{2}\Psi'' f^2,$$

где

$$\Psi_0 = \Psi(\nu_0); \quad \Psi' = \left.\frac{\mathrm{d}\Psi}{\mathrm{d}\nu_0}\right|_{\nu_0} \qquad \Psi'' = \left.\frac{\mathrm{d}^2\Psi}{\mathrm{d}\nu^2}\right|_{\nu_0}$$

получаем

$$H(\nu_0 + f) \approx |H_0| \exp\left[-j\left(\Psi_0 + \Psi' f + \frac{1}{2}\Psi'' f^2 \right) \right].$$

Тогда из (22.28) следует, что передаточная функция для огибающей может быть аппроксимирована следующим образом:

$$H_e(f) \approx |H_0| \exp\left[-j\left(\Psi_0 + \Psi' f + \frac{1}{2}\Psi'' f^2\right)\right]. \tag{22.34}$$

Такой фильтр эквивалентен каскаду идеального фильтра и фильтра с чирпом (рис. 22.10). Действие идеального фильтра состоит в умножении на постоянный множитель $H_0 = |H_0| \exp(-j\Psi_0)$, которое не меняет форму импульса и может быть проигнорировано, и внесении фазового сдвига $\exp(-j2\pi\tau_з f)$, эквивалентного временно́й задержке

$$\boxed{\tau_з = \frac{\Psi'}{2\pi}.} \tag{22.35}$$

Групповая задержка

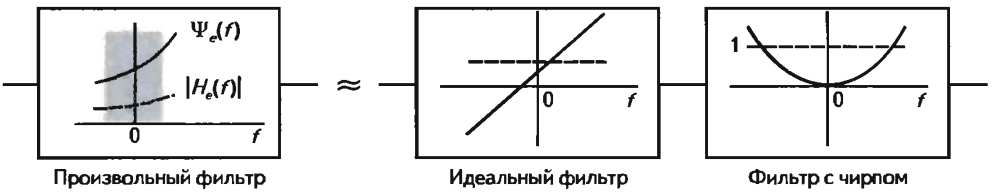


Рис. 22.10. Аппроксимация произвольного фильтра с медленно меняющейся передаточной функцией каскадом из идеального фильтра (включающего временну́ю задержку) и фильтра с чирпом

Фильтр с чирпом имеет передаточную функцию $\exp(-jb\pi^2 f^2)$ с коэффициентом чирпа

$$\boxed{b = \frac{\Psi''}{2\pi^2}.} \tag{22.36}$$

Коэффициент чирпа

Мы приходим к выводу, что главный источник искажения импульса в диспергирующей системе с медленно меняющейся фазой можно представить как фильтр с чирпом. Примеры подобных систем на основе угловой дисперсии и брэгговских решеток будут рассмотрены далее в этом разделе. Диспергирующие среды также описываются фильтрами с чирпом, как будет показано в разд. 22.3. Более точное приближение для описания фазового фильтра требует включения дополнительных членов в разложение Тейлора для фазы $\Psi(\nu)$. Член третьего порядка соответствует фазовому фильтру $\exp[-j(1/6)\Psi''' f^3]$; аналогично можно определить и другие члены высших порядков.

Прохождение спектрально-ограниченного гауссова импульса через фильтр с чирпом

Теперь рассмотрим действие фильтра с чирпом, передаточная функция которого выражается уравнением (22.31), а коэффициент чирпа равен b, на спектрально ограниченный (не имеющий чирпа) гауссов импульс с комплексной огибающей

$$\mathcal{A}_1(t) = A_{10}\exp\left(-\frac{t^2}{\tau_1^2}\right).$$

Поскольку Фурье-образ $\mathcal{A}_1(t)$ равен

$$A_1(f) = \frac{A_{10}\tau_1}{2\sqrt{\pi}}\exp\left(-\pi^2\tau_1^2 f^2\right),$$

из (22.9) следует, что прошедший через фильтр импульс имеет комплексную огибающую с Фурье-образом

$$A_2(f) = A_0\frac{\tau_1}{2\sqrt{\pi}}\exp\left[-\pi^2\left(\tau_1^2 + jb\right)f^2\right]. \tag{22.37}$$

Это выражение является Фурье-образом гауссова импульса с чирпом, у которого ширина равна τ_2, а параметр чирпа a_2. В соответствии с (22.18), Фурье-образ такого импульса равен

$$A_2(f) = A_{20}\frac{\tau_2}{2\sqrt{\pi(1-ja_2)}}\exp\left(-\frac{\pi^2\tau_2^2 f^2}{1-ja_2}\right). \tag{22.38}$$

Приравнивая показатели экспонент в (22.37) и (22.38), получаем

$$\tau_1^2 + jb = \frac{\tau_2^2}{1-ja_2}, \tag{22.39}$$

а приравнивая амплитуды, получаем

$$A_{20} = A_{10}\sqrt{1-ja_2}\,\frac{\tau_1}{\tau_2}.$$

Приравнивая действительные и мнимые части в (22.39), получаем выражения, связывающие параметры входного и выходного импульсов:

Ширина:	$\tau_2 = \tau_1\sqrt{1+\frac{b^2}{\tau_1^4}}$;	(22.40)
Параметр чирпа:	$a_2 = \frac{b}{\tau_1^2}$;	(22.41)
Амплитуда:	$A_{20} = \frac{A_{10}}{\sqrt{1+jb/\tau_1^2}}$.	(22.42)

Мы приходим к заключению, что при прохождении через фильтр с чирпом не имевший такового гауссов импульс остается гауссовым, а его свойства меняются следующим образом:

• *Ширина импульса увеличивается* на множитель

$$\sqrt{1+a_2^2} = \sqrt{1+\frac{b^2}{\tau_1^4}}.$$

Для $|b| = \tau_1^2$ этот множитель составляет $\sqrt{2}$. Таким образом, фильтр начинает оказывать заметное действие, когда его коэффициент чирпа достигает порядка квадрата ширины исходного импульса. Для $|b| \gg \tau_1^2$, т. е. при большом коэффициенте чирпа фильтра или малой длительности исходного импульса, $\tau_2 \approx |b|/\tau_1$. Это означает, что длительность импульса после прохождения фильтра прямо пропорциональна $|b|$ и обратно пропорциональна τ_1, так что более короткие импульсы претерпевают большее уширение.

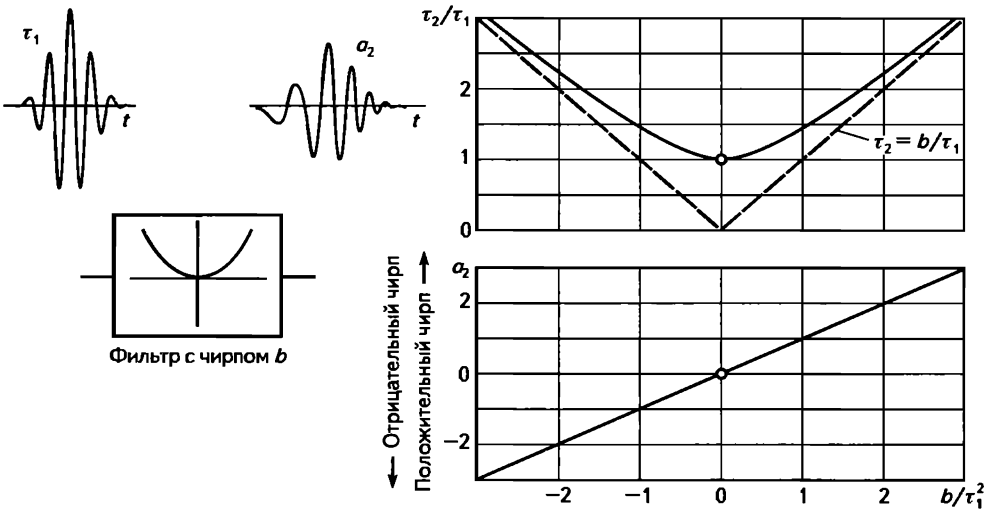


Рис. 22.11. Фильтр с коэффициентом чирпа b преобразует не имевший чирпа гауссов импульс длительностью τ_1, помеченный светлым кружком, в гауссов импульс длительностью τ_1 с параметром чирпа a_2. Длительность импульса растет с ростом $|b|$ тем сильнее, чем меньше τ_1. Параметр чирпа прямо пропорционален b и имеет большую величину для малых τ_1

• *Спектрально ограниченный импульс приобретает чирп* с параметром a_2, который прямо пропорционален коэффициенту чирпа фильтра b и обратно пропорционален квадрату исходной длительности импульса. При фильтрации импульс приобретает положительный чирп, если b положительно, и наоборот. При $b = \tau_1^2$ $a_2 = 1$.

• *Спектральная ширина импульса остается неизменной.* Исходный импульс имеет спектральную ширину $\Delta\nu = 0{,}375/\tau_1$, фильтрованный — такую же спектральную ширину

$$\frac{0{,}375}{\tau_2}\sqrt{1+a_2^2} = \frac{0{,}375}{\tau_1} = \Delta\nu.$$

Это не удивительно, поскольку фильтр с чирпом является фазовым фильтром, который не меняет спектральную интенсивность исходного импульса. Инвариантность спектральной ширины можно объяснить следующим образом. Длительность импульса увеличивается на множитель $\sqrt{1+a_2^2}$, так что связанная с ним спектральная ширина уменьшается на тот же множитель. Однако, поскольку после фильтрации импульс приобретает чирп, его спектр уширяется на тот же множитель. В результате спектральная ширина импульса не меняется.

Зависимость отношения τ_2/τ_1, характеризующего удлинение импульса, и параметра чирпа a_2 от отношения b/τ_1^2 показана на рис. 22.11.

Прохождение через фильтр с чирпом гауссова импульса, уже имеющего чирп

Если гауссов импульс, уже имеющий чирп, проходит через фильтр, вносящий чирп, то на выходе получается гауссов импульс с чирпом и измененными параметрами. Импульс будет либо уширен, либо сжат, а его параметр чирпа изменится, а при некоторых условиях может обратиться в нуль, так что получится импульс без чирпа (спектрально ограниченный). Это свойство компрессии дает основу для метода генерации пико- и фемтосекундных импульсов.

Если исходный импульс имеет длительность τ_1, параметр чирпа a_1 и комплексную огибающую

$$\mathcal{A}_1(t) = A_{10}\exp\left[-(1-ja_1)\frac{t^2}{\tau_1^2}\right],$$

то в результате прохождения через фильтр с чирпом (22.31) получается гауссов импульс с чирпом

$$\mathcal{A}_2(t) = A_{20}\exp\left[-(1-ja_2)\frac{t^2}{\tau_2^2}\right],$$

где

$$\frac{\tau_2^2}{1-ja_2} = \frac{\tau_1^2}{1-ja_1} + jb. \qquad (22.43)$$

Приравнивая действительные и мнимые части в (22.43), получаем следующие выражения для длительности τ_2 и параметра чирпа a_2:

$$\tau_2 = \tau_1\sqrt{1 + 2a_1\frac{b}{\tau_1^2} + \left(1 + a_1^2\right)\frac{b^2}{\tau_1^4}}; \tag{22.44}$$

$$a_2 = a_1 + \left(1 + a_1^2\right)\frac{b}{\tau_1^2}. \tag{22.45}$$

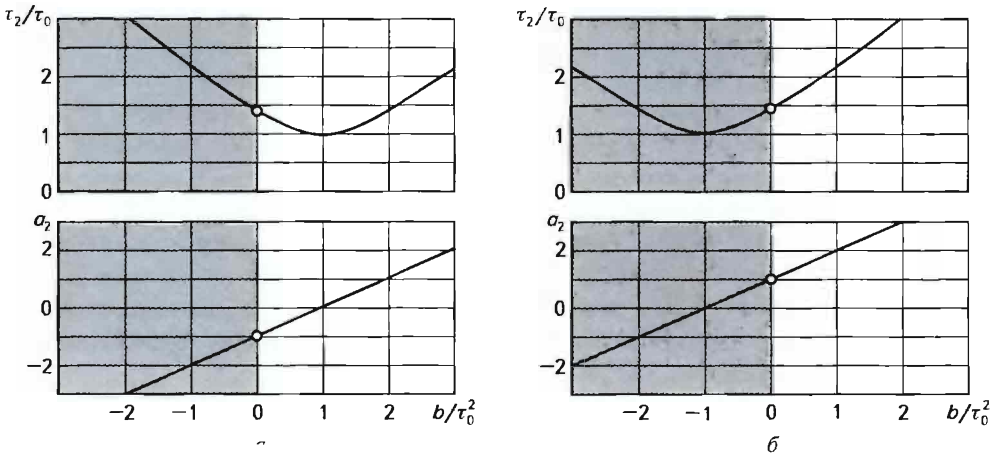


Рис. 22.12. Фильтрация гауссова импульса длительностью τ_1 с параметром чирпа a_1. Фильтр имеет коэффициент чирпа b, положительный в светлых и отрицательный в темных областях. Фильтрованный импульс имеет длительность τ_2 и параметр чирпа a_2. Параметры исходного импульса отмечены светлыми кружками. Минимальная длительность импульса $\tau_0 = \tau_1/\sqrt{1 + a_1^2}$ использована для нормировки. Верхние графики показывают зависимость нормированной длительности импульса от отношения b/τ_0^2, нижние — зависимость параметра чирпа a_2 от b/τ_0^2. Рассмотрены два значения исходного параметра чирпа:
а — $a_1 = -1$; *б* — $a_1 = 1$

Схематическая зависимость коэффициента уширения импульса τ_2/τ_1 и параметра чирпа a_2 от отношения b/τ_1^2 показана на рис. 22.12. Для определения значения b_{min}, при котором выходной импульс имеет минимальную длительность τ_0, приравняем производную от выражения (22.44) для τ_2 по b к нулю. В результате получим:

Минимальная длительность:	$\tau_0 = \dfrac{\tau_1}{\sqrt{1 + a_1^2}};$	(22.46)
Коэффициент чирпа:	$b_{min} = -a_1\tau_0^2 = -\dfrac{a_1}{1 + a_1^2}\tau_1^2.$	(22.47)

С помощью (22.46) и (22.47) выразим (22.44) и (22.45) через b_{min} и τ_0

Длительность:	$\tau_2 = \tau_0\sqrt{1+\dfrac{b-b_{min}^2}{\tau_0^4}}$	(22.48)
Параметр чирпа:	$a_2 = \dfrac{b-b_{min}}{\tau_0^2}.$	(22.49)

Когда $b = b_{min}$, (22.48) и (22.49) дают $\tau_2 = \tau_0$ и $a_2 = 0$, так что импульс одновременно максимально сжат и лишен чирпа. Как следует из (22.47) для максимального сжатия исходного импульса с положительным чирпом ($a_1 > 0$) требуется фильтр с отрицательным чирпом ($\tau_2 = \tau_0 < 0$) и наоборот. Если исходный импульс не имеет чирпа ($a_1 = 0$), то никакой фильтр, вносящий чирп, не способен обеспечить его дальнейшее сжатие, поскольку он уже имеет минимальную возможную ширину ($b_{min} = 0$ и $\tau_0 = \tau_1$).

Заметим, что (22.48) и (22.49) идентичны формулам (22.40) и (22.41), которые были выведены для исходного импульса без чирпа, за исключением того, что b заменено на $b - b_{min}$. Таким образом, графики на рис. 22.11 применимы к случаю исходного импульса с чирпом, если сдвинуть их по горизонтали на величину b_{min}, определенную из (22.47).

Пример 22.1

Сжатие и расширение чирпированного импульса с помощью фильтра, вносящего чирп

а. Гауссов импульс длительностью τ_1 с отрицательным параметром чирпа $a_1 = -1$ пропускается через фильтр с коэффициентом чирпа b. Прошедший импульс также гауссов, имеет длительность τ_1 и параметр чирпа a_2. В этом случае фильтрованный импульс получится максимально сжатым и лишенным чирпа, если $b = b_{min} = \tau_1^2/2$, а фактор сжатия равен $\sqrt{1+a_1^2} = \sqrt{2}$, так что ширина сжатого импульса $\tau_0 = \tau_1\sqrt{2}$. Нормированная ширина импульса τ_2/τ_0 показана на рис. 22.12, *а* в зависимости от отношения b/τ_0^2. При малых положительных значениях b импульс сжимается и приобретает положительный чирп. Максимальное сжатие (и отсутствие чирпа) достигается при $b/\tau_0^2 = 1$ (т. е. $b/\tau_1^2 = 0{,}5$). По мере роста b импульс расширяется и приобретает дополнительный отрицательный чирп.

б. Импульс с положительным чирпом $a_1 = 1$ расширяется с помощью повышающего чирп фильтра ($b > 0$) и его параметр чирпа становится $a_2 > 1$. Применение понижающего чирп фильтра ($b < 0$) приводит к сжатию. Максимальное сжатие достигается при $b/\tau_0^2 = -1$ (или $b/\tau_1^2 = -0{,}5$), как показано на рис. 22.12, *б*.

Приложение: усилитель импульса с чирпом

Возможности усиления ультракороткого импульса высокой пиковой мощности часто ограничены нелинейными эффектами, такими как насыщение и самофокусировка в оптическом усилителе. Такие ограничения можно ослабить, если перед усилением расширить импульс с помощью фильтра с чирпом, а затем сжать уже усиленный импульс, как показано на рис. 22.13. Первый фильтр снижает пиковую интенсивность импульса, сохраняя его полную энергию. Второй фильтр с параметром чирпа равной величины, но противоположенного знака возвращает импульсу его первоначальную длительность. В результате процесс усиления распределяется на более длительное время, а пиковая мощность не превышает допустимых для усилителя пределов.

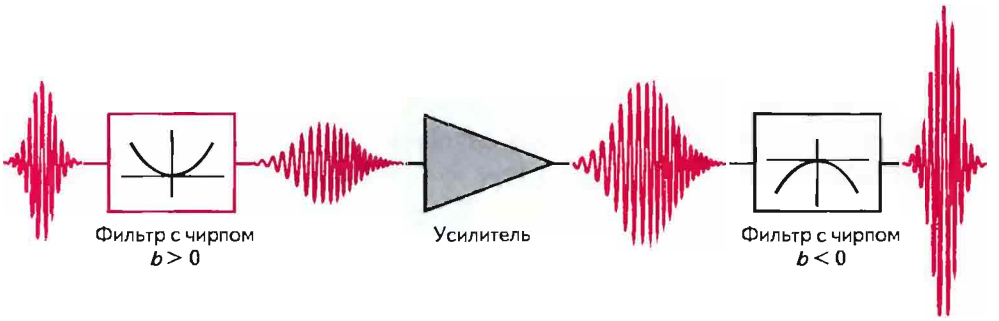


Рис. 22.13. Усиление импульса с чирпом

22.2.2. Осуществление фильтрации с изменением чирпа

Фильтры с чирпом реализуются с использованием диспергирующих оптических систем. Ниже перечислены различные механизмы возникновения дисперсии у оптических элементов.

- **Материальная дисперсия** возникает из-за зависимости показателя преломления и/или коэффициента поглощения оптических материалов от частоты/длины волны.
- **Пространственная дисперсия** имеет множество форм:
  - ***угловая дисперсия*** возникает из-за зависимости угла отклонения от частоты/длины волны у некоторых оптических элементов. Она наиболее выражена у дифракционных оптических элементов, таких как дифракционные решетки и голографические элементы. Преломляющие элементы, такие как призмы, обладают угловой дисперсией из-за материальной дисперсии их вещества;
  - ***дисперсия при многолучевом распространении*** связана с существованием множества путей с различными оптическими длинами. Примером является **модовая дисперсия** в оптически волноводах, причина которой состоит в различии постоянных распространения у разных направляемых мод;

— оптические системы, в которых преобладающая роль принадлежит интерференционным эффектам, зависят от длины волны и, следовательно, обладают ***интерферометрической дисперсией.*** Например, слоистые среды и периодические структуры, такие как решетки Брэгга, имеют зависящие от частоты коэффициенты отражения и пропускания. Оптические резонаторы имеют высокую частотную селективность и, следовательно, обладают сильной дисперсией;

— аналогично, дифракция на малых отверстиях зависит от длины волны и, следовательно, может вызывать значительные изменения профиля коротких оптических импульсов; это разновидность ***дифракционной дисперсии.*** Вообще говоря, распространение через пространственные структуры или отражение от них всегда дает вклад в этот вид дисперсии, если размеры неоднородностей сравнимы с длиной волны. Даже одномодовые волноводы обладают ***волноводной дисперсией,*** которая связана с локализацией света в мелкомасштабной структуре (см. разд. 8.2).

• **Поляризационная дисперсия** — результат зависимости анизотропных свойств оптического материала, элемента или системы от длины волны.

• **Нелинейная дисперсия** также играет важную роль в изменении формы интенсивных оптических импульсов. Она связана с частотной зависимостью нелинейных оптических эффектов, таких как самомодуляция фазы и параметрические взаимодействия, подчиняющиеся частотно-зависимым условиям сохранения энергии и фазового синхронизма.

Каждый из этих дисперсионных эффектов можно использовать для реализации фильтра с чирпом, что иллюстрируют следующие примеры.

Фильтры с чирпом на основе угловой дисперсии

Оптические элементы, обладающие угловой дисперсией, такие как призмы и дифракционные решетки, можно использовать в качестве фильтров, вносящих чирп. Такой элемент, схематически показанный на рис. 22.14, *а*, разлагает по направлениям монохроматические компоненты, составляющие импульс. Предположим, что компонента с частотой ν направляется под углом $\theta(\nu)$ к направлению распространения центральной компоненты с частотой ν_0, т. е. $\theta(\nu_0) = 0$. Если ℓ_0 — оптическая длина пути центральной компоненты, то оптический путь компоненты с частотой ν равен $\ell_0 \cos\theta(\nu)$, как легко видеть из рис. 22.14, *а*. Сдвиг фазы, приобретаемый спектральной компонентой ν, равен

$$\Psi(\nu) = \frac{2\pi\nu}{c}\ell_0 \cos\theta(\nu), \tag{22.50}$$

а соответствующий фазовый фильтр имеет передаточную функцию

$$H(\nu) = \exp[-j\Psi(\nu)].$$

Импульсный пучок, как правило, фильтруется с помощью четырех идентичных оптических элементов, расположенных как показано на рис. 22.14, *б*.

Один элемент разделяет спектральные компоненты импульса по направлениям. Другой обращенный элемент возвращает лучам параллельность, как показано в левом квадрате рис. 22.14, *б*. Обратный процесс происходит в оставшихся двух элементах, показанных в правом квадрате рис. 22.14, *б*. Полная система представляет собой фазовый фильтр с

$$\Psi(\nu) = \frac{2\pi\nu}{c}\ell_0 \cos\theta(\nu),$$

где ℓ_0 — полная длина оптического пути для компоненты с центральной частотой.

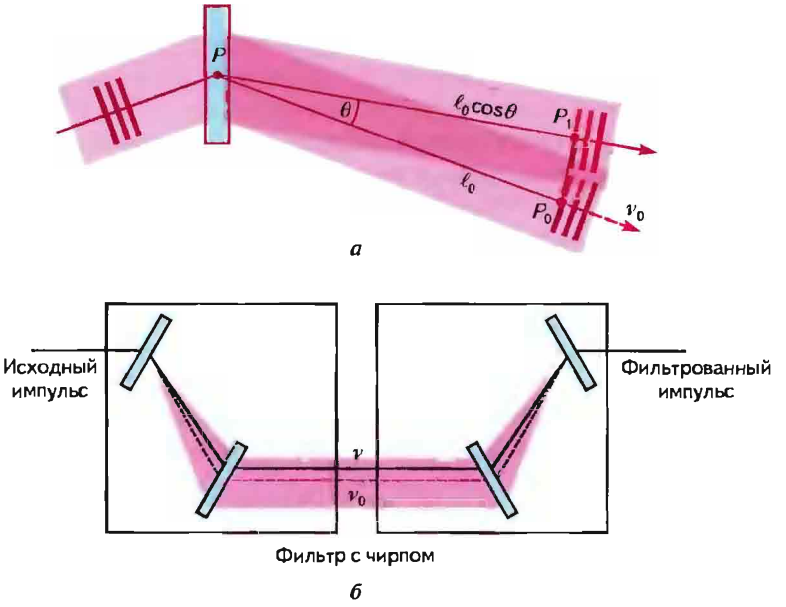


Рис. 22.14. Оптический элемент, обладающий угловой дисперсией (*а*). Компонента с частотой ν отклоняется от компоненты с центральной частотой ν_0 на угол $\theta(\nu)$. В точке наблюдения P_0 оптическая длина пути для центральной компоненты равна ℓ_0 (расстояние $\overline{PP_0}$). Оптическая длина пути для компоненты с частотой ν — это расстояние $\overline{PP_1}$, где P_1 определяется волновым фронтом, проходящим через точку наблюдения P_0. Следовательно, расстояние $\overline{PP_1}$ в треугольнике PP_1P_0 равно $\ell_0 \cos\theta(\nu)$. Фильтр с чирпом на основе комбинации четырех элементов (*б*), показанных на верхнем рисунке (*а*)

Функция $\theta(\nu)$ зависит от используемого диспергирующего элемента, как будет показано в нижеследующих примерах. В типичном случае $\theta(\nu)$ достаточно мало, так что

$$\cos\theta(\nu) \approx 1 - \frac{\theta^2(\nu)}{2};$$

$$\Psi(\nu) \approx \frac{2\pi\nu}{c}\ell_0\left[1 - \frac{1}{2}\theta^2(\nu)\right]. \qquad (22.51)$$

Если $\theta(\nu)$ меняется медленно в пределах спектральной ширины импульса, то ее можно аппроксимировать несколькими членами разложения в ряд Тейлора в точке ν_0. Производные функции $\Psi(\nu)$, взятые в точке $\nu = \nu_0$, где $\theta(\nu_0) = 0$, выражаются как

$$\Psi' \approx \frac{2\pi}{c}\ell_0; \quad \Psi'' \approx -\frac{2\pi\nu}{c}\ell_0\left(\frac{\mathrm{d}\theta}{\mathrm{d}\nu}\right)^2 \tag{22.52}$$

Из (22.35) и (22.36) следует, что такой фильтр эквивалентен временной задержке $\tau_{\text{вр}} = \ell_0/c$ и фильтру с коэффициентом чирпа

$$\boxed{b \approx \frac{\ell_0}{\pi\lambda_0}\alpha_\nu^2,} \tag{22.53}$$

Коэффициент чирпа при угловой дисперсии

где $\alpha_\nu = \mathrm{d}\theta/\mathrm{d}\nu$ — **коэффициент угловой дисперсии**. Поскольку b всегда отрицательно в этом приближении, независимо от знака α_ν, такие фильтры всегда вносят отрицательный чирп. Члены высшего порядка в разложении фазы, конечно, тоже вносят свой вклад в формирование импульса.

Пример 22.2

Призменный фильтр с чирпом

Угол отклонения $\theta_{\text{вр}}(\nu)$ луча, падающего на призму, зависит от ее геометрической формы и показателя преломления $n(\nu)$ (рис. 22.15). Поскольку $\theta(\nu) = \theta_3(\nu) - \theta_3(\nu_0)$, коэффициент угловой дисперсии

$$\alpha_\nu = \frac{\mathrm{d}\theta}{\mathrm{d}\nu} = \frac{\mathrm{d}\theta_3}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}\nu}.$$

С помощью соотношений

$$\frac{\mathrm{d}n}{\mathrm{d}\nu} = -\frac{\lambda_0}{\nu_0}\frac{\mathrm{d}n}{\mathrm{d}\lambda_0} = \frac{n-N}{\nu_0},$$

где N — групповой показатель преломления материала (см. разд. 5.6),

$$N = n - \lambda_0\frac{\mathrm{d}n}{\mathrm{d}\lambda_0},$$

получаем

$$\alpha_\nu = \frac{n-N}{\nu_0}\frac{\mathrm{d}\theta_3}{\mathrm{d}n}. \tag{22.54}$$

Для тонкой призмы с углом при вершине α угол отклонения равен $\theta_3 = (n-1)\alpha$ [см. (1.12)], так что

$$\frac{\mathrm{d}\theta_3}{\mathrm{d}n} = \alpha;$$

$$\alpha_\nu = \frac{n-N}{\nu_0}\alpha. \tag{22.55}$$

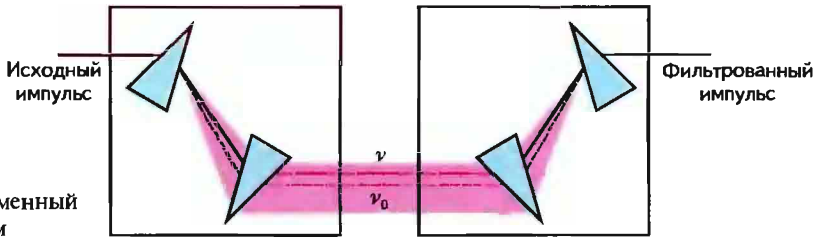


Рис. 22.15. Призменный фильтр с чирпом

Например, для стекла ВК7 на длине волны $\lambda_0 = 800$ нм $n = 1{,}51$ и $N = 3$, 11. Для призмы с $\alpha = 15°$ $\alpha_\nu = 1{,}11\ \ 10^{-15}$ с = 1,11 фс. Для $\ell_0 = 1$ см коэффициент чирпа по формуле (22.53) получается $b = -5 \cdot 10^{-27}$ с$^2 \approx -(71$ фс$)^2$. В соответствии с (22.40) и (22.41) импульс без чирпа длительностью $\tau_1 = 50$ фс, проходя через это устройство, уширяется в $(1 + b^2/\tau_1^4)^{1/2} \approx 2{,}23$ раза и приобретает чирп с параметром $a_2 = b/\tau_1^2 = 2$.

Пример 22.3

Фильтр с чирпом на основе дифракционной решетки

В системе с дифракционной решеткой с периодом Λ (рис. 22.16) углы падения и дифракции θ_1 и θ_2 связаны условием дифракции (2.44). Если $\theta_2 = \theta_{20} + \theta(\nu)$, где θ_{20} — угол отклонения компоненты с центральной частотой, то для дифракции первого порядка

$$\sin\theta_1 + \sin\left[\theta_{20} + \theta(\nu)\right] = \frac{\lambda}{\Lambda} = \frac{c}{\nu\Lambda}. \tag{22.56}$$

Вычисляя производные от обеих частей при $\nu = \nu_0$, получим

$$\alpha_\nu = \frac{\mathrm{d}\theta}{\mathrm{d}n} = \frac{-c}{\nu_0^2 \Lambda \cos\theta_{20}} = \frac{-\lambda_0^2}{c\Lambda\cos\theta_{20}}. \tag{22.57}$$

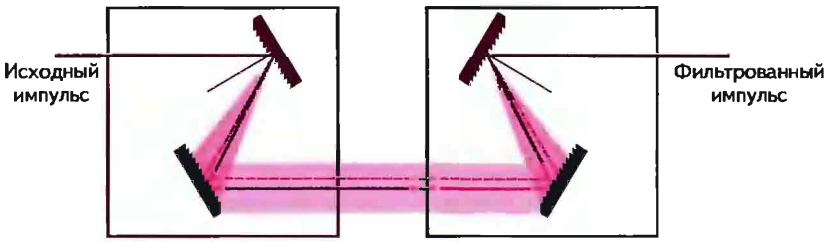


Рис. 22.16. Дифракционная решетка как фильтр с отрицательным чирпом

В симметричном случае, когда $\theta_1 = \theta_{20}$, $\sin\theta_{20} = \lambda_0/2\Lambda$ и, следовательно,

$$\alpha_\nu = -\frac{1}{\nu_0}\frac{\lambda_0^2}{\sqrt{\Lambda^2 - \left(\lambda_0/2\right)^2}}, \tag{22.58}$$

так что

$$b = -\frac{\lambda_0 \ell_0}{\pi c^2} \frac{\lambda_0^2}{\Lambda^2 - (\lambda_0/2)^2}. \qquad (22.59)$$

При $\lambda_0 = 800$ нм и $\Lambda = 1{,}6$ мкм $\alpha_{\nu} = -2{,}72 \cdot 10^{-15}$с $= -2{,}72$ фс. Для $\ell_0 = 10$ см $b = -2{,}94 \cdot 10^{-25}$ с$^2 \approx -(542$ фс$)^2$.

Фильтры с чирпом на решетках Брэгга

Решетки Брэгга с переменным шагом (чирпом) (рис. 22.17) часто используются в качестве фильтров, вносящих чирп. Как описано в подразд. 7.1.3, решетка Брэгга — это периодическая структура, которая селективно отражает оптические волны. Решетка с периодом Λ отражает только волны, у которых длина волны λ удовлетворяет условию Брэгга $\Lambda = m\lambda/2$, где m — целое число; на других длинах волн происходит пропускание света без изменений. Таким образом, решетка является узкополосным фильтром. Если шаг решетки меняется в пространстве, то каждый сегмент отражает свет с длиной волны, удовлетворяющей условию Брэгга с локальным значением Λ. Отраженные волны проходят разное расстояние в зависимости от положения отражающей части решетки, так что система действует как частотно-чувствительный фазовый фильтр. Если частота периодической структуры меняется линейно с расстоянием, то говорят, что решетка имеет линейный чирп. Такая решетка функционирует как фильтр с линейным чирпом.

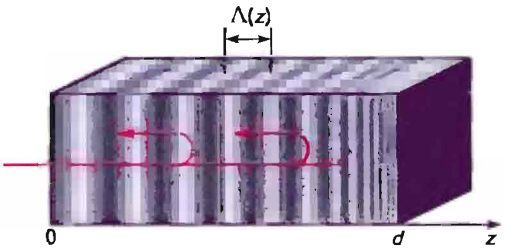


Рис. 22.17. Решетка Брэгга с убывающим периодом действует как фильтр с положительным чирпом

Предположим, что период решетки Брэгга есть функция $\Lambda(z)$ положения z, выбранная так, чтобы частота линейно зависела от z, т. е.

$$\Lambda^{-1}(z) = \Lambda_0^{-1} + \xi z,$$

где Λ_0 — период при $z = 0$; ξ — постоянная. Для определения влияния решетки на оптический импульс, разложим импульс на спектральные компоненты и изучим действие решетки на каждую компоненту. Компонента с частотой ν отражается от решетки в точке z, для которой $\Lambda = m\lambda/2$, т. е.

$$\Lambda(z) = \frac{m\lambda}{2} = \frac{mc}{2\nu} \quad \text{или} \quad z = \frac{2\nu}{mc\xi} - \frac{1}{\xi\Lambda_0}.$$

Эта компонента проходит расстояние $2z$ и претерпевает фазовый сдвиг

$$\Psi = \frac{2\pi\nu}{c}(2z),$$

так что

$$\Psi = \frac{8\pi}{mc^2\xi}\nu^2 + \frac{4\pi}{c\xi\Lambda_0}\nu. \tag{22.60}$$

Из (22.35) и (22.36) следует, что решетка Брэгга с чирпом эквивалентна временной задержке $\tau_з = 2/c\xi\Lambda_0$ и фильтру с коэффициентом чирпа

$$\boxed{b = \frac{8}{m\pi\xi c^2}.} \tag{22.61}$$

Коэффициент чирпа для решетки Брэгга

Если $\xi > 0$, то решетка имеет увеличивающуюся частоту, как показано на рис. 22.17, и коэффициент чирпа $b > 0$, т. е. фильтр вносит положительный чирп. Если частота решетки убывает, фильтр вносит отрицательный чирп.

22.2.3. Сжатие импульсов

Спектрально ограниченный импульс не может быть сжат фильтром, вносящим чирп. Такой фильтр расширяет импульс и сообщает ему чирп, но не меняет его спектральной ширины. Однако сжатия можно достичь при помощи сочетания фазовой модуляции с последующим пропусканием через фильтр с чирпом. Фазовый модулятор *умножает* импульс на зависящий от времени фазовый множитель, что вносит чирп и спектральное уширение, но не меняет длительность импульса. Затем чирпированный импульс подвергается сжатию с помощью фильтра с чирпом, который сохраняет новую спектральную ширину, но сжимает импульс во времени.

Для сжатия нечирпированного импульса

$$\mathcal{A}(t) = A_0 \exp\left(-\frac{t^2}{\tau_1^2}\right)$$

сначала преобразуем его в импульс с чирпом путем *умножения* на квадратичный фазовый множитель $\exp(j\zeta t^2)$, где ζ — постоянная, с помощью **квадратичного фазового модулятора** (КФМ). В результате получается импульс с чирпом

$$A_1(t) = A_{10} \exp\left[-(1 - ja_1)\frac{t^2}{\tau_1^2}\right],$$

с параметром чирпа

$$a_1 = \zeta\tau_1^2. \tag{22.62}$$

Если $\zeta > 0$, то импульс приобретает положительный чирп и последующее его фильтрование с отрицательным чирпом приведет к сжатию. Наоборот, при $\zeta < 0$ импульс приобретает отрицательный чирп, и для его сжатия нужен фильтр с положительным чирпом. В любом случае импульс сжимается в

$$\sqrt{1+a_1^2} = \sqrt{1+\zeta^2\tau_1^4} \text{ раз.}$$

Система показана на рис. 22.18.

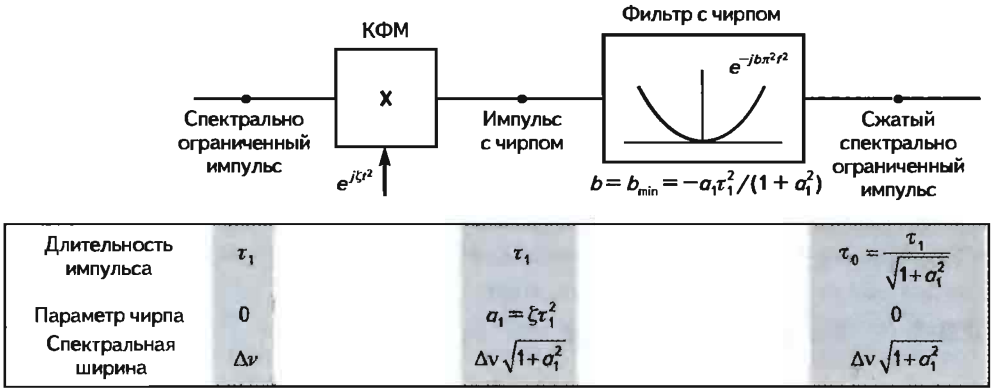


Рис. 22.18. Сжатие спектрально ограниченного импульса с помощью квадратичного модулятора фазы (КФМ) и фильтра с чирпом

Если исходный импульс имеет чирп

$$A_1(t) = A_{10}\exp\left[-(1-ja_1)\frac{t^2}{\tau_1^2}\right],$$

то модуляция фазы $\exp(j\zeta t^2)$ преобразует его в другой чирпированный импульс

$$A_2(t) = A_{10}\exp\left[-(1-ja_2)\frac{t^2}{\tau_1^2}\right],$$

с той же длительностью, но измененным параметром чирпа

$$\boxed{a_2 = a_1 + \zeta\tau_1^2.} \qquad (22.63)$$

Влияние КФМ на параметр чирпа

Таким образом, квадратичный фазовый модулятор, у которого знак ζ противоположен знаку a_1, может свести чирп исходного импульса к нулю или даже обратить его знак.

Резюме

Квадратичный фазовый модулятор (КФМ) и фильтр с чирпом выполняют двойную функцию. Одна операция представляет собой Фурье-аналог другой:

КФМ	= умножение на квадратичную фазовую функцию	Меняет спектральную ширину	Сохраняет временную ширину
Фильтр с чирпом	= свертка с квадратичной фазовой функцией	Сохраняет спектральную ширину	Меняет временную ширину

КФМ можно осуществить с помощью электрооптических модуляторов (см. подразд. 20.1.2), хотя произвести соответствующий сигнал $\exp(j\zeta t^2)$ не просто. Пассивная фазовая модуляция происходит, когда интенсивные импульсы распространяются в нелинейной среде с керровской нелинейностью, как будет описано в подразд. 22.5.3 в связи с самомодуляцией фазы; этот эффект можно использовать для сознания КФМ.

22.2.4. Формирование импульсов

Методы формирования импульсов, обсужденные до сих пор, основаны на фильтрации с приданием чирпа, осуществляемой диспергирующими оптическими элементами. Хотя фильтры с чирпом могут расширять и сжимать импульсы, они не могут менять форму импульса произвольным образом или генерировать импульсы заранее заданной формы, как это часто бывает необходимо в приложениях, касающихся оптической связи и обработки сигналов. Общее формирование ультракоротких импульсов можно осуществить с помощью оптического частотно-пространственного или время-пространственного отображения в сочетании с пространственной модуляцией, чему и посвящен настоящий раздел.

Частотно-пространственное отображение

Частотно-пространственное отображение оптического импульса достигается с помощью дифракционной решетки и линзы, которая направляет каждую спектральную компоненту в единственную точку фокальной плоскости линзы, как показано в левой части рис. 22.19. Эта система проектирует Фурье-образ временного профиля импульса на фокальную плоскость в виде пространственной картины. Модулятор меняет модуль и фазу в соответствии с передаточной функцией желаемого линейного фильтра, формирующего импульс. Это достигается с применением микролитографической или голографической маски или программируемого пространственного модулятора света (ПМС) (см. подразд. 20.3.2). Обратная операция пространственно-спектраль-

ного отображения выполняется следующей линзой и дифракционной решеткой, формирующими импульс измененной формы. Это сводится к обратному преобразованию Фурье, вся операция совпадает с пространственной фильтрацией в Фурье-оптике (см. подразд. 4.4.2). Этот метод стал признанным средством формирования ультракоротких импульсов общего вида.

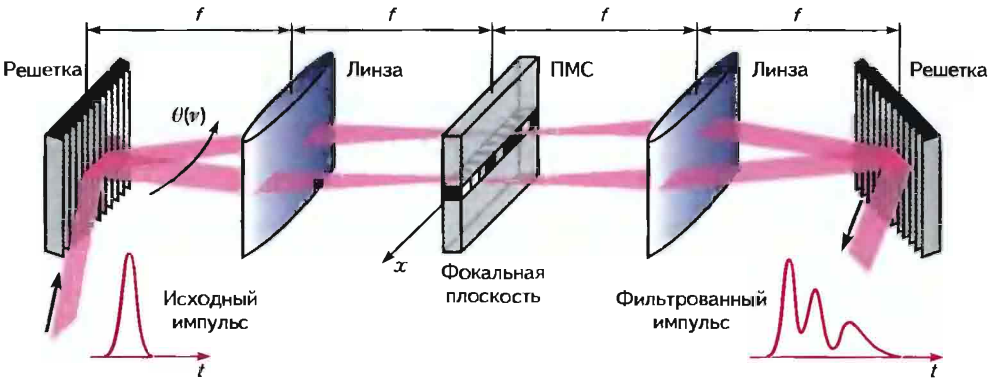


Рис. 22.19. Система формирования импульса включает:
1) частотно-пространственное отображение — решетка и линза отображают Фурье-образ импульса в пространственную картину в Фурье-плоскости;
2) модуляцию при помощи пространственного модулятора света (ПМС);
3) пространственно-частотное отображение с помощью линзы и решетки, осуществляющих обратное преобразование Фурье

Система, изображенная на рис 22.19, количественно описывается следующим образом. Если $\theta(\nu)$ — угол отклонения, вносимый решеткой на частоте ν, то Фурье-компонента на этой частоте будет фокусироваться в точку с $x = \theta(\nu)f$ в фокальной плоскости линзы (Фурье-плоскости), где f — фокусное расстояние линзы, а угол $\theta(\nu)$ предполагается малым. Следовательно, маска с амплитудным коэффициентом пропускания $p(x)$ эквивалентна фильтру с передаточной функцией $H(\nu) = p[\theta(\nu)f]$. Если аппроксимировать $\theta(\nu)$ линейной функцией частоты, $\theta(\nu) \approx \alpha_\nu \nu$, где $\alpha_\nu = \mathrm{d}\theta/\mathrm{d}\nu$ — коэффициент угловой дисперсии решетки [даваемый формулой (22.57)], то форма передаточной функции фильтра $H(\nu)$ — масштабированная версия профиля маски $p(x)$, т. е.

$$H(\nu) = p(\alpha_\nu f \nu). \tag{22.64}$$

При таком частотно-пространственном отображении координата x на Фурье-плоскости соответствует частоте $\nu = x/\alpha_\nu f$, а спектральная ширина $\Delta\nu$ соответствует координатному интервалу $X = \alpha_\nu f \Delta\nu$.

Предыдущий упрощенный анализ был основан на предположении, что исходный импульс является плоской волной, так что дифракция не играет роли. Для исходного пучка конечной ширины W в плоскости решетки спектральная компонента частоты ν отклоняется на угол $\theta(\nu) \approx \alpha_\nu \nu$, но имеет при этом угло-

вую расходимость, пропорциональную $\lambda/W = c\nu/W$, что соответствует пространственному расплыванию

$$\delta x = \frac{f\lambda_0}{W} = \frac{cf}{\nu W}.$$

Это частотно-зависимое расплывание ограничивает пространственное разрешение системы. Маска полной ширины X имеет примерно

$$M = \frac{X}{\delta x} = \frac{X}{\lambda_0 f/W}$$

независимых точек, где λ_0 — центральная длина волны. Пространственное расплывание δx соответствует спектральному интервалу

$$\delta\nu = \frac{\lambda_0 f/W}{\alpha_\nu f} = \frac{\lambda_0}{\alpha_\nu W}.$$

Это ограничивает спектральное разрешение фильтрации импульса $M = XW/\lambda_0 f$ независимыми точками.

Изменение формы пико- и фемтосекундных импульсов успешно осуществлялось с помощью множества ПМС-технологий, включая деформируемые зеркала, многоэлементные матрицы жидкокристаллических модуляторов (время отклика от миллисекунд до субмиллисекунд, низкая производительность) и матрицы полупроводниковых оптоэлектронных модуляторов (наносекундные времена перепрограммирования).

Время-пространственное отображение

В другой схеме для придания импульсу произвольной формы используется пространственный модулятор света (ПМС), состыкованный с дифракционной решеткой, за которым следует линзовая $2f$-система с точечным отверстием на оси в Фурье-плоскости, как показано на рис. 22.20. Решетка умножает спектральную компоненту частоты ν, имеющую комплексную огибающую $A_1(\nu)$, на фазовый множитель $\exp(j2\pi\gamma\nu x)$, зависящий от частоты и координаты, где γ — постоянная. ПМС модулирует ее управляемым пространственным шаблоном $p(x)$, а линзовая система действует как пространственный интегратор, формирующий амплитуду

$$A_2(\nu) \propto A_1(\nu)\int p(x)\exp(j2\pi\gamma\nu x)\,dx \propto A_1(\nu)P(-\gamma\nu), \qquad (22.65)$$

где $P(\nu_x)$ — результат пространственного Фурье-преобразования функции $p(x)$. Следовательно, полная система действует как линейная система с передаточной функцией $H(\nu) \propto P(-\gamma\nu)$, которая соответствует функции отклика на импульсное воздействие

$$h(t) \propto p\left(-\frac{t}{\gamma}\right). \qquad (22.66)$$

Отсюда следует, что коэффициент пропускания ПМС в точке x управляет значением функции отклика на импульсное воздействие в один и только один момент времени $t = -\gamma x$. Таким образом, система осуществляет прямое отображение временно́й области на пространственную, что можно использовать для изменения формы или синтеза фемтосекундных импульсов с произвольно заданным временны́м профилем.

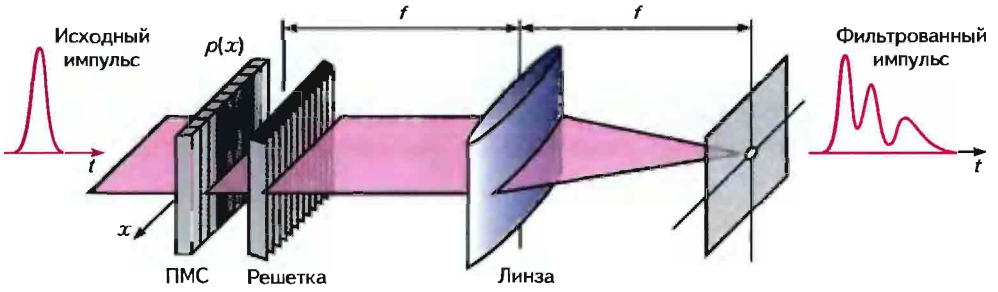


Рис. 22.20. Формирование импульса на основе время-пространственного отображения. Система имеет функцию отклика на импульсное воздействие $h(t)$, представляющую собой масштабированную функцию пропускания ПМС $p(x)$

22.3. РАСПРОСТРАНЕНИЕ ИМПУЛЬСОВ В ОПТИЧЕСКИХ ВОЛНОВОДАХ

В этом разделе изучается распространение оптического импульса в протяженной диспергирующей линейной среде, такой как оптическое волокно. Процесс распространения рассматривается как линейный фильтр с передаточной функцией, определяемой частотно-зависимой постоянной распространения. Для импульсов с медленно меняющейся огибающей (например, пикосекундных оптических импульсов) такой фильтр можно аппроксимировать комбинацией временно́й задержки и фильтра с чирпом. Поэтому математический аппарат для описания распространения импульса будет основан на анализе, проведенном в подразд. 22.2.1. Будет выведено дифференциальное уравнение, описывающее эволюцию огибающей импульса по мере прохождения среды, и установлена аналогия между дисперсией и обычной оптической дифракцией.

22.3.1. Оптическое волокно как фильтр с чирпом

Диспергирующая среда как фильтр

При прохождении расстояния z в линейной диспергирующей среде монохроматическая плоская волна частоты ν (рис. 22.21) приобретает фазовый сдвиг $\beta(\nu)z$, где $\beta(\nu) = 2\pi\nu n(\nu)/c$ — постоянная распространения; $n(\nu)$ — показатель преломления. Таким образом, распространение эквивалентно умноже-

нию на фазовый множитель $\exp[-j\beta(\nu)z]$. Поскольку импульсная волна с волновой функцией $U(z, t)$ представляет собой суперпозицию множества монохроматических волн, фазовый множитель $H(\nu) = \exp[-j\beta(\nu)z]$ есть передаточная функций линейной системы, представляющей распространение, т. е. $V(z, \nu) = H(\nu)V(0, \nu)$, где $V(z, \nu)$ — Фурье-образ $U(z, t)$.

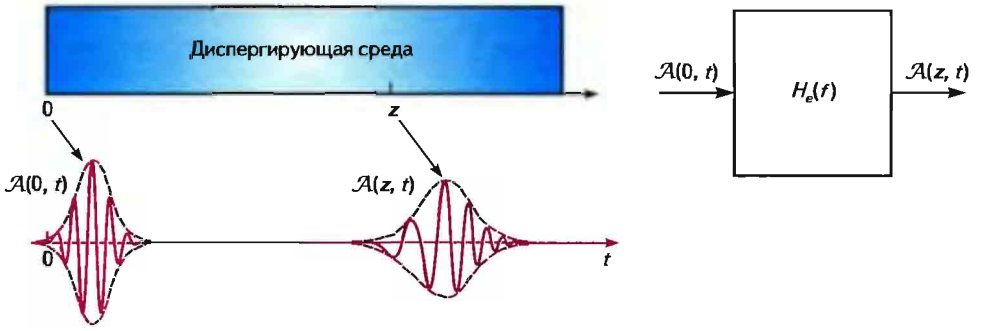


Рис. 22.21. Прохождение оптического импульса через диспергирующую среду эквивалентно действию фазового фильтра

Для импульсов с узким спектральным распределением комплексная волновая функция выражается через медленную огибающую

$$U(z, t) = \mathcal{A}(z, t)\exp(-j\beta_0 z)\exp(j2\pi\nu_0 t),$$

где ν_0 — центральная частота; $\beta_0 = \beta(\nu_0)$. В Фурье-представлении это превращается в

$$V(z, \nu) = A(z, \nu - \nu_0)\exp(-j\beta_0 z),$$

и, следовательно, соотношение

$$V(z, \nu) = \exp[-j\beta(\nu)z]V(0, \nu),$$

в соотношение

$$A(z, \nu - \nu_0) = A(0, \nu - \nu_0)\exp\{-j[\beta(\nu) - \beta(\nu_0)]z\}.$$

Введем частотную разность $f = \nu - \nu_0$, тогда

$$A(z, f) = H_e(f)A(0, f), \tag{22.67}$$

где $H_e(f)$ — **передаточная функция для огибающей**,

$$\boxed{H_e(f) = \exp\{-j[\beta(\nu_0 + f) - \beta(\nu_0)]z\}.} \tag{22.68}$$

Передаточная функция для огибающей

Таким образом, влияние диспергирующей среды на огибающую импульса моделируется фазовым фильтром $H_e(f) = \exp[-j\Psi(f)]$ с фазой $\Psi(f) = [\beta(\nu_0 + f) - \beta(\nu_0)]z$.

Аппроксимация диспергирующей среды временной задержкой и фильтром, вносящим чирп

Если постоянная распространения $\beta(\nu)$ медленно меняется в пределах спектральной ширины импульса, можно использовать результаты разложения в ряд Тейлора $\Psi(f) \approx \Psi' f + (1/2)\Psi'' f^2$, описанные в подразд. 22.2.1, где Ψ' и Ψ'' — первая и вторая производные функции $\Psi(\nu)$ по ν в точке ν_0, а $\Psi(0) = 0$. Передаточная функция для огибающей тогда может быть аппроксимирована выражением

$$H_e(f) \approx \exp\left[-j\left(\Psi' f + \frac{1}{2}\Psi'' f^2\right)\right] = \exp(-j2\pi\tau_з f)\exp\left(-jb\pi^2 f^2\right), \quad (22.69)$$

где

$$\tau_з = \frac{\Psi'}{2\pi}; \quad b = \frac{\Psi''}{2\pi^2}.$$

Отсюда следует, что процесс распространения импульса эквивалентен комбинации временной задержки и фильтра, вносящего чирп.

Множитель $\exp(-j2\pi\tau_з f)$ описывает задержку на время

$$\tau_з = \frac{\Psi'}{2\pi} = \frac{1}{2}\pi\frac{d\beta}{d\nu}z = \frac{d\beta}{d\omega}z = \frac{z}{v},$$

называемую **групповой задержкой**, где v — групповая скорость,

$$\boxed{v = \frac{1}{\beta'} = \frac{c_0}{N};} \quad (22.70)$$

Групповая скорость ($\beta' = d\beta/d\omega$)

N — групповой показатель преломления,

$$N = n - \lambda_0\frac{dn}{d\lambda_0}.$$

Эти параметры были ранее введены при упрощенном анализе, проведенном в разд. 5.6.

Множитель $\exp(-jb\pi^2 f^2)$ описывает фильтр, вносящий чирп, с коэффициентом чирпа

$$b = \frac{\Psi''}{2\pi^2} = \frac{1}{2}\pi^2\frac{d^2\beta}{d\nu^2}z = 2\beta''z,$$

где

$$\beta'' = \frac{d^2\beta}{d\omega^2}.$$

Коэффициент чирпа пропорционален расстоянию z и обычно записывается в виде

$$\boxed{b = 2\beta'' z = \frac{D_\nu}{\pi} z,} \qquad (22.71)$$

Коэффициент чирпа

где D_ν — коэффициент дисперсии групповой скорости (ДГС),

$$\boxed{D_\nu = 2\pi\beta'' = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{1}{v}\right) = \frac{\lambda_0^3}{c_0^2}\frac{\mathrm{d}^2 n}{\mathrm{d}\lambda_0^2}.} \qquad (22.72)$$

Коэффициент ДГС ($\beta'' = \mathrm{d}^2\beta/\mathrm{d}\omega^2$)

Это производная групповой задержки на единице длины по частоте ν, как говорилось ранее в разд. 5.6. Если $\beta'' > 0$ (или $D_\nu > 0$), то дисперсия групповой скорости среды называется нормальной, или положительной, и среда действует как фильтр с положительным чирпом. При $\beta'' < 0$ (или $D_\nu < 0$) дисперсия групповой скорости среды называется аномальной, или отрицательной, и среда подобна фильтру с отрицательным чирпом ($b < 0$).

Пример 22.4

Настраиваемый фильтр с чирпом на основе сочетания угловой и материальной дисперсии в призме

В примере 22.1 было показано, что при преломлении импульсного пучка призмой вносится чирп в результате угловой дисперсии. В этом примере мы рассмотрим влияние материальной дисперсии, которым до сих пор пренебрегали. Если центральный луч, идущий через призму на рис. 22.16, проходит расстояние внутри нее, то материальная дисперсия создает чирпирующий фильтр с коэффициентом чирпа $b = 2\beta''L = D_\nu L/\pi$ [см. (22.71)]. Для призмы из стекла BK7 при $\lambda = 800$ нм коэффициент дисперсии $D_\nu = 0{,}284 \cdot 10^{-24}$ с^2/м, так что для $L = 1$ см коэффициент чирпа $b = D_\nu L/\pi = +9 \cdot 10^{-28}$ с$^2 = +(30$ фс$)^2$. В примере 22.1 было показано, что коэффициент чирпа из-за угловой дисперсии для тонкой призмы с углом 15° при вершине равен $b \approx -(71$ фс$)^2$. Полный коэффициент чирпа равен сумме вкладов материальной и угловой дисперсии. В данном случае значение полного b отрицательно. Расстояние L можно подбирать путем смещения призмы в направлении, ортогональном ее основанию, как показано на рис. 22.22.

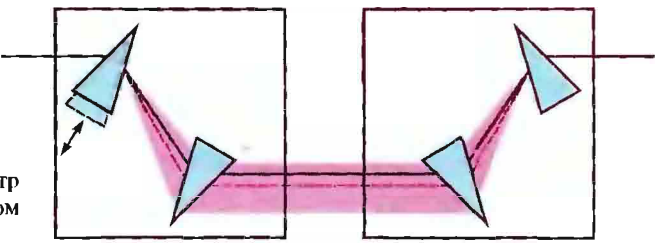


Рис. 22.22. Призменный фильтр с настраиваемым коэффициентом чирпа

Резюме

Распространение импульса в диспергирующей среде можно аппроксимировать двумя эффектами: задержкой во времени, связанной с групповой скоростью $v = 1/\beta' = c_0/N$, и внесением чирпа фильтром с коэффициентом чирпа $b = 2\beta''z = D_\nu z/\pi$, пропорциональным пройденному расстоянию z. Параметры β' и β'' — производные постоянной распространения β по угловой частоте ω, а D_ν — коэффициент ДГС.

22.3.2. Распространение гауссова импульса в оптическом волокне

Поскольку линейная диспергирующая среда может быть аппроксимирована временной задержкой и фильтром с чирпом, мы можем легко описать распространение гауссова импульса в такой среде с помощью общих результатов подразд. 22.2.1.

Спектрально ограниченный гауссов импульс на входе

Рассмотрим сначала спектрально ограниченный (без чирпа) гауссов импульс шириной τ_0 при $z = 0$. На расстоянии z импульс задерживается на время $\tau_з = z/v$ и преобразуется фильтром с коэффициентом чирпа $b = D_\nu z/\pi$. В соответствии с (22.40)—(22.42), импульс остается гауссовым, но его ширина возрастает до

$$\tau(z) = \tau_0\left(1 + \frac{D_\nu^2}{\pi^2\tau_0^4}z^2\right)^{1/2}$$

появляется чирп с параметром

$$a(z) = \frac{D_\nu}{\pi\tau_0^2}z,$$

а амплитуда становится равной

$$A_0(z) = A_0\left(1 - j\frac{D_\nu}{\pi\tau_0^2}z\right)^{-1/2}$$

Определяя параметр так, что

$$\frac{D_\nu}{\pi\tau_0^2} = \frac{1}{z_0},$$

можно записать эти выражения в более простом виде, приведенном в табл. 22.2, включающей также выражение для комплексной огибающей, основанное на (22.12). Величина z_0 называется **дисперсионной длиной** и характеризует среду и начальную длительность импульса.

Таблица 22.2. **Характеристики гауссова импульса, распространяющегося в диспергирующей среде с групповой скоростью v, коэффициентом дисперсии D_ν и дисперсионной длиной z_0. При $z = 0$ импульс спектрально ограничен, имеет ширину τ_0, амплитуду A_0 и интенсивность $I_0 = |A_0|^2$**

$\mathcal{A}(z, t) = A_0\sqrt{\frac{-jz_0}{z-jz_0}}\exp\left[j\frac{\pi}{D_\nu}\frac{(t-z/v)^2}{z-jz_0}\right]$	Комплексная огибающая	(22.73)
$I(z, t) = I_0\frac{\tau_0}{\tau(z)}\exp\left[-2\frac{(t-z/v)^2}{\tau^2(z)}\right]$	Интенсивность	(22.74)
$\int I(t)\,dt = \sqrt{\frac{\pi}{2}}I_0\tau_0$	Плотность энергии	(22.75)
$\tau(z) = \tau_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}$	Ширина импульса	(22.76)
$a(z) = \frac{z}{z_0}$	Параметр чирпа	(22.77)
$z_0 = \pi\frac{\tau_0^2}{D_\nu} = \frac{\tau_0^2}{2\beta''}$	Дисперсионная длина $\|z_0\|$	(22.78)
$\Delta\nu = \frac{0,375}{\tau_0}$	Спектральная ширина	(22.79)

Можно сделать следующие наблюдения.

• Центр импульса задерживается на время z/v, т. е. импульс распространяется с групповой скоростью $v = 1/\beta'$

• Ширина импульса $\tau(z)$ достигает минимального значения τ_0 при $z = 0$ и растет с ростом $|z|$, как показывает рис. 22.23. При $z = |z_0|$ импульс расширяется в $\sqrt{2}$ раз, а при $z = \sqrt{3}|z_0|$ его ширина удваивается. Для $z \gg |z_0|$

$$\tau(z) \approx \frac{\tau_0 z}{|z_0|} = \frac{D_\nu}{\pi\tau_0}z,$$

т. е. импульс линейно уширяется со скоростью, обратно пропорциональной начальной ширине импульса τ_0. Через спектральную ширину $\Delta\nu = 0,375/\tau_0$ длительность импульса выражается как

$$\tau(z) \approx \frac{1}{0,375\pi}|D_\nu|\Delta\nu z = 0,85|D_\nu|\Delta\nu z,$$

что согласуется с тем фактом, что D_ν — скорость уширения импульса на единицу длины и на единицу спектральной ширины [с/(м · Гц)]. Это соотношение можно записать также через коэффициент дисперсии D_λ [пс/(км · нм)] как $\tau(z) \approx 0,85|D_\lambda|\Delta\lambda z$, что представляет собой приближенный вариант (5.131).

• Параметр чирпа $a(z)$ равен 0 при $z = 0$ по определению и растет линейно с расстоянием z, достигая значения единица при $z = |z_0|$, как видно на рис 22.23. Знак чирпа такой же, как и знак D_ν. Для нормальной дисперсии $D_\nu > 0$ и $a(z) > 0$ при $z > 0$, т. е. импульс приобретает положительный чирп. В видимом диапазоне нормальная дисперсия означает, что «синие» волны медленнее «красных», что согласуется с положительным знаком чирпа. Противоположное имеет место при аномальной дисперсии.

• Дисперсионная длина $|z_0|$ зависит от модуля коэффициента дисперсии среды D_ν и начальной длительности импульса τ_0. Это расстояние, на котором длительность импульса возрастает в $\sqrt{2}$ раз, а параметр чирпа достигает единицы.

• Спектральная ширина $\Delta\nu = 0{,}375/\tau_0$ остается неизменной при распространении импульса. Спектральное сжатие, которое сопровождает временное расширение импульса, полностью компенсируется равным спектральным расширением, сопровождающим возникновение чирпа. Этого следовало ожидать, поскольку распространение в диспергирующей среде моделируется фазовым фильтром, который не меняет спектральную интенсивность.

• Плотность энергии, переносимой импульсом, не зависит от z, чего и следовало ожидать в среде без потерь.

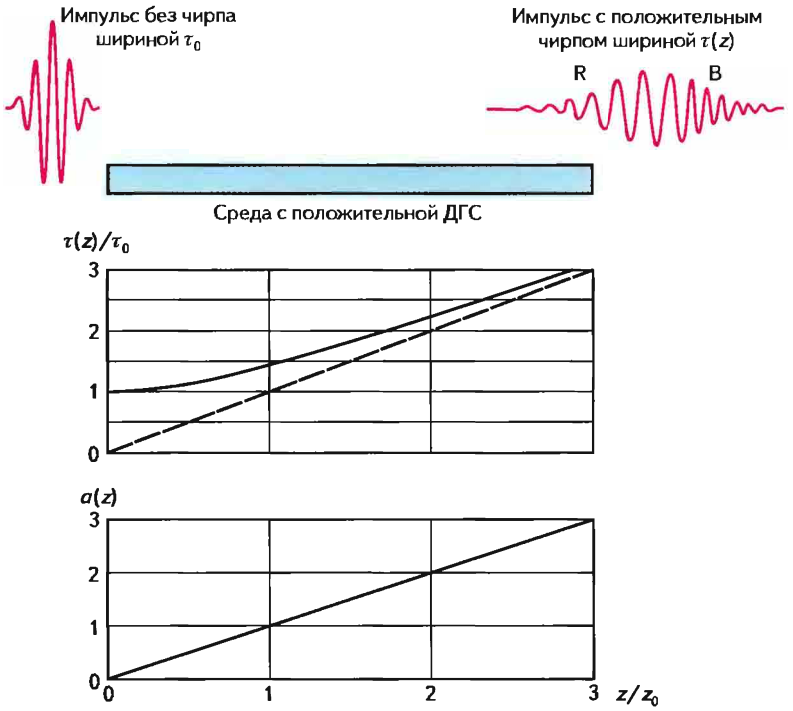


Рис. 22.23. Распространение первоначально не имевшего чирпа гауссова импульса через диспергирующую среду. Импульс остается гауссовым, но его ширина $\tau(z)$ увеличивается и он приобретает чирп с увеличивающимся параметром $a(z)$

Пример 22.5

Уширение импульса в стекле ВК7

Коэффициент дисперсии стекла ВК7 при $\lambda = 620$ нм равен $\beta'' = 1{,}02 \cdot 10^{-25}$ с^2/м. Для слоя толщиной 1 см это соответствует коэффициенту чирпа $b = 2\beta''z = 2{,}04 \cdot 10^{-23}$ с^2 = (4,5 пс)^2. Это значит, что когда гауссов импульс шириной 4,5 пс проходит через слой, его ширина увеличивается в $\sqrt{2}$ раз. Для более коротких импульсов с $\tau_0 = 100$ фс и центральной длиной волны $\lambda_0 = 620$ нм дисперсионная длина

$$|z_0| = \frac{\tau_0^2}{2|\beta''|} = 0{,}5 \text{ мм}.$$

Импульс удваивает свою длительность при прохождении слоя толщиной $\sqrt{3}z_0 = 0{,}87$ мм.

Входной гауссов импульс с чирпом

Согласно (22.46), при распространении через диспергирующую среду чирпированный гауссов импульс длительностью τ_1 с параметром чирпа a_1 при $z = 0$ достигает минимальной длительности

$$\boxed{\tau_0 = \frac{\tau}{\sqrt{1 + a_1^2}}} \qquad (22.80)$$

Минимальная длительность

на расстоянии $z_{\min}$, для которого $(D_\nu/\pi)z_{\min} = b_{\min}$. Из (22.47)

$$z_{\min} = -a_1 \frac{\pi}{D_\nu}\tau_0^2, \qquad (22.81)$$

что можно переписать через дисперсионную длину $z_0 = \pi\tau_0^2/D_\nu$ как

$$\boxed{z_{\min} = -a_1 z_0.} \qquad (22.82)$$

Положение минимальной длительности импульса

В конце концов (22.48) и (22.49) превращаются в следующие выражения для длительности импульса и параметра его чирпа как функций расстояния z:

$$\boxed{\tau(z) = \tau_0\sqrt{1 + \frac{(z - z_{\min})^2}{z_0^2}};} \qquad (22.83)$$

$$\boxed{a(z) = \frac{z - z_{\min}}{z_0}.} \qquad (22.84)$$

Уравнения (22.83) и (22.84) идентичны (22.76) и (22.77) для исходного импульса без чирпа, за исключением сдвига на расстояние z_{min} до точки, где длительность импульса минимальна.

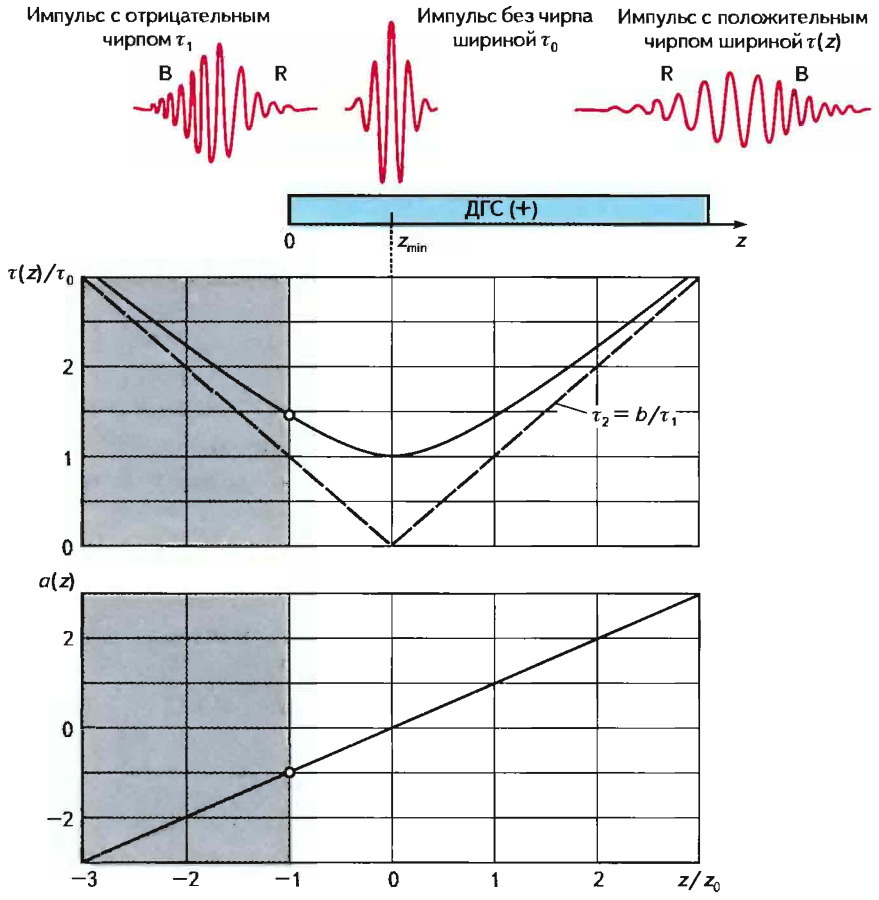


Рис. 22.24. Распространение изначально имевшего отрицательный чирп гауссова импульса ($a_1 = -1$) через среду с нормальной дисперсией. Длительность импульса $\tau(z)$ убывает от начального значения τ_1 до минимального $\tau_0 = \tau_1/\sqrt{2}$ и затем снова растет. Первоначально отрицательный параметр чирпа растет линейно и меняет знак при $z > z_{min}$. В данном примере $z_{min} = z_0$

Следовательно, выражения в табл. 22.2 справедливы всегда как для положительных, так и для отрицательных z и могут использоваться для изначально чирпированных импульсов, если начало среды поместить в точку z, в которой параметр чирпа равен начальному значению. Это иллюстрирует рис. 22.24 — еще один график $\tau(z)$ и $a(z)$, основанный на (22.76) и (22.77) для положительных и отрицательных значений z. Например, для среды с положительным z_0 (положительная ДГС, или нормальная дисперсия), когда начальный параметр

чирпа равен $a_1 = -1$, $z_{min} = z_0$, так что среда начинается при $z = -z_0$. Процесс сжатия импульса и последующего расширения теперь ясен. Импульс максимально сжат в $\sqrt{2}$ раз и лишается чирпа на расстоянии $z_{min} = z_0$. При дальнейшем распространении импульс расширяется и приобретает положительный чирп.

Поскольку начальный параметр чирпа a_1 и коэффициент дисперсии D_ν могут быть положительными или отрицательными, имеется ряд возможностей.

- Для среды с нормальной дисперсией ($D_\nu > 0$) фильтр повышает чирп и параметр z_0 положителен. Если первоначально импульс имел отрицательный чирп ($a_1 < 0$), то z_{min} положительно, так что импульс действительно сжимается по мере распространения в положительном направлении z. Для первоначально положительного чирпа импульса ($a_1 > 0$) z_{min} отрицательно, и импульс не будет сжиматься.
- Для среды с аномальной дисперсией ($D_\nu < 0$) фильтр понижает чирп и параметр z_0 отрицателен. Для первоначально положительного чирпа импульса ($a_1 > 0$) z_{min} положительно, так что импульс сжимается по мере распространения в положительном направлении z. Для первоначально отрицательного чирпа импульса ($a_1 < 0$) z_{min} отрицательно, и импульс не будет сжиматься.

Итак, сжатие может происходить при прохождении импульса с положительным чирпом через понижающую чирп (аномальную) среду или импульса с отрицательным чирпом через повышающую чирп (нормальную) среду.

Сжатие импульса с использованием КФМ и диспергирующей среды

Как описано в подразд. 22.2.1, спектрально ограниченный импульс можно сжать с помощью комбинации квадратичного фазового модулятора (КФМ) и фильтра, вносящего чирп. В качестве такого фильтра может быть использована диспергирующая среда, как показано на рис. 22.25. Если ширина исходного импульса равна τ_1, то модуляция фазовым множителем $\exp(j\zeta t^2)$ эквивалентна внесению чирпа с коэффициентом $a_1 = \zeta\tau_1^2$. Спектральная ширина чирпированного импульса расширяется в $\sqrt{1 + a_1^2}$ раз. Если ζ отрицательно, то чирп импульса понижается, и дальнейшее распространение в среде с положительной ДГС (нормальной дисперсией) приводит к сжатию импульса до минимальной ширины

$$\tau_0 = \frac{\tau_1}{\sqrt{1 + a_1^2}} = \frac{\tau_1}{\sqrt{1 + \zeta^2\tau_1^4}}. \tag{22.85}$$

Импульс также будет сжиматься, если ζ положительно, а среда имеет отрицательную ДГС. Пользуясь (22.82) и (22.85), мы приходим к заключению, что минимум ширины достигается на расстоянии

$$z_{min} = -a_1 z_0 = -\frac{\pi\tau_0^2}{D_\nu} a_1 = -\frac{\pi}{D_\nu}\frac{\tau_1^2 a_1}{1 + a_1^2} = -\frac{\pi\zeta}{D_\nu}\frac{\tau_1^4}{1 + \zeta^2\tau_1^4}, \tag{22.86}$$

которое положительно, если ζ и D_ν имеют противоположные знаки.

В пределе, когда $a_1 = \zeta\tau_1^2 \gg 1$,

$$\tau_0 \approx \frac{\tau_1}{a_1} = \frac{1}{\zeta\tau_1} \tag{22.87}$$

и $z_{min} \approx f$, где

$$f = \frac{\pi}{-\zeta D_\nu}. \tag{22.88}$$

Это расстояние можно трактовать как **фокусное расстояние** данной системы сжатия импульса.

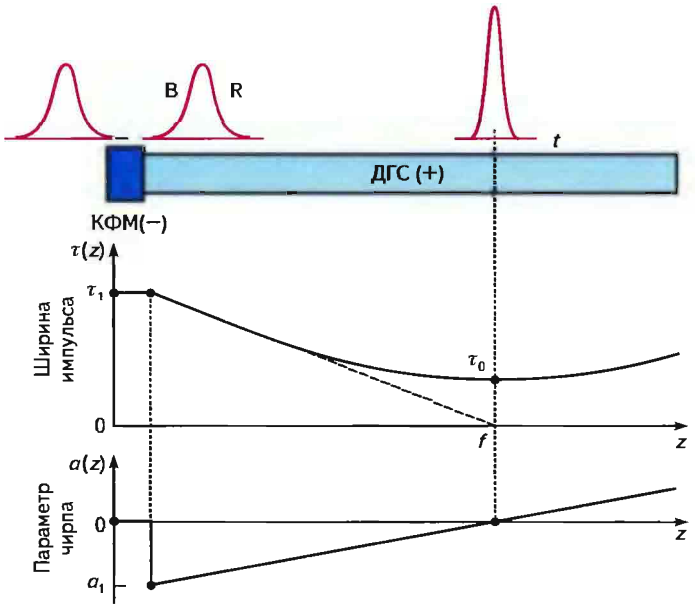


Рис. 22.25. Сжатие импульса с помощью квадратичного фазового модулятора (КФМ) и среды с дисперсией групповой скорости (ДГС)

Пример 22.6

Сжатие импульса в оптическом волокне из кварцевого стекла

а. Гауссов импульс с постоянной времени $\tau = 100$ фс и центральной длиной волны $\lambda_0 = 850$ нм (например, от титан-сапфирового лазера) распространяется по оптическому волокну из кварцевого стекла. На данной длине волны кварцевое стекло имеет нормальную дисперсию (положительную ДГС) с $D_\lambda = -200$ пс/(км · нм) (см. рис. 5.28), что соответствует

$$D_\nu = -\frac{\lambda_0^2}{c_0} D_\lambda = +4{,}82 \cdot 10^{25} \ \text{с}^2/\text{м}.$$

Если импульс первоначально не имеет чирпа, то $\tau_0 = 100$ фс и, следовательно, дисперсионная длина равна $z_0 = \pi\tau_0^2/D_\nu = 6{,}52$ см. На этом расстоянии импульс уширяется в $\sqrt{2}$ раз и имеет коэффициент чирпа $a = 1$. На расстоянии $z = 6{,}52$ м ширина импульса увеличится примерно в $z/z_0 = 100$ раз и станет равной 10 пс, а параметр чирпа достигнет значения $a = 100$.

б. Если исходный импульс промодулирован по фазе множителем $\exp(j\zeta t^2)$, то $a_1 = \zeta\tau^2$. Для $\zeta = -10^{-2}$ (фс)^2 импульс приобретает отрицательный чирп с параметром $a = -1$. В результате распространения по волокну первоначальный 100-фемтосекундный импульс сжимается до $\tau_0 = 100/\sqrt{2} = 77$ фс на расстоянии $z_{\min} = -a_1 z_0 = \pi\tau_0^2/D_\nu = 3{,}26$ см. Поскольку импульс стал уже, он расширяется более быстро при дальнейшем распространении по волокну. На расстоянии $z = 6{,}52$ м его ширина увеличивается примерно в $z/z_0 \approx 200$ раз и достигает 14,1 пс.

Упражнение 22.1

Компенсация дисперсии в оптических волокнах

Расплывание импульсов в оптических волокнах можно исключить с помощью баланса нормальной и аномальной дисперсии.

а. Импульс без чирпа с центральной длиной волны $\lambda_0 = 1{,}55$ мкм и шириной $\tau_1 = 10$ пс передается по оптическому волокну из кварцевого стекла. На этой длине волны кварцевое стекло имеет аномальную дисперсию с $D_\lambda = +20$ пс/(км · нм). Определите длительность импульса τ и параметр чирпа a на расстоянии $d_1 = 100$ км.

б. Импульс нужно сжать до первоначальной ширины 10 пс с помощью другого волокна длины d_2 (рис. 22.26), сделанного из некоторого материала с нормальной дисперсией $D_\lambda = -100$ пс/(км · нм). Определите d_2.

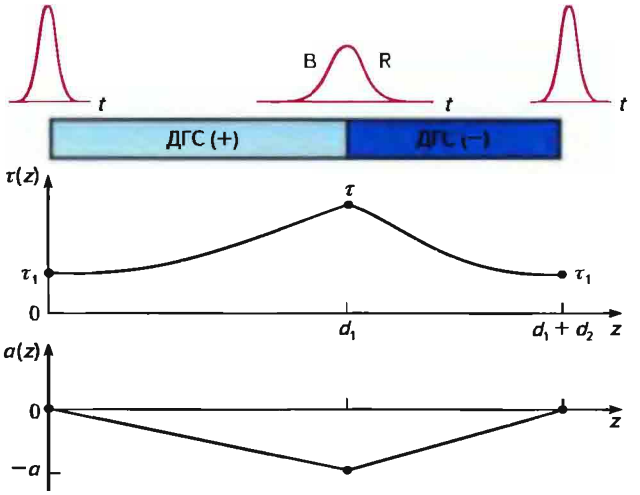


Рис. 22.26. Компенсация дисперсии в оптических волокнах

Упражнение 22.2

Компенсация дисперсии с помощью периодической последовательности фазовых модуляторов

Уширение импульсов в оптическом волокне можно уменьшить с помощью периодического набора квадратичных фазовых модуляторов, расположенных на расстоянии $2d$. Каждый модулятор вносит квадратичную фазу $\exp(j\zeta t^2)$. Если коэффициент дисперсии модулятора ζ положителен, а материал волокна имеет отрицательную ДГС, то ширина импульса и параметр его чирпа возрастают и убывают периодически, как показано на рис. 22.27. Покажите, что условие такого периодического поведения имеет вид

$$\zeta = -\frac{2a}{\tau^2} = -\frac{2\mathrm{d}/z_0}{\tau_0^2\left[1+(\mathrm{d}/z_0)^2\right]} = -\frac{2D_\nu \mathrm{d}}{\pi\tau_0^4\left[1+\left(D_\nu \mathrm{d}/\pi\tau_0^2\right)^2\right]}, \qquad (22.89)$$

где τ_0 и τ — минимальное и максимальное значения ширины импульса; a — параметр чирпа; $z_0 = \pi\tau_0^2/D_\nu$.

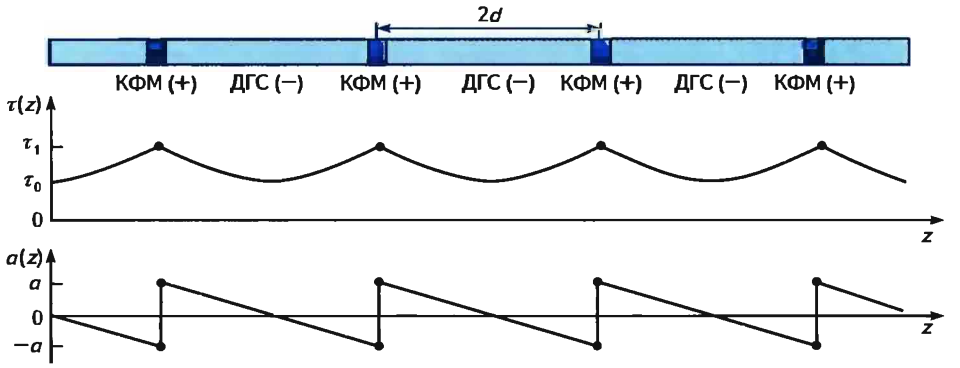


Рис. 22.27. Компенсация дисперсии с помощью периодической последовательности положительной КФМ и отрицательной ДГС

*22.3.3. Уравнение диффузии для медленной огибающей

В подразд. 22.3.1 было показано, что диспергирующая среда с постоянной распространения, аппроксимированной разложением Тейлора до квадратичных членов, эквивалентна фильтру огибающей импульса с передаточной функцией

$$H_e(f) = \exp\left(-\frac{j2\pi zf}{v}\right)\exp\left(-j\pi D_\nu z f^2\right),$$

где v — групповая скорость; D_v — коэффициент дисперсии. Покажем теперь, что при определенных условиях огибающая $\mathcal{A}(z, t)$ удовлетворяет дифференциальному уравнению в частных производных

$$\frac{\partial^2 \mathcal{A}}{\partial t^2} + j\frac{4\pi}{D_v}\left(\frac{\partial \mathcal{A}}{\partial z} + \frac{1}{v}\frac{\partial \mathcal{A}}{\partial t}\right) = 0. \tag{22.90}$$

Если время задержки z/v не учитывать (или использовать координатную систему, движущуюся со скоростью импульса), то (22.90) упрощается к виду

$$\boxed{\frac{\partial^2 \mathcal{A}}{\partial t^2} + j\frac{4\pi}{D_v}\frac{\partial \mathcal{A}}{\partial z} = 0,} \tag{22.91}$$

Уравнение диффузии для медленной огибающей

в котором легко узнать **уравнение диффузии**.

Доказательство справедливости уравнения диффузии для огибающей импульса в диспергирующей среде

Начнем с уравнения фильтра $A(z, f) = A(0, f)H_e(z, f)$, откуда

$$A(z, f) \approx A(0, f)\exp\left(-j2\pi\frac{z}{v}f - j\pi D_v z f^2\right),$$

где $A(z, f)$ — Фурье-образ $\mathcal{A}(z, t)$. Беря производную по z, получаем дифференциальное уравнение

$$\frac{\mathrm{d}}{\mathrm{d}z}A(z, f) \approx \left(-j2\pi\frac{f}{v} - j\pi D_v f^2\right)A(z, f).$$

Произведем обратное преобразование Фурье по f в обеих его частях и заметим, что обратные Фурье-образы функций $A(z, f)$, $j2\pi fA(z, f)$ и $(j2\pi f)^2A(z, f)$ — это $\mathcal{A}(z, t)$, $\partial\mathcal{A}(z, t)/\partial t$ и $\partial^2\mathcal{A}(z, t)/\partial t^2$ соответственно. Получается уравнение (22.90).

Функция отклика на импульсное воздействие, связанная с уравнением диффузии, имеет вид

$$h_e(t) = \frac{1}{\sqrt{jD_v z}}\exp\left(j\frac{\pi t^2}{D_v z}\right) \tag{22.92}$$

и идентична соответствующей функции для фильтра с чирпом (22.32) при $b = D_v z/\pi$.

Известно, что для первоначально гауссова распределения $\mathcal{A}(0, t) = A_0\exp(-t^2/\tau_0^2)$ уравнение диффузии имеет гауссово решение

$$\mathcal{A}(z, t) = A_0\sqrt{\frac{jz_0}{z - jz_0}}\exp\left(j\frac{\pi}{D_v}\frac{t^2}{z - jz_0}\right),$$

где $z_0 = \pi\tau_0^2/D_v$. Учитывая задержку во времени, заменим t на $t - z/v$ и получим (22.73).

Вывод уравнения диффузии для огибающей из уравнения Гельмгольца

Уравнение (22.90) можно вывести непосредственно из уравнения Гельмгольца

$$\left[\frac{d^2}{dz^2} + \beta^2(\nu)\right] V(z,\ \nu) = 0.$$

Поскольку

$$U(z,\ t) = \mathcal{A}(z,\ t)\exp(-j\beta_0 z)\exp(j2\pi\nu_0 t),$$

ее Фурье-образ:

$$V(z,\ \nu) = A(z,\ \nu - \nu_0)\exp(-j\beta_0 z),$$

где $A(z,\ \nu)$ — Фурье-образ $\mathcal{A}(z,\ t)$. Подставим $\nu = \nu_0 + f$ в уравнение Гельмгольца:

$$\left[\frac{d^2}{dz^2} + \beta^2(\nu_0 + f)\right] A(z,\ f)\exp(-j\beta_0 z) = 0.$$

В приближении медленной огибающей:

$$\frac{d^2}{dz^2}\left[A\exp(-j\beta_0 z)\right] \approx \left(-j2\beta_0\frac{dA}{dz} - \beta_0^2 A\right)\exp(-j\beta_0 z)$$

уравнение Гельмгольца принимает вид

$$-j2\beta_0\frac{dA}{dz} + \left[\beta^2(\nu_0 + f) - \beta_0^2\right]A = 0.$$

При слабой дисперсии

$$\beta^2(\nu_0 + f) - \beta_0^2 \approx 2\beta_0\left[\beta(\nu_0 + f) - \beta_0\right].$$

Как и раньше, аппроксимируем постоянную распространения $\beta(\nu)$ трехчленным разложением в ряд Тейлора

$$\beta(\nu_0 + f) \approx \beta_0 + \pi f\beta' + 2\pi^2 f^2\beta'',$$

где

$$\beta_0 = \beta(\omega_0);\quad \beta' = \left.\frac{d\beta}{d\omega}\right|_{\omega_0} \qquad \beta'' = \left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_0}$$

С учетом этого уравнение Гельмгольца приобретает вид

$$-j\frac{dA}{dz} + (2\pi f\beta' + 2\pi^2 f^2\beta'')A = 0.$$

Выполняя обратное преобразование Фурье и замечая, что множители $j2\pi f$ и $-4\pi^2 f^2$ эквивалентны производным $\partial/\partial t$ и $\partial^2/\partial t^2$ соответственно, мы получаем

$$-j\frac{\partial A}{\partial z} - j\beta'\frac{\partial A}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 A}{\partial t^2} = 0.$$

Наконец, подставляя $\beta' = 1/v$ и $\beta'' = D_\nu/2\pi$, получаем (22.90).

*22.3.4. Аналогия между дисперсией и дифракцией

Удивительная аналогия обнаруживается между уравнением диффузии для огибающей

$$\frac{\partial^2 \mathcal{A}}{\partial t^2} + j\frac{4\pi}{D_\nu}\frac{\partial \mathcal{A}}{\partial z} = 0,$$

которое описывает распространение импульса $\mathcal{A}(z, t)$ в диспергирующей среде (в системе, движущейся со скоростью v и без учета дисперсионных членов выше квадратичного), и параксиальным уравнением Гельмгольца

$$\nabla_T^2 A - j\frac{4\pi}{\lambda}\frac{\partial A}{\partial z} = 0,$$

описывающим дифракцию оптического пучка $A(x, y, z)$ в свободном пространстве в параксиальном приближении. Оба они являются уравнениями диффузии (первое — одномерным, второе — двумерным). Данная аналогия указывает, что временное расплывание (дисперсия) импульса по мере его распространения в диспергирующей среде подчиняется тому же математическому закону, что и пространственное расплывание (дифракция) пучка в поперечной плоскости по мере его распространения в свободном пространстве, причем время t играет роль поперечной координаты $\rho = (x, y)$, а коэффициент дисперсии $-D_\nu$ — роль длины волны λ. Различные свойства этой аналогии сведены в табл. 22.3.

Аналогию между коэффициентом дисперсии $-D_\nu$ и длиной волны λ можно более полно оценить, если время t выразить в единицах расстояния ct, проходимого со скоростью света. В этих единицах величина $c^2 D_\nu$ имеет размерность длины и ее роль в определении масштаба дисперсии импульса количественно совпадает с ролью, которую играет длина волны в определении пространственного масштаба дифракции. Например, если $D_\nu = 10^{-23}$ с^2/м, то $c^2 D_\nu \approx 0{,}9$ мкм, что эквивалентно 3 фс.

Другая интересная аналогия касается роли линзы в изменении кривизны волнового фронта и роли квадратичного фазового модулятора (КФМ), придающего импульсу чирп. Тонкая линза вносит фазовый множитель $\exp(j\pi\rho^2/\lambda f)$ [см. (2.40)], в то время как КФМ — фазовый множитель $\exp(j\zeta t^2)$ (см. подразд. 22.2.3). Записывая

$$\exp(j\zeta t^2) = \exp\left(\frac{\pi j t^2}{-D_\nu f}\right),$$

Таблица 22.3. **Сравнение дифракции в пространстве (в параксиальном приближении) и дисперсии в диспергирующей среде (в приближении второго порядка). Коэффициент дисперсии $-D_v$ при дисперсии импульса играет роль длины волны λ при дифракции. Квадратичный фазовый модулятор играет роль временной линзы**

Дифракция		Дисперсия	
Комплексная огибающая	$A(\rho,\ z)$	Комплексная огибающая	$A(z,\ t)$
Поперечная координата	$\rho = \sqrt{x^2 + y^2}$	Время	t
Аксиальная координата	z	Аксиальная координата	z
Параксиальное уравнение Гельмгольца	$\nabla_T^2 A - j\frac{4\pi}{\lambda}\frac{\partial A}{\partial z} = 0$	Уравнение диффузии для медленной огибающей (движущаяся система координат)	$\frac{\partial^2 \mathcal{A}}{\partial t^2} + j\frac{4\pi}{D_v}\frac{\partial \mathcal{A}}{\partial z} = 0$
Длина волны	λ	Коэффициент дисперсии	$-D_v$
Функция отклика на импульсное воздействие $h_e(\rho)$	$\frac{j}{\lambda z}\exp\left(-j\frac{\pi\rho^2}{\lambda z}\right)$	Функция отклика на импульсное воздействие $h_e(\rho)$	$\frac{1}{\sqrt{jD_v z}}\exp\left(j\frac{\pi t^2}{D_v z}\right)$
Линза	$\exp\left(\frac{j\pi\rho^2}{\lambda f}\right)$	КФМ	$\exp\left(j\zeta t^2\right)$
Фокусное расстояние	f	Фокусное расстояние	$f = \frac{\pi}{-D_v\zeta}$

где $\zeta = -\pi D_\nu f$, мы видим, что КФМ эквивалентен временной линзе, которая сжимает импульс до минимальной ширины при $z = f$, где $f = \pi/(-D_\nu \zeta)$ — фокусное расстояние, что подтверждает (22.88).

Математическая аналогия между временны́м расплыванием гауссова импульса в диспергирующей среде (см. подразд. 22.3.2) и пространственным расплыванием гауссова пучка в свободном пространстве (см. гл. 3) продемонстрирована в табл. 22.4. Дисперсионная длина z_0 аналогична дифракционной (рэлеевской) длине z_0. Хотя последняя всегда положительна, первая из них определена так, что она положительна при нормальной и отрицательна при аномальной дисперсии. Этим объясняется отрицательный знак у параметра z_0, который появляется в выражении для комплексной огибающей гауссова импульса.

Из-за математической аналогии между пространственной дифракцией и временной дисперсией, а также между линзой и квадратичным фазовым модулятором (КФМ) каждая оптическая система, составленная из участков свободного пространства и линз, имеет свой временной аналог, состоящий из комбинации диспергирующих сред и КФМ. На рис. 22.28 показан ряд примеров.

- Временно́е расплывание импульса в диспергирующей среде аналогично пространственной дифракции пучка или волны, прошедшей через апертуру.
- Временно́е сжатие импульса КФМ аналогично пространственной фокусировке пучка линзой. Например, поскольку гауссов пучок фокусируется линзой с фокусным расстоянием f в пятно шириной

$$W_1 = \frac{\lambda}{\pi W_0} f,$$

по аналогии следует, что гауссов импульс сжимается КФМ до длительности

$$\tau_1 = \frac{1}{\zeta \tau_0} = -\frac{D_\nu}{\pi \tau_0} f,$$

где $f = \pi/(-D_\nu \zeta)$ — фокусное расстояние КФМ. Другим примером временно́й фокусировки с помощью КФМ является превращение двух отдельных импульсов при $z = 0$ в один импульс при $z = f$.

- Аналогом получения изображения одной линзой в обычной оптике является система, в которой КФМ играет роль временно́й линзы, генерирующей увеличенную или уменьшенную копию временного профиля импульса, т. е. временно́е изображение (см. задачу 2 к разд. 22.3).
- Периодическая последовательность КФМ, поддерживающая ширину импульса, аналогична периодическому набору ретранслирующих линз.
- Аналог $2f$-системы преобразования Фурье (см. разд. 4.2) — это временна́я $2f$-система преобразования Фурье с помощью КФМ. Например, такая система преобразует оптический импульс с фазовой модуляцией в импульс с амплитудной модуляцией, профиль которого представляет собой Фурье-образ исходного импульса.

Таблица 22.4. **Сравнение дифракции гауссова пучка в свободном пространстве и дисперсии гауссова импульса в диспергирующей среде**

Гауссов пучок		Гауссов импульс	
Ширина	$W(z) = W_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}$	Ширина	$\tau(z) = \tau_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}$
Дифракционная длина	$z_0 = \frac{\pi W_0^2}{\lambda}$	Дисперсионная длина	$\lvert z_0\rvert = \frac{\pi\tau_0^2}{\lvert D_\nu\rvert}$
Угол расходимости	$\theta_0 = \frac{\lambda}{\pi W_0}$	Скорость расплывания (с/м)	$\frac{\lvert D_\nu\rvert}{\pi\tau_0}$
Кривизна волнового фронта	$\frac{1}{R(z)} = \frac{2\pi}{\lambda}\frac{z}{z^2+z_0^2}$	Скорость приобретения чирпа	$\varphi'' = \frac{2\pi}{D_\nu}\frac{z}{z^2+z_0^2}$
Пространственный чирп	$a(z) = \frac{W^2(z)}{2R(z)} = \frac{z}{z_0}$	Параметр чирпа	$a(z) = \frac{1}{2}\varphi''\tau^2(z) = \frac{z}{z_0}$

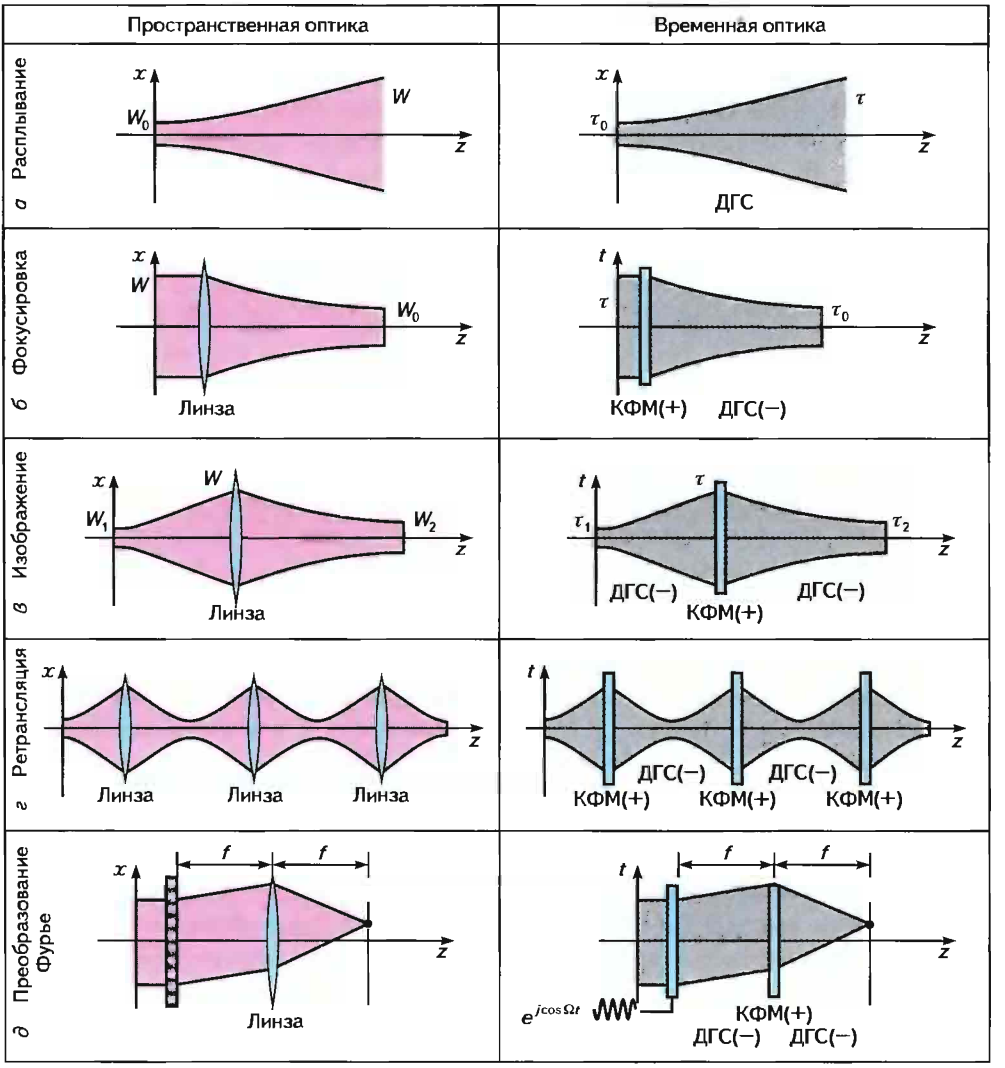


Рис. 22.28. Аналогия между пространственной (*слева*) и временно́й (*справа*) оптикой. Квадратичный фазовый модулятор (КФМ) играет роль линзы. Затененные области представляют пространственную ширину пучка (*слева*) и временну́ю ширину (длительность) импульса (*справа*) как функции z. В правом столбце временны́е задержки игнорируются и показано только изменение временной протяженности импульса. Предполагается, что оптический импульс (*справа*) распространяется в среде с отрицательной ДГС. Рисунки правого столбца применимы и к среде с положительной ДГС, но в этом случае КФМ должен быть отрицательным

Главное различие между пространственной дифракцией и временной дисперсией состоит в том, что длина волны λ всегда положительна, тогда как ее аналог $-D_\nu$ может быть и положительным, и отрицательным. Следствия этого различия можно оценить, исследуя функции отклика на импульсное воздействие в табл. 22.3. Положительность длины волны λ имеет своим следствием то, что свет от точечного источника образует расходящийся фронт (сферическую волну). По аналогии, в среде с отрицательным D_ν (положительным $-D_\nu$) импульс света расплывается в более широкий импульс с отрицательным чирпом. В противном случае положительного D_ν (т. е. нормальной дисперсии) исходный импульс расплывается в более широкий импульс с положительным чирпом. Возможны оба знака чирпа, в то время как при пространственной дифракции существуют только расходящиеся волны.

22.4. ЛИНЕЙНАЯ ОПТИКА УЛЬТРАКОРОТКИХ ИМПУЛЬСОВ

Пространственные и временные характеристики импульсных волн неразрывно связаны. Пространственное расплывание или фокусировка зависят от начального профиля во времени, а форма импульса во времени — от его начального пространственного распределения. Эти эффекты особенно выражены у ультракоротких импульсов и в оптических системах с угловой дисперсией. Лишь в очень особых случаях импульсная волна сохраняет точно свой временной профиль в процессе распространения (например, плоская и сферическая волны, см. подразд. 22.1.3). Для импульсов с медленно меняющейся огибающей применимо квазистационарное приближение, в котором временные и пространственные изменения *приближенно* разделяются. Это приближение, однако, неприменимо к ультракоротким импульсам.

В данном разделе мы рассмотрим распространение ультракоротких импульсов в пучках, проходящих через простые отображающие системы. Начнем с упрощенного анализа на основе оптики лучей, а затем перейдем к теории, основанной на оптике волн с использованием методов Фурье-оптики.

22.4.1. Оптика лучей

Оптика лучей основана на описании света лучами, которые отражаются и преломляются на оптических границах в соответствии с законом Снелла (см. разд. 1.1). Временные эффекты включаются в эту теорию, поскольку считается, что лучи распространяются с зависящей от среды скоростью $c = c_0/n$. Мы пользовались этой теорией в подразд. 9.3.2 для оценки разброса времен прихода оптических лучей в оптическом волокне путем определения времени распространения для каждого оптического пути и оценки разности между наибольшей и наименьшей задержкой.

Если некоторые компоненты оптической системы изготовлены из диспергирующего материала, то задержка, вносимая этими элементами, должна

рассчитываться на основе групповой скорости $v = c_0/N$, а не фазовой скорости $c = c_0/n$, где $N = n - \lambda \mathrm{d}n/\mathrm{d}\lambda$ — групповой показатель преломления. Оценка уширения оптического импульса в процессе распространения через оптическую систему, следовательно, является упражнением по определению разности между наибольшим и наименьшим запаздыванием для всех возможных оптических путей.

Уширение импульса в однолинзовой изображающей системе

В однолинзовой изображающей системе, показанной на рис. 22.29, оптический импульс испускается из точки P_1 в виде множества лучей, которые затем встречаются в точке P_2. Каждый луч проходит через воздух и стекло и испытывает соответствующую задержку. Если стекло не вносит дисперсии, то в соответствии с принципом Ферма (см. разд. 1.1) все лучи приходят в одно и то же время и импульс не уширяется. Для учета влияния дисперсии удобно ввести дифференциальную задержку как разность между групповой задержкой (связанной с групповой скоростью v) и фазовой задержкой (за счет фазовой скорости c). Разность между наибольшей и наименьшей дифференциальной задержкой и составляет уширение импульса. Конечно, дифференциальная задержка равна нулю на не вносящих дисперсии частях оптического пути, так что вклад в нее даст только прохождение луча в материале линзы.

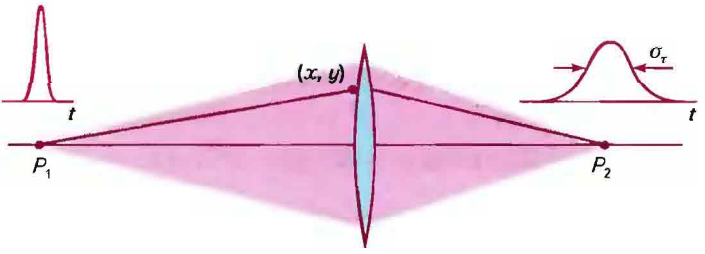


Рис. 22.29. Уширение импульса в однолинзовой изображающей системе обусловлено материальной (хроматической) дисперсией

Характеризуя каждый луч координатами точки (x, y) в плоскости линзы, где ее толщина равна $d(x, y)$, получаем для дифференциальной задержки выражение

$$\Delta\tau(x, y) = \left|\frac{1}{c} - \frac{1}{v}\right| d(x, y) = \frac{|n - N|}{c_0} d(x, y). \qquad (22.93)$$

Увеличенная ширина импульса равна разности между максимальным и минимальным значениями $\tau(x, y)$, так что

$$\Delta\tau(x, y) = |n - N| \frac{\Delta d}{c_0}, \qquad (22.94)$$

где Δd — разность между максимальной и минимальной толщиной линзы. Для тонкой линзы с фокусным расстоянием f и максимальной толшиной d_0 используем (2.37) и (2.40) и получим

$$d(x,\ y) = d_0 - \frac{x^2 + y^2}{2R} = d_0 - \frac{x^2 + y^2}{2(n-1)} f,$$

так что

$$\Delta d = \frac{(D/2)^2}{2(n-1)} f,$$

где D — диаметр линзы. Следовательно,

$$\Delta\tau(x,\ y) = \frac{|n-N|}{n-1} \frac{(D/2)^2}{2c_0 f},$$

откуда

$$\boxed{\Delta\tau(x,\ y) = \frac{|n-N|}{n-1} \frac{1}{8F_{\#}^2} \frac{f}{c_0},} \qquad (22.95)$$

Уширение импульса

где $F_{\#} = f/D$ — относительная апертура линзы.

Например, для линзы из стекла BK7 при $\lambda = 400$ нм $n = 1{,}53$ и $n - N = \lambda\, dn/d\lambda = -0{,}052$. Если $f = 30$ мм и $F_{\#} = 2$, то импульс расширяется на $\Delta\tau \approx 307$ фс.

В этой системе причиной уширения импульса является дифференциальная материальная дисперсия, связанная с множеством пространственных траекторий лучей. Без материальной дисперсии существование множества различных путей не приводит к уширению импульса, благодаря принципу Ферма.

*22.4.2. Волновая и Фурье-оптика

Волновая природа света заставляет тонкий монохроматический пучок расходиться в широкий конус, причем угол расходимости прямо пропорционален длине волны и обратно пропорционален исходной ширине пучка. Если пучок модулируется ультракоротким импульсом с широким спектром, то каждая из монохроматических компонент расходится в свой собственный конус, причем компоненты с меньшими длинами волны занимают конусы с меньшими углами. Следовательно, спектральный состав распространяющегося света в каждой точке пространства меняется, причем точки, более удаленные от оси, имеют меньшую энергию на более коротких волнах, как показано на рис. 22.30. В периферийной области спектр, следовательно, сдвинут в сторону более низкой частоты (красный сдвиг), а спектральная ширина уменьшена, что сопровождается уширением импульса. Этот пример показывает, что пространственные и временные характеристики света связаны самим процессом распространения волны, особенно если пучок предельно узкий, а импульс предельно короткий.

Хотя распространение ультракоротких световых импульсов через произвольные оптические системы усложнено неустранимой пространственно-временно́й связью, анализ концептуально прост, если система линейна, поскольку можно использовать метод Фурье и свести задачу к суперпозиции решений для каждой из монохроматических компонент поля. Произвольная импульсная волна $U(\boldsymbol{r}, t)$ раскладывается в сумму монохроматических компонент, амплитуды которых даются преобразованием Фурье

$$V(\boldsymbol{r}, \nu) = \int U(\boldsymbol{r}, t) \exp(j2\pi\nu t)\, \mathrm{d}t.$$

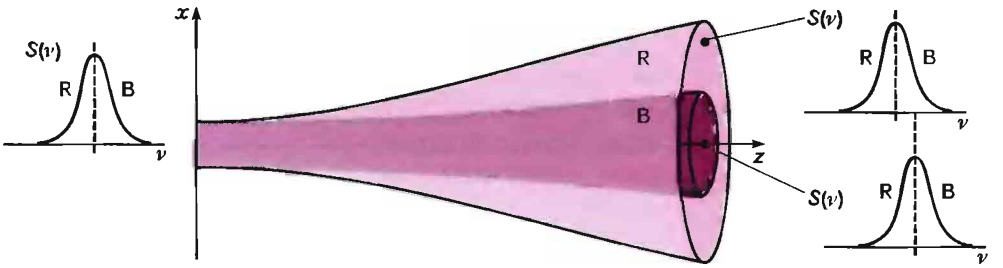


Рис. 22.30. Расходимость импульсного пучка. Длинноволновые компоненты (R) расходятся в конусы с большими углами, чем коротковолновые (B). Это приводит к подавлению коротковолновых компонент в периферийных точках и, следовательно, к красному сдвигу и сжатию спектральной полосы, сопровождающемуся уширением импульса

Распространение каждой монохроматической компоненты через систему определяется с помощью средств, разработанных в гл. 3—10, а полное решение затем получается с помощью их суперпозиции, т. е. посредством обратного преобразования Фурье

$$U(\boldsymbol{r}, t) = \int V(\boldsymbol{r}, \nu) \exp(-j2\pi\nu t)\, \mathrm{d}\nu.$$

Фурье-оптика импульсных волн

Распространение монохроматического света между параллельными плоскостями (1, 2) с произвольной линейной оптической системой между ними в общем случае описывается линейным преобразованием

$$V_2(x, y, \nu) = \iint h(x, x', y, y', \nu) V_1(x', y', \nu)\, \mathrm{d}x'\mathrm{d}y', \qquad (22.96)$$

где h — функция отклика на импульсное воздействие системы на частоте ν (см. гл. 4). Для импульсной волновой функции $U_1(x, y, t)$ на входе волновая функция на выходе $U_2(x, y, t)$ может быть легко найдена вычислением Фурье-образа $V_2(x, y, \nu)$ с помощью (22.96) с последующим обратным преобразованием Фурье.

Функция отклика на импульсное воздействие была определена в гл. 4 для различных оптических элементов. Воспроизведем эти результаты с явным выделением зависимости от частоты ν.

• **Свободное пространство.** В соответствии с (4.18) в приближении Френеля распространение на расстояние в свободном пространстве эквивалентно системе с функцией отклика на импульсное воздействие

$$h(x,\ x',\ y,\ y',\ \nu) \approx \frac{j\nu}{cz}\exp\left[-j\frac{2\pi\nu}{c}\frac{(x-x')^2+(y-y')^2}{2z}\right]. \tag{22.97}$$

Мы не учитываем множитель $\exp(-j2\pi\nu z/c)$, поскольку он представляет тривиальное не зависящее от x, y запаздывание на z/c.

• **Апертура.** Прохождение через плоскую апертуру эквивалентно умножению на апертурную функцию (единица внутри апертуры и 0 вне ее).

• **Линза.** Прохождение через линзу с фокусным расстоянием f эквивалентно умножению на квадратичный фазовый множитель

$$t(x,\ y,\ \nu) \approx \exp\left(-jn\frac{2\pi\nu}{c_0}d_0\right)\exp\left(j\frac{2\pi\nu}{c_0}\frac{\rho^2}{2f}\right), \tag{22.98}$$

где ρ — радиальное расстояние, $\rho = \sqrt{x^2+y^2}$; фокусное расстояние дается выражением

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right), \tag{22.99}$$

где R_1 и R_2 — радиусы кривизны сферической линзы [см. (2.40) и (2.42)]. Если показатель преломления материала линзы n зависит от длины волны, то f будет зависеть от частоты ν. Материальная дисперсия приводит к хроматической аберрации, которая играет важную роль в искажении ультракоротких оптических импульсов.

С помощью этих уравнений можно в принципе определить пространственно-временну́ю зависимость выходной волны для любой входной импульсной волны и для любой системы, являющейся комбинацией свободного пространства, линз и апертур.

Система оптического преобразования Фурье

Возьмем, например, систему с дифракцией Фраунгофера, включающую распространение монохроматической волны между передней и задней фокальными плоскостями линзы. Эта система описывается функцией отклика на импульсное воздействие

$$h(x,\ x',\ \nu) \approx \frac{j\nu}{cz}\exp\left(-j\frac{2\pi\nu}{cf}xx'\right), \tag{22.100}$$

которая соответствует операции пространственного преобразования Фурье для монохроматического света (см. разд. 4.2). Для простоты не будем рассматривать зависимость от y. Для этой системы характерна сильная пространственно-временна́я связь, т. е. временна́я зависимость в данной точке плоскости выхода сильно зависит от пространственного распределения волны на входной плоскости. Аналогично, пространственное распределение на выходной плоскости чувствительно к временно́й зависимости входного поля.

Для иллюстрации рассмотрим частный случай, когда волновая функция на входе имеет раздельную зависимость от координаты и времени, $U_1(x, y) = g(t)p(x)$. Для получения такого поля можно пропустить плоскую импульсную волну с амплитудой $g(t)$ через пространственный модулятор света (ПМС) с частотно-независимым пропусканием $p(x)$, как показано на рис. 22.31. Подставляя $V_1(x, \nu) = G(\nu)p(x)$, где $G(\nu)$ — Фурье-образ функции $g(t)$, в (22.100) и (22.96), получаем поле на выходной плоскости

$$V_2(x, \nu) \propto j\nu G(\nu) P\left(\frac{\nu x}{cf}\right), \tag{22.101}$$

где $P(\nu_x)$ — пространственное преобразование Фурье функции $p(x)$,

$$P(\nu_x) = \int p(x) \exp(j2\pi\nu_x x)\,dx.$$

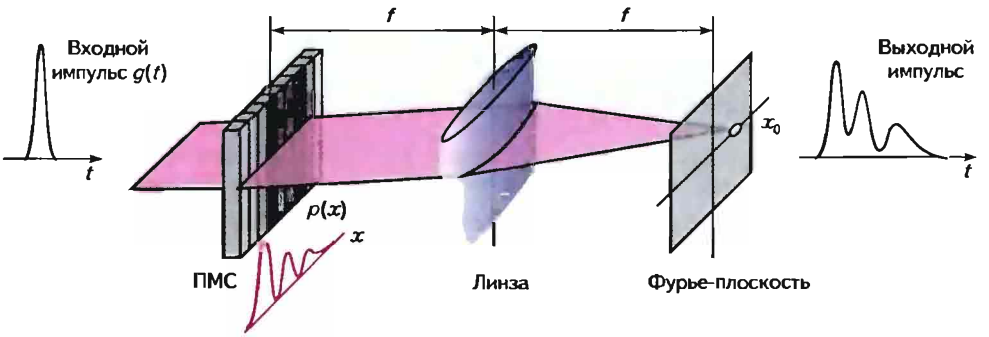


Рис. 22.31. Система пространственного преобразования Фурье связывает временны́е и пространственные распределения входного импульсного света. Форма выходного импульса в фиксированной точке определяется пространственным распределением на входе, которое контролируется ПМС

Из (22.101) очевидно, что у выходного поля пространственная и временна́я зависимости не разделяются. Временна́я форма поля волны в фиксированной точке x_0 выходной плоскости дается выражением (22.101), так что передаточная функция линейной системы, связывающей $U_2(x_0, t)$ с входным импульсом $g(t)$:

$$H(\nu) \propto j\nu P\left(\frac{\nu x_0}{cf}\right). \tag{22.102}$$

Эта *временна́я* передаточная функция является масштабированной версией пространственного Фурье-образа входного *пространственного* распределения $p(x)$. Соответствующая временна́я функция отклика на импульсное воздействие получается взятием обратного преобразования Фурье от обеих частей (22.102):

$$\boxed{h(t) \propto p\left(\frac{tcf}{x_0}\right),} \qquad (22.103)$$

Пространственно-временна́я конверсия

так что величина функции $h(t)$ в момент времени t контролируется коэффициентом пропускания ПМС в одной и только одной точке $x = (cf/x_0)t$. И наоборот, коэффициент пропускания маски в точке x управляет значением функции отклика на импульсное воздействие в один и только один момент времени $t = (x_0/cf)x$. Система осуществляет прямую пространственно-временну́ю конверсию, которую можно использовать для формирования импульсов. Аналогичная система формирования импульсов с помощью комбинации дифракционной решетки и ПМС обсуждалась в подразд. 22.2.4.

*22.4.3. Оптика пучков

Фурье-подход, описанный в предыдущем подразд. 22.4.2, можно применить в исследованию импульсных гауссовых пучков. Рассмотрим вначале гауссов пучок, промодулированный в плоскости перетяжки импульсом $g(t)$, т. е.

$$U_1(x,\ y,\ t) = g(t)\exp\left(-\frac{\rho^2}{W_0^2}\right) \quad \text{или} \quad V_1(x,\ y,\ t) \propto G(\nu)\exp\left(-\frac{\rho^2}{W_0^2}\right),$$

где W_0 — радиус пучка; $G(\nu)$ — Фурье-образ функции $g(t)$. На произвольном расстоянии z пространственно-временна́я волновая функция определяется с помощью (22.96) и (22.97),

$$V_2(x,\ y,\ \nu) \propto \nu G(\nu)\frac{jz_0(\nu)}{z + jz_0(\nu)}\exp\left(-j\frac{\pi\nu}{c}\frac{\rho^2}{x + jz_0(\nu)}\right), \qquad (22.104)$$

где $z_0(\nu)$ — дифракционная (рэлеевская) длина на частоте,

$$z_0(\nu) = \frac{\pi W_0^2}{c}\nu. \qquad (22.105)$$

Формула (22.104) представляет собой стандартное выражение для волновой функции гауссова пучка [см. (3.5)] с явным указанием частотной зависимости дифракционной длины. Радиус пучка и радиус кривизны, определяемые формулами (3.8) и (3.9), также зависят от частоты.

Если спектральная полоса достаточно узкая, то в соответствии с квазистационарным приближением пространственное распределение гауссова пучка

можно аппроксимировать его величиной на центральной частоте $\nu \approx \nu_0$ и, следовательно, зависимости от координат и времени разделяются, как описано ранее формулой (22.25). Для ультракоротких (широкополосных) импульсов это приближение неприменимо.

Временной профиль импульса можно определить в произвольной точке (ρ, z) путем обратного преобразования Фурье (22.104). В общем случае задача требует численного решения.

Гауссов пучок, промодулированный гауссовым импульсом

Если исходная волна промодулирована гауссовым импульсом

$$g(t) = \exp\left(-\frac{t^2}{\tau^2}\right)\exp(j2\pi\nu_0 t),$$

то ее Фурье-образ

$$G(\nu) \propto \exp\left[-\pi^2\tau_0^2(\nu-\nu_0)^2\right]$$

также гауссов. Приближенное аналитическое выражение для $V_2(x, y, \nu)$ в дальней зоне [$z \gg z_0(\nu)$ для всех ν] можно получить следующим образом. Множитель

$$\frac{1}{z + jz_0(\nu)} = \frac{1}{z}\frac{1}{1 + jz_0(\nu)/z}$$

в показателе экспоненты (22.104) аппроксимируется выражением $z^{-1}[1 - jz_0(\nu)/z]$, а такой же множитель в амплитуде аппроксимируется как z^{-1}. Используя (22.105), мы получаем выражение, справедливое в дальней зоне

$$V_2(x, y, \nu) \propto j\nu\exp\left[-\pi^2\tau_0^2(\nu-\nu_0)^2\right]\exp\left(-\pi^2\frac{W_0^2\rho^2}{c^2z^2}\nu^2\right)\exp\left(-j2\pi\nu\frac{\rho^2}{2cz}\right). \quad (22.106)$$

Теперь можно определить обратное преобразование Фурье (22.106). Фазовый множитель в экспоненте эквивалентен временной задержке $\rho^2/2cz$. Множитель $j\nu$ в амплитуде эквивалентен производной $\partial/\partial t$. Стоящие в середине две гауссовы функции объединяются в одну гауссову функцию ν, результат обратного преобразования Фурье которой является новой гауссовой функцией. Результат можно представить в нормированном виде

$$\boxed{U_2(x, y, t) \propto \frac{\exp\left[-\pi N\rho^2/\left(\rho^2+\rho_0^2\right)\right]}{1+\rho^2/\rho_0^2}\frac{\exp\left(-t_\rho^2/\tau_\rho^2\right)}{1+jt_\rho/\pi N\tau_0}\exp\left(-j2\pi\nu_\rho t_\rho\right),} \quad (22.107)$$

Гауссов пучок, промодулированный гауссовым импульсом

где t_ρ — зависящее от координат время задержки,

$$t_\rho = t - \frac{\rho^2}{2cz} = t - \pi N\tau_0\left(\frac{z}{z_0}\right)\left(\frac{\rho^2}{2\rho_0^2}\right); \quad (22.108)$$

τ_ρ — зависящая от координат постоянная времени,

$$\tau_\rho = \tau_0 \sqrt{1 + \frac{\rho^2}{\rho_0^2}}; \tag{22.109}$$

ν_ρ — зависящая от координат центральная частота,

$$\nu_\rho = \frac{\nu_0}{1 + \rho^2/\rho_0^2} = \frac{N/\tau_0}{1 + \rho^2/\rho_0^2}; \tag{22.110}$$

N — число оптических периодов в пределах ширины исходного импульса τ_0,

$$N = \nu_0 \tau_0 \tag{22.111}$$

и

$$\rho_0 = \pi N W(z) = \pi N W_0 \frac{z}{z_0}, \tag{22.112}$$

где $W(z) = W_0 z/z_0$ — радиус пучка в дальней зоне для стационарной волны на центральной частоте ν_0, и z_0 — соответствующая дифракционная длина,

$$z_0 = \frac{\pi W_0^2}{\lambda_0}.$$

Как функция нормированного поперечного расстояния ρ/ρ_0 и нормированного времени t/τ_0 волновая функция в дальней зоне полностью описывается двумя свободными параметрами: N и отношением z/z_0.

Интенсивность $I_2(x, y, t) = |U_2(x, y, t)|^2$ равна

$$I_2(x, y, t) \propto \frac{\exp\left[-2\pi N \rho^2/\left(\rho^2 + \rho_0^2\right)\right]}{1 + \rho^2/\rho_0^2} \frac{\exp\left(-2t_\rho^2/\tau_\rho^2\right)}{1 + t_\rho^2/\pi^2 N^2 \tau_0^2}. \tag{22.113}$$

Это универсальная функция t/τ_0 и ρ/ρ_0, характеризуемая единственным свободным параметром N. Спектральная интенсивность $S_2(x, y, \nu) = |V_2(x, y, \nu)|^2$ равна

$$S_2(x, y, \nu) \propto \frac{\nu^2}{\nu_0^2} \exp\left[-2\pi^2 N^2 \frac{\rho^2}{\rho^2 + \rho_0^2}\right] \exp\left[-2\pi^2 N^2 \frac{(\nu - \nu_\rho)^2}{\nu_\rho^2}\right], \tag{22.114}$$

что также является универсальной функцией t/τ_0 и ρ/ρ_0, характеризуемой свободным параметром N.

На основе (22.107)—(22.114) можно сделать заключение, что импульс в точке (ρ, z) в дальней зоне имеет следующие характеристики (рис. 22.32).

- Импульс задержан на время $\rho^2/2cz$, представляющее собой время распространения от центра пучка (0, 0) до точки (ρ, z).

• Временно́й профиль импульса — произведение гауссовой функции шириной

$$\tau_\rho = \tau_0 \left(1 + \frac{\rho^2}{\rho_0^2}\right)^{1/2}$$

и лоренцевой функции шириной $\pi N \tau_0$. Ширина гауссовой функции равна τ_0 при $\rho = 0$ и растет с увеличением поперечного расстояния ρ, достигая значения $\sqrt{2}\tau_0$ при $\rho = \rho_0$. Сдвиг фазы $\operatorname{arctg}(t_p/\pi N \tau_0)$, вносимый лоренцевой функцией, — проявление эффекта Гюи (см. подразд. 3.1.2) для импульсных гауссовых пучков.

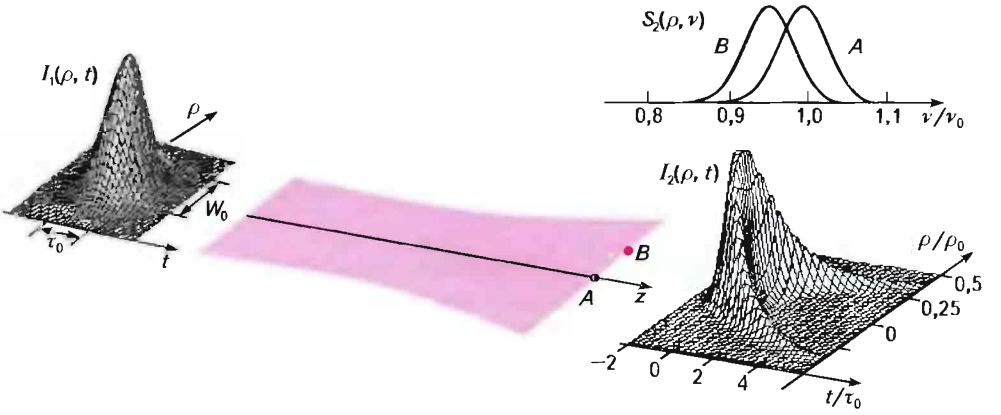


Рис. 22.32. Временно́е и пространственное расплывание гауссова пучка, промодулированного гауссовым импульсом. Вначале пучок имеет радиус W_0 и временну́ю длительность τ_0 (см. левый поверхностный график). Интенсивность в дальней зоне $I_2(\rho, t)$ показана на правом поверхностном графике. Время нормировано на начальную длительность импульса τ_0, поперечное расстояние — на ρ_0, а интенсивность выражена в произвольных единицах. В фиксированный момент времени t функция $I_2(\rho, t)$ дает мгновенный снимок интенсивности как функции положения точки. Она меняется от одногорбого распределения при $t = 0$ до двугорбой функции при $t = \tau_0$ и, в конце концов, превращается в два отдельных пика при $t = 2\tau_0$ и далее. Временно́й профиль при фиксированном положении также изображается этой поверхностью. В центре пучка импульс имеет наименьшую длительность. Во внеосевых точках импульс ослабляется, задерживается и удлиняется. Спектральная интенсивность $S_2(\rho, \nu)$ показана (справа наверху) как функция нормированной частоты ν/ν_0 в двух точках A и B и нормирована так, что пиковое значение равно единице для каждой точки. На этом графике $N = \nu_0\tau_0 = 5$, т. е. импульс состоит из пяти оптических периодов. Например, $N = 5$ для импульса с центральной частотой $\nu_0 = 750$ ТГц ($\lambda_0 = 400$ нм) и длительностью $\tau_0 = 6{,}67$ фс. Если $W_0 = 1$ мм, то $z_0 = 7{,}85$ м, $W_0 z/z_0 = 5$ мм и $\rho_0 \approx 8$ см

• Центральная частота импульса ν_ρ зависит от поперечного расстояния ρ. Начиная со значения ν_0 на оси ($\rho = 0$), она монотонно убывает с ростом ρ, достигая $\nu_0/2$ при $\rho = \rho_0$. Это следствие того факта, что длинноволновые (низ-

кочастотные) компоненты импульса расходятся более широкими конусами, как показано на рис. 22.32. По этой же причине чем дальше точка от оси пучка, тем меньше спектральная ширина и больше длительность импульса.

• Первоначальное гауссово пространственное распределение резко меняется с ростом t. Имевшее вначале один пик распределение расширяется, затем его вершина становится плоской и в конечном счете распределение становится двугорбым по мере затухания (см. рис. 22.32).

Фокусировка импульсного пучка

Если пучок с произвольным пространственным распределением $p(x, y)$, промодулированный импульсом произвольной временно́й формы $g(t)$, пропускается через линзу с фокусным расстоянием f, за которой следует участок свободного пространства длиной z, то подставляя $U_1(x, y, t) = g(t)p(x, y)$ в (22.96) и (22.97), получаем

$$V_2(x, y, \nu) \propto G(\nu)\iint \mathrm{d}x'\mathrm{d}y' p(x', y')\exp\left(-j\frac{2\pi\nu}{c}\mathrm{d}_0\right)\times$$
$$\times\exp\left(j\frac{2\pi\nu}{c}\frac{x'^2 + y'^2}{2f}\right)\exp\left(-j\frac{2\pi\nu}{c}\frac{(x-x')^2 + (y-y)^2}{2z}\right), \qquad (22.115)$$

где $G(\nu)$ — Фурье-образ. Здесь мы предположили, что линза имеет апертуру, перекрывающую пучок.

Если материал линзы не имеет дисперсии, так что n и f не зависят от ν, то в точках фокальной плоскости $z = f$ (22.115) упрощается:

$$V_2(x, y, \nu) \propto \nu G(\nu) P\left(\frac{\nu x}{cf}, \frac{\nu y}{cf}\right)\exp\left(-j\frac{2\pi\nu}{c}\frac{\rho^2}{2f}\right), \qquad (22.116)$$

где $P(\nu_x, \nu_y)$ — пространственное преобразование Фурье функции $p(x, y)$,

$$P(\nu_x, \nu_y) = \iint \mathrm{d}x\mathrm{d}y\, p(x, y)\exp\left[j2\pi(\nu_x x + \nu_y y)\right].$$

Множитель $\exp(-j2\pi\nu\mathrm{d}_0/c)$ опущен, поскольку он теперь представляет простое запаздывание по времени. Волновая функция в фокальной плоскости является результатом обратного временно́го преобразования Фурье функции $V_2(x, y, \nu)$, так что

$$U_2(x, y, t) \propto \int \nu G(\nu) P\left(\frac{\nu x}{cf}, \frac{\nu y}{cf}\right)\exp\left[j2\pi\nu\left(t - \frac{\rho^2}{2cf}\right)\right]\mathrm{d}\nu. \qquad (22.117)$$

Связь временны́х и пространственных характеристик импульсного пучка очевидна из (22.117). Кроме того, что временна́я задержка $t - \rho^2/2cf$ зависит от координат, Фурье-образ исходного пространственного профиля масштабируется частотно-зависимым множителем cf/ν, прежде чем быть усредненным по спектральному распределению импульса.

Например, для гауссова пучка

$$p(x,\ y) = \exp\left(-\frac{\rho^2}{W_0^2}\right),$$

промодулированного гауссовым импульсом

$$g(t) = \exp\left(-\frac{t^2}{\tau_0^2}\right)\exp(j2\pi\nu_0 t),$$

т. е.

$$G(\nu) \propto \exp\left[-\pi^2\tau_0^2(\nu-\nu_0)^2\right];\quad P(\nu_x,\ \nu_y) = \exp\left[-\pi^2 W_0^2\left(\nu_x^2+\nu_y^2\right)\right]$$

формула (22.117) дает

$$V_2(x,\ y,\ \nu) \propto \nu\exp\left[-\pi^2\tau_0^2(\nu-\nu_0)^2\right]\exp\left[-\left(\frac{\pi W_0}{cf}\right)^2 v^2\rho^2\right]\exp\left(-j2\pi\nu\frac{\rho^2}{2cf}\right). \quad (22.118)$$

Это выражение совпадает с выражением для гауссова пучка в дальней зоне (22.106) при $z = f$. Следовательно, соответствующая волновая функция $U_2(x, y, t)$ дается выражениями (22.107)—(22.114) при $z = f$. Графики, приведенные на рис. 22.32, применимы здесь при $z = f$, причем в качестве z_0 нужно брать дифракционную длину исходного (а не сфокусированного) пучка и

$$\rho_0 = \frac{\pi N W_0 f}{z_0} = \frac{N\lambda_0 f}{W_0} = \pi N W_0', \quad (22.119)$$

где W_0' — радиус в фокальной плоскости для стационарного пучка с длиной волны λ_0 [см. (3.45) и (3.47)],

$$W_0' = \frac{\lambda_0 f}{\pi W_0}.$$

Как и ранее, $N = \nu_0\tau_0$ — число оптических периодов на протяжении исходного импульса. Следовательно, характеристический радиус ρ_0 в πN раз больше, чем W_0'. На рис. 22.33 показано пространственно-временно́е распределение импульса в фокальной плоскости.

*Импульсные пучки в диспергирующих средах

Процесс дифракции импульсного света в диспергирующей среде может быть сложным. Если среда линейная и однородная, то уравнение Гельмгольца $[\nabla^2 + \beta^2(\nu)]V(\boldsymbol{r},\ \nu) = 0$ описывает этот процесс при произвольных дисперсионных свойствах, характеризуемых постоянной распространения $\beta(\nu)$,

и для импульса с произвольным пространственно-спектральным профилем $V(\boldsymbol{r}, \nu)$. Как только $V(\boldsymbol{r}, \nu)$ найдено путем решения этого уравнения, соответствующая волновая функция $U(\boldsymbol{r}, t)$ легко определяется посредством обратного преобразования Фурье. Этот подход в принципе применим независимо от того, насколько сильно диспергирующей является среда и насколько коротким импульс.

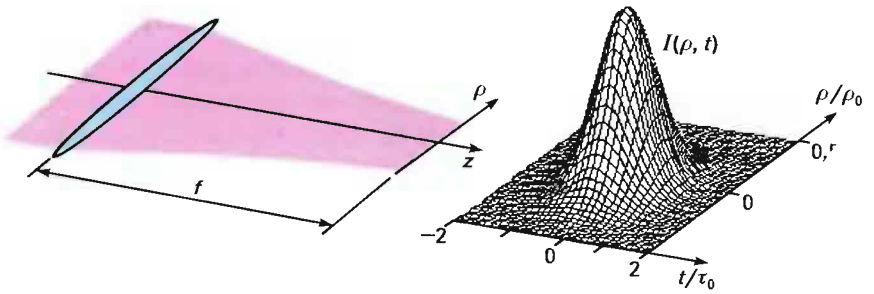


Рис. 22.33. Пространственно-временно́й профиль интенсивности в фокальной плоскости для гауссова пучка, промодулированного гауссовым импульсом и сфокусированного линзой с фокусным расстоянием f. На этом графике начальный импульс состоял из $N = 5$ оптических периодов, а начальный пучок имел дифракционную длину $z_0 \gg f$. Разница между этим профилем и профилем, показанным на рис. 22.32, обусловлена тем, что здесь в силу $f \ll z_0$ временна́я задержка $\dfrac{\rho^2}{2cf} = \tau_0 \dfrac{\pi N}{2} \dfrac{f}{z_0} \dfrac{\rho^2}{\rho_0^2}$ пренебрежимо мала в периферийных точках с $\rho < \rho_0$

Приближения, аналогичные тем, которые независимо приводят к параксиальному уравнению Гельмгольца, описывающему дифракцию, и уравнению для медленной огибающей, описывающему дисперсию импульса (см. табл. 22.3), можно сочетать для получения дифференциального уравнения в частных производных для огибающей $\mathcal{A}(\boldsymbol{r}, t)$ импульса с узким спектральным распределением. Процедура, аналогичная описанной в подразд. 22.2.3, приводит к обобщенному параксиальному волновому уравнению

$$\boxed{-\lambda_0 \nabla_T^2 \mathcal{A} + D_\nu \frac{\partial^2 \mathcal{A}}{\partial t^2} + j\left(\frac{\partial \mathcal{A}}{\partial z} + \frac{1}{v}\frac{\partial \mathcal{A}}{\partial t}\right) = 0.} \qquad (22.120)$$

Обобщенное параксиальное волновое уравнение

Это уравнение обобщает (22.24), которое применимо к недиспергирующим средам ($D_\nu = 0$), и (2.27), применимое в стационарном случае, когда

$$\frac{\partial^2 A}{\partial t^2} = \frac{\partial A}{\partial t} = 0.$$

Доказательство обобщенного параксиального волнового уравнения

Волновая функция и ее Фурье-образ связаны с огибающей и ее Фурье-образом соотношениями

$$U(\boldsymbol{r},\ t) = \mathcal{A}(\boldsymbol{r},\ t)\exp(-j\beta_0 z)\exp(j2\pi\nu_0 t);\quad V(\boldsymbol{r},\ \nu) = A(\boldsymbol{r},\ \nu-\nu_0)\exp(-j\beta_0 z).$$

Параксиальное приближение

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\left[A\exp(-j\beta_0 z)\right] \approx \left(-j2\beta_0\frac{\mathrm{d}A}{\mathrm{d}z} - \beta_0^2 A\right)\exp(-j\beta_0 z)$$

можно использовать для сведения уравнения Гельмгольца к виду

$$\left[\nabla_T^2 - j2\beta_0\frac{\mathrm{d}}{\mathrm{d}z}\right]A + \left[\beta^2(\nu_0+f) - \beta_0^2\right]A = 0. \qquad (22.121)$$

Для слабой дисперсии используем приближение

$$\beta^2(\nu_0+f) - \beta_0^2 \approx 2\beta_0\left[\beta(\nu_0+f) - \beta_0\right]$$

вместе с трехчленным разложением Тейлора

$$\beta(\nu_0+f) \approx \beta_0 + 2\pi\beta' f + 2\pi^2\beta'' f^2$$

Тогда уравнение Гельмгольца сводится к уравнению

$$\nabla_T^2 A - j2\beta_0\frac{\partial A}{\partial z} + 2\beta_0\left[2\pi f\beta' + 2\pi^2 f^2\beta''\right]A = 0. \qquad (22.122)$$

Выполняя обратное преобразование Фурье и замечая, что множители $j2\pi f$ и $-4\pi^2 f^2$ эквивалентны производным $\partial/\partial t$ и $\partial^2/\partial t^2$ соответственно, мы получаем

$$\nabla_T^2 A - 2\beta_0\left[j\frac{\partial A}{\partial z} + j\beta'\frac{\partial A}{\partial t} + \frac{1}{2}\beta''\frac{\partial^2 A}{\partial t^2}\right] = 0. \qquad (22.123)$$

Наконец, подставляя

$$\beta' = \frac{1}{v};\quad \beta'' = \frac{D_\nu}{2\pi};\quad \beta_0 = \frac{2\pi}{\lambda_0},$$

получаем (22.120).

Параксиальное уравнение для медленной огибающей имеет пространственно-временно́е гауссово решение

$$A(x,\ y,\ z,\ t) = A_0\sqrt{\frac{-jz_0'}{z-jz_0'}}\exp\left[-j\frac{\pi}{D_\nu}\frac{t-z/v}{z-jz_0'}\right]\frac{jz_0}{z+jz_0}\exp\left(-j\frac{\pi}{\lambda}\frac{\rho^2}{z+jz_0}\right), \qquad (22.124)$$

которое имеет начальную огибающую

$$A(x,\ y,\ 0,\ t) = A_0 \exp\left(-\frac{t^2}{\tau_0^2}\right)\exp\left(-\frac{\rho^2}{W_0^2}\right),$$

гауссову в пространстве и времени, где z_0' и z_0 — дисперсионная длина, связанная с начальной длительностью импульса τ_0, и дифракционная длина, связанная с начальным радиусом пучка W_0 соответственно,

$$z_0' = \frac{\pi\tau_0^2}{D_\nu};\quad z_0 = \frac{\pi W_0^2}{\lambda}.$$

Это решение сочетает в себе дифракцию гауссова пучка (см. гл. 3) и дисперсию гауссова импульса (см. разд. 22.3) с разделением пространственной и временно́й зависимости, как показано на рис. 22.34.

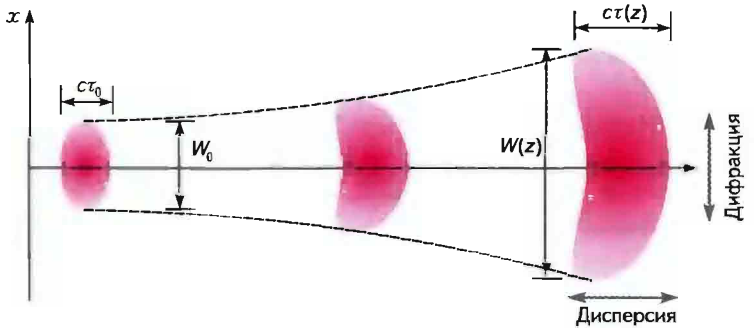


Рис. 22.34. Три снимка пространственного распределения импульса в процессе его распространения в линейной диспергирующей среде. Из-за дифракции импульс расплывается в поперечном направлении x. Из-за дисперсии он расплывается во времени (что показано здесь как пространственное расплывание в направлении распространения z)

Поскольку (22.120) и (22.124) допускают разделение пространственных и временны́х переменных, мы приходим к заключению, что приближения, использованные при получении этих уравнений, фактически равносильны квазистационарному приближению, описанному в подразд. 22.1.3.

*Уравнение для огибающей предельно узкого импульсного пучка

Когда условия приближения медленной огибающей не выполняются (т. е. импульс очень короткий, а пучок очень тонкий), разделение пространственных и временны́х переменных становится невозможным. Дифференциальное уравнение, описывающее огибающую импульса, принимает более сложную форму, да и сама концепция огибающей в этом случае менее информативна. Исходя из уравнения Гельмгольца

$$\left[\nabla^2 + \beta^2(\nu)\right]V(\boldsymbol{r},\ \nu) = 0$$

и подставляя

$$V(\boldsymbol{r},\ \nu) = A(\boldsymbol{r},\ \nu - \nu_0)\exp(-j\beta_0 z);\quad \nu = \nu_0 + f,$$

получаем

$$\left(\nabla_T^2 + \frac{\partial^2}{\partial z^2} - j2\beta_0\frac{\partial}{\partial z}\right)A + \left[\beta^2(\nu_0 + f) - \beta_0^2\right]A = 0.$$

Разлагая функцию $[\beta^2(\nu_0 + f) - \beta_0^2]$ в ряд Тейлора до второго порядка, получаем

$$\left[\beta^2(\nu_0 + f) - \beta_0^2\right] \approx (2\beta_0\beta')2\pi f + \frac{1}{2}(2\beta'^2 + 2\beta_0\beta'')(2\pi f)^2$$

Производя обратный переход во временну́ю область и перегруппировывая члены, получаем

$$-\lambda_0\nabla_T^2\mathcal{A} + D_\nu\frac{\partial^2\mathcal{A}}{\partial t^2} + j4\pi\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right)\mathcal{A} - \lambda_0\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\mathcal{A} = 0, \qquad (22.125)$$

где $v = 1/\beta'$ и $D_\nu = 2\pi\beta''$. Уравнение (22.125) является более общим, чем (22.120), поскольку приближения параксиальности и слабой дисперсии не использовались при его получении. Если

$$\beta'^2 \ll \beta_0\beta''\ \left(\text{или}\ \frac{\lambda_0}{v^2} \ll D_\nu\right);\quad \frac{\partial^2\mathcal{A}}{\partial z^2} \ll \frac{4\pi}{\lambda_0}\frac{\partial\mathcal{A}}{\partial z},$$

то четвертый член в (22.125) пренебрежимо мал, и (22.125) превращается в (22.120).

Уравнение (22.125) можно записать в системе координат, движущейся со скоростью импульса, с помощью преобразования $t' \to t - z/v$ и $z' = z$. В результате получается дифференциальное уравнение

$$-\lambda_0\nabla_T^2\mathcal{A} + D_\nu\frac{\partial^2\mathcal{A}}{\partial t'^2} + j4\pi\frac{\partial\mathcal{A}}{\partial z'} - \lambda_0\left(\frac{\partial^2\mathcal{A}}{\partial z'^2} - \frac{2}{v}\frac{\partial^2\mathcal{A}}{\partial t'\partial z'}\right) = 0, \qquad (22.126)$$

в котором явно проявляется неразделимость пространственных и временны́х переменных.

22.5. НЕЛИНЕЙНАЯ ОПТИКА УЛЬТРАКОРОТКИХ ИМПУЛЬСОВ

Предыдущие разделы настоящей главы касались распространения оптических импульсов в *линейных* средах, причем акцент был сделан на роли дисперсии групповой скорости (ДГС) в изменении формы коротких импульсов. В данном разделе мы рассмотрим распространение оптических импульсов

в *нелинейных* средах. Нелинейные эффекты чаще встречаются в оптике ультракоротких импульсов из-за их высокой интенсивности. Нелинейные оптические явления были введены в рассмотрение в гл. 21, в частности, рассматривались трехволновое смешение в средах с квадратичной нелинейностью, двух- и четырехволновое смешение в средах с нелинейностью третьего порядка. В этом разделе некоторые из этих явлений будут рассмотрены в импульсном режиме. Раздел 22.5 посвящен импульсным параметрическим процессам, включая трехволновое смешение, оптическое выпрямление и самомодуляцию фазы. В подразд. 22.5.2 рассматриваются оптические солитоны, а разд. 22.5 посвящен генерации суперконтинуума.

22.5.1. Импульсные параметрические процессы

Трехволновое смещение в среде с нелинейностью второго порядка обсуждалось в подразд. 21.2.3 для стационарных волн, а теория связанных волн была развита в разд. 21.4. Принципиальные условия смешения волн определяются законами сохранения энергии и импульса. Для импульсных волн с центральными угловыми частотами ω_1, ω_2, ω_3 и центральными волновыми векторами $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, $\boldsymbol{k}_3$ эти условия имеют вид $\omega_1 + \omega_2 = \omega_3$ и $\boldsymbol{k}_1 + \boldsymbol{k}_2 = \boldsymbol{k}_3$. Если дисперсией можно пренебречь, то стационарная теория применима и к импульсному случаю, т. е. импульс рассматривается как квазистационарный в каждый момент времени своего существования, а огибающие трех волн подчиняются тем же связанным уравнениям (21.112).

Эффект сноса

Если среда обладает дисперсией первого порядка, но не имеет дисперсии второго (ДГС) или высших порядков, то все три импульсные волны распространяются со своими групповыми скоростями без изменения своей формы (только их амплитуды меняются в результате процесса смешения). Поскольку эти скорости, вообще говоря, различны, импульсы в конце концов разделяются и параметрический процесс, отвечающий за смешение волн, исчезает. Этот процесс известен как **эффект сноса**. Следовательно, для эффективного смешения импульсных волн требуется выполнение дополнительного условия — равенство групповых скоростей $v_1 = v_2 = v_3$. Эффект сноса иллюстрируется рис. 22.35 в вырожденном случае коллинеарной генерации второй гармоники ($\omega_1 = \omega_2 = \omega$ и $\omega_3 = 2\omega$).

Трудно одновременно удовлетворить и условия согласования частот и фаз, и условие равенства групповых скоростей одновременно. В подразд. 21.2.4 и 21.4.1 было показано, что при наличии фазовой расстройки Δk вторая гармоника заметно уменьшается на расстоянии $L_к = 2\pi/|\Delta k|$, которое называется длиной когерентности [см. (21.35)]. При рассогласовании групповых скоростей, характеризуемых величиной $\Delta\beta' = 1/v_3 - 1/v_1$, импульсы расходятся во времени на $\Delta\beta' z = z/v_3 - z/v_1$ после прохождения расстояния z. Если эта задержка равна

длительности импульса τ, то импульсы более не перекрываются и нелинейное взаимодействие исчезает. Это происходит на расстоянии

$$\boxed{L_g = \frac{\tau}{|\Delta\beta'|}}, \qquad (22.127)$$

Длина сноса

называемом длиной сноса. Наименьшая из длин $L_к$ и L_g диктует, какой из двух эффектов — расстройка синхронизма или расстройка групповых скоростей — преобладает.

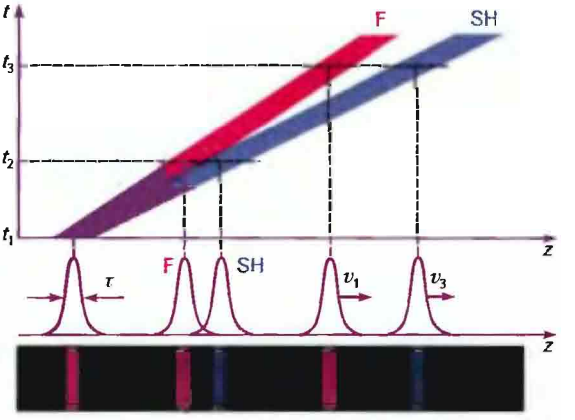


Рис. 22.35. Импульсная волна на основной частоте (F) и связанная с ней волна второй гармоники (SH) разделяются в процессе распространения с разными скоростями (в данном примере волна SH быстрее). Верхний график — пространственно-временна́я диаграмма для импульсов длительности τ. Нижняя схема показывает три мгновенных снимка бегущих импульсов в моменты времени $t_1 < t_2 < t_3$

Например, для кристалла KDP при использовании обыкновенной основной волны с $\lambda_1 = 2{,}06$ мкм и необыкновенной волны второй гармоники с $\lambda_3 = 0{,}53$ мкм в о-н-о-конфигурации типа II расстройка из-за групповой скорости

$$\Delta\beta' = 2\left(\frac{1}{v_3} - \frac{1}{v_1}\right) \approx 5{,}2 \cdot 10^{-10} \text{ с/м}.$$

Для импульса длительностью 100 фс длина сноса

$$L_g = \frac{\tau}{|\Delta\beta'|} \approx 0{,}2 \text{ мм}.$$

*Уравнения связанных волн для импульсного трехволнового смешения

Уравнения для связанных волн, которые были выведены в разд. 21.4 для стационарных волн, можно легко обобщить на импульсные волны. Для коллинеарных плоских волн, бегущих в направлении z, электрические поля выражаются через комплексные огибающие как

$$\mathcal{E}_q = \mathrm{Re}\left\{\sqrt{2\eta\hbar\omega_q}\, a_q(z,\ t)\exp\left[j\left(\omega_q t - \beta_q z\right)\right]\right\}, \quad q = 1,\ 2,\ 3,$$

где a_1, a_2, a_3 — нормированные комплексные огибающие трех импульсов; β_1, β_2, β_3 — постоянные распространения на центральных частотах ω_1, ω_2 и ω_3. Используя приближение медленно меняющейся огибающей и двучленное разложение Тейлора для постоянной распространения $\beta(\omega)$ вблизи каждой из трех центральных частот

$$\beta(\omega_q + \Omega) \approx \beta_q + \Omega\beta_q',$$

где β_q' — производная $\partial\beta/\partial\omega$ в точке ω_q, мы получим связанные уравнения

$$\begin{aligned}
\left(\frac{\partial}{\partial z} + \frac{1}{v_1}\frac{\partial}{\partial t}\right)a_1 &= -jga_3a_2^*;\\
\left(\frac{\partial}{\partial z} + \frac{1}{v_2}\frac{\partial}{\partial t}\right)a_2 &= -jga_3a_1^*;\\
\left(\frac{\partial}{\partial z} + \frac{1}{v_3}\frac{\partial}{\partial t}\right)a_3 &= -jga_1a_2,
\end{aligned} \qquad (22.128)$$

где $v_q = 1/\beta_q'$ — групповая скорость волны ω_q; g — постоянная, определяемая формулой (21.93). Эти уравнения аналогичны связанным уравнениям в стационарном случае (21.92). Если групповые скорости равны, $v_1 = v_2 = v_3 = v$, то, переходя в координатную систему, движущуюся со скоростью v, можно свести связанные уравнения (22.128) к стационарному виду (21.92) и использовать решения, представленные в разд. 21.4, с заменой z на $z - vt$. Если групповые скорости не равны, решение (22.128) становится более сложным.

Если среда имеет также ДГС (см. задачу 2 к разд. 22.5), то трехчленное разложение Тейлора

$$\beta(\omega_q + \Omega) \approx \beta_q + \Omega\beta_q' + \frac{1}{2}\Omega\beta_q''$$

приводит к связанным уравнениям

$$\begin{aligned}
\left(\frac{\partial}{\partial z} + \frac{1}{v_1}\frac{\partial}{\partial t} - j\frac{\beta_1''}{2}\frac{\partial^2}{\partial t^2}\right)a_1 &= -jga_3a_2^*;\\
\left(\frac{\partial}{\partial z} + \frac{1}{v_2}\frac{\partial}{\partial t} - j\frac{\beta_2''}{2}\frac{\partial^2}{\partial t^2}\right)a_2 &= -jga_3a_1^*;\\
\left(\frac{\partial}{\partial z} + \frac{1}{v_3}\frac{\partial}{\partial t} - j\frac{\beta_3''}{2}\frac{\partial^2}{\partial t^2}\right)a_3 &= -jga_1a_2.
\end{aligned} \qquad (22.129)$$

Оптическое выпрямление в импульсном режиме: генерация терагерцовых импульсов

Импульсная волна с центральной частотой в оптическом диапазоне и со спектральной шириной порядка терагерца может быть преобразована в импульс терагерцевого излучения. По существу импульс смещается в терагерцевую область из оптической, как если бы он подвергался выпрямлению. На рис. 22.36 дана схематическая иллюстрация процесса.

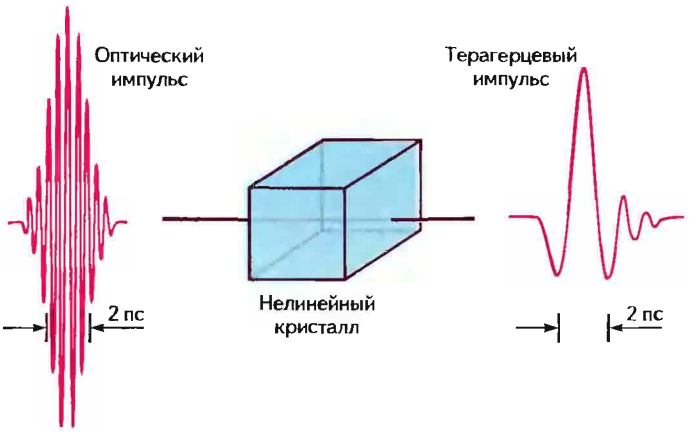


Рис. 22.36. Генерация терагерцового импульса путем преобразования частоты оптической волны вниз

При распространении оптического импульса $\mathcal{E}(t) = \mathrm{Re}\{\mathcal{A}(t)\exp(j\omega_0 t)\}$ с медленно меняющейся амплитудой $\mathcal{A}(t)$ через среду с коэффициентом нелинейности второго порядка d он индуцирует плотность поляризации $2d\mathcal{E}^2(t)$, в которой содержится компонента на частоте $2\omega_0$, отвечающая на генерацию второй гармоники, и другая компонента

$$\mathcal{P}_{\text{ТГц}} = d\,|\mathcal{A}(t)|^2 \qquad (22.130)$$

представляющая оптическое выпрямление (см. подразд. 21.2.1, 21.2.3 и 21.4.2).

Чтобы определить необходимые условия синхронизма для этого параметрического процесса, используем метод Фурье. Импульсная оптическая волна может рассматриваться как сумма монохроматических волн с частотами, занимающими спектральную полосу вокруг центральной частоты ω_0. При прохождении нелинейной среды эти монохроматические компоненты смешиваются парами, каждая из которых генерирует преобразованную по частоте вниз монохроматическую волну на разностной частоте. В соответствии с (21.20д) пара волн с частотами $\omega_1 = \omega$ и $\omega_2 = \omega + \Omega$ генерирует компоненту нелинейной плотности поляризации $P_{\text{ТГц}}(\Omega) = 2dE^*(\omega)E(\omega + \Omega)$ на терагерцевой частоте Ω, и сумма таких компонент для всех пар равна

$$P_{\text{ТГц}}(\Omega) = \int 2d\,E^*(\omega)\,E(\omega + \Omega)\,\mathrm{d}\omega. \qquad (22.131)$$

Во временно́м представлении это эквивалентно (22.131) Для учета влияния нелинейной дисперсии коэффициент нелинейности d в (22.30) нужно заменить на частотно-зависимый коэффициент $d(\Omega, \omega, \omega + \Omega)$ (см. разд. 21.7).

Такой процесс преобразования частоты вниз должен удовлетворять условиям синхронизма для всех частот ω и Ω. Это условие не может быть выполнено точно, и возникает расстройка

$$\Delta k = k(\omega + \Omega) - k(\omega) - k(\Omega). \tag{22.132}$$

Если $\Omega \ll \omega$, это выражение можно записать приближенно

$$\Delta k \approx \Omega \frac{\mathrm{d}k}{\mathrm{d}\omega} - k(\Omega) = \Omega\left[\frac{1}{v(\omega)} - \frac{1}{c(\Omega)}\right] = \left[N(\omega) - n(\Omega)\right]\frac{\Omega}{c_0}, \tag{22.133}$$

где $v(\omega)$ — групповая скорость,

$$v(\omega) = \left(\frac{\mathrm{d}k}{\mathrm{d}\omega}\right)^{-1}$$

$N(\omega)$ — групповой показатель преломления на частоте ω; $c(\Omega)$ и $n(\Omega)$ — фазовая скорость и показатель преломления на терагерцевой частоте Ω. Таким образом, прибор надо проектировать так, чтобы групповой показатель преломления на оптических частотах был равен фазовому показателю преломления на терагерцевых частотах.

Как было показано в подразд. 21.2.4, для кристалла длиной L эта фазовая расстройка мала, если $L < L_к$, где $L_к$ — длина когерентности [см. (21.53)]. Для учета этой расстройки под интеграл (22.131) нужно включить множитель

$$\int_0^L \exp(j\Delta k z)\,\mathrm{d}z = \frac{\exp(j\Delta k L) - 1}{j\Delta k}.$$

Самомодуляция фазы импульсов

Самомодуляция фазы (СМФ) происходит в нелинейных средах с эффектом Керра (см. подразд. 21.3.1). Фаза $\Delta\varphi$, вносимая этим эффектом, для волны, проходящей расстояние z в среде с оптическим коэффициентом Керра n_2, равна $\Delta\varphi = -n_2 I k_0 z$, где I — оптическая интенсивность; k_0 — волновое число. Для оптического импульса интенсивность является функцией времени $I(t)$, и, следовательно, фаза меняется во времени

$$\Delta\varphi(t) = -n_2 I(t) k_0 z. \tag{22.134}$$

Это соответствует изменению мгновенной частоты [см. (22.4)]

$$\Delta\omega_i = -n_2 \frac{\mathrm{d}I}{\mathrm{d}t} k_0 z. \tag{22.135}$$

Для импульса простой формы, такого как показан на рис. 22.37, если n_2 положительно, частота задней (правой) половины импульса возрастает (сдвигается в синюю сторону), поскольку $dI/dt < 0$, в то время как частота передней (левой) половины импульса уменьшается (сдвиг в красную сторону), поскольку $dI/dt > 0$. В результате импульс приобретает положительный чирп (т. е. его мгновенная частота растет) вблизи своего центра. Отсюда следует, что СМФ можно использовать для внесения чирпа, а следовательно — для формирования импульсов (см. подразд. 22.2.3).

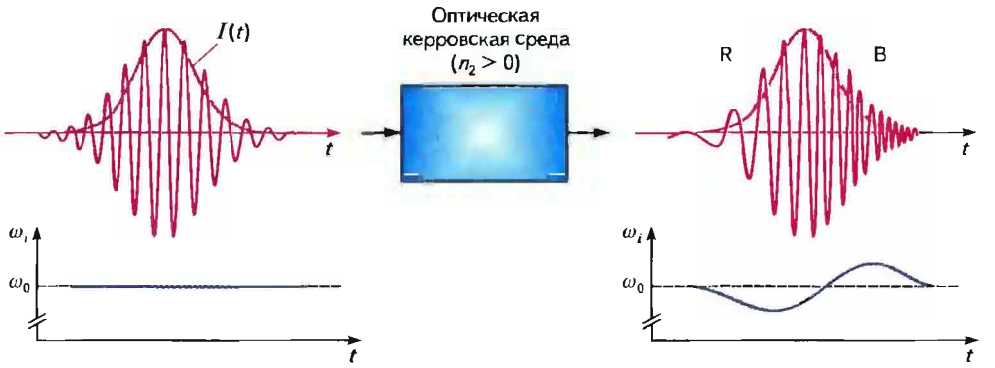


Рис. 22.37. Чирпирование оптического импульса при распространении в нелинейной среде с оптическим эффектом Керра

Например, вблизи центра гауссов импульс можно аппроксимировать параболической функцией

$$I(t) = I_0 \exp\left(-\frac{2t^2}{\tau^2}\right) \approx I_0\left(1 - \frac{2t^2}{\tau^2}\right),$$

так что зависящая от времени составляющая фазы примерно квадратично зависит от времени,

$$\Delta\varphi = 2n_2 I_0 k_0 z \frac{t^2}{\tau^2},$$

что соответствует линейному чирпу с коэффициентом чирпа $a = 2n_2 I_0 k_0 z$ того же знака, что и коэффициент Керра n_2. Самомодуляция фазы, следовательно, вносит квадратичную модуляцию фазы, где

$$\zeta = \frac{2n_2 I_0 k_0 z}{\tau^2}. \qquad (22.136)$$

Удобно записать параметр чирпа, вносимого СМФ, в виде

$$a = \frac{z}{z_{\text{нл}}}; \quad z_{\text{нл}} = \frac{1}{2n_2 I_0 k_0}, \qquad (22.137)$$

Параметр чирпа СМФ

где $|z_{нл}|$ называется **нелинейной характеристической длиной** керровской среды. Фаза, вносимая распространением через нелинейный материал на расстояние $2|z_{нл}|$ в максимуме интенсивности I_0, равна единице, т. е. $n_2 I_0 k_0 2|z_{нл}| = 1$.

В ходе проведенных рассуждений мы неявно предполагали, что среда является слабо диспергирующей, так что уширение импульса пренебрежимо мало, т. е. ДГС можно пренебречь по сравнению с СМФ. Это условие получается, если $|z_0| \gg |z_{нл}|$. Анализ распространения импульсов в материалах, обладающих как ДГС, так и СМФ, сложен, как будет видно в следующем разделе.

Квадратичная модуляция фазы, вносимая нелинейной СМФ, может быть использована в сочетании с линейным диспергирующим устройством, как описано в подразд. 22.2.3 и проиллюстрировано примером 22.7. Такая комбинация обеспечивает сжатие импульса в

$$\sqrt{1+a^2} = \sqrt{1+\left(\frac{z}{z_{нл}}\right)^2} \text{ раз.}$$

Пример 22.7

Сжатие импульса с помощью СМФ в волокне и ДГС на дифракционной решетке

Импульс длительностью 65 фс с пиковой мощностью $P_0 = 300$ кВт на центральной длине волны $\lambda_0 = 620$ нм приобретает чирп в оптическом волокне из кварцевого стекла длиной 9 мм с площадью поперечного сечения $A = 100$ мкм^2, как показано на рис. 22.38. На данной длине волны $n_2 \approx 3{,}2 \cdot 10^{-20}$ м^2/Вт, так что нелинейная характеристическая длина равна

$$|z_{нл}| = \frac{1}{2n_2 I_0 k_0} = \frac{\lambda_0 A}{4\pi |n_2| P_0} \approx 0{,}5 \text{ мм.}$$

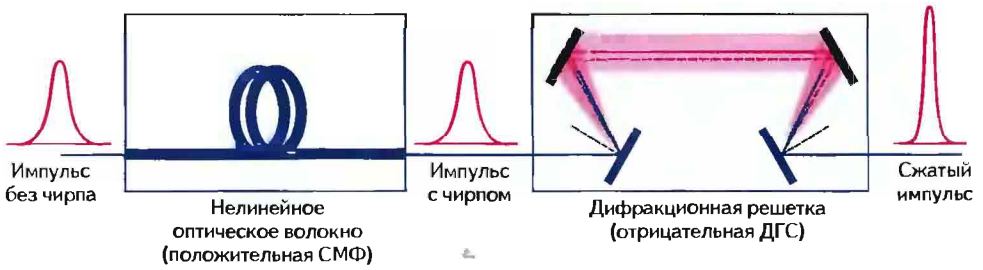


Рис. 22.38. Сжатие импульса с помощью комбинации квадратичной фазовой модуляции (КФМ), вносимой за счет СМФ, и фильтра с чирпом. Фазовый модулятор выполнен из оптического волокна, обеспечивающего СМФ за счет эффекта Керра. Фильтр с чирпом — дифракционная решетка, вносящая ДГС

Поскольку длина волокна $z = 9$ мм, параметр чирпа, вносимого СМФ, равен $a = z/z_{нл} = 18$. Это соответствует максимальному сжатию импульса в $\sqrt{1+a^2} \approx 18$ раз,

т. е. до 3,6 фс. Волокно вносит также ДГС. На 620 нм $\beta'' = 6\ \ 10^{-26}$ с^2/м, так что дисперсионная длина для импульса длительностью $\tau_0 = 65$ фс составляет

$$z_0 = \frac{\tau_0^2}{2\beta''} = 3{,}5 \text{ см}.$$

Поскольку $z_0 \gg z_{нл}$, СМФ преобладает над ДГС. Для достижения максимального сжатия с помощью дифракционной решетки должен вносится коэффициент чирпа

$$b = \frac{a}{1+a^2}\tau_0^2 = 2{,}35 \cdot 10^{-28} \text{ с}^2 = (3{,}6 \text{ фс})^2$$

22.5.2. Оптические солитоны

Сочетание самомодуляции фазы (СМФ) и дисперсии групповой скорости (ДГС) в среде, обладающей как нелинейным оптическим эффектом Керра, так и линейной дисперсией, может приводить как к расплыванию, так и к сжатию импульса, в зависимости от величины и знака этих двух эффектов. При определенных условиях оптический импульс особой формы и интенсивности может распространяться в такой среде без изменения своей формы, как если бы он распространялся в *линейной недиспергирующей* среде. Это происходит, когда ДГС полностью компенсирует влияние СМФ, как показано на рис. 22.39, *в*. Такие импульсные стационарные волны называются **уединенными волнами.** Оптические **солитоны** — особый вид уединенных волн, которые являются ортогональными в том смысле, что при пересечении в среде их профили интенсивности не меняются (в результате взаимодействия появляется только фазовый сдвиг), и каждая волна продолжает распространяться как независимый объект.

Солитонный процесс можно проиллюстрировать с помощью механической аналогии, показанной на шуточном рис. 22.40. Здесь тяжелый автомобиль представляет центральную часть оптического импульса. Он меняет поверхность земли, которая предполагается упругой, подобно тому, как интенсивный пик импульса меняет показатель преломления среды. Быстрый спортивный автомобиль, аналогичный задней части импульса, замедляется на подъеме за счет возникшего наклона поверхности. Медленный велосипед, аналогичный передней части импульса, ускоряется на спуске за счет наклона поверхности вниз. В результате этого самоподдерживающегося процесса три члена команды движутся с одинаковой скоростью, сохраняя дистанцию между собой.

Солитоны имеют специфический профиль и уровень интенсивности, при которых эффекты ДГС и СМФ находятся в балансе. Для таких импульсов эффект чирпа из-за СМФ точно компенсируется естественным расплыванием импульса за счет ДГС. Любое слабое расширение импульса усиливает процесс сжатия, а любое сужение приводит к уменьшению сжатия, так что ширина импульса поддерживается постоянной. Солитоны можно представлять себе как

моды (собственные функции) нелинейной диспергирующей системы. Математический анализ этого явления основан на решении нелинейного волнового уравнения, описывающего распространение огибающей импульса, которое будет представлено ниже. Однако сначала мы остановимся на простом выводе условия существования солитона.

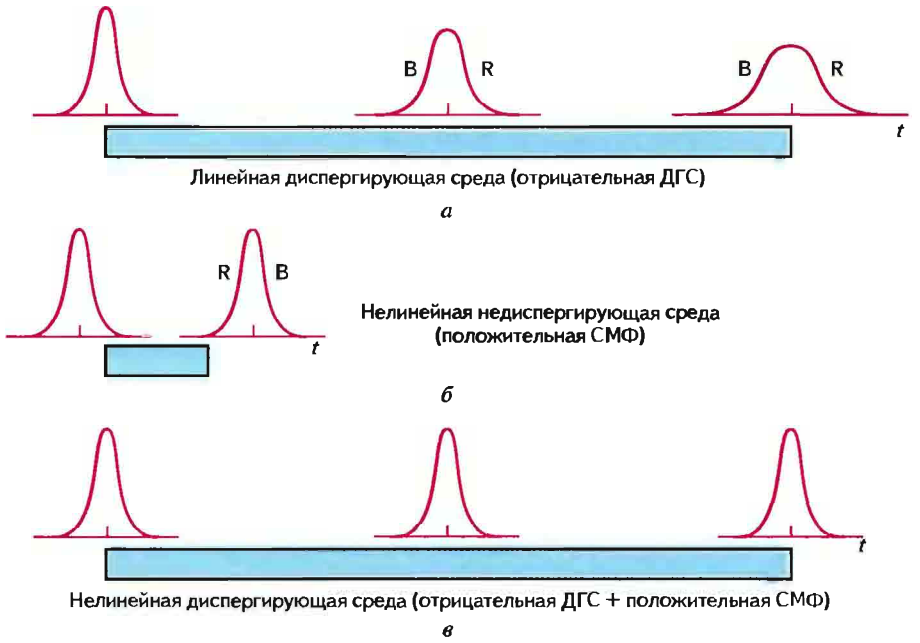


Рис. 22.39. В линейной среде с отрицательной ДГС (аномальная дисперсия) коротковолновая составляющая B имеет большую групповую скорость и, следовательно, распространяется быстрее, чем длинноволновая составляющая R (*а*). В результате импульс расплывается. В нелинейной среде с положительным оптическим эффектом Керра ($n_2 < 0$) СМФ вносит отрицательный частотный сдвиг в переднюю часть импульса (R) и положительный — в заднюю часть (B) (*б*). Импульс приобретает чирп, но его форма не меняется. При распространении волны с чирпом через линейную диспергирующую среду (*б*) импульс будет сжиматься (*а*), поскольку часть, сдвинутая в синюю сторону, догоняет ту, которая сдвинута в красную сторону. Если среда обладает как дисперсией, так и нелинейностью, то импульс может сжиматься, расширяться либо поддерживаться (образуя уединенную волну) в зависимости от величины и знака ДГС и СМФ (*в*). Эта иллюстрация показывает уединенную волну, образованную за счет баланса между отрицательной ДГС и положительной СМФ



Рис. 22.40. Транспортная аналогия солитона

Условие существования солитона

Корректное выражение условия существования солитона получается путем приравнивания к нулю суммы фаз, вносимых ДГС и СМФ на малом участке пути Δz. Как говорилось ранее в данном разделе, импульс, распространяющийся через нелинейную среду с оптическим эффектом Керра, претерпевает СМФ, вносящую квадратичную модуляцию фазы $\exp(j\zeta t^2)$ при

$$\zeta = \frac{2n_2 I_0 k_0 \Delta z}{\tau_0^2};$$

где I_0 и τ_0 — пиковая интенсивность импульса и его длительность соответственно; n_2 — оптический коэффициент Керра. Кроме того, как описано в разд. 22.3, ДГС в линейной диспергирующей среде вносит фазовый сдвиг at^2/τ_0^2, где параметр чирпа

$$a = \frac{\Delta z}{z_0} = 2\beta'' \frac{\Delta z}{\tau_0^2},$$

где β'' — коэффициент материальной дисперсии; $|z_0|$ — дисперсионная длина (см. табл. 22.2). Условие солитонного распространения импульса состоит в том, что два эти вклада в фазу должны быть равны по величине и противоположны по знаку, т. е.

$$\zeta = -\frac{a}{\tau_0^2}, \tag{22.138}$$

или

$$\boxed{k_0 n_2 I_0 = -\frac{\beta''}{\tau_0^2},} \tag{22.139}$$

Условие существования солитона (фаза)

или, что эквивалентно,

$$\boxed{z_{нл} = -z_0,} \tag{22.140}$$

Условие существования солитона (длина)

т. е. дисперсионная длина ДГС равна нелинейно характеристической длине. Другими словами, фазовый сдвиг, вносимый СМФ при длине распространения, равной удвоенной дисперсионной длине ДГС $|z_0|$, равен единице ($-k_0 n_2 I_0 2z_0 = 1$).

По-другому это условие можно вывести, представляя среду как периодическую последовательность локализованных элементов СМФ, между которыми расположены элементы, расширяющие импульс (ДГС) шириной Δz, как показано на рис. 22.41. Схема идентична системе ретрансляции импульса, описанной в упражнении 22.5. Формула (22.138) может быть выведена из (22.89) в пределе $\Delta z \to 0$.

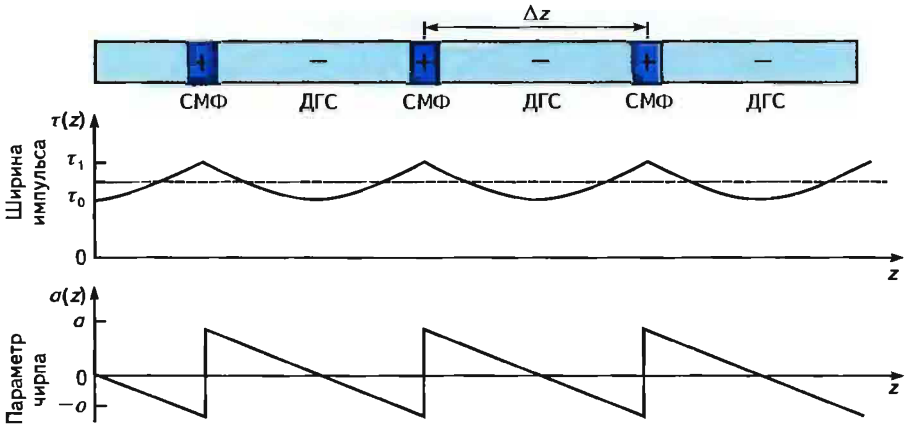


Рис. 22.41. Простая модель среды с отрицательной ДГС и положительной СМФ

Другое выражение условия существования солитона включает амплитуду импульса A_0, где $I_0 = |A_0|^2/2\eta$; η — электромагнитный импеданс среды. Результат записывается через произведение пиковой амплитуды импульса A_0 и его длительности τ_0

$$\boxed{A_0\tau_0 = \sqrt{-\frac{\beta''}{\gamma}},} \qquad (22.141)$$

Условие существования солитона (площадь)

где

$$\gamma = \frac{k_0 n_2}{2\eta} = \frac{\pi n_2}{\lambda_0 \eta} \qquad (22.142)$$

еще один материальный параметр. Предполагается, что γ и β'' имеют противоположные знаки. Таким образом, произведение пиковой амплитуды и ширины импульса $A_0\tau_0$ — постоянная величина, определяемая отношением параметра β'', описывающего ДГС, и параметра γ, описывающего СМФ. Для данного материала произведение $A_0\tau_0$ фиксировано, откуда вытекает ряд следствий.

- Пиковая амплитуда импульса A_0 обратно пропорциональна его длительности τ_0.
- Пиковая мощность обратно пропорциональна τ_0^2.
- Плотность энергии импульса $\int I(t)\,dt$ обратно пропорциональна τ_0, так что солитон меньшей длительности должен нести большую энергию.

Ниже будет показано, что одно из солитонных решений нелинейного волнового уравнения, описывающего распространение импульса в среде с ДГС и СМФ, имеет вид

$$\boxed{\left|\mathcal{A}(t)\right| = |A_0|\,\mathrm{sech}\left(\frac{t}{\tau_0}\right),} \qquad (22.143)$$

Огибающая солитона

где sech (·) = 1/ch (·) — гиперболический секанс, график которого показан на рис. 22.42. Это симметричная колоколообразная кривая со следующими характеристиками:

- пиковая амплитуда равна A_0;
- ширина профиля амплитуды на уровне половины максимума — $2{,}63\tau_0$;
- площадь под графиком амплитуды — $2\pi A_0\tau_0$;
- интенсивность $I(t) \propto |A_0|^2 \operatorname{sech}^2(t/\tau_0)$; ширина ее профиля на уровне половины максимума — $1{,}76\tau_0$.

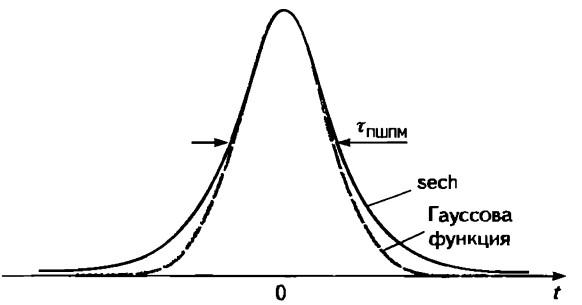


Рис. 22.42. Функция sech по сравнению с гауссовой функцией такой же высоты и ширины на уровне половины максимума

Нелинейное уравнение для медленно меняющейся амплитуды

Для описания распространения оптического импульса в нелинейной диспергирующей среде с ДГС и СМФ начнем с волнового уравнения (21.3)

$$\left[\nabla^2 - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right]\mathcal{E} = \mu_0 \frac{\partial^2}{\partial t^2}\left(\mathcal{P}_{\text{л}} + \mathcal{P}_{\text{нл}}\right), \tag{22.144}$$

где $\mathcal{E}(\boldsymbol{r}, t)$ — электрическое поле; $\mathcal{P}_{\text{л}}(\boldsymbol{r}, t)$ — линейная часть плотности поляризации, подверженная дисперсии; $\mathcal{P}_{\text{нл}}(\boldsymbol{r}, t) = 4\chi^{(3)}\mathcal{E}^3$ — нелинейная часть плотности поляризации, дисперсия которой не учитывается. Перенося линейный член из правой части (22.144) в левую часть и переходя к Фурье-представлению, получаем

$$\left[\nabla^2 + \beta^2(\omega)\right]\mathcal{E} = -\mu_0\omega^2\mathcal{P}_{\text{нл}}, \tag{22.145}$$

где $\beta(\omega)$ — постоянная распространения в линейной среде; $E = E(\boldsymbol{r}, \omega)$ и $P_{\text{нл}} = P_{\text{нл}}(\boldsymbol{r}, \omega)$ — Фурье-образы $\mathcal{E}(\boldsymbol{r}, t)$ и $\mathcal{P}_{\text{нл}}(\boldsymbol{r}, t)$ соответственно. В отсутствие нелинейности (22.145) воспроизводит уравнение Гельмгольца (2.11).

Рассмотрим импульс в виде плоской волны, распространяющейся в направлении z, с центральной угловой частотой ω_0 и центральным волновым числом $\beta_0 = \beta(\omega_0) = \omega_0/c$:

$$\mathcal{E} = \operatorname{Re}\left\{\mathcal{A}(z, t)\exp\left[j(\omega_0 t - \beta_0 z)\right]\right\} \tag{22.146}$$

и предположим, что комплексная огибающая $\mathcal{A}$ — медленно меняющаяся функция t и z (по сравнению с периодом $2\pi/\omega_0$ и длиной волны $2\pi/\beta_0$ соответственно).

Используем три предположения:

1) медленность изменения огибающей;

2) слабость дисперсии;

3) малость нелинейности.

Покажем, что огибающая $\mathcal{A}(z, t)$ удовлетворяет дифференциальному уравнению

$$\boxed{\frac{D_\nu}{4\pi}\frac{\partial^2\mathcal{A}}{\partial t^2}+\gamma|\mathcal{A}|^2\mathcal{A}+j\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}\right)\mathcal{A}=0,} \qquad (22.147)$$

Нелинейное волновое уравнение для медленной огибающей

где $v = 1/\beta'$ — групповая скорость; $D_\nu = 2\pi\beta''$ — коэффициент дисперсии; β' и β'' — первая и вторая производные $\beta(\omega)$ по ω в точке ω_0, γ дается формулой (22.142). Для линейной среды $\gamma = 0$ и получается линейное уравнение для медленной огибающей (22.86).

---

*Вывод нелинейного волнового уравнения для медленной огибающей

Для вывода (22.147) начнем с нелинейного уравнения Гельмгольца (22.145) и используем некоторые приближения. Подставим

$$E = A(z,\ \omega-\omega_0)\exp(-j\beta_0 z);\quad P_{\text{нл}} = A_{\text{нл}}(z,\ \omega-\omega_0)\exp(-j\beta_0 z)$$

в (22.145) и введем обозначение $\Omega = \omega - \omega_0$, тогда

$$\left[\frac{\partial^2}{\partial z^2}+\beta^2(\omega)\right]\left[A(z,\ \omega)\exp(-j\beta_0 z)\right] = -\mu_0\omega^2 A_{\text{нл}}(z,\ \omega)\exp(-j\beta_0 z). \qquad (22.148)$$

Теперь упростим (22.148) с помощью ряда приближений.

- Поскольку $\omega \approx \omega_0$, множитель ω^2 в правой части (22.148) заменим на ω_0^2.
- Применим приближение медленной огибающей

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\left[A\exp(-j\beta_0 z)\right] \approx \left(-j2\beta_0\frac{\mathrm{d}A}{\mathrm{d}}-\beta_0^2 A\right)\exp(-j\beta_0 z),$$

тогда (22.148) принимает вид

$$\left[-j2\beta_0\frac{\mathrm{d}}{\mathrm{d}z}\right]A+\left[\beta^2(\omega+\Omega)-\beta_0^2\right]A = -\mu_0\omega_0^2 A_{\text{нл}}. \qquad (22.149)$$

- Предполагая дисперсию слабой, имеем

$$\beta^2(\omega_0+\Omega)-\beta_0^2 \approx 2\beta_0\left[\beta(\omega_0+\Omega)-\beta_0\right].$$

В результате разложения Тейлора

$$\beta(\omega_0 + \Omega) = \beta_0 + \beta'\Omega + \frac{1}{2}\beta''\Omega^2$$

(22.149) превращается в уравнение

$$-j2\beta_0 \frac{dA}{dz} + 2\beta_0\left(\Omega\beta' + \frac{1}{2}\Omega^2\beta''\right)A = -\mu_0\omega_0^2 A_{\text{нл}}. \qquad (22.150)$$

• Поскольку $\mathcal{P}_{\text{нл}} = 4\chi^{(3)}\mathcal{E}^3$, $\mathcal{P}_{\text{нл}}$ содержит компоненты вблизи частот ω_0 и $3\omega_0$. Оставляя только член вблизи ω_0, запишем

$$\mathcal{P}_{\text{нл}} = \text{Re}\{\mathcal{A}_{\text{нл}}(z, t)\exp[j(\omega_0 t - \beta_0 z)]\},$$

где $\mathcal{A}_{\text{нл}}(z, t)$ — медленно меняющаяся огибающая. С помощью (22.146) находим [см. (21.44а)], что

$$A_{\text{нл}} = 3\chi^{(3)}|A|^2 A. \qquad (22.151)$$

Наконец, преобразуем (22.150) обратно во временное представление, используя тот факт, что $j\Omega A(z, \Omega)$ и $-\Omega^2 A(z, \Omega)$ соответствуют при этом $\partial A/\partial t$ и $\partial^2 A/\partial t^2$. С учетом (22.151) получаем нелинейное уравнение для медленной огибающей (22.147).

Уравнение (22.135) можно получить также, если предположить, что нелинейная среда приблизительно линейна с постоянной распространения $\beta(\omega) + \Delta\beta$, где $\Delta\beta = (\omega_0/c_0)n_2 I$. Интенсивность $I = |A|^2/2\eta$ предполагается достаточно медленно меняющейся, так что можно считать ее не зависящей от времени. Фурье-анализ, который привел к дифференциальному уравнению (22.90) для линейной среды, теперь несложно модифицировать путем добавления члена, пропорционального $\Delta\beta A$. Этот член порождает дополнительный член $\gamma|A|^2 A$, что приводит к уравнению (22.147).

Нелинейное уравнение Шредингера

Уравнению (22.147) должна удовлетворять комплексная огибающая $\mathcal{A}(z, t)$ плоской импульсной волны, распространяющейся в направлении z в протяженной нелинейной диспергирующей среде с групповой скоростью v, дисперсионным параметром β'' и коэффициентом нелинейности γ. Как упоминалось выше, решение в виде уединенной волны возможно, если $\beta'' < 0$ (т. е. среда характеризуется отрицательной ДГС) и $\gamma > 0$ (т. е. оптический коэффициент Керра $n_2 > 0$).

Удобно переписать (22.147) в безразмерных переменных путем нормировки времени, расстояния и амплитуды на τ_0, $2z_0$ и A_0 соответственно:

• τ_0 — длительность импульса;

• $z_0 = \tau_0^2/2\beta''$ — дисперсионная длина в линейной диспергирующей среде при данной длительности импульса;

• $A_0 = (-\beta''/\gamma)^{1/2}/\tau_0$ — пиковая амплитуда импульса, удовлетворяющего условию образования солитона (22.141).

Используя систему отсчета с запаздыванием и вводя безразмерные переменные

$$t = \frac{t - z/v}{\tau_0}; \quad z = \frac{z}{2z_0}; \quad \psi = \frac{A}{A_0}, \tag{22.152}$$

преобразуем нелинейное уравнение для медленной огибающей (22.147) к виду

$$\boxed{\frac{1}{2}\frac{\partial^2\psi}{\partial t^2} + |\psi|^2\psi + j\frac{\partial\psi}{\partial z} = 0.} \tag{22.153}$$

Нелинейное уравнение Шредингера

Это уравнение представляет собой нелинейное уравнение Шредингера.

Фундаментальный солитон

Простейшее решение (22.153) в виде уединенной волны получается, если искать его в виде с разделенными переменными $\psi(z, t) = \mathcal{T}(t)\exp[jZ(z)]$, где $\mathcal{T}(t)$ и $Z(z)$ — действительные функции. Прямой подстановкой этого выражения в (22.153) с использованием метода разделения переменных получаем два дифференциальных уравнения: $Z'(z) = \vartheta$ и $\mathcal{T}''(t) = 2(\vartheta - \mathcal{T}^2)\mathcal{T}$, где ϑ — постоянная. Предполагая, что $\mathcal{T} = \mathcal{T}' = 0$ при $|t| \to \infty$ и $\mathcal{T} = 1$, $\mathcal{T}' = 0$ при $t = 0$ (на пике импульса), можно решить эти обыкновенные дифференциальные уравнения путем прямого интегрирования, что дает $\mathcal{T}(t) = \operatorname{sech}(t)$ и $Z(z) = z/2$. Следовательно,

$$\psi(z, t) = \operatorname{sech}(t)\exp\left(\frac{jz}{2}\right). \tag{22.154}$$

Это решение называется **фундаментальным солитоном**. Оно соответствует огибающей

$$\boxed{\mathcal{A}(z, t) = A_0 \operatorname{sech}\left(\frac{t - z/v}{\tau_0}\right)\exp\left(\frac{jz}{4z_0}\right),} \tag{22.155}$$

Фундаментальный солитон

которая распространяется со скоростью v без изменения формы. Такой солитон достигается, если начальный импульс при $z = 0$

$$\mathcal{A}(0, t) = A_0\operatorname{sech}\left(\frac{t}{\tau_0}\right). \tag{22.156}$$

Солитон высшего порядка

Фундаментальный солитон — один из представителей семейства решений нелинейного уравнения Шредингера со свойствами уединенной волны. С начальным импульсом $\psi(0, t) = N\operatorname{sech}(t)$, где N — целое число, согласу-

ется решение, которое называется N — **солитонной волной**. Такая волна распространяется как периодическая функция z с периодом $z_п = \pi/2$, который называется **солитонным периодом**. Он соответствует физическому расстоянию

$$z_п = \pi |z_0| = \frac{\pi}{2} \frac{\tau_0^2}{|\beta''|}.$$

При $z = 0$ огибающая $\mathcal{A}(0, t)$ является функцией гиперболический секанс с пиковой амплитудой NA_0, т. е. в N раз больше, чем у фундаментального солитона. По мере распространения импульса он сначала сжимается, а затем рассыпается на отдельные импульсы, которые затем сливаются и в конечном счете воспроизводят начальный импульс при $z = z_п$. Такая картина повторяется периодически. Например, солитон с $N = 2$ имеет волновую функцию

$$\psi(z, t) = 4 \frac{\cosh 3t + 3e^{4jz} \cosh t}{\cosh 4t + 4\cosh 2t + 3\cos 4z} e^{jz/2}, \qquad (22.157)$$

модуль которой изображен на рис. 22.43.

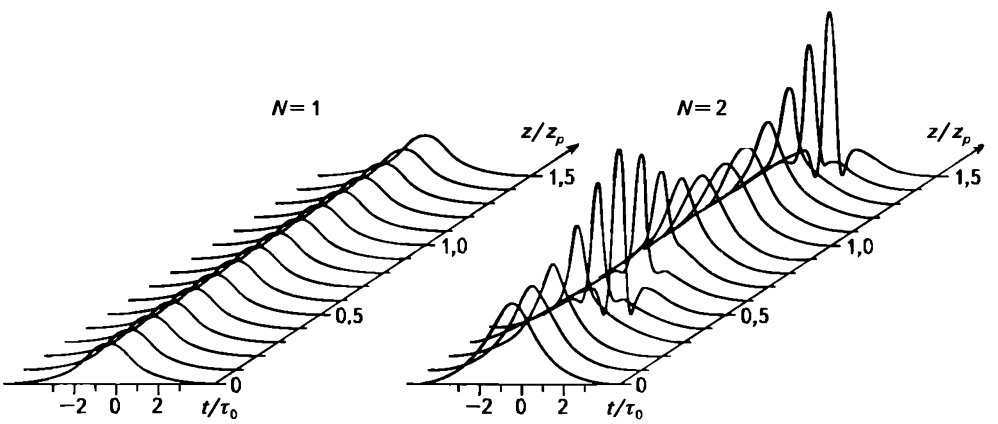


Рис. 22.43. Распространение фундаментального солитона ($N = 1$) и солитона с $N = 2$

Периодическое сжатие и расплывание многосолитонной волны учитывается периолическим дисбалансом между сжатием импульса из-за чирпа, вносимого самомодуляцией фазы, и его расплыванием из-за дисперсии групповой скорости. Начальное сжатие использовалось для генерации субпикосекундных импульсов.

Солитон-солитонное взаимодействие

Когда два солитона, разделенные некоторым запаздыванием по времени, запускаются в нелинейную среду, их форма и разделяющий их промежуток меняются так, как будто они испытывают действие сил притяжения и отталкивания. Например, два одинаковых отдельных фундаментальных соли-

тона сначала притягиваются в процессе распространения через среду; интервал времени, разделяющий их, уменьшается до тех пор, пока оба импульса не сольются в один импульс, после чего они испытывают действие сил отталкивания, которые снова разделяют их на два импульса. Процесс повторяется с периодом

$$L_{\text{п}} = \pi \exp\left(\frac{T}{2\tau_0}\right) z_0, \tag{22.158}$$

где T — начальное расстояние между центрами импульсов; τ_0 — длительность отдельного солитона; z_0 — дисперсионная длина ДГС. Это можно показать путем решения нелинейного уравнения Шредингера с соответствующими граничными условиями. Например, если $T = 10\tau_0$, так что импульсы хорошо разделены и только их хвосты взаимодействуют, то $L_{\text{п}} \approx 466 z_0$ весьма велико. Однако этот эффект может быть существенным в длинных оптических волокнах, поскольку он может устанавливать предел представления битов солитонами в волоконных системах связи, как описывается в подразд. 24.2.5.

Пример 22.8

Солитоны в оптических волокнах

Ультракороткие солитоны были получены в стеклянных волокнах в областях аномальной дисперсии ($\lambda_0 > 1{,}3$ мкм), где ДГС отрицательна. Впервые они наблюдались в 700-метровом волокне из кварцевого стекла от лазера с синхронизацией мод, работающего на длине волны $\lambda_0 = 1{,}55$ мкм. Форма импульса была близка к гиперболическому секансу с $\tau_0 = 4$ пс (что соответствует $\tau_{\text{пшпм}} = 7$ пс). На этой длине волны коэффициент дисперсии $D_\lambda = 16$ пс/нм км (см. рис. 9.27), что соответствует

$$\beta'' = \frac{D_\nu}{2\pi} = -\frac{\lambda_0^2}{c_0}\frac{D_\lambda}{2\pi} \approx -20 \text{ пс}^2/\text{км}.$$

Показатель преломления $n = 1{,}45$, коэффициент нелинейности $n_2 = 3{,}19 \cdot 10^{-20}$ м^2/Вт, что соответствует

$$\gamma = \frac{\pi}{\lambda_0}\frac{n_2}{\eta} = 2{,}48 \cdot 10^{-16} \text{ м/В}^2,$$

где $\eta = \eta_0/n = 260$ Ом. Амплитуда

$$A_0 = \frac{(|\beta''|/\gamma)^{1/2}}{\tau_0} \approx 2{,}25 \cdot 10^6 \text{ В/м}$$

соответствует интенсивности

$$I_0 = \frac{A_0^2}{2\eta} \approx 10^6 \text{ Вт/см}^2$$

При площади сечения волокна 10 мкм^2 это соответствует мощности около 100 мВт. Период солитона

$$z_{\text{п}} = \pi z_0 = \frac{\pi \tau_0^2}{2|\beta''|} = 1{,}26 \text{ км}.$$

Генерация и поддержание солитонов

Для возбуждения фундаментального солитона входной импульс должен иметь профиль гиперболического секанса и точное значение произведения амплитуды на длительность $A_0\tau_0$ (22.141). При более низком значении этого произведения возбудится обычный оптический импульс, а при более высоком — фундаментальный солитон и, возможно, солитонные волны высших порядков, а остаточная энергия превратится в паразитный обыкновенный импульс. Если входной импульс имеет другой профиль или в нем присутствует чирп, то результирующий импульс при определенных условиях может развиться в фундаментальный (или высших порядков) солитон после прохождения расстояния в несколько солитонных периодов.

Если среда имеет потери, импульс постепенно теряет энергию, так что нелинейный эффект становится слабее дисперсионного, что приводит к уширению импульса и утрате его солитонной природы. В оптических волокнах эта проблема может быть решена с помощью распределенного рамановского усиления (см. подразд. 21.3.1), компенсирующего потери за счет поглощения и рассеяния. Сосредоточенное усиление также годится, если расстояние между усилителями намного меньше солитонного периода $z_{\text{п}}$.

Благодаря уникальному свойству сохранения формы и длительности на больших расстояниях, оптические солитоны имеют потенциальные приложения в области передачи цифровых данных по оптическим волокнам с более высокой скоростью и на большие расстояния, чем позволяют современные возможности линейной оптики (см. подразд. 22.1.4). Оптические солитоны длительностью в несколько десятков пикосекунд успешно передавались на много сотен километров по оптическому волокну.

Солитонные лазеры

Оптоволоконные лазеры также применялись для генерации пикосекундных солитонов. Лазер представляет собой одномодовое волокно в виде кольцевого резонатора (рис. 22.44). Волокно является комбинацией легированного эрбием волоконного усилителя (см. подразд. 14.3.3) и нелегированного волокна, обеспечивающего формирование импульса и его солитонное действие. Импульсы получаются с помощью фазового модулятора для достижения синхронизации мод. Полностью интегральная система была разработана с применением накачки от лазерного диода на InGaAsP и интегрально-оптического фазового модулятора.

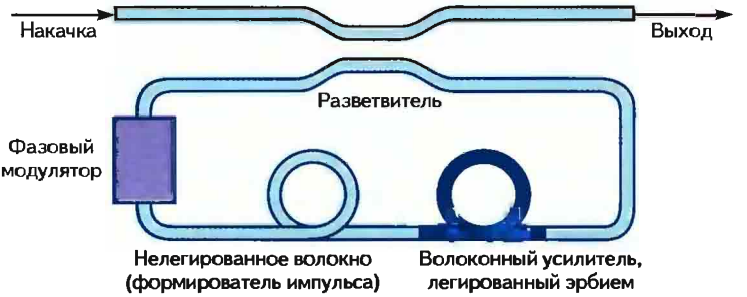


Рис. 22.44. Солитонный лазер на оптическом волокне

Темновые солитоны

Темновой солитон — это кратковременный провал в интенсивности стационарной световой волны. Темновые солитоны имеют свойства, аналогичные описанным выше свойствам «светлых» солитонов, но могут генерироваться в области нормальной дисперсии ($\lambda_0 < 1{,}3$ мкм в кварцевых оптических волокнах). Они имеют устойчивые характеристики и могут быть использованы в оптических переключателях.

Аналогия между временны́ми и пространственными солитонами

Оптические солитоны, описанные в разд. 22.5, аналогичны пространственным солитонам (самонаправляющимся пучкам), описанным в подразд. 21.3.2. Пространственные солитоны — это монохроматические волны, пространственно локализованные в поперечной плоскости. Они распространяются через нелинейную среду без изменения своего пространственного распределения в результате баланса между дифракцией и пространственной самомодуляцией фазы в соответствии с нелинейным уравнением Шредингера

$$-\frac{\lambda}{4\pi}\frac{\partial^2 \mathcal{A}}{\partial x^2} + \gamma|\mathcal{A}|^2\,\mathcal{A} + j\frac{\partial \mathcal{A}}{\partial z} = 0, \tag{22.159}$$

Нелинейная дифракция пучка

где $\gamma = \pi n_2/\lambda\eta_0$ и n_2 — оптический коэффициент Керра. Уравнение (22.159) эквивалентно (21.52).

Нелинейное уравнение Шредингера (22.151), описывающее временны́е солитоны в нелинейной диспергирующей среде, можно переписать в движущейся системе координат ($t' = t - z/v$, $z' = z$) как

$$\frac{D_\nu}{4\pi}\frac{\partial^2 \mathcal{A}}{\partial t^2} + \gamma|\mathcal{A}|^2\,\mathcal{A} + j\frac{\partial \mathcal{A}}{\partial z} = 0, \tag{22.160}$$

Нелинейная дисперсия импульса

где $\gamma = \pi n_2/\lambda\eta_0$. Это совпадает с (22.159), причем время t играет роль поперечной пространственной координаты x, а коэффициент дисперсии — D_ν (определяющий дисперсионное расплывание импульса) играет роль длины волны l (определяющей дифракционную расходимость пучка). Следовательно, временны́е солитоны являются формальными аналогами пространственных солитонов. Фактически термин *солитон* относится к определенному классу решений нелинейного уравнения Шредингера, описывающих импульсы, распространяющиеся без изменения; они могут быть пространственными или временны́ми.

Пространственно-временны́е солитоны и световые пули

Пространственно-временно́й солитон — это комбинация временно́го и пространственного солитона, т. е. импульсный пучок, сохраняющий и пространственный, и временно́й профиль в процессе распространения в керровской среде (рис. 22.45). В этом случае временно́е уширение, связанное с отрицательной (аномальной) дисперсией, и пространственная расходимость из-за дифракции одновременно компенсируются самомодуляцией фазы и самофокусировкой, возникающими за счет положительного нелинейного оптического эффекта Керра. Дифференциальное уравнение в частных производных, описывающее эти три явления, является комбинацией уравнений (22.159) и (22.160):

$$-\frac{\lambda}{4\pi}\nabla_T^2\mathcal{A} + \frac{D_\nu}{4\pi}\frac{\partial^2\mathcal{A}}{\partial x^2} + \gamma|\mathcal{A}|^2\mathcal{A} + j\frac{\partial\mathcal{A}}{\partial z} = 0. \qquad (22.161)$$

Нелинейность, дифракция и дисперсия

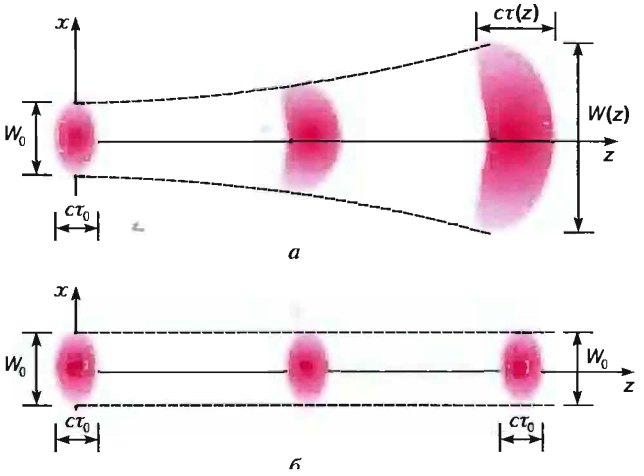


Рис. 22.45. Пространственное и временно́е расплывание импульсного пучка в результате распространения в линейной диспергирующей среде (*а*). Пространственно-временно́й солитон — импульсный пучок, сохраняющий пространственный и временно́й профиль при распространении в нелинейной среде (*б*)

Необходимое условие для пространственно-временны́х солитонов — это равенство дисперсионной длины

$$|z_0| = \frac{\pi\tau_0^2}{|D_\nu|}$$

и дифракционной длины

$$z_0 = \frac{\pi W_0^2}{\lambda},$$

так что

$$\frac{\tau_0}{W_0} = \left(\frac{\lambda}{|D_\nu|}\right)^{1/2}$$

*22.5.3. Суперконтинуум

Суперконтинуум — это свет высокой яркости со сверхшироким непрерывным спектром. **Генерация суперконтинуума (ГСК)** осуществляется путем пропускания ультракороткого импульса с высокой пиковой мощностью (накачки) через нелинейную среду со специфическими дисперсионными свойствами; примерами являются оптические волокна со смещением дисперсии, волокна со сглаживанием дисперсии и фотонно-кристаллические волокна (ФКВ). Были продемонстрированы источники суперконтинуума со спектром, простирающимся от 400 до 3000 нм.

Несколько нелинейных механизмов, включая самомодуляцию фазы (СМФ), вынужденное комбинационное рассеяние (ВКР), четырехволновое смешение (ЧВС) и самоиндуцированный сдвиг частоты солитона (ССЧС) могут давать вклад в ГСК вместе или по отдельности. Эти нелинейные эффекты чувствительны к знаку дисперсии среды и центральной длине волны λ_0 импульса накачки и относительному положению длины волны $\lambda_{д0}$, на которой дисперсия среды равна нулю. Именно появление нелинейных ФКВ с $\lambda_{д0}$, близкой к длине волны титан-сапфирового лазера, впервые сделало возможным практическую ГСК.

Ниже следует краткое описание основных нелинейных механизмов, дающих вклад в ГСК. На рис. 22.46 дана схематическая иллюстрация этих процессов.

• **Самомодуляция фазы (СМФ)** — основной механизм ГСК в нелинейных волокнах с нормальной дисперсией ($D_\lambda < 0$) на центральной длине волны накачки λ_0, поскольку в этом случае солитоны не могут образоваться. Как обсуждалось в подразд. 22.5.1, СМФ приводит к появлению чирпа у импульса, что вызывает спектральное уширение. Коэффициент чирпа a соответствует спектральному уширению в $\sqrt{1+a^2}$ раз. Для среды длиной L с оптическим коэффициентом Керра n_2 параметр чирпа равен $a = L/z_{нл}$, где $z_{нл} = (2n_2 I_0 k_0)^{-1}$ — нелинейная характеристическая длина; I_0 — пиковая интенсивность импульса.

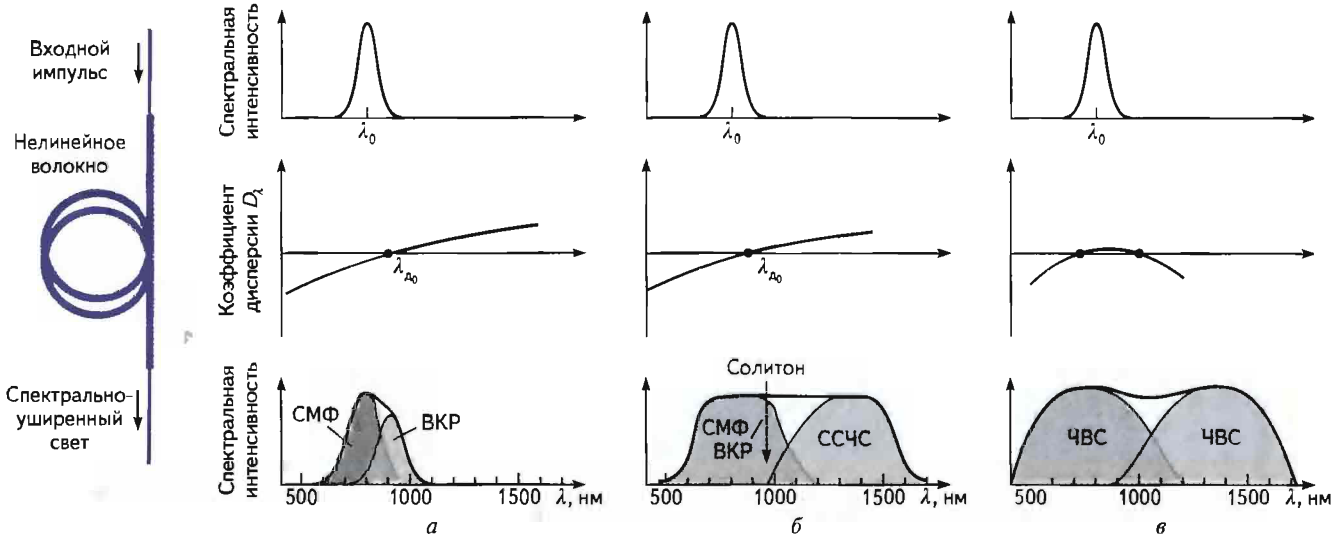


Рис. 22.46. Основные нелинейные механизмы генерации суперконтинуума (ГСК) путем спектрального уширения ультракороткого импульса, пропускаемого через нелинейное диспергирующее волокно:

а — самомодуляция фазы (СМФ) в сочетании с вынужденным комбинационным рассеянием (ВКР); *б* — самоиндуцированный сдвиг частоты солитона (ССЧС); *в* — четырехволновое смешение (ЧВС)

• **Вынужденное комбинационное рассеяние (ВКР)** уширяет спектральное распределение далее в длинноволновую сторону, поскольку его результатом является понижение частоты.

• Когда λ_0 близко к $\lambda_{д0}$, комбинированный эффект СМФ/ВКР уширяет спектр в область аномальной дисперсии, создавая условия для формирования солитона. Оптические солитоны обычно испытывают сдвиг их несущей частоты вниз, что соответствует сдвигам в длинноволновую область, увеличивающимся с ростом мощности накачки. Этот так называемый **самоиндуцированный сдвиг частоты солитона (ССЧС)** возникает из-за ВКР внутри импульса.

• В микроструктурном волокне, имеющем две широко разнесенные длины волны, на которых дисперсия равна нулю, при λ_0, лежащей между ними, преобладающим нелинейным механизмом спектрального уширения является СМФ и ЧВС Процесс СМФ уширяет импульс накачки, что способствует созданию условий фазового синхронизма для **четырехволнового смешения (ЧВС)**. За счет этого генерируется свет как на более низких, так и на более высоких частотах, что соответствует ГСК со спектром, имеющим два пика. При существенном уширении два пика ЧВС сливаются в одно плоское распределение.

22.6. ДЕТЕКТИРОВАНИЕ ИМПУЛЬСОВ

Измерение ультракороткого оптического импульса — сложная задача, поскольку наиболее быстрые существующие фотоприемники обычно для этого слишком медленны. Методы решения этой задачи опираются прежде всего на использование сверхбыстрого оптического затвора (вентиля), управляемого другим более коротким опорным импульсом, и механизма, вносящего управляемую временную задержку между двумя импульсами. Измерение повторяется при различных задержках в процессе прохождения света через затвор, что обеспечивает оценку профиля интенсивности импульса $I(t)$. Для измерения фазы импульса $\phi(t)$ была искусно адаптирована интерферометрическая методика в сочетании с нелинейными оптическими процессами. В спектральном представлении импульс полностью характеризуется своей спектральной интенсивностью $S(\nu)$ и спектральной фазой $\psi(\nu)$. Эти функции можно измерить с помощью оптических спектроанализаторов и интерферометрических методов, как будет описано в данном разделе.

Еще один сложный аспект проблемы детектирования ультракоротких импульсов — то, что используемые оптические элементы неизбежно меняют импульс до того, как его характеристики будут измерены. Такие эффекты должны быть минимизированы путем тщательного проектирования системы или компенсированы соответствующими методами обработки сигнала после детектирования.

22.6.1. Измерение интенсивности

Профиль интенсивности короткого оптического импульса можно непосредственно измерить с помощью фотоприемника, время отклика которого намного меньше длительности импульса. Измеряемый фототок

$$i(t) = \Re A I(t) \qquad (22.162)$$

Быстрый приемник

пропорционален интенсивности импульса $I(t)$, A — площадь активной зоны приемника; $\Re$ — токовая чувствительность фотоприемника (А/Вт) (см. подразд. 18.1.3). Здесь предполагается, что A достаточно мало, так что оптическая интенсивность измеряется локально.

Когда время отклика приемника становится значительным, импульс фототока становится уширенной и искаженной копией оптического импульса. Тогда для определения истинной формы импульса требуются другие средства. Если $h_D(t)$ — функция отклика приемника на импульсное воздействие, где

$$\int I(t) h_D(t) \mathrm{d}t = \Re,$$

то фототок представляет собой свертку

$$i(t) = A \int I(\tau) h_D(t - \tau) \mathrm{d}\tau, \qquad (22.163)$$

Произвольный приемник

которая является импульсом большей длительности. Если время отклика [ширина функции $h_D(t)$] намного короче, чем длительность импульса [ширина функции $I(t)$], то свертка (22.163) имеет форму функции большей длительности, и идеальное соотношение (22.162) восстанавливается.

В другом предельном случае, когда длительность оптического импульса много меньше времени отклика приемника,

$$i(t) \approx h_D(t) A \int I(t) \mathrm{d}\tau,$$

так что фототок имеет временной профиль функции отклика приемника на импульсное воздействие, а не профиль оптического импульса. Если схема приемника имеет постоянную времени $\tau_к$, превышающую малое время отклика самого приемника, финальный отклик равен

$$i(t) \approx \tau_к^{-1} \int h_D(t) \mathrm{d}t A \int I(\tau) \mathrm{d}\tau$$

или

$$i(t) \approx \tau_к^{-1} \Re A \int I(\tau) \mathrm{d}\tau. \qquad (22.164)$$

Медленный приемник

При таких условиях приемник измеряет площадь под оптическим импульсом, или оптическую энергию; приемник утрачивает временное разрешение и является интегратором. Эти три случая схематически изображены на рис. 22.47.

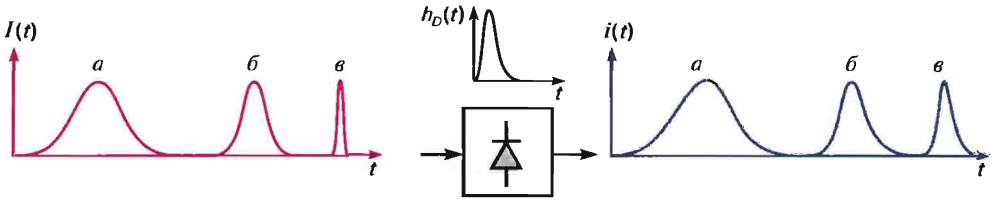


Рис. 22.47. Отклик фотоприемника с функцией отклика на импульсное воздействие $h_D(t)$ на три импульса:

а — длинный; *б* — промежуточный; *в* — короткий

Как можно измерить временной профиль ультракороткого импульса пикосекундной или фемтосекундной длительности с использованием «медленного» приемника, время отклика которого, в лучшем случае, несколько десятков наносекунд?

Измерение короткого импульса медленным приемником и быстрым затвором

Временной профиль короткого оптического импульса можно измерить медленным приемником с использованием быстрого затвора (переключателя или вентиля). Как показано на рис. 22.48, затвор открывается только на короткое время (окно), в течение которого медленный приемник регистрирует только часть импульса. Измерение повторяется с открыванием окна в разные моменты времени, и полученный набор данных используется для восстановления профиля импульса. Поскольку затворы с электронным управлением не достигают скоростей переключения, пригодных в пико- и фемтосекундном диапазоне, можно использовать оптический затвор, управляемый опорным (стробирующим) оптическим импульсом с длительностью, намного меньшей, чем у измеряемого импульса (см. гл. 23).

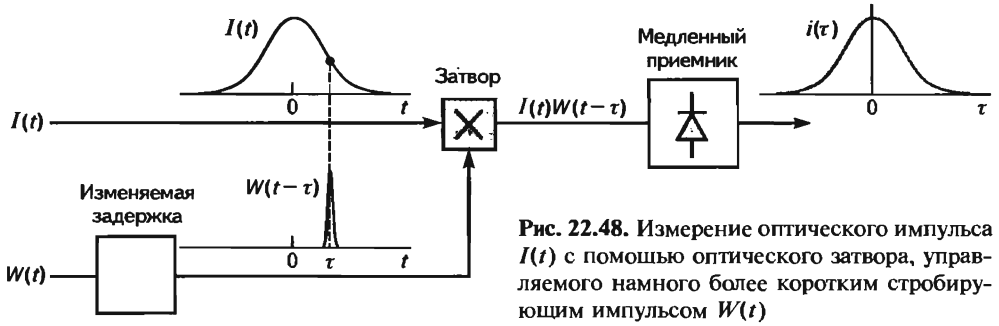
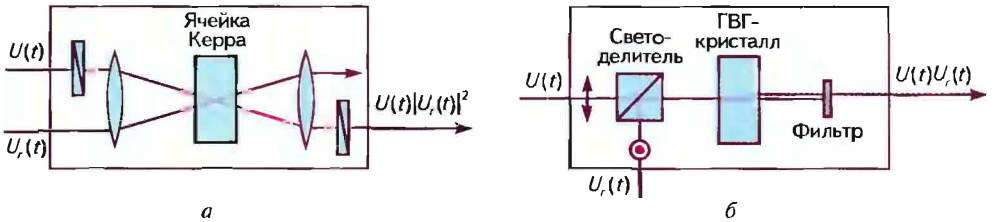


Рис. 22.48. Измерение оптического импульса $I(t)$ с помощью оптического затвора, управляемого намного более коротким стробирующим импульсом $W(t)$

Два примера оптических затворов, используемых для измерения ультракоротких импульсов, показаны на рис. 22.49.



а *б*

Рис. 22.49. Оптический затвор с ячейкой Керра (*а*). Опорный импульс с интенсивностью $I_r(t) = |U_r(t)|^2$ изменяет фазовую задержку в керровской среде. Поскольку измеряемый импульс пропускается через два скрещенных поляризатора, между которыми находится керровская среда, он модулируется вырезающей функцией $W(t) \propto I_r(t) = |U_r(t)|^2$. Затвор с использованием генерации второй гармоники (ГВГ) (*б*). Измеряемый импульс $U(t)$ и стробирующий импульс $U_r(t)$, имеющие ортогональные поляризации, взаимодействуют в коллинеарной конфигурации типа II (см. подразд. 21.2.4) и генерируют импульс второй гармоники с амплитудой $\propto U(t)U_r(t)$, так что вырезающая функция $W(t) \propto U_r(t)$

Теперь остановимся на влиянии конечного времени переключения на разрешение измерительной схемы. Если $W(t)$ — коэффициент пропускания затвора, приводимого в действие стробирующим импульсом в момент $t = 0$, то при условии, что затвор срабатывает с запаздыванием τ, прошедший оптический импульс есть $I(t)W(t - \tau)$. При регистрации медленным приемником результирующий фототок пропорционален площади под пропущенным импульсом

$$i(t) \propto \int I(\tau)W(t-\tau)\mathrm{d}\tau. \qquad (22.165)$$

В идеальном случае вырезающая функция $W(t)$ — это дельта-функция $\delta(t)$, и фототок пропорционален $I(\tau)$, т. е. выборочному значению интенсивности импульса при $t = \tau$. В реальном случае измеренный фототок пропорционален свертке оптического импульса и вырезающей функции. Временное разрешение измерительной схемы тогда равно ширине вырезающей функции $W(t)$, которая определяется скоростью затвора. Время задержки можно менять за счет вырезающей функции $W(t)$ или самого оптического импульса $I(t)$.

Одиночный импульс в сравнении с цугом импульсов

Описанный метод измерения формы короткого импульса медленным приемником легко осуществить при наличии периодического цуга одинаковых импульсов. Затвор устанавливается на различную временну́ю задержку τ для каждого из следующих друг за другом импульсов, как показано на рис. 22.50, и показания приемника записываются последовательно. Частота следования импульсов, конечно, должна быть достаточно низкой, чтобы медленный приемник успевал восстановиться перед измерением следующего импульса.

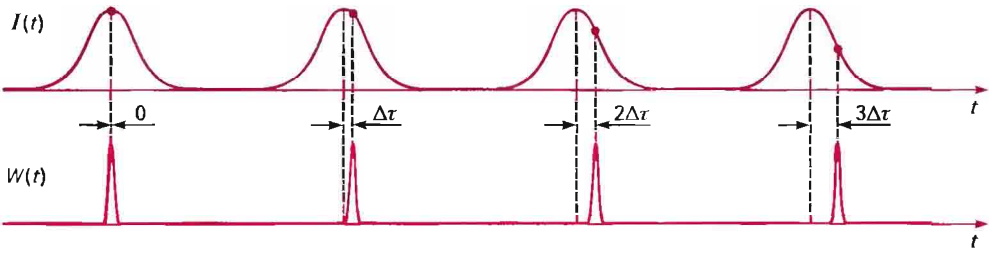


Рис. 22.50. Измерение профиля импульса путем стробирования отдельных импульсов периодического цуга с временны́ми задержками $\tau = m\Delta\tau$, $m = 0, 1, 2,$

Как быть, если нужно измерить профиль одиночного импульса? Это можно осуществить с помощью нескольких приемников. Копии импульса генерируются разветвляющим оптическим элементом, и каждая копия подвергается различной временно́й задержке, прежде чем направляется на затвор, управляемый вырезающей функцией, как показано на рис. 22.51.

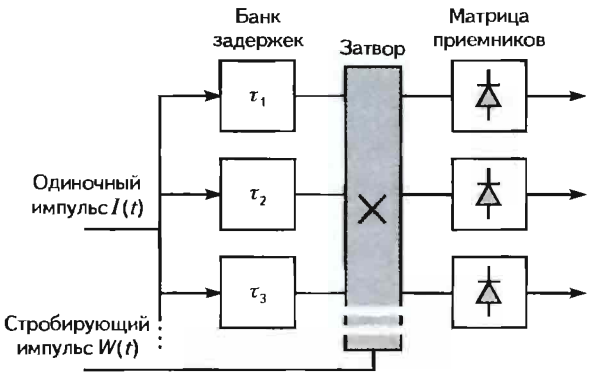


Рис. 22.51. Измерение профиля одиночного импульса с использованием банка задержек, затвора и матрицы приемников

Временно-пространственное преобразование: принцип стрик-камеры (фотохронографа)

Принцип разделения импульса с различной задержкой, показанный на рис. 22.51, можно реализовать оптически с использованием расширенного пучка, пересекаемого под углом плоским пространственно-разрешающим приемником (матрицей детекторов или ПЗС-камерой), как показано на рис. 22.52. Импульсная плоская волна, распространяющаяся вдоль оси z, имеет интенсивность $I(t - z/c)$. Волна, бегущая под углом θ, имеет интенсивность $I(t - [x\sin\theta + z\cos\theta]/c)$. Если пучок пересекается детектирующей плоскостью $z = 0$, то регистрируемая интенсивность равна $I(t - x\sin\theta/c)$, так что в точке x задержка импульса составляет $\tau_x = x\sin\theta/c$. Следовательно, каждый детектирующий элемент получает свою задержку, что реализует схему рис. 22.51. Если затвор делает снимок в момент $t = 0$, то показание приемника в точке x пропорционально $I(-x\sin\theta/c)$.

Таким образом, форма импульса записывается в виде пространственной картинки с инвертированным профилем в масштабе, где импульс шириной τ_0 создает изображение с поперечной шириной $c\tau_0/\sin\theta$. Например, импульс шириной $\tau_0 = 10$ пс дает изображение шириной $c\tau_0 = 3$ мм вдоль направления распространения. Под углом $\theta = 30°$ это соответствует ширине 6 мм в детектирующей плоскости.

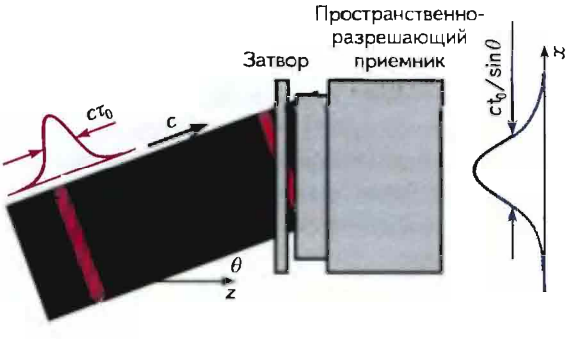


Рис. 22.52. Временно-пространственное преобразование импульса с использованием наклонной волны

В этом заключается основная идея стрик-камеры (фотохронографа, электронно-оптической камеры). Импульс света отражается от поверхности (в более старых технологиях — от вращающегося барабана) и развертывается таким образом, что лучи, попадающие в разные точки приемника, проходят различные расстояния и, следовательно, испытывают различную задержку. Такую зависящую от координат временну́ю задержку можно внести также путем пропускания пучка через стеклянный клин.

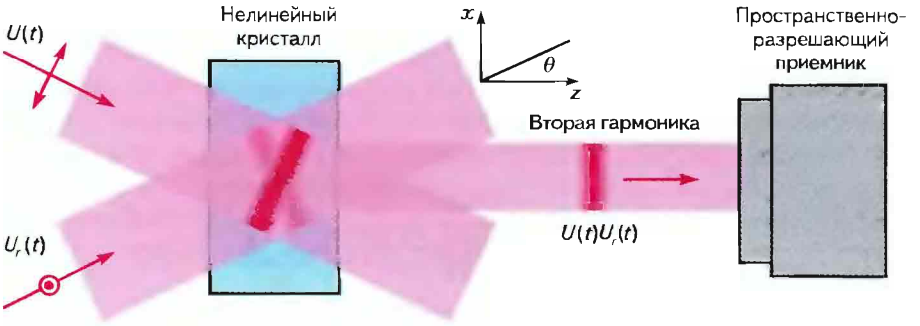


Рис. 22.53. Измерение одиночного импульса с использованием ГВГ типа II и временно-пространственного преобразования

В системе, показанной на рис. 22.52, может использоваться оптический керровский затвор или затвор на основе ГВГ, управляемой стробирующим импульсом. Один из вариантов использования ГВГ-затвора в такой схеме показан на рис. 22.53. Исследуемый и стробирующий импульсы имеют вид ортогонально поляризованных наклонных волн под углами θ и $-\theta$ к оси z. Их волновые функции $U(t - [x\sin\theta + z\cos\theta]/c)$ и $U_r(t - [-x\sin\theta + z\cos\theta]/c)$, так что относительная задержка равна $\tau_x = (2\sin\theta/c)x$ в точке x. Затвор основан на

неколлинеарном процессе ГВГ типа II. Генерируемая волна на частоте второй гармоники имеет волновую функцию, пропорциональную произведению U_r, так что измеренная интенсивность пропорциональна U_r. В результате принятый сигнал пропорционален автокорреляционной функции интенсивности $G_I(\tau_x)$.

Измерение автокорреляции интенсивности

Как упоминалось выше, основной принцип измерения ультракороткого оптического импульса $I(t)$ медленным приемником основан на использовании более короткого стробирующего импульса, который открывает и закрывает оптический затвор. Если такого импульса не существует, то исследуемый импульс можно сжать и использовать для этой цели; в этом случае $W(t)$ есть сжатая версия. Другая возможность — получение квадрата $I^2(t)$ интенсивности импульса путем генерации второй гармоники. Квадрат функции $I^2(t)$ уже, чем она сама. Нелинейные процессы высших порядков также можно использовать для генерации еще более коротких импульсов, хотя и меньшей интенсивности.

Если такое сжатие невозможно или нежелательно, исследуемый импульс можно непосредственно использовать как вырезающую функцию, как схематически показано на рис. 22.54. Фототок тогда пропорционален автокорреляционной функции интенсивности

$$G_I(\tau) = \int I(t)\, I(t-\tau)\, \mathrm{d}t. \qquad (22.166)$$

Автокорреляционная функция интенсивности

Поскольку $I(t)$ — действительная функция конечной протяженности, $G_I(\tau)$ — симметричная функция, спадающая от максимального значения $G_I(0)$ при $\tau = 0$ до нуля при $\tau = \infty$.

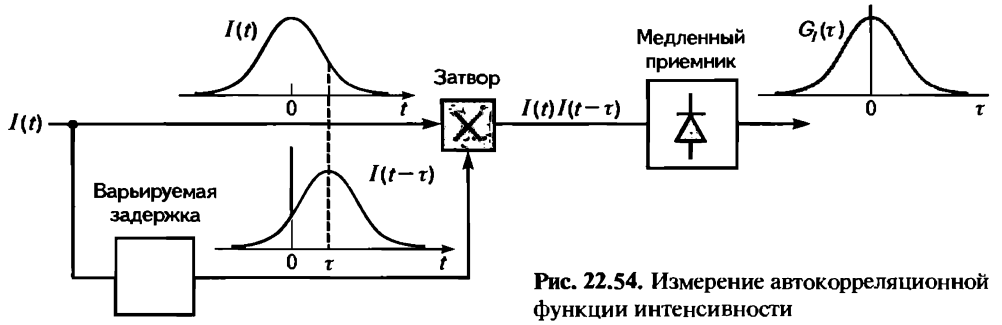


Рис. 22.54. Измерение автокорреляционной функции интенсивности

Автокорреляционная функция импульса произвольной формы в общем случае представляет собой более широкий симметричный импульс. Например, гауссов импульс с интенсивностью

$$I(t) = \exp\left(-\frac{2\tau^2}{\tau_0^2}\right)$$

и шириной τ_0 имеет гауссову автокорреляционную функцию

$$G_I(\tau) \propto \exp\left(-\frac{\tau^2}{\tau_0^2}\right),$$

которую можно записать как $\exp[-2(\tau^2/\sqrt{2}\tau_0)^2]$, т. е. импульс шириной $\sqrt{2}\tau_0$.

Знание автокорреляционной функции в общем случае недостаточно для определения самой функции. Это можно увидеть, если заметить, что Фурье-образ $G_I(\tau)$ равен $|\tilde{I}(\nu)|^2$, где $\tilde{I}(\nu)$ — Фурье-образ $I(t)$. Измерение $G_I(\tau)$ позволяет определить модуль $|\tilde{I}(\nu)|$, но не дает информации о фазе и, следовательно, не может быть использовано для полного восстановления комплексной огибающей. Исключение составляет случай симметричного импульса, т. е. $I(-t) = I(t)$, поскольку в этом случае $\tilde{I}(\nu)$ действительно, т. е. имеет нулевую фазу. Если математическое выражении профиля несимметричной функции известно, то измерение автокорреляционной функции достаточно для оценки ее параметров, например ее ширины.

22.6.2. Измерение спектральной интенсивности

Оптический анализатор спектра

Спектральную интенсивность $S(\nu) = |A(\nu)|^2$ оптического импульса с комплексной огибающей $\mathcal{A}(t)$ можно измерить с помощью оптического анализатора спектра. Анализатор представляет собой просто группу спектральных фильтров, настроенных на набор частот/длин волн. Если используется группа «медленных» приемников, регистрирующих энергию в каждой из спектральных компонент, то результатом измерения будет спектральная интенсивность $S(\nu)$. В общем случае невозможно восстановить функцию $\mathcal{A}(t)$ из модуля ее Фурье-образа $A(\nu)$ в отсутствие информации о фазе. Исключением является симметричный импульс, Фурье-образ которого действителен. Оптическая реализация этой идеи показана схематически на рис. 22.55.

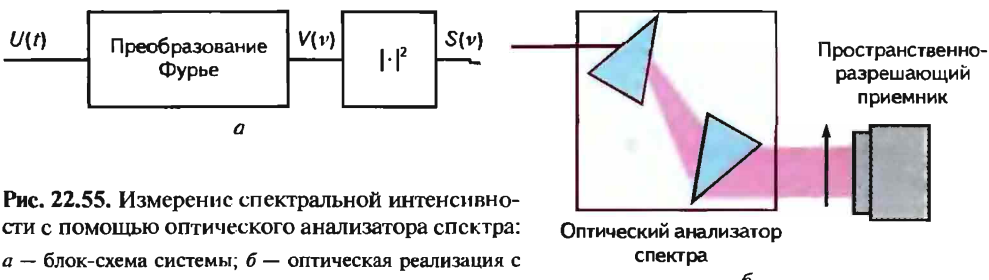


Рис. 22.55. Измерение спектральной интенсивности с помощью оптического анализатора спектра:
а — блок-схема системы; *б* — оптическая реализация с использованием призм

Интерферометрический анализатор спектра

Спектральную интенсивность $S(\nu)$ оптического импульса можно также измерить с помощью интерферометра (рис. 22.56). Напомним (см. разд. 11.2), что интерферометр Майкельсона можно использовать как Фурье-преобразующий спектрометр. Когда импульсный оптический пучок с комплексной волновой функцией $U(t)$ расщепляется на два пучка с помощью светоделителя и один пучок задерживается на время τ по отношению к другому, результирующее оптическое поле $\frac{1}{\sqrt{2}}\left[U(t)+U(t-\tau)\right]$ имеет интенсивность $\frac{1}{2}\left|U(t)+U(t-\tau)\right|^2$ При детектировании «медленным приемником» результат есть функция оптической задержки

$$R_U(\tau)=\frac{1}{2}\int\left|U(t)+U(t-\tau)\right|^2 \mathrm{d}t =$$

$$=\frac{1}{2}\int\left|U(t)\right|^2 \mathrm{d}t+\frac{1}{2}\int\left|U(t-\tau)\right|^2 \mathrm{d}t+\mathrm{Re}\int U^*(t)U(t-\tau)\,\mathrm{d}t.$$

Подставляя $U(t) = \mathcal{A}(t)\exp(j2\pi\nu_0 t)$, получаем

$$\begin{aligned} R_U(\tau) &= G_A(0)+\mathrm{Re}\left\{G_A(\tau)\exp(-j2\pi\nu_0\tau)\right\} = \\ &= G_A(0)+\left|G_A(\tau)\right|\cos\left[2\pi\nu_0\tau-\arg\left\{G_A(\tau)\right\}\right], \end{aligned} \tag{22.167}$$

где $G_A(\tau)$ — автокорреляционная функция комплексной огибающей,

$$G_A(\tau)=\int A^*(t)\,A(t-\tau)\,\mathrm{d}t. \tag{22.168}$$

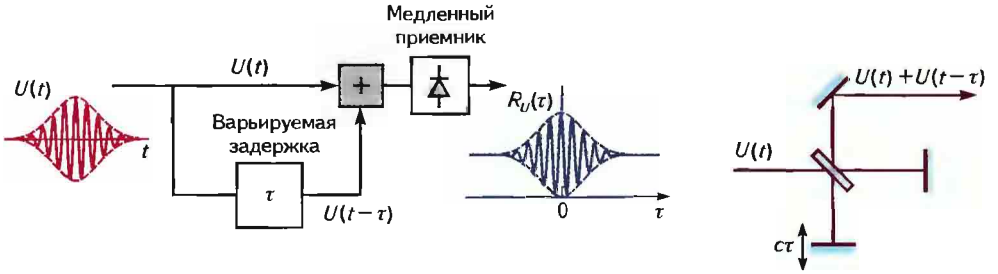


Рис. 22.56. Интерферометрическое измерение спектральной интенсивности импульса. Интерферограмма используется для определения автокорреляционной функции огибающей импульса $G_A(\tau)$, Фурье-образ которой и есть спектральная интенсивность

Функция $G_A(\tau)$ равна результату обратного преобразования Фурье спектральной интенсивности $S(\nu) = |A(\nu)|^2$. Измерение $R_U(\tau)$ дает картину интерференционных полос с видностью $|G_A(\tau)|/G_A(0)$. Схема позволяет определить $G_A(\tau)$ путем тщательного анализа видности и положения полос. Таким образом, интерферометр дает ту же информацию, что и обычный анализатор спектра.

22.6.3. Измерение фазы

Полная характеристика оптического импульса включает измерение комплексной огибающей, т. е. модуля и фазы волновой функции

$$U(t) = \sqrt{I(t)}\exp\left[j2\pi\nu_0 t + \varphi(t)\right]$$

или, что эквивалентно, модуля и фазы ее Фурье-образа

$$V(\nu) = \sqrt{S(\nu)}\exp\left[j\psi(\nu)\right].$$

Методика, представленная в подразд. 22.6.1, обеспечивает измерение интенсивности $I(t)$, но не дает информации о фазе $\varphi(t)$. Методики, описанные в подразд. 22.6.2, обеспечивают измерение спектральной интенсивности $S(\nu)$, но не дают информации о спектральной фазе $\psi(\nu)$. При некоторых условиях комплексная функция может быть полностью определена по ее модулю и модулю ее Фурье-образа, т. е. из $I(t)$ и $S(\nu)$. В этом разделе мы познакомимся с другими измерениями, которые непосредственно чувствительны к фазе $\varphi(t)$ или спектральной фазе $\psi(\nu)$. Измерение фазы часто основано на интерферометрии, поскольку интенсивность на выходе интерферометра имеет высокую чувствительность к разности фаз интерферирующих волн.

Спектральная интерферометрия

Общепринятым методом измерения фазы является гетеродинирование, которое является разновидностью временной интерферометрии (см. подразд. 2.6.2). Импульс

$$U(t) = \sqrt{I(t)}\exp\left[j2\pi\nu_0 t + \varphi(t)\right]$$

смешивается с известным опорным импульсом

$$U_r(t) = \sqrt{I_r(t)}\exp\left[j2\pi\nu_r t + \varphi_r(t)\right]$$

с отличающейся центральной частотой $\nu_r = \nu_0 + f$. Интенсивность суммы

$$\left|U(t) + U_r(t)\right|^2 = I(t) + I_r(t) + 2\sqrt{I(t)\,I_r(t)}\cos\left[2\pi ft + \varphi_r(t) - \varphi(t)\right] \quad (22.169)$$

является интерферограммой с частотой биений f (полос в секунду), равной разности между центральными частотами, и зависящей от времени фазой $\varphi_r(t) - \varphi(t)$, которую можно легко извлечь из интерферограммы. Однако для сверхкоротких импульсов приемник слишком медленный, и все временные особенности интерферограммы смазываются, так что метод гетеродинирования, или временной интерферометрии, неприменим.

Однако интерферометрия работает, если ее осуществлять в Фурье-области, и в этом случае называется **спектральной интерферометрией**. Импульс $U(t)$ задерживается на фиксированное время τ и складывается с известным опорным

импульсом $U_r(t)$ той же частоты. Фурье-образ суммы $U(t-\tau) + U_r(t)$ затем измеряется медленным приемником, создающим интерферограмму, как показано на рис. 22.57. Если Фурье-образы $U(t)$ и $U_r(t)$ равны

$$V(\nu) = \sqrt{S(\nu)}\exp\left[j\psi(\nu)\right]; \quad V_r(\nu) = \sqrt{S_r(\nu)}\exp\left[j\psi_r(\nu)\right]$$

соответственно, то спектральный интерферометр измеряет интерферограмму

$$\begin{aligned}&\left|V(\nu)e^{-j2\pi\tau\nu} + V_r(\nu)\right|^2 = \\ &= S(\nu) + S_r(\nu) + 2\sqrt{S(\nu)S_r(\nu)}\cos\left[2\pi\tau\nu + \psi_r(\nu) - \psi(\nu)\right],\end{aligned} \tag{22.170}$$

которая представляет собой картину полос (по частоте) с видностью, определяемой спектральной интенсивностью $S(\nu)$, и положением полос, определяемым разностью фаз $\psi(\nu) - \psi_r(\nu)$. Следовательно, такое измерение дает полную информацию о $V(\nu)$, а значит о $U(t)$. Дуализм временной и спектральной интерферометрии виден из того, что (22.169) и (22.170) одинаковы во форме, причем t и ν меняются ролями, а задержка τ играет роль частотной разности f. Главная трудность спектральной интерферометрии — необходимость известного опорного импульса.

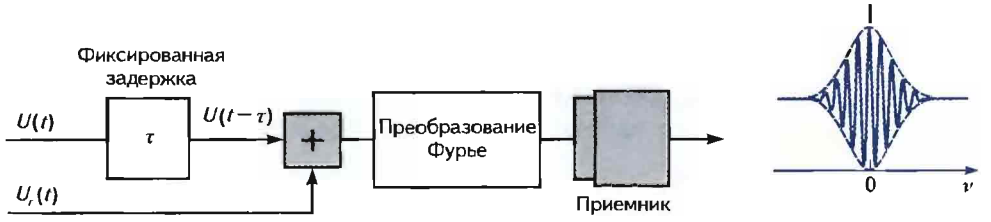


Рис. 22.57. Спектральный интерферометр генерирует интерферограмму в Фурье-области

Спектральная интерферометрия с использованием исследуемого импульса в качестве опорного

Исследуемый импульс не может быть использован как опорный для самого себя, поскольку фазовый член в (22.170) исчезает, если $\psi_r(\nu) = \psi(\nu)$. Один из методов решения этой проблемы — использовать сдвинутую по частоте копию исходного импульса в качестве опорного, т. е. $V_r(\nu) = V(\nu + f)$. В результате получается интерферограмма

$$\begin{aligned}&\left|V(\nu)e^{-j2\pi\tau\nu} + V(\nu+f)\right|^2 = \\ &= S(\nu) + S(\nu+f) + 2\sqrt{S(\nu)S(\nu+f)}\cos\left[2\pi\tau\nu + \psi(\nu) - \psi(\nu+f)\right],\end{aligned} \tag{22.171}$$

и система схематически иллюстрируется рис. 22.58. Из такой частотной интерферограммы можно найти разность фаз $\psi(\nu+f) - \psi(\nu)$. Если сдвиг частоты f

мал, то разность фаз можно использовать для аппроксимации производной $d\psi/d\nu$, интегрируя которую получим фазу $\psi(\nu)$.

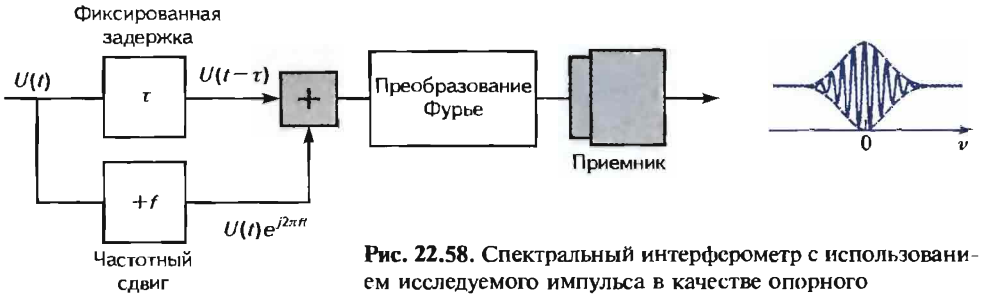


Рис. 22.58. Спектральный интерферометр с использованием исследуемого импульса в качестве опорного

Нелинейная интерферометрия

Как было показано ранее, обычный (временной) интерферометр, измеряющий площадь под графиком функции $|U(t) + U(t - \tau)|^2$, дает полную информацию о спектральной интенсивности, но не дает информации о спектральной фазе. Один из подходов к решению проблемы извлечения фазовой информации из интерферометра или ее проверки состоит в преобразовании суммы $U(t) + U(t - \tau)$ в ее квадрат $[U(t) + U(t - \tau)]^2$ перед подачей на приемник, т. е. измерение площади под графиком функции $|[U(t) + U(t - \tau)]^2|^2$. Это приводит к

$$R_U^{(2)}(\tau) = \frac{1}{2}\int \left| \left[U(t) + U(t-\tau) \right]^2 \right|^2 dt, \qquad (22.172)$$

операции, иллюстрируемой блок-схемой на рис 22.59. Операцию возведения в квадрат можно легко осуществить с помощью процесса генерации второй гармоники в нелинейном оптическом кристалле.

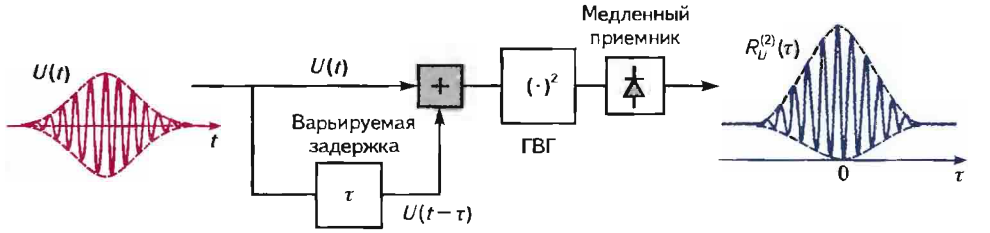


Рис. 22.59. Нелинейный интерферометр

Чтобы показать, что новая функция $R_U^{(2)}(\tau)$ содержит фазовую информацию, подставим $U(t) = \mathcal{A}(t)\exp(j2\pi\nu_0 t)$ в (22.172) и выделим члены с частотами 0, ν_0 и $2\nu_0$:

$$R_U^{(2)}(\tau) \propto C_0(\tau) + 4\,\mathrm{Re}\left\{C_1(\tau)e^{j2\pi\nu_0\tau}\right\} + 2\,\mathrm{Re}\left\{C_2(\tau)e^{j4\pi\nu_0\tau}\right\}, \qquad (22.173)$$

где

$$C_0(\tau) = \int I^2(t)\,dt + \int I^2(t-\tau)\,dt + 4\int I(t)\,I(t-\tau)\,dt = 2G_I(0) + 4G_I(\tau); \quad (22.174)$$

$$C_1(\tau) = \int \mathcal{A}^*(t)\,\mathcal{A}(t-\tau)\left[I(t) + I(t-\tau)\right]dt; \quad (22.175)$$

$$C_2(\tau) = \int \left[\mathcal{A}^*(t)\,\mathcal{A}(t-\tau)\right]^2 dt; \quad (22.176)$$

$G_I(\tau)$ — автокорреляционная функция интенсивности, определяемая выражением (22.166). Следовательно, функция $R_U^{(2)}(\tau)$ представляет собой сумму трех членов: неосциллирующего члена $C_0(\tau)$ и двух осциллирующих членов на частотах ν_0 и $2\nu_0$. Эти члены можно выделить посредством Фурье-анализа $R_U^{(2)}(\tau)$. Первый член зависит только от автокорреляционной функции интенсивности $G_I(\tau)$ и не имеет фазовой зависимости. Два других члена зависят как от интенсивности, так и от фазы импульса. Полная функция ограничена сверху огибающей с максимальным значением

$$R_U^{(2)}(0) = 16\int I^2(t)\,dt = 16G_I(0)$$

и минимальным значением $R_U^{(2)}(0) = 0$. Значение функции $R_U^{(2)}(\infty)$ равно:

$$R_U^{(2)}(\infty) = C_0(\infty) = 2\int I^2(t)\,dt = 2G_I(0).$$

Отношение $R_U^{(2)}(\infty)/R_U^{(2)}(0)$ меняется от пикового значения 8 при $\tau = 0$ до асимптотического значения единица при $\tau = \infty$. Например, для гауссова импульса длительностью τ_0 с параметром линейного чирпа a

$$C_0(\tau) = 2G_I(0)\left[1 + 2\exp\left(-\frac{\tau^2}{\tau_0^2}\right)\right]; \quad (22.177)$$

$$C_1(\tau) = 2G_I(0)\exp\left[-\left(3 + a^2\right)\frac{\tau^2}{4\tau_0^2}\right]\cos\left(\frac{a\tau^2}{2\tau_0^2}\right); \quad (22.178)$$

$$C_2(\tau) = G_I(0)\exp\left[-\left(1 + a^2\right)\frac{\tau^2}{\tau_0^2}\right]. \quad (22.179)$$

График нормированной функции $R_U^{(2)}(\tau)/R_U^{(2)}(\infty)$ изображен на рис. 22.60 для трех значений параметра чирпа a. Очевидно, что профиль интерферограммы, особенно точка, в которой осциллирующие члены исчезают, обладает высокой чувствительностью к a и, следовательно, может быть использован для определения a из экспериментальных данных.

Не существует общей процедуры вычисления фазы по данным измерения $R_U^{(2)}(\tau)$. Однако такое измерение можно использовать для проверки известных моделей амплитуды и фазы импульса или оценки неизвестных параметров.

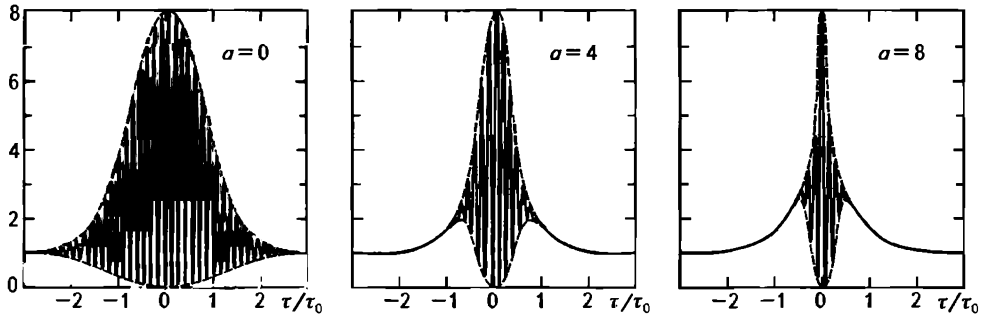


Рис. 22.60. Нормированная автокорреляционная функция интенсивности

$$R_U^{\text{ГВГ}}(\infty) = C_0(\infty) = 2\int I^2(t)\,\mathrm{d}t$$

в зависимости от нормированной временно́й задержки τ/τ_0 для чирпированного гауссова импульса при трех значениях параметра чирпа a

Нелинейная интерферометрия с нелинейными приемниками

В альтернативной реализации нелинейного интерферометра, изображенного на рис. 22.59, операция возведения в квадрат выполняется самим приемником. Это достигается в приемниках с двухфотонным поглощением, например фотодиодах с шириной запрещенной зоны больше энергии фотона, но меньше удвоенной энергии фотона. В таком приемнике фототок пропорционален квадрату интенсивности (поскольку он поглощает пары фотонов). В результате нелинейный интерферометр измеряет функцию

$$R_U^{\prime(2)}(\tau) = \int \left|U(t) + U(t-\tau)\right|^4 \mathrm{d}t, \qquad (22.180)$$

которая, как и (22.172), содержит информацию о форме и ширине импульса.

*22.6.4. Измерение спектрограмм

Как упоминалось в подразд. 22.1.1, спектрограмма оптического импульса $U(t)$ представляет собой функцию времени и частоты, равную квадрату модуля Фурье-образа импульса, если рассматривать его через движущееся окно, создаваемое вырезающей функцией $W(t)$:

$$S(\nu,\ \tau) = \left|\Phi(\nu,\ \tau)\right|^2 \qquad \Phi(\nu,\tau) = \int U(t)W(t-\tau)\exp(-j2\pi\nu t)\,\mathrm{d}t. \quad (22.181)$$

Спектрограмму можно измерить, пропуская импульс $U(t)$ через оптический затвор, управляемый задержанной вырезающей функцией $W(t-\tau)$, и измеряя спектр произведения $U(t)W(t-\tau)$ анализатором спектра при каждой временной задержке τ, как схематически показано на рис. 22.61. Оптическое воплощение этой идеи достигается с использованием движущегося зеркала для внесения временной задержки, оптического анализатора спектра, такого как

показанный на рис. 22.55, и подходящего оптического затвора. Метод известен как **оптическое стробирование с частотным разрешением (frequency resolved optical gating — FROG)**.

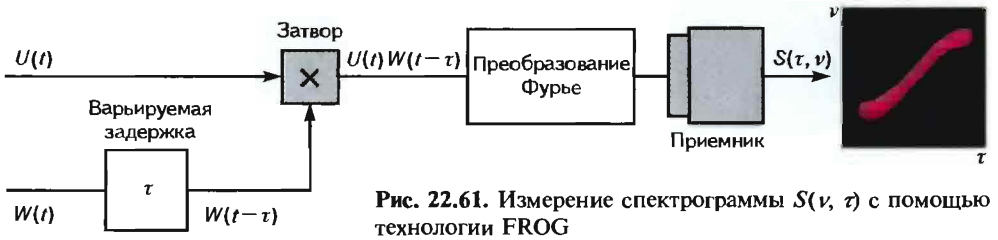


Рис. 22.61. Измерение спектрограммы $S(\nu, \tau)$ с помощью технологии FROG

За неимением достаточно короткой вырезающей функции $W(t)$, для этой цели можно использовать сам импульс $U(t)$ или другой связанный с ним импульс. Связь между $W(t)$ и $U(t)$ зависит от природы используемого оптического затвора, как видно из следующих примеров.

• Для затвора на основе генерации второй гармоники (ГВГ) (см. рис. 22.53) с входными импульсами $U(t)$ и $U(t - \tau)$ на основной частоте волна на частоте второй гармоники пропорциональна произведению $U(t)U(t - \tau)$, так что $W(t) \propto U(t)$ и

$$\Phi(\nu, \tau) = \int U(t)U(t-\tau)\exp(-j2\pi\nu t)\,\mathrm{d}t. \qquad (22.182)$$

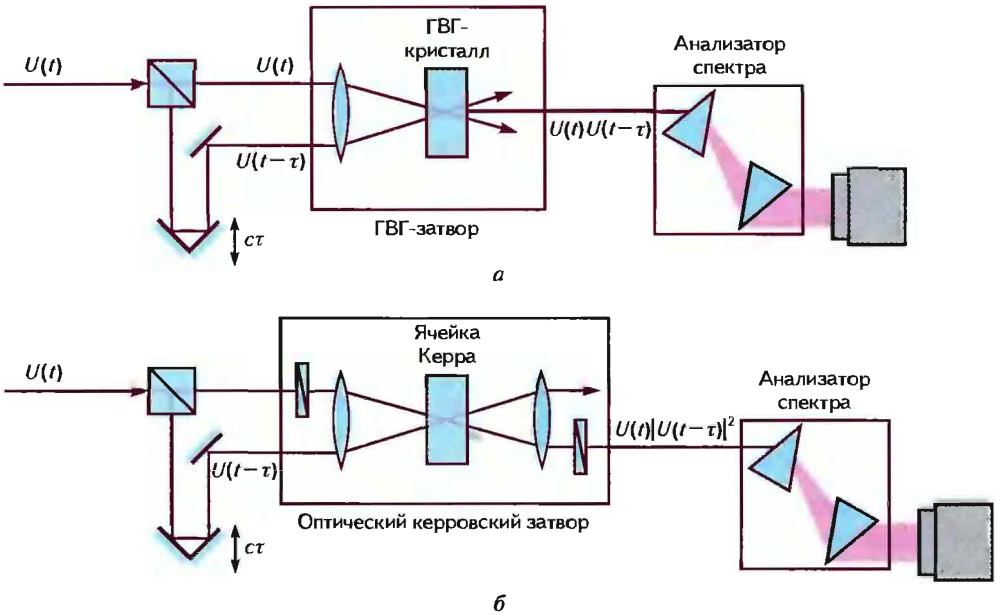


Рис. 22.62. Две реализации оптического стробирования с разрешением по частоте (FROG):
а — ГВГ-FROG; *б* — PG-FROG

Функция времени и частоты (22.182) известна как **функция распределения Вигнера**. Полная оптическая система, которая реализует блок-схему рис. 22.61, изображена на рис. 22.62, *а* и называется **ГВГ-FROG**. Эта система хорошо подходит для измерения одиночных импульсов, как обсуждалось ранее.

• Для поляризационного затвора на оптическом эффекте Керра (см. рис. 22.49, *а*) $W(t)$ пропорциональна интенсивности импульса $I(t)$, так что $W(t) \propto I(t) = |U(t)|^2$ и

$$\Phi(\nu,\ \tau) = \int U(t)\left|U(t-\tau)\right|^2 \exp(-j2\pi\nu t)\,\mathrm{d}t. \qquad (22.183)$$

При использовании этого затвора для реализации блок-схемы рис. 22.61 система называется FROG с поляризационным затвором (polarization-gated FROG, PG-FROG) и показана на рис. 22.62, *б*.

Предлагались и другие нелинейные оптические схемы, включающие затвор на основе генерации третьей гармоники, которому соответствует вырезающая функция $W(t) \propto U(t)^2$, и затвор на основе самодифракции, которому соответствует $W(t) \propto [U^*(t)]^2$.

Нахождение волновой функции импульса из спектрограммы

В любом из своих многочисленных вариантов спектрограмма $S(\nu, \tau)$ представляет собой двумерную «картину», которую можно использовать для характеристики оптического импульса или выявления его основных отличительных признаков. Ее также можно использовать для восстановления волновой функции импульса $U(t)$, ее абсолютной величины и фазы.

Вычисление $U(t)$ по измеренной спектрограмме $S(\nu, \tau)$ не является непосредственным. Общее выражение для $S(\nu, \tau)$, измеренной по любой из упомянутых выше схем с использованием затворов, можно записать в виде

$$S(\nu,\ \tau) = \left|\Phi(\nu,\tau)\right|^2 \qquad \Phi(\nu,\ \tau) = \int g(t,\ \tau)\exp(-j2\pi\nu t)\,\mathrm{d}t, \qquad (22.184)$$

где $g(t, \tau) = U(t)W(t - \tau)$; $W(t)$ связано с $U(t)$; например, $W(t) = U(t)$ для ГВГ-FROG и $W(t) = |U(t)|^2$ для PG-FROG.

Если бы комплексная функция $\Phi(\nu, \tau)$ была известна, то $U(t)$ можно было бы легко вычислить следующим образом. Посредством обратного преобразования Фурье функции $\Phi(\nu, \tau)$ по отношению к ν при каждом τ получаем

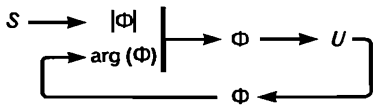
$$g(t,\ \tau) = \int \Phi(\nu,\ \tau)\exp(j2\pi\nu t)\,\mathrm{d}\nu. \qquad (22.185)$$

Зная $g(t, \tau) = U(t)W(t - \tau)$, можно вычислить волновую функцию $U(t)$ путем интегрирования по τ:

$$\int g(t,\ \tau)\,\mathrm{d}\tau = \int U(t)W(t-\tau)\,\mathrm{d}\tau = U(t)\int W(t-\tau)\,\mathrm{d}\tau \propto U(t). \qquad (22.186)$$

Коэффициент пропорциональности равен площади вырезающей функции, которая неизвестна.

Задача нахождения $\Phi(\nu, \tau)$ из измеренной $S(\nu, \tau) = |\Phi(\nu, \tau)|^2$ — это «задача недостающей фазы». Для решения этой и других задач такого рода было предложено много алгоритмов. Один из них — итерационный — содержит последовательность шагов, показанных на схеме:



1. По измеренной спектрограмме $S(\nu, \tau)$ определяется модуль $|\Phi(\nu, \tau)| = [S(\nu, \tau)]^{1/2}$. Начальное предположение о недостающей фазе $\arg\{\Phi(\nu, \tau)\}$ используется для вычисления $U(t)$ по описанной выше схеме [обратное преобразование Фурье $\Phi(\nu, \tau)$ по ν и интегрирование по τ] с точностью до неизвестного коэффициента пропорциональности.

2. Зная $U(t)$, вычисляем $\Phi(\nu, \tau)$ и определяем новое приближение для неизвестной фазы $\arg\{\Phi(\nu, \tau)\}$, которое в сочетании с измеренным модулем $|\Phi(\nu, \tau)|$ используется для получения нового уточненного приближения для $U(t)$.

3. Процесс повторяется до достижения сходимости и получения волновой функции $U(t)$, согласующейся с измеренной спектрограммой.

Пример показан на рис. 22.63 (см. цв. вклейку).

Рекомендуемая литература

ОБЩАЯ

См. также список литературы к гл. 21.

Diels J.-C., Rudolph W. Ultrashort Laser Pulse Phenomena. Elsevier, 2nd ed. 2006.

Agrawal G.P. Nonlinear Fiber Optics. Academic Press, 1991; 4th ed. 2006.

Gibbon P. Short Pulse Laser Interactions with Matter: An Introduction. Imperial College Press (London), 2005.

Rullière C., ed. Femtosecond Laser Pulses: Principles and Experiments. Springer-Verlag, 2nd ed. 2005.

Uesaka M., ed. Femtosecond Beam Science. Imperial College Press (London), 2005.

Kärtner F.X., ed. Few-Cycle Laser Pulse Generation and Its Applications. Springer-Verlag, 2004.

Andreev A.A. Generation and Application of Ultrahigh Laser Fields. Nova Science, 2002.

Trebino R., ed., Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses. Kluwer, 2000.

Kamiya T., Saito F., Wada O., Yajima H., eds. Femtosecond Technology: From Basic Research to Application Prospects. Springer-Verlag, 1999.

Shvartsburg A.B. Time-Domain Optics of Ultrashort Waveforms. Oxford University Press, 1996.

Sueta T., Okoshi T., eds. Ultrafast and Ultra-Parallel Optoelectronics. Wiley, 1996.

Trebino R., Walmsley I.A., eds. Generation, amplification, and measurement of ultrashort laser pulses. SPIE Proceedings. Vol. 2116, 1994.

Kaiser W., ed. Ultrashort Laser Pulses: Generation and Applications. Springer-Verlag, 1993.

Shvartsburg A.B. Non-Linear Pulses in Integrated and Waveguide Optics. Oxford University Press, 1993.

Akhmanov S.A., Vysloukh V.A., Chirkin A.S. Optics of Femtosecond Laser Pulses. American Institute of Physics, 1992.

Dianov E.M., Mamyshev P.V., Prokhorov A.M., Serkin V.N. Nonlinear Effects in Optical Fibers. Harwood, 1989.

Rudolph W., Wilhelmi B. Light Pulse Compression. Harwood, 1989.

Herrmann J., Wilhelmi B. Lasersfor Ultrashort Light Pulses. North-Holland, 1987.

КНИГИ ПО СОЛИТОНАМ

Mollenauer L., Gordon J. Solitons in Optical Fibers: Fundamentals and Applications. Academic Press, 2006.

Dauxois T., Peyrard M. Physics of Solitons. Cambridge University Press, 2006.

Malomed B.A. Soliton Management in Periodic System. Springer-Verlag, 2006.

Taylor J.R., ed. Optical Solitons: Theory and Experiment. Cambridge University Press, 1992; reprinted 2005.

Hasegawa A., Matsumoto M. Optical Solitons in Fibers. Springer-Verlag, 3rd ed. 2003.

Kivshar Y.S., Agrawal G.P. Optical Solitons: From Fibers to Photonic Crystals. Academic Press, 2003.

Akhmediev N.N., Ankiewicz A. Solitons, Nonlinear Pulses and Beams. Chapman & Hall, 1997.

Drazin P.G., Johnson R.S. Solitons: An Introduction. Cambridge University Press, 1989; reprinted 1993.

Olver P.J., Sattinger D.H., eds. Solitons in Physics, Mathematics, and Nonlinear Optics. Springer-Verlag, 1990.

Dodd R.K., Elbeck J.C., Gibson J.D., Morris H.C. Solitons and Nonlinear Wave Equations. Academic Press, 1982; reprinted 1984.

Lamb, Jr. G.L. Elements of Soliton Theory. Wiley, 1980.

Lonngren K., Scott A., eds. Solitons in Action. Academic Press, 1978.

СТАТЬИ

Issue on ultrafast science and technology. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 2, 2006.

Goulielmakis E., Uiberacker M., Kienberger R., Baltuška A., Yakovlev V., Scrinzi A., Westerwalbesloh Th., Kleineberg U., Heinzmann U., Drescher M., Krausz F. Direct Measurement of Light Waves. Science. Vol. 305, 2004. P. 1267—1270.

Mourou G.A., Umstadter D. Extreme Light. Scientific American. Vol. 286, № 5, 2002. P. 81—86.

Issue on ultrafast phenomena and their applications. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, № 4, 2001.

Leaird D.E., Weiner A.M. Femtosecond Direct Space-to-Time Pulse Shaping. IEEE Journal of Quantum Electronics. Vol. 37, 2001. P. 494—504.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Baltuška A., Pshenichnikov M.S., Wiersma D.A. Second-Harmonic Generation Frequency-Resolved Optical Gating in the Single-Cycle Regime. IEEE Journal of Quantum Electronics. Vol. 35, 1999. P. 459—478.

Walmsley I.A. Measuring Ultrafast Optical Pulses Using Spectral Interferometry. Optics & Photonics News. Vol. 10, № 4, 1999. P. 29—33.

Weiner A.M. Femtosecond Fourier Optics: Shaping and Processing of Ultrashort Optical Pulses. In International Trends in Optics and Photonics. T. Asakura, ed. Springer-Verlag, 1999. P. 233—246.

Issue on ultrafast optics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 4, № 2, 1998.

Segev M., Stegeman G.I.A. Self-Trapping of Optical Beams: Spatial Solitons. Physics Today. Vol. 51, № 8, 1998. P. 42—48.

Binjrajka V., Chang C.-C., Emanuel A.W.R., Leaird D.E., Weiner A.M. Pulse Shaping of Incoherent Light by Use of a Liquid-Crystal Modulator Array. Optics Letters. Vol. 21, 1996. P. 1756—1758.

Wefers M.M., Nelson K.A., Weiner A.M. Multi-Dimensional Femtosecond Pulse Shaping. In Ultrafast Phenomena X. P.F. Barbara, J.G. Fujimoto, W.H. Knox, W. Zinth, eds. Springer-Verlag, 1996. P. 159—160.

Weiner A.M. Femtosecond Optical Pulse Shaping and Processing. Progress in Quantum Electronics. Vol. 19, 1995. P. 161—238.

Haus H.A. Optical Fiber Solitons, Their Properties and Uses. Proceedings of the IEEE. Vol. 81, 1993. P. 970—983.

Trebino R., Kane D.J. Using Phase Retrieval to Measure the Intensity and Phase of Ultrashort Pulses: Frequency-Resolved Optical Gating. Journal of the Optical Society of America A. Vol. 10, 1993. P. 1101—1111.

Kempe M., Rudolph W. Femtosecond Pulses in the Focal Region of Lenses. Physical Review A. Vol. 48, 1993. P 4721—4729.

Kempe M., Rudolph W. The Impact of Chromatic and Spherical Aberration on the Focusing of Ultrashort Light Pulses by Lenses. Optics Letters. Vol. 18, 1993. P. 137—139.

Kempe M., Rudolph W., Stamm U., Wilhelmi B. Spatial and Temporal Transformation of Femtosecond Laser Pulses by Lenses and Lens Systems. Journal of the Optical Society of America B. Vol. 9, 1992. P. 1158—1165.

Weiner A.M. Dark Optical Solitons, in Optical Solitons: Theory and Experiment. J.R. Taylor, ed. Cambridge University Press, 1992. P. 378—408.

Gosnell T.R., Taylor A.J., eds. Selected Papers on Ultrafast Laser Technology. SPIE Optical Engineering Press (Milestone Series. Vol. 44), 1991.

Christov I. Generation and Propagation of Ultrashort Optical Pulses. In Progress in Optics. E. Wolf, ed. Elsevier. Vol. 29, 1991. P. 199—291.

Oughstun K.E. Pulse Propagation in a Linear, Causally Dispersive Medium. Proceedings of the IEEE. Vol. 79, 1991. P. 1379—1490.

Bell T.E. Light That Acts Like Natural Bits. IEEE Spectrum. Vol. 27, № 8, 1990. P. 56—57.

Kolner B.H., Nazarathy M. Temporal Imaging with a Time Lens. Optics Letters. Vol. 14, 1989. P. 630—632.

Zewail A.H. Laser Femtochemistry. Science. Vol. 242, 1988. P. 1645—1653.

Haus H.A., Islam M.N. Theory of the Soliton Laser. IEEE Journal of Quantum Electronics. Vol. 21, 1985. P. 1172—1188.

Задачи

К РАЗДЕЛУ 22.1

1. **Суперпозиция двух гауссовых импульсов.** Спектрально-ограниченный гауссов импульс прибавляется с гауссову импульсу с параметром чирпа a, остальные параметры у обоих импульсов одинаковы. Найдите выражения для интен-

сивности, фазы, спектральной интенсивности, спектральной фазы и параметра чирпа суммарного импульса.

2. **Импульс в виде гиперболического секанса.** Импульс имеет комплексную огибающую $\operatorname{sech}(t/\tau) = 1/\operatorname{ch}(t/\tau)$, где τ — постоянная. Покажите, что ширина распределения интенсивности на уровне половины максимума $\tau_{\text{пшпм}} = 1{,}76\tau$, спектральная интенсивность $S(\nu) = \operatorname{sech}^2(\pi^2\tau\nu)$, а ее ширина на уровне половины максимума $\Delta\nu = 0{,}993/\tau$. Сравните с гауссовым импульсом (таблицу преобразований Фурье см. в приложении А).

К РАЗДЕЛУ 22.2

1. **Толстая призма как фильтр с чирпом.** Толстая призма используется как фильтр с чирпом. Угол падения выбран так, чтобы выполнялось условие Брюстера, с целью минимизировать потери на отражение. Угол при вершине α выбран таким, что падающий луч и центральный отклоненный луч симметричны относительно призмы. При этих двух условиях покажите, что угол отклонения θ_3 удовлетворяет условию $\mathrm{d}\theta_3/\mathrm{d}\nu = -2$, а коэффициент чирпа дается формулой $b \approx -4(n - N)^2\ell_0\lambda_0/\pi c^2$. Покажите, что при равных параметрах коэффициент чирпа больше, чем у тонкой призмы (см. пример 22.2) в $4/\alpha^2$ раз.

2. **Фильтр с чирпом на основе решетки Брэгга.** Спроектируйте фильтр с чирпом на основе решетки Брэгга для импульсов с центральной частотой $\nu_0 = 300$ Гц (длина волны 1 мкм) и шириной на высоте половины максимума $\tau_{\text{пшпм}} = 0{,}44$ пс. Фильтр должен иметь коэффициент чирпа $b = (2\text{ пс})^2$. Укажите размеры решетки и максимальный и минимальный шаги ее периодической структуры, гарантирующие отражение решеткой всех спектральных компонент импульса.

К РАЗДЕЛУ 22.3

1. **Распространение прямоугольного импульса через оптическое волокно.** Прямоугольный импульс шириной τ распространяется по оптическому волокну, которое действует как чирпирующий фильтр с параметром чирпа $b = D_\nu z/\pi$ [см. (22.71)]. Покажите, что на достаточно больших расстояниях такой импульс меняет свою форму от прямоугольной до функции sinc. Выведите выражение для новой ширины импульса.

2. **Временно́е изображение с помощью «временно́й линзы».** Оптический импульс с шириной τ_1 и произвольной формой проходит расстояние d_1 в волокне с положительной ДГС, после чего он модулируется фазовым множителем $\exp(j\zeta t^2)$, а затем проходит расстояние d_2 в волокне из того же материала. Ширина выходного импульса τ_2. Предполагая, что d_1 и d_2 намного больше дисперсионной длины волокна z_0, покажите, что новый импульс будет задержанной копией исходного с временны́м увеличением $\tau_2/\tau_1 = d_2/d_1$, если выполнено условие $1/d_1 + 1/d_2 = 1/f$, где $f = -\pi/\zeta D_\nu$ — фокусное расстояние фазового модулятора для данной среды (ζ отрицательно, а f положительно). Это означает, что данная система эквивалентна временно́й изображающей системе.

К РАЗДЕЛУ 22.5

1. **Смешение чирпированных волн и усиление чирпа.**

а. Три импульсные коллинеарные плоские волны с центральными угловыми частотами ω_1, ω_2 и $\omega_3 = \omega_1 + \omega_2$ смешиваются в среде с коэффициентом нелинейности второго порядка d. Среда обладает дисперсией и имеет показатели преломления n_1, n_2 и n_3 и групповые скорости v_1, v_2 и v_3 на трех центральных частотах. Все три импульса имеют чирп с параметрами a_1, a_2 и a_3. Каким должно быть соотношение между a_1, a_2 и a_3 для эффективного трехволнового смешения?

Указание. Считайте, что законы сохранения энергии и импульса (условия синхронизма) выполняются в каждый момент времени.

б. Покажите, что параметр чирпа сигнальной и/или холостой волны может быть больше, чем у накачки. Обсудите возможные применения такого «усиления чирпа».

2. **Импульсное трехволновое смешение в среде с ДГС.** Выведите связанные уравнения для трех волн (22.129) для среды с ДГС. Вы можете воспользоваться следующей процедурой. Начните с уравнения Гельмгольца с источником, равным Фурье-образу

$$S = \mu_0 \frac{\partial^2 \mathcal{P}_{\text{нл}}}{\partial t^2},$$

где $\mathcal{P}_{\text{нл}} = 2d\mathcal{E}^2$. Выразите поле $\mathcal{E}$ как суперпозицию трех волн с различными центральными частотами и медленно меняющимися огибающими и преобразуйте уравнение Гельмгольца в три отдельных уравнения на трех частотах. Упростите эти уравнения, используя приближения медленной огибающей, слабой дисперсии и трехчленного тейлоровского разложения постоянных распространения. Используйте обратное преобразование Фурье для перехода во временно́е представление.

3. **Зависимость характеристик солитона от ДГС.** Сравните характеристики двух фундаментальных солитонов с одинаковой энергией, распространяющихся в двух протяженных средах (например, оптических волокнах) с коэффициентами ДГС $D_\lambda = 20$ пс/(км · нм) и $D_\lambda = 10$ пс/(км · нм), а в остальном совершенно одинаковых (равны показатели преломления и коэффициенты Керра n_2). Сравните ширину солитонов, их пиковую амплитуду, площадь под профилем амплитуды и солитонное расстояние.

4. **Солитоны в оптическом волокне.** Покажите, что произведение пиковой интенсивности и дисперсионной длины для фундаментального солитона — постоянная величина $I_0|z_0| = \lambda_0/4\pi n_2$. Для волокна из плавленого кварца с коэффициентом Керра $n_2 = 3{,}19 \cdot 10^{-20}$ м^2/Вт определите пиковую интенсивность I_0, если дисперсионная длина $|z_0| = 30$ км.

К РАЗДЕЛУ 22.6

1. **Измерение гауссова импульса.** Гауссов спектрально ограниченный импульс с длительностью (по половине высоты максимума) 50 фс и центральной частотой, соответствующей длине волны 800 нм, измеряется с помощью коррелятора интенсивности, показанного на рис. 22.54.

   а. Определите форму и ширину (по половине высоты максимума) измеряемой автокорреляционной функции.

   б. Есть предположение, что измерение можно улучшить, если один из импульсов, например, распространяющийся в верхнем плече, специально растянуть путем пропускания через волокно из кварцевого стекла. Какой должна быть длина волокна, если импульс нужно растянуть в 5 раз? (Кварцевое стекло имеет коэффициент дисперсии $D_\lambda = -110$ пс/(км · нм) при 800 нм.) Какова будет ширина новой корреляционной функции после введения волокна?

   в. Если эту идею применить к нелинейному интерферометру, показанному на рис. 22.59, и волокно также поместить в верхнее плечо, то какие положительные моменты и какие проблемы возникнут при измерениях импульсов?

2. **Интерферометр с приемником на двухфотонном поглощении.** Интерферометр с приемником на двухфотонном поглотителе измеряет функцию (22.180). Сравните этот интерферометр с нелинейным интерферометром на основе генерации второй гармоники с последующим однофотонным приемником, измеряющим (22.172). Разложите (22.180) и приведите к виду, аналогичному (22.173), чтобы провести сравнение различных членов.

ГЛАВА 23 ОПТИЧЕСКИЕ МЕЖСОЕДИНЕНИЯ И КОММУТАТОРЫ

Межсоединения и коммутаторы — существенные компоненты распределенных систем, таких как системы связи, обработки информации и вычислительной техники. Появление оптических волокон как перспективной основы систем связи и сетей стимулировало разработку разнообразных фотонных коммутаторов, а внедрение уплотнения с разделением по длине волны (wavelength division multiplexing — WDM) (см. подразд. 24.4.2) добавило новое измерение в технику коммутации, что, в свою очередь, мотивировало развитие специальных фотонных коммутаторов. С другой стороны, несколько десятилетий исследовательской работы в области цифровых оптических вычислений до сих пор не дали промышленных продуктов, способных конкурировать с электронными компьютерами. Тем не менее побочным продуктом этих усилий явилась разработка ряда технологий создания оптических логических элементов, что позволило оптическим межсоединениям занять важное место в электронных компьютерных системах.

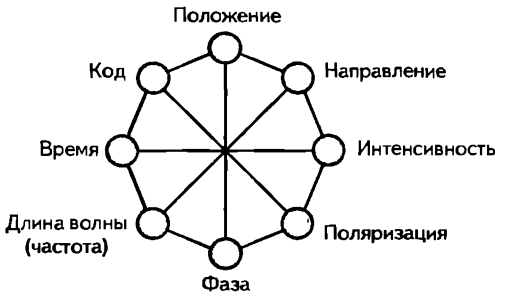


Рис. 23.1. Признаки оптического пучка, которые могут быть использованы для модуляции, мультиплексирования, маршрутизации и коммутации

Оптический пучок характеризуется несколькими характерными свойствами — положением, направлением, длиной волны (или частотой), интенсивностью, фазой (для когерентных волн), поляризацией, временем (для оптических импульсов) или кодом (на основе последовательности оптических импульсов), как схематически показано на рис. 23.1. Один из соответствующих параметров, например интенсивность, можно модулировать и использовать для передачи сигнала между двумя точками. Другой параметр, например длина волны, может служить меткой различных сигналов, передаваемых одним пучком. Такой процесс называется мультиплексированием (уплотнением). Для разделения различных сигналов (демультиплексирования) требуется оптический маршрутизатор, чувствительный к длине

волны. Коммутаторы используются для направления оптического сигнала от одной точки к другой по одному из возможных адресов назначения. Такие переключатели называются **пространственными**. Кроме этого коммутатор может передавать сигнал из одного временного окна в другое или с одной длины волны на другую. Такой коммутатор называется **коммутатором с разделением по времени или по длине волны**. Коммутаторы, управляемые адресом, включенным в каждый пакет приходящих данных, называются **коммутаторами пакетов**.

О данной главе

В этой главе дается введение в общие принципы устройства оптических межсоединений, пассивных оптических маршрутизаторов и фотонных коммутаторов (рис. 23.2). Многие фундаментальные принципы фотоники, введенные в предшествующих главах (Фурье-оптика и голография, оптика волноводов, электрооптика, оптика полупроводников, акустооптика, нелинейная оптика и оптика сверхбыстрых процессов), найдут свое применение здесь.

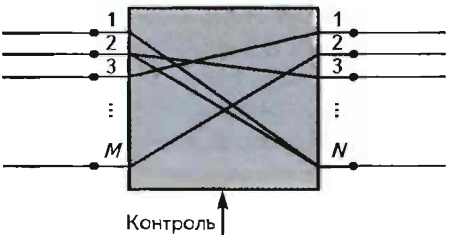


Рис. 23.2. Система, направляющая каждый из M оптических пучков, входящих в M точках, в один из N выходных портов. Если соединения фиксированы и не зависят от природы входящих пучков, система называется **межсоединением**. Если оптический параметр входящих пучков, такой как длина волны, диктует, на какой из выходных портов пучок должен быть направлен, система называется **пассивным маршрутизатором**. Если конфигурация соединений может меняться под действием внешнего управляющего сигнала, система называется **пространственным коммутатором**

Раздел 23.1 охватывает оптические межсоединения через свободное пространство, а также на основе планарных волноводных схем и оптических волокон. В этих межсоединениях входные пучки направляются в назначенные выходные порты независимо от их свойств или переносимой ими информации.

Пассивные оптические маршрутизаторы описаны в разд. 23.2. Здесь каждый входной оптический пучок направляется в один или более портов, исходя из таких его свойств, как длина волны, поляризация или интенсивность. Например, компоненты с разными длинами волн одного и того же пучка могут направляться в разные порты. В этом случае прибор действует как демультиплексор с разделением по длине волны. Операция, обратная данной, когда пучки с различными длинами волн объединяются в один пучок, называется мультиплексированием (уплотнением) с разделением по длине волны (WDM). Эти операции играют важную роль в современных оптических системах связи (см. разд. 24.4).

Фотонные пространственные коммутаторы описаны в разд. 23.3. Простейшими примерами таких устройств является переключатель с двумя положениями «вкл.—выкл.», который соединяет или разъединяет два порта (т. е. передает или

блокирует пучок), или переключатель, который селективно направляет пучок по одному из двух возможных путей, независимо от параметров пучка или переносимых им данных. За вводными сведениями об общих типах и свойствах коммутаторов следует краткий обзор различных технологий их исполнения, включая электрооптические, полупроводниковые лазерные усилители, жидкокристаллические, микромеханические системы и полностью оптические устройства. Описаны также коммутаторы с разделением по времени и коммутаторы пакетов.

Раздел 23.4 посвящен **оптическим логическим элементам** на основе бистабильных оптических устройств. Это переключатели с памятью, т. е. системы, у которых выходной сигнал принимает одно из двух (или нескольких) значений, в зависимости от текущего значения и предыдущей истории сигнала на входе.

23.1. ОПТИЧЕСКИЕ МЕЖСОЕДИНЕНИЯ

Цифровые системы для обработки сигналов и вычислений содержат большое число связанных между собой логических элементов, переключателей и элементов памяти. В электронных системах такие межсоединения осуществляются посредством проводов, коаксиальных кабелей или проводящих каналов внутри полупроводниковых интегральных схем. В фотонике межсоединения могут аналогичным образом осуществляться с использованием оптических волноводов с интегрально-оптическими или волоконно-оптическими разветвителями. Для межсоединений можно использовать также пучки света в свободном пространстве, направляемые микролинзами или дифракционными оптическими элементами. Такая возможность отсутствует в электронных системах, поскольку электронные пучки должны быть в вакууме и не могут пересекаться друг с другом без взаимного отталкивания.

На рис. 23.3 продемонстрировано несколько конфигураций межсоединений (называемых также **разветвителями**). Каждый входной порт связан с одним или многими выходными портами и наоборот. Например, в выходном Т-разветвителе входной порт соединяется с каждым из выходных портов. В межсоединении типа «звезда» или направленном 3-дБ-разветвителе каждый входной порт соединен с каждым выходным портом.

Матрица межсоединения

Схемы, показанные на рис 23.3, указывают лишь наличие связи между элементами и ничего не говорят о количественных соотношениях между оптическими полями или интенсивностями в соединяемых портах. Для линейных когерентных оптических межсоединений оптическое поле $U_\ell^{(o)}$ в выходном порте ℓ ($\ell = 1, 2, \quad N$) связано с оптическим полем $U_m^{(i)}$ во входных портах $m = 1, 2, \ldots, M$ линейной суперпозицией

$$U_\ell^{(o)} = \sum_{m=1}^{M} T_{\ell m} U_m^{(i)}, \tag{23.1}$$

где весовые коэффициенты $\{T_{lm}\}$ — комплексные числа, определяющие **матрицу межсоединения T**. Например, 3-дБ-делитель 2 × 2 на рис. 23.3, *в* описывается матрицей межсоединения

$$\mathbf{T} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}, \tag{23.2}$$

которая идентична матрице идеального светоделителя (см. подразд. 7.1.1). Для этого прибора оптическая мощность, приносимая одним пучком (в отсутствие другого) делится поровну между двумя выходящими пучками. Аналогично описываются и другие межсоединения. Матрица для каскада межсоединений определяется посредством перемножения матриц каждого из них, как описано в подразд. 7.1.1.

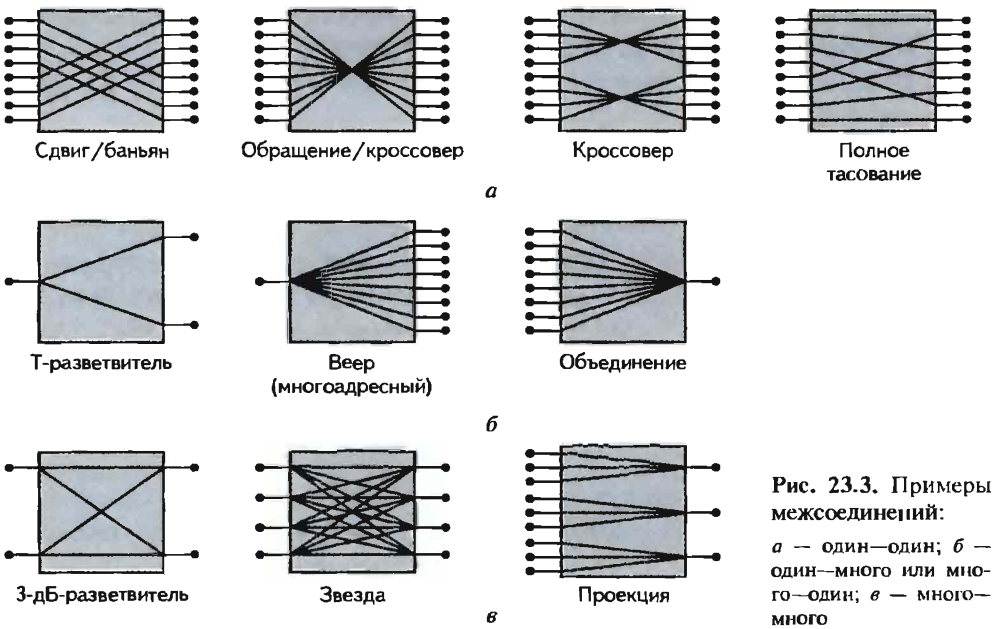


Рис. 23.3. Примеры межсоединений:
а — один—один; *б* — один—много или много—один; *в* — много—много

Поскольку свет предполагается когерентным, фазовые соотношение между входными пучками, а также фазы, вносимые элементами соединительного устройства, играют важную роль. Действительно, интерферометрические эффекты часто используются для перераспределения входной мощности между выходными портами заранее заданным образом. Если свет некогерентный, то интенсивность (а значит и мощность) в каждом выходном порте представляет собой взвешенную суперпозицию интенсивностей (мощностей) на входных портах (см. подразд. 11.3.2):

$$P_l^{(o)} = \sum_{m=1}^{M} |T_{lm}|^2 P_m^{(i)}. \tag{23.3}$$

Например, мощности на входе и выходе 3-дБ-разветвителя связаны матрицей, все элементы которой равны 1/2.

Основными рабочими параметрами практических соединительных устройств являются следующие мощностные отношения, обычно выражаемые в дБ = [−10 log (1/отношение)]:

- **Вносимые потери** определяются коэффициентом пропускания мощности из порта в порт, в идеале это 0 дБ для канала без потерь.
- Для устройства, распределяющего мощность между несколькими выходными портами, **коэффициент деления мощности** определяется отношением мощности в выходном порте к суммарной мощности всех выходных портов. Например, для идеального 3-дБ- (симметричного) разветвителя коэффициент деления мощности равен 3 дБ.
- **Перекрестные помехи** определяются отношением мощности нежелательного сигнала в выходном порте к входной мощности, подлежащей направлению в другой выходной порт (или несколько портов).
- **Избыточные потери** определяются отношением полной выходной мощности к полной входной мощности.

Невзаимные соединительные устройства: изоляторы и циркуляторы

Указание портов межсоединения в качестве входных и выходных подразумевает направленность передачи — от входа в выходу (слева направо в примерах на рис. 23.3). Некоторые межсоединения взаимны, т. е. если передача направлена от выходных портов к входным матрица межсоединения остается неизменной. В противном случае устройство называется **невзаимным**.

Изоляторы

Простейшим примером невзаимного межсоединения является однонаправленная связь с матрицей 1×1, пропускающая сигнал только в одном направлении, как показано на рис. 23.4, *а*. Это часто реализуется с помощью оптического изолятора, действующего подобно диоду или однонаправленному клапану (см. подразд. 6.4.2). Действие изолятора описывается вносимыми потерями [коэффициентом пропускания по мощности в прямом направлении, дБ] и **обратной изоляцией** [коэффициентом пропускания по мощности в обратном направлении, дБ].

Многопортовые невзаимные межсоединения

Обозначение «вход/выход» неприменимо, если порт играет двойную роль как передатчик и приемник. В этом случае соединение характеризуется просто количеством портов. На рис. 23.4, *б* и *в* представлены примеры трехпортовых межсоединений с использованием однонаправленных связей. Такие соединения используются в двухсторонних системах связи, как показано для четырехпортового межсоединения на рис. 23.4, *д*. В другой четырехпортовой системе,

показанной на рис. 23.4, *г*, связь между левыми и правыми портами имеет параллельную конфигурацию в прямом направлении (слева направо) и перекрестную — в обратном направлении (справа налево).

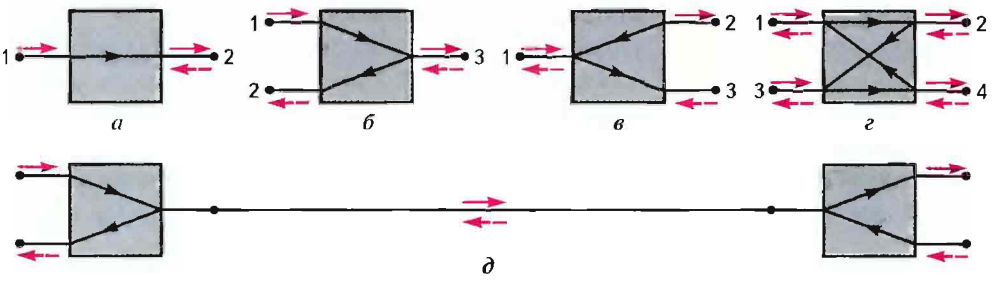


Рис. 23.4. Двухпортовый однонаправленный соединитель (изолятор) (*а*). Трехпортовое межсоединение с использованием двух однонаправленных связей (*б*) и (*в*). Двухсторонняя связь (*г*) и (*д*)

Циркуляторы

Еще одним примером невзаимного соединения является оптический циркулятор. Это соединительное устройство с тремя или более портами, связанными однонаправленными связями, указывающими в одном направлении. Как показано на рис. 23.5, четырехпортовый циркулятор эквивалентен межсоединению, показанному на рис. 23.4, *г*. Циркуляторы находят множество приложений в системах связи и сетях. Пример использования циркуляторов в оптических мультиплексорах с заменой данных в выделенном канале (оптических мультиплексорах ввода/вывода, OMBB) описан в подразд. 23.2.1 (см. рис. 23.19).

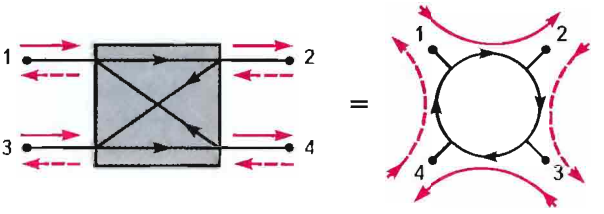


Рис. 23.5. Четырехпортовый циркулятор. Обе конфигурации эквивалентны

23.1.1. Межсоединения в свободном пространстве на основе рефракции и дифракции

Обычные оптические элементы (зеркала, линзы, призмы и т. д.) стандартным образом используются в оптических системах в качестве соединений. Один из примеров — изображающая система, в которой линза используется для связи точек в плоскости предмета с точками в плоскости изображения. Чтобы оценить порядок плотности такого соединения, заметим, что в

хорошей изображающей системе примерно 1000 × 1000 независимых точек на 1 мм^2 в плоскости предмета оптически связаны посредством линзы с соответственными 1000 × 1000 точками на 1 мм^2 в плоскости изображения. Для того чтобы осуществить это с помощью электротехники, требуется миллион непересекающихся и хорошо изолированных проводящих каналов на 1 мм^2!

Стандартные оптические элементы можно использовать для создания специальных межсоединений, таких как сдвиг, обращение, кроссовер, тасование, объединение, разветвление, звезда и проекция, как показано на рис. 23.6 (см. также рис. 23.3). Их миниатюризация возможна с применением компонентов **микрооптики**, таких как миниатюрные светоделители, линзы, стрежни с градиентом показателя преломления, призмы, фильтры и решетки. Такие компоненты совместимы с оптическими волокнами, которые часто используются для передачи света.

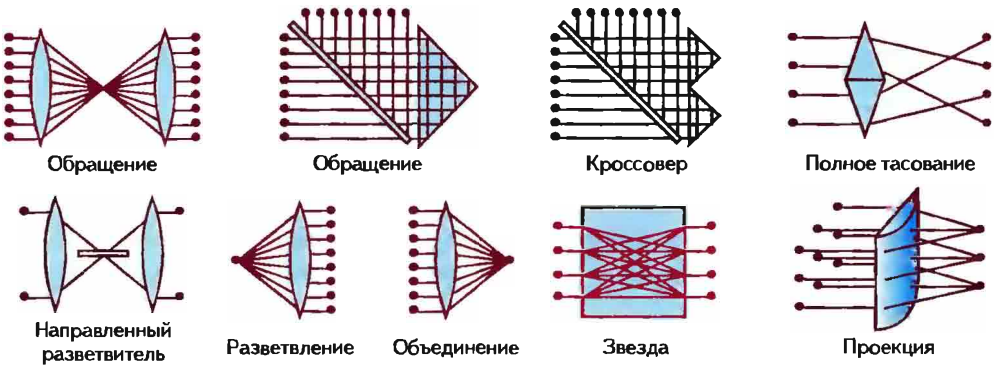


Рис. 23.6. Примеры простых оптических межсоединений, создаваемых обычными оптическими элементами. Призма изгибает параллельные лучи света предпочтительно и задает упорядоченную карту межсоединения, соответствующую обращению или кроссоверу. Две призмы, ориентированные соответствующим образом, обеспечивают полное тасование — операцию, используемую в алгоритмах сортировки и быстрого преобразования Фурье (БПФ). Линза задает разветвление, объединение или обращение. Светоделитель с двумя линзами создает направленный разветвитель. Стеклянный стержень служит звездообразным соединением. Астигматическая оптическая система, такая как цилиндрическая линза, осуществляет проекцию, при которой точки каждого ряда во входной плоскости связываются с одной точкой в плоскости выхода

Реализация карт произвольных оптических межсоединений требует создания специальных оптических элементов, которые могут быть весьма сложными и непрактичными. Однако для создания оптических межсоединений высокой плотности были успешно использованы голограммы, генерируемые с помощью компьютера и состоящие из большого числа сегментов фазовых решеток с различными пространственными частотами и ориентациями.

Фазовая решетка — тонкий оптический элемент, комплексный амплитудный коэффициент пропускания которого представляет собой двумерную пери-

одическую функцию с единичным модулем. Простейшая фазовая решетка имеет комплексный амплитудный коэффициент пропускания

$$t(x,\ y) = \exp\left[-j2\pi\left(\nu_x x + \nu_y y\right)\right],$$

где ν_x и ν_y — пространственные частоты в направлениях x и y; они определяют период и ориентацию решетки. В подразд. 2.4.2 и 4.1.1 было показано, что когда когерентный оптический пучок с длиной волны λ проходит через такую решетку, он приобретает фазовый сдвиг, заставляющий его отклоняться на углы $\arcsin \lambda\nu_x \approx \lambda\nu_x$ и $\arcsin \lambda\nu_y \approx \lambda\nu_y$, где $\lambda\nu_x \ll 1$ и $\lambda\nu_y \ll 1$, как показано на рис. 23.7. Меняя пространственные частоты ν_x и ν_y, т. е. периодичность и ориентацию решетки, можно менять углы отклонения пучка.

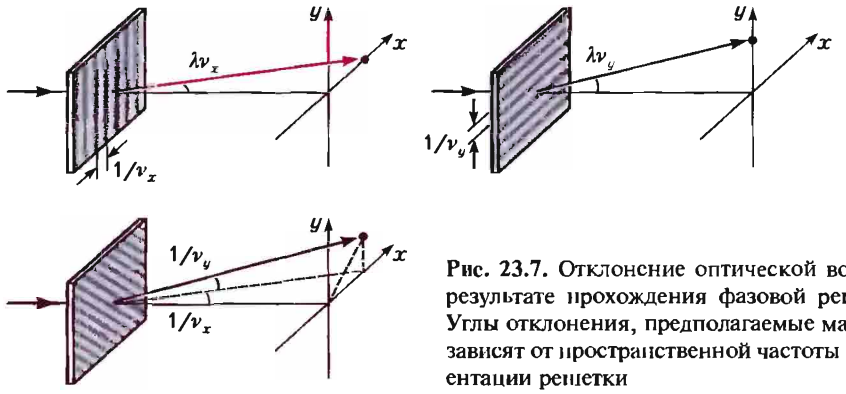


Рис. 23.7. Отклонение оптической волны в результате прохождения фазовой решетки. Углы отклонения, предполагаемые малыми, зависят от пространственной частоты и ориентации решетки

Как описано в подразд. 4.1.1, этот принцип можно использовать для изготовления межсоединения с произвольной картой, если решетку составить из сегментов с различными пространственными частотами. Оптические пучки, проходящие через различные сегменты решетки, претерпевают различный изгиб, в соответствии с требуемой картой межсоединения (рис. 23.8).

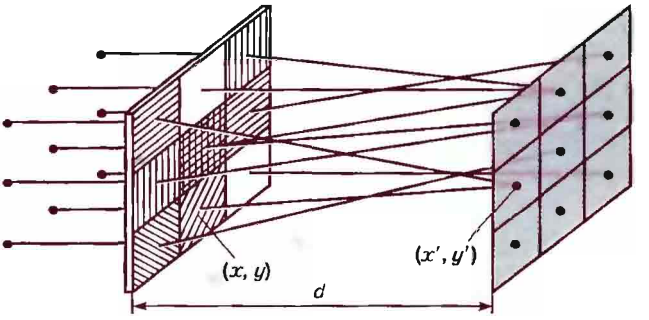


Рис. 23.8. Голографическая карта межсоединения, созданная матрицей фазовых решеток с различной периодичностью и ориентацией

Если сегмент решетки находится в точке (x, y) и имеет частоты $\nu_x = \nu_x(x, y)$ и $\nu_y = \nu_y(x, y)$, то углы отклонения равны приблизительно $\lambda\nu_x$ и $\lambda\nu_y$, и пучок

приходит в точку выходной плоскости с координатами (x', y'), удовлетворяющими соотношениям

$$\frac{x'-x}{d} \approx \lambda \nu_x; \quad \frac{y'-y}{d} \approx \lambda \nu_y, \tag{23.4}$$

где d — расстояние между голограммой и выходной плоскостью, а все углы считаются малыми. При заданном соотношении между (x', y') и (x, y), т. е. при заданной карте соединения, необходимые для изготовления голограммы пространственные частоты ν_x и ν_y могут быть определены с помощью (23.4) для каждой точки.

Голографические соединительные устройства могут устанавливать соединение одного порта с многими или многих портов с одним путем отображения одной точки в несколько разных точек или нескольких точек в одну. Например, на рис. 23.8 центральный элемент решетки является суперпозицией двух гармонических функций, так что его комплексный амплитудный коэффициент пропускания

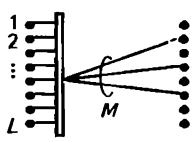
$$t(x, y) = \exp\left[-j2\pi\left(\nu_{x_1} x + \nu_{y_1} y\right)\right] + \exp\left[-j2\pi\left(\nu_{x_2} x + \nu_{y_2} y\right)\right];$$

падающий пучок расщепляется на две компоненты одинаковой интенсивности, одна из которых отклоняется на углы $(\lambda\nu_{x_1}, \lambda\nu_{y_1})$, а другая — $(\lambda\nu_{x_2}, \lambda\nu_{y_2})$, причем все углы малые. Взвешенные межсоединения могут быть реализованы путем приписывания различных весов различным решеткам. Следовательно, можно создать произвольное межсоединение путем соответствующего выбора пространственных частот решетки в каждой точке голограммы.

Упражнение 23.1

Пропускная способность соединения

Произведение размера на ширину полосы для квадратной голограммы размером $a \times a$ равно $(Ba)^2$, где B — наибольшая пространственная частота (число штрихов на 1 мм), которая может быть реализована при печати голограммы. Покажите, что если голограмма используется для отправки L входящих пучков по M направлениям, то произведение ML не может превышать $(Ba)^2$



$$\boxed{ML \le (Bd)^2.}$$

Указание. Используйте анализ, аналогичный представленному в подразд. 19.2.2 в связи с акустооптическими соединительными устройствами [см. (19.51)].

Каково максимальное число соединений на 1 мм^2, если наибольшая пространственная частота равна 1000 штрихов/мм и каждая точка входной плоскости связана с каждой точкой выходной плоскости?

В пределе, в котором элементы решетки имеют бесконечно малые площади, мы имеем непрерывную (а не дискретную) карту соединения, т. е. правило геометрического преобразования координат, преобразующее каждую точку (x, y) в плоскости входа в соответствующую точку (x', y') в плоскости выхода. Если желаемое преобразование определяется двумя непрерывными функциями

$$x' = \psi_x(x, y); \quad y' = \psi_y(x, y), \tag{23.5}$$

то частоты решетки должны непрерывно меняться по x и y, как у частотно-модулированного (ЧМ) сигнала (рис. 23.9). Предполагая, что решетка имеет коэффициент пропускания $t(x, y) = \exp[-j\varphi(x, y)]$, находим для локальных (или мгновенных) частот

$$2\pi\nu_x = \frac{\partial\varphi}{\partial x}; \quad 2\pi\nu_y = \frac{\partial\varphi}{\partial y}. \tag{23.6}$$

(это аналог мгновенной частот ЧМ-сигнала). Подставляя это в (23.4), получаем

$$\boxed{\frac{\psi_x(x, y) - x}{d} = \frac{\lambda}{2\pi}\frac{\partial\varphi}{\partial x}; \quad \frac{\psi_y(x, y) - y}{d} = \frac{\lambda}{2\pi}\frac{\partial\varphi}{\partial y}.} \tag{23.7}$$

Эти два дифференциальных уравнения в частных производных можно решить и определить фазовую функцию решетки $\varphi(x, y)$.

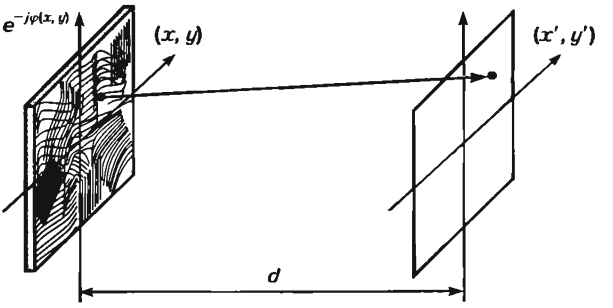


Рис. 23.9. Дифракция на фазовой голограмме в качестве непрерывной системы межсоединения

Пример 23.1

Карта объединения

Предположим, что все точки (x, y) плоскости входа должны отобразиться в точку $(x', y') = (0, 0)$ плоскости выхода, так что получается карта объединения. Подставляя $\psi_x(x, y) = \psi_y(x, y) = 0$ в (23.7) и решая два дифференциальных уравнения в частных производных, получаем

$$\varphi(x, y) = -\frac{\pi(x^2 + y^2)}{\lambda d}.$$

Не удивительно, что это в точности фазовый сдвиг, вносимый линзой с фокусным расстоянием d (см. подразд. 2.4.2).

Упражнение 23.2

Логарифмическая карта

Покажите, что логарифмическое преобразование координат

$$x' = \psi_x(x,\ y) = \ln x; \tag{23.8}$$

$$y' = \psi_y(x,\ y) = \ln y \tag{23.9}$$

реализуется голограммой с фазовой функцией

$$\varphi(x,\ y) = \frac{2\pi}{\lambda d}\left(x \ln x - x - \frac{1}{2}x^2 + y \ln y - y - \frac{1}{2}y^2\right). \tag{23.10}$$

Как только соответствующая фаза $\varphi(x, y)$ определена, оптически элемент можно изготовить с помощью **компьютерной голографии**. Этот подход позволяет кодировать комплексную функцию $\exp[-j\varphi(x, y)]$ с помощью бинарной функции, принимающей только два значения, например 1 и 0 или −1 и 1. Это то же самое, что кодировать изображение с помощью черных точек, размер или плотность которых меняются пропорционально локальной степени почернения изображения (примером является полутоновая печать изображений в газетах). С помощью компьютера бинарное изображение печатается на маске (транспаранте), которая играет роль голограммы. Бинарное изображение можно также напечатать с помощью гравирования штрихов на подложке, которые модулируют фазу падающей когерентной волны. Такая техника известна как **поверхностно-рельефная голография**. Ссылки по обсуждению компьютерной голографии приведены в списке литературы для чтения.

Динамические (с изменяемой конфигурацией) межсоединения можно сконструировать с использованием акустооптических или магнитооптических приборов. Однако число соединяемых точек намного меньше, чем достигается в случае голографических решеток. Динамические голографические соединения можно построить на основе нелинейно-оптических процессов, таких как четырехволновое смешение в фоторефрактивных материалах. Две волны интерферируют и создают решетку, от которой третья волна отражается. Угол между двумя волнами определяет пространственную частоту решетки, которая определяет наклон отраженной волны (см. разд. 20.4 и подразд. 21.3.5).

23.1.2. Волноводные межсоединения

В планарных оптических схемах межсоединения создаются путем нанесения оптических волноводов на подложку из $LiNbO_3$ или кремния (см. подразд. 8.5.2) примерно так же, как металлические проводящие элементы в электронных печатных платах или интегральных схемах. Примеры изображены на рис. 23.10, а комбинации и каскады этих основных соединений можно использовать для создания более сложных межсоединений.

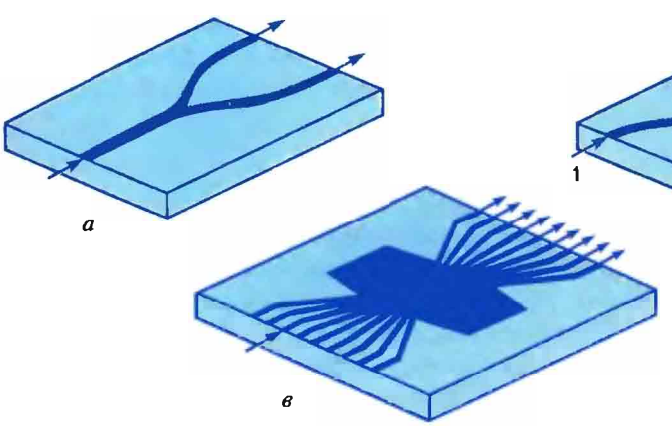


Рис. 23.10. Интегрально-оптические схемы, реализующие некоторые из межсоединений рис. 23.3:

а — Т-развертвитель; *б* — 3-дБ-разветвитель; *в* — звездообразный разветвитель

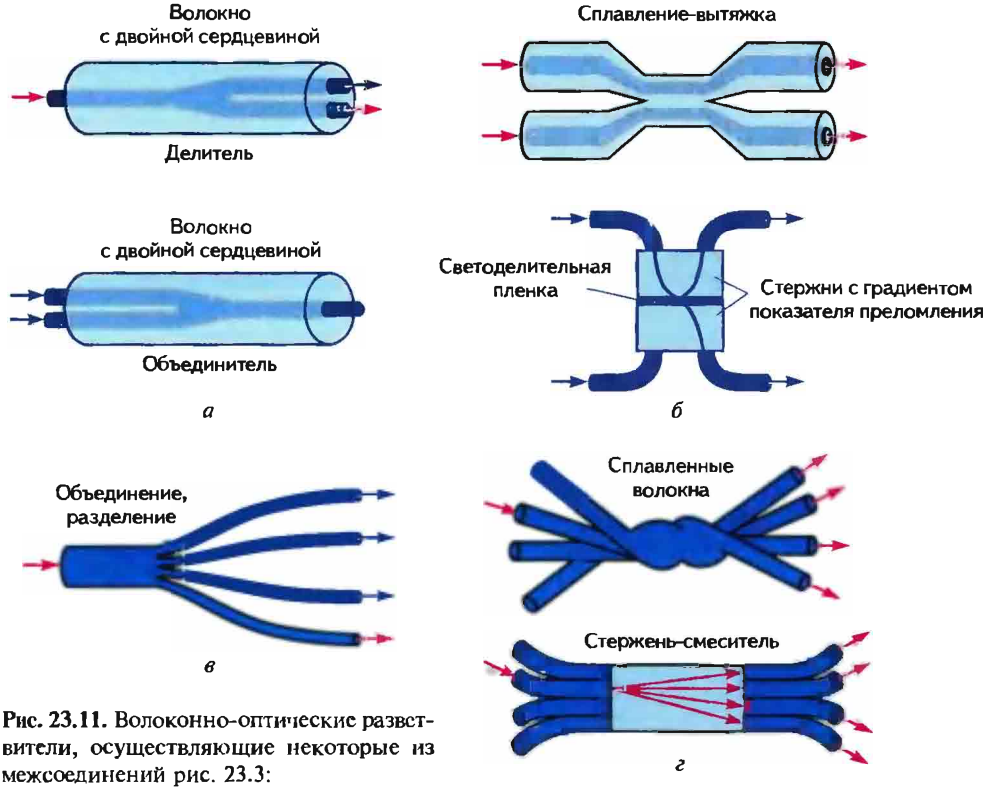


Рис. 23.11. Волоконно-оптические разветвители, осуществляющие некоторые из межсоединений рис. 23.3:

а — волокно с двойной сердцевиной, используемое как Т-разветвитель, делитель или объединитель; *б* — 3-дБ-разветвители: первый — из двух сплавленных волокон, второй — с двумя линзами в виде стержней с градиентом показателя преломления, разделенными светоделительной пленкой; *в* — объединитель или делитель; *г* — звездообразные соединения: одно с использованием сплавленных волокон, другое — со слоем стекла, проходя через который свет из одного волокна расходится и достигает всех других волокон

Волноводные разветвители используются для распределения оптической мощности в заданных количествах. Например, разветвитель, показанный на рис. 23.10, *б*, описывается матрицей межсоединения (см. подразд. 8.5.2)

$$T = \begin{bmatrix} \cos CL & -j\sin CL \\ -j\sin CL & \cos CL \end{bmatrix}, \tag{23.11}$$

где C — коэффициент связи; L — длина взаимодействия. Мощность, приходящая в порт 1, делится между выходными портами 1 и 2 пропорционально $\cos^2 CL$ и $\sin^2 CL$ соответственно. При $CL = \pi/4$ мощность делится поровну и разветвитель становится 3 дб-делителем.

Приложения оптоволоконной технологии, особенно в телекоммуникациях, стимулировали развитие многих волоконно-оптических соединительных устройств. Примеры, показанные на рис. 23.11, соответствуют приведенным на рис. 23.10 для планарной интегральной оптики, а разветвители, показанные на рис. 23.11, *б*, описываются той же самой матрицей межсоединения (23.11).

23.1.3. Невзаимные оптические межсоединения

Оптические исполнения невзаимных межсоединений основаны, прежде всего, на использовании вращателя Фарадея. Как было продемонстрировано в подразд. 6.4.2, оптический вентиль можно создать с использованием 45-градусного фарадеевского вращателя, помещенного между двумя поляризаторами, ориентированными под углом 45° друг к другу. Линейно- поляризованный свет проходит в прямом направлении и блокируется в обратном.

В подразд. 6.4.2 было также показано, что сочетание 45-градусного фарадеевского вращателя с последующей полуволновой пластинкой дает полезное невзаимное устройство. Состояние поляризации линейно-поляризованного света, идущего в прямом направлении с плоскостью поляризации, ориентированной под углом 22,5° к быстрой оси пластинки, не меняется. Однако плоскость поляризации обратной волны поворачивается на 90° Следовательно, данное устройство можно использовать вместе с поляризующими светоделителями для получения невзаимных соединений, что иллюстрируется примером на рис. 23.12, *а*, а также **оптических циркуляторов**, как показано на рис. 23.12, *б*.

23.1.4. Оптические межсоединения в микроэлектронике

Возможность использования оптических межсоединений вместо обычных электрических в микроэлектронике и компьютерных системах привела к интенсивным исследованиям и развитию этой области в течение последних десятилетий. Успешное применение волоконной оптики для связи между компьютерами (например, в локальных оптических сетях, см. разд. 24.4) при-

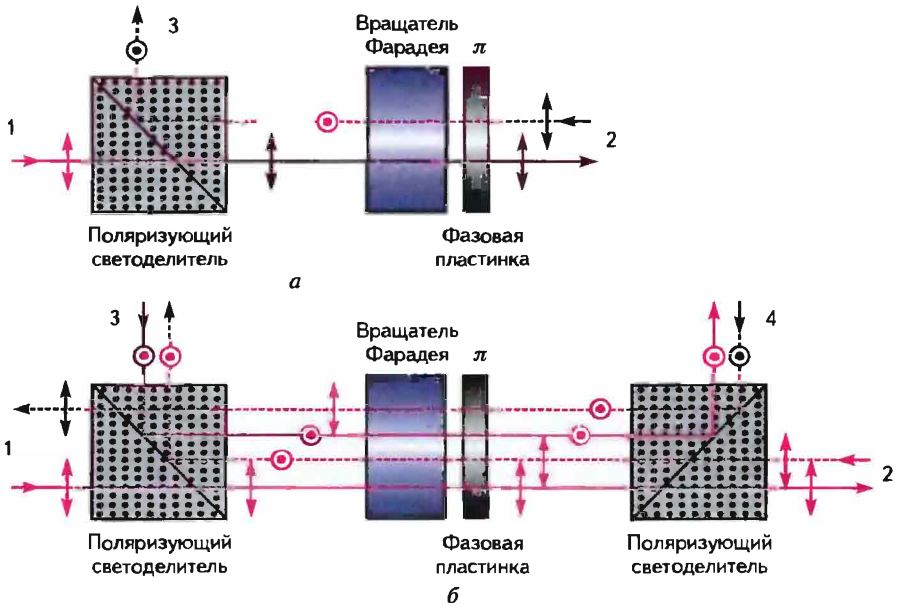


Рис. 23.12. Исполнение 3-портового невзаимного межсоединения, показанного на рис. 23.4, *б*, с использованием поляризующего светоделителя в сочетании с фарадеевским вращателем и полуволновой пластинкой (*а*). Свет проходит из порта 1 в порт 2 и из порта 2 в порт 3. Оптический циркулятор, реализующий четырехпортовое межсоединение 1-2-3-4-1, показанное на рис. 23.5, с применением комбинации двух поляризующих светоделителей, вращателя Фарадея и полуволновой пластинки (*б*)

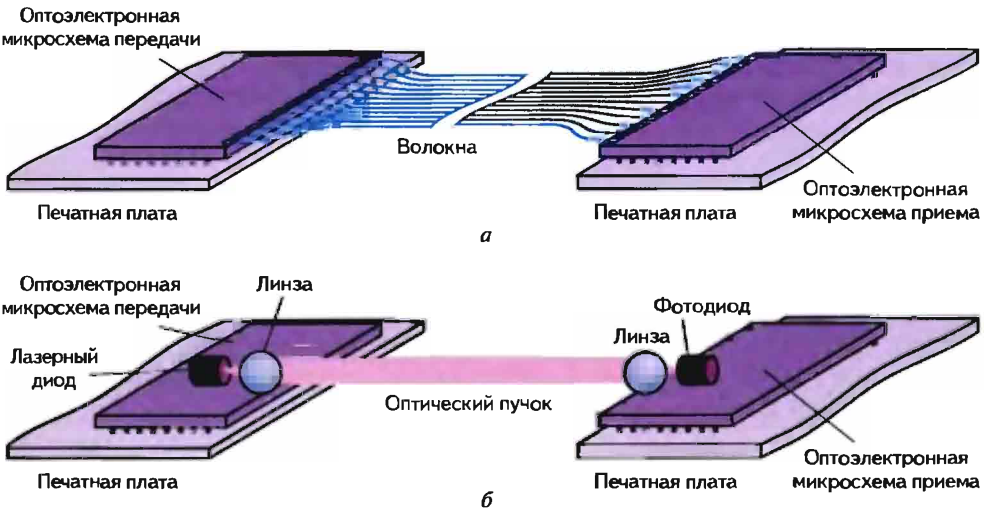


Рис. 23.13. Соединение между платами с использованием набора оптических волокон (*а*) микрооптической связи через свободное пространство (*б*)

вело к разработке систем, в которых волокно используется для связи между процессорами, соединительными и материнскими платами. Примеры межсоединений между платами с использованием оптического волокна показаны на рис. 23.13, *а*. Такие короткие волоконно-оптические соединения могут работать при скоростях передачи данных намного больше, чем у электрических, а технология их изготовления хорошо разработана, как описано в гл. 24. Были созданы также прототипы оптических соединений между платами через свободное пространство (см. пример на рис. 23.13, *б*). Каждая плата имеет приемно-передающий электронный чип, включающий источники света, например матрицу из поверхностно-излучающих лазеров с вертикальным резонатором (VCSEL) и матрицу фотоприемников, а также связанные с ними электрические схемы.

Разработаны также прототипы связи между микросхемами посредством оптических межсоединений. В примере, изображенном на рис. 23.14, каждая микросхема сопряжена с приемно-передающей оптоэлектронной микросхемой, а две оптоэлектронные микросхемы соединены планарным диэлектрическим волноводом. Рассматривались также оптические связи через свободное пространство при помощи отражающих зеркал.

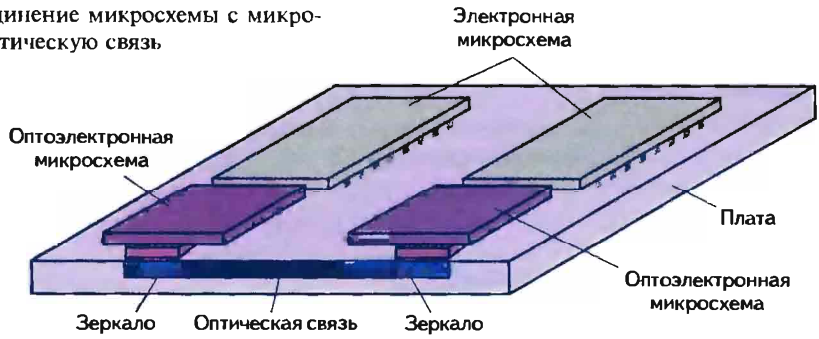


Рис. 23.14. Соединение микросхемы с микросхемой через оптическую связь

Использование ультракоротких оптических соединений внутри электронной микросхемы, очевидно, является намного более сложной задачей. Внутрисхемные оптические соединения привлекли интерес в связи с прогрессом в сознании высокоскоростных и высокоплотных схем и появлением архитектуры параллельных вычислений, что создало коммуникационные проблемы узких мест и сделало повышение пропускной способности межсоединений задачей первостепенной важности. В сверхбольших интегральных схемах (СБИС) межсоединения занимают большую часть доступной площади чипа. Для минимизации влияния задержки в межсоединениях, которая достигает или даже превышает времена переключения, прилагаются значительные усилия по выравниванию длин связей. Оптические межсоединения потенциально способны облегчить решение некоторых таких проблем, однако их коммерческая жизнеспособность пока не доказана.

Обоснование целесообразности оптических схемных соединений

Оптические соединения обладают рядом преимуществ для целей меж- и внутрисхемной связи, принципиально вытекающих из малости длины волны света и, соответственно, высокой частоты (например, 20—50 ТГц), которая существенно больше ширины полосы передаваемых сигналов. Электронные соединения используют видеосигналы на сравнительно много более низких частотах (например, в гигагерцовом режиме).

• *Плотность.* Наибольшая плотность свободных от интерференции межсоединений достигается на ненаправляемых пучках, каждый из которых имеет малую ширину и угол расходимости, ограниченный только дифракцией (произведение ширины и угла расходимости узкого пучка имеет порядок длины волны, которая мала на оптических частотах). Поскольку такие пучки могут пересекаться (проходить сквозь друг друга) без взаимного влияния (предполагается, что среда линейна), их можно использовать в трехмерной конфигурации для создания межсоединений с плотностями, недостижимыми с помощью электрических проводов. Свет можно также направлять планарными или квазипланарными *диэлектрическими* волноводами с низкими потерями и шириной, величина которой ограничена снизу длиной волны. Они могут быть плотно упакованы, при этом перекрестные помехи минимальны. Напротив, электрические соединения требуют металлических проводников, таких как полосковые линии, которые служат линиями передачи или волноводами для электромагнитных волн, связанных с осциллирующими электрическими зарядами. Металлические проводники вносят потери и не могут быть плотно упакованы, так как при близком расположении между ними возникает сильная электромагнитная связь.

• *Ширина полосы.* Ширина полосы электронной полосковой линии длины ℓ с площадью сечения A, размещенной на плоскости подложки, пропорциональна отношению A/ℓ^2. Это видно из следующих соображений: если полоса ограничена RC-эффектами, то сопротивление $R \propto \ell/A$, а емкость $C \propto \ell$, так что постоянная времени $RC \propto \ell^2/A$. Аналогичное рассуждение применимо к линии, полоса которой ограничена LC-эффектами. Следовательно, полоса частот определяется характеристическим отношением $\ell/\sqrt{A}$ и не может быть изменена путем миниатюризации устройства или увеличения его размеров. Оптические межсоединения не имеют такого ограничения, поскольку их полоса пропускания определяется другими физическими эффектами и в общем случае гораздо шире. Кроме того, в оптических межсоединениях максимальная полоса частот сигнала, передаваемого по каждому каналу, не зависит от плотности соседних каналов. Иначе говоря, увеличение скорости передачи данных не влияет на перекрестные помехи между соседними линиями. Это вытекает из малости отношения ширины полосы к несущей частоте модулированного света. Не так обстоит дело с электронными соединениями, у которых плотность электронных связывающих элементов требует резкого сокращения при высоких частотах модуляции, чтобы исключить емкостные и индуктивные связи между

соседними элементами. Следовательно, оптические межсоединения имеют большее произведение плотности на ширину полосы по сравнению с их электронными аналогами.

• *Задержка.* Фотоны в свободном пространстве распространяются со скоростью 0,3 мм/пс, а в кремнии — со скоростью ≈0,086 мм/пс. Соответствующая задержка из-за конечности времени распространения составляет ≈3,3 пс/мм и ≈9,4 пс/мм соответственно. Задержки электрических сигналов в полосковых линиях из керамик и полиимидов составляют примерно 10,2 и 6,8 пс/мм соответственно. Следовательно, дело не в задержке, как таковой. Однако в то время как скорость света не зависит от числа соединений в разветвляющемся оптическом соединительном устройстве, в электронных линиях передачи скорость обратно пропорциональна емкости на единицу длины, так что она зависит от полной емкостной «нагрузки». Следовательно, время задержки увеличивается с ростом числа каналов ветвления. Оптика дает большую гибкость при создании разветвителей и соединителей, где ограничение числа каналов связано только с доступной оптической мощностью.

• *Мощность.* Во избежание отражений электрические соединения должны заканчиваться с обеспечением согласования импедансов. Это обычно требует большего расхода мощности. В оптических соединениях отражение можно значительно уменьшить с помощью антиотражательных покрытий, так что требования к мощности ограничены чувствительностью фотоприемников и эффективностью преобразования электрического сигнала в оптический и обратно, а также коэффициентами пропускания проходимых светом элементов.

Конструктивная реализация схемных оптических межсоединений

Внутрисхемные оптические соединения — ультракороткие оптические связи, соединяющие точки внутри микросхемы. Каждая связь включает три компонента: электронно-оптический преобразователь (передатчик), модулируемый электрическим сигналом в точке внутри микросхемы, оптический пучок и оптоэлектронный преобразователь (приемник), подающий сигнал в другую точку той же микросхемы.

Оптические пучки могут направляться волноводами на основе кремния и могут маршрутизироваться внешним устройством, таким как голограмма, как показано на рис. 23.15, *а*. Особый случай конфигурации с внешней маршрутизации — однонаправленное межсоединение между одной или несколькими внешними точками и точками микросхемы. Такую систему легче изготовить, поскольку она не требует передатчиков, размещенных на чипе. Одно из полезных применений — оптическое распределение задающих импульсов. В этом случае сигнал от внешнего генератора задающих импульсов модулирует внешний источник света, который передает сигнал на множество фотоприемников на микросхеме с помощью отражательной голограммы, как показано на рис. 23.15, *б*. Таким образом, гарантируется точная синхронизация высокоскоростных синхронных схем и снимается проблема рассогласования синхронизирующих им-

пульсов из-за различия задержек в разных каналах. Конечно, голограмма может быть убрана, и свет может передаваться непосредственно на все точки микросхемы. Это дает устойчивую систему, нечувствительную к разъюстировке, но ее эффективность низка, так как большая часть оптической мощности пропадает.

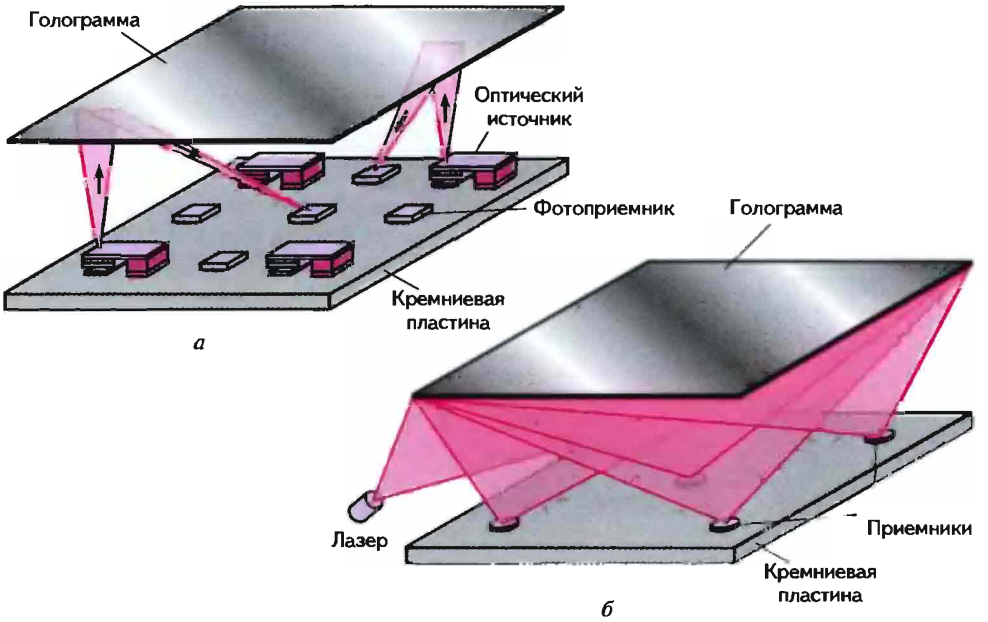


Рис. 23.15. Соединения между источниками и приемниками на микросхеме через внешнюю отражательную голограмму, используемую как маршрутизирующий элемент (*а*). Однонаправленное межсоединение, направляющее задающие импульсы от внешнего источника на фотоприемники на кремниевой пластине (*б*)

В идеале все три компонента оптической связи должны быть монолитно интегрированы с кремниевой подложкой микросхемы и совместимы с технологией КМОП (комплементарных металлооксидных полупроводников). Кремниевые фотодиоды легко встраиваются в кремниевые чипы, а оптические волноводы типа «кремний на диэлектрике» (КНД) (см. разд. 8.3) можно использовать в качестве оптических соединений, хотя полезная площадь на поверхности микросхемы может быть недостаточна для таких волноводов. Основная трудность связана с передатчиками, поскольку источники света невозможно эффективно выполнить из кремния, так как он является непрямозонным материалом (см. подразд. 16.1.4). Эффективные светоизлучающие материалы, такие как AlGaAs/GaAs, выращенные на кремниевой подложке с помощью гетероэпитаксии, не вполне надежны из-за несогласованности двух материалов по параметру решетки и тепловому расширению. Другие идеи прямой генерации света в кремнии с использованием фотонно-кристаллических структур остаются в области продолжающихся исследований.

Другой подход к решению проблемы передатчика — замена оптических источников электрооптическими модуляторами с освещением от внешнего источника света и модуляцией от местных электрических сигналов внутри микросхемы. Однако полностью кремниевые модуляторы остаются либо слишком большими, либо слишком медленными для применения в оптических межсоединениях.

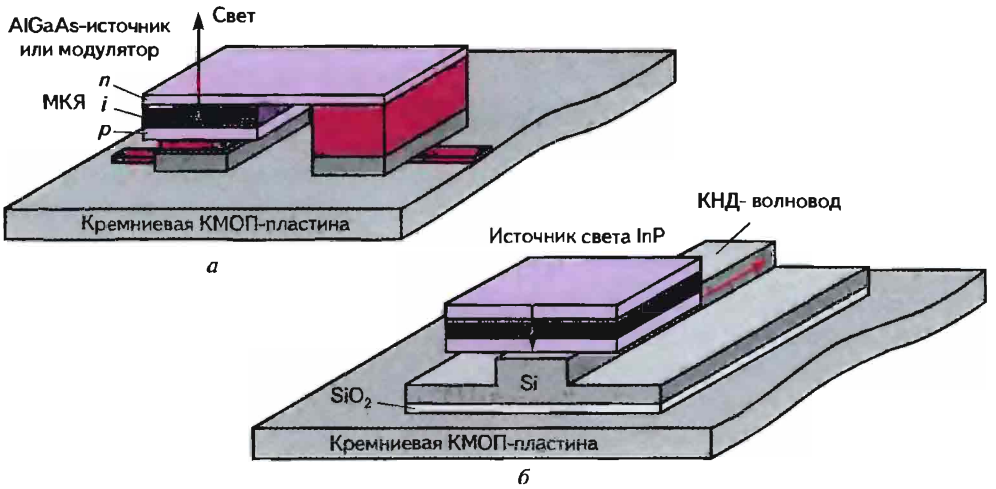


Рис. 23.16. Гибридная интеграция. Передатчики монтируются в оптоэлектронных схемах, интегрированных с кремниевой пластиной с использованием крепления методом перевернутого кристалла:

а — оптический источник на AlGaAs или модулятор, присоединенный к кремниевой пластине в поверхностно-нормальной архитектуре; *б* — источник света на InP, присоединенный к кремниевой пластине. Свет вводится в находящийся непосредственно на пластине гребневой волновод «кремний на диэлектрике» (КНД)

Один из практических подходов к преодолению проблем несогласованности между оптоэлектронной технологией сложных полупроводников, применяемой для изготовления оптических источников и модуляторов, и кремниевой КМОП-технологией, являющейся основой современной электроники, — это **гибридная интеграция**. Этот подход основан на соединении независимо изготовленных оптоэлектронных и электронных микросхем. Процесс гибридной интеграции, известный как **крепление методом перевернутого кристалла**, позволяет объединить тысячи оптоэлектронных устройств на единой кремниевой пластине с поперечным выстраиванием лучше 1 мкм. Пользуясь этой технологией упаковки, можно присоединить к кремниевой пластине источники света или оптические модуляторы в виде двумерных матриц поверхностно-нормальной архитектуры, как показано на рис. 23.16, *а*. Примером является матрица поверхностно-излучающих лазеров с вертикальным резонатором (VCSEL) на основе GaAs, работающих на длине волны 850 нм при скоростях передачи более 10 ГБ/с с низковольтным управлением (примерно 1 В). Такая же конфигурация может

быть использована в матрицах электрооптических модуляторов на основе электропоглощающих полупроводников (см. разд. 20.5). Эти матрицы источников или модуляторов света можно использовать в конфигурации с внешней маршрутизацией (см. рис. 23.15, *а*). На рис. 23.16, *б* показана другая конфигурация с внутрипроцессорной маршрутизацией с использованием КНД-волновода и источника света на InP, присоединенного к кремниевой пластине.

23.2. ПАССИВНЫЕ ОПТИЧЕСКИЕ МАРШРУТИЗАТОРЫ

Маршрутизация пучков света может производиться на основе физических свойств, таких как длина волны, интенсивность, фаза, поляризация или время. Как показывает пример, приведенный на рис. 23.17, *а*, компонента с признаком X_ℓ в каждом пучке направляется в ℓ-й выходной порт, где $\ell = 1, 2, ..., N$; N — количество выходных портов, равное количеству признаков. Далее следует несколько примеров.

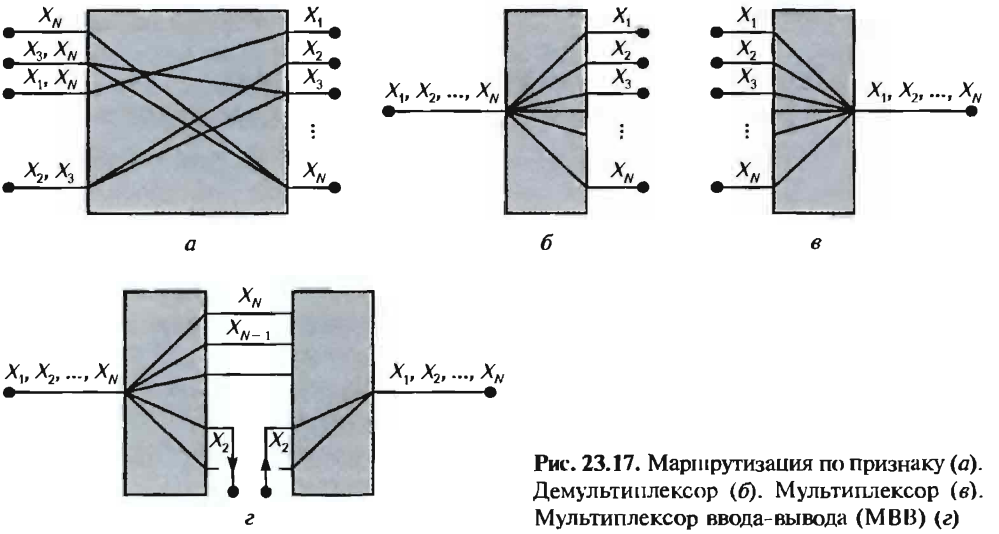


Рис. 23.17. Маршрутизация по признаку (*а*). Демультиплексор (*б*). Мультиплексор (*в*). Мультиплексор ввода-вывода (МВВ) (*г*)

• **Демультиплексор** — $1 \times N$-маршрутизатор, сортирующий компоненты одного входного пучка по признакам $X_1, X_2, ..., X_N$ и направляющий их в отдельные выходные порты, как показано на рис. 23.17, *б*. Демультиплексор можно реализовать с помощью операции «передача—селекция»: $1 \times N$-разветвитель передает копии входного пучка на все выходные порты через фильтры, пропускающие только нужную компоненту и блокирующие остальные.

• **Мультиплексор** — устройство, обратное демультиплексору. Как показано на рис. 23.17, *в*, входные пучки с различными оптическими признаками $X_1, X_2, ..., X_N$ сливаются в один пучок, который впоследствии может быть снова разложен на компоненты с помощью демультиплексора. Мультиплексирование и демуль-

типлексирование по частоте, длине волны и времени широко используются в оптических системах связи.

• **Оптический мультиплексор ввода/вывода (OMBB)**, показанный на рис. 23.17, *г* — еще одно важное устройство маршрутизации, используемое в сетях связи. Здесь демультиплексор сортирует компоненты, выделяет одну из них по избранному признаку, например X_2, сбрасывает переносимые этой компонентой данные и вместо них добавляет новые, затем снова сливает все компоненты в один пучок с помощью мультиплексора.

23.2.1. Маршрутизаторы с разделением по длине волны

Маршрутизаторы с разделением по длине волны обычно используются в оптоволоконных сетях со спектральным уплотнением. Как описано в подразд. 24.3.3, в таких системах используется несколько каналов связи на разных длинах волн в одном оптическом волокне. В них применяются маршрутизаторы, которые объединяют каналы на входе в волокно и разделяют их на выходе, называемые **мультиплексорами** и **демультиплексорами с разделением по длине волны** соответственно.

Техническое исполнение мультиплексоров и демультиплексоров с разделением по длине волны

Для демультиплексирования с разделением по длине волны применяются следующие схемы, показанные на рис. 23.18 (см. цв. вклейку).

• Оптический элемент, обладающий угловой дисперсией, разделяет компоненты одного оптического пучка, имеющие различную длину волны, по направлению, т. е. для каждой длины волны получается отдельный пучок. Простейшими оптическими элементами, обеспечивающими угловую дисперсию, являются призма (рис. 23.18, *а*) и дифракционная решетка (рис. 23.18, *в*). Угловая дисперсия призмы ограничена скоростью изменения показателя преломления с длиной волны $dn/d\lambda$, которая обычно недостаточно велика, чтобы должным образом разделить компоненты со слегка отличающимися длинами волн. Призмы из фотонно-кристаллических материалов (см. гл. 7), называемые **суперпризмами**, могут иметь диспергирующую способность на два-три порядка выше. Дифракционные решетки (подразд. 2.4.2) имеют более высокую диспергирующую способность, чем обычные призмы. С их помощью можно разрешить компоненты, разность частот которых составляет несколько гигагерц.

• Разделение по длине волны можно осуществить также с помощью набора фильтров, настроенных на различные длины волн. Входящий свет передается на различные фильтры, каждый из которых пропускает только канал с одной длиной волны и блокирует остальные. В другом варианте пучок может направляться через последовательность фильтров с малой спектральной шириной, каждый из которых пропускает одну длину волны и отражает все другие на следующий фильтр, как показано на рис. 23.18, *б*. Для направления лучей между фильтрами используется стержневая градиентная линза (СГЛ).

• В аналогичном исполнении для разделения компонент с разными длинами волн используется зависимость от длины волны коэффициента отражения **волоконной решетки Брэгга** (ВРБ) (подразд. 7.1.3). Компонента с брэгговской длиной волны ($\lambda_B = \Lambda/2$, где Λ — период решетки) отражается, а все остальные компоненты проходят. Для разделения нескольких длин волн используется соответствующий набор брэгговских решеток (рис. 23.18, *г*).

• Еще один вариант — последовательность кольцевых микрорезонаторов, каждый из которых настроен на свою длину волны (рис. 23.18, *д*).

• В других конструкциях используются интерферометры, такие как **интерферометр Маха—Цендера**, а также **маршрутизаторы с волноводными решетками**, которые будут описаны далее.

Оптический мультиплексор ввода-вывода (OMBB)

Оптический мультиплексор ввода-вывода (OMBB) извлекает данные из каналов, выбранных по длине волны, и заменяет их новыми данными в оптическом пучке многоканальной связи. Доступ к отдельному каналу может быть осуществлен с помощью демультиплексора, за которым следует мультиплексор, как на схеме, приведенной на рис. 23.17. Данные выделяются (убираются) из выбранных каналов с использованием приемников, а новые данные добавляются с помощью модулированных источников света. В другой схеме выбранные по длине волны каналы отделяются от других каналов с помощью чувствительного к длине волны оптического элемента. Примеры OMBB такого типа с применением волоконной решетки Брэгга (ВРБ) и набора кольцевых микрорезонаторов показаны на рис. 23.19 и 23.20 (см. цв. вклейку) соответственно.

Интерферометр Маха—Цендера как демультиплексор

Так как интерферометры чувствительны к длине волны, они подходят для осуществления маршрутизации с разделением по длинам волн. Например, интерферометр Маха—Цендера (ИМЦ), показанный на рис. 23.21 (см. цв. вклейку), можно использовать как двухволновый демультиплексор. Для направления компонент с длинами волн λ_1 и λ_2 на разные выходные порты разность хода Δd выбирается так, чтобы разность фаз $\phi = 2\pi d/\lambda$ была целым четным числом p для λ_1 и целым нечетным числом p для λ_2, т. е. $\Delta d = q_1\lambda_1/2$ и $\Delta d = q_2\lambda_2/2$, где q_1 — четное целое; q_2 — нечетное целое число.

Разрешение маршрутизатора, т. е. ближайшие длины волн, которые можно разделить, можно определить из соотношения

$$\frac{1}{\lambda_1} - \frac{1}{\lambda_2} = \frac{q_1 - q_2}{2\Delta d},$$

полагая в нем $|q_1 - q_2| = 1$, так что

$$\left|\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right| = \frac{1}{2\Delta d}.$$

Соответствующая разность частот $\Delta\nu = |\nu_1 - \nu_2|$ тогда равна

$$\Delta\nu = \frac{c}{2\Delta d}. \tag{23.12}$$

Например, если $\Delta d = 1$ мм и $n = 1{,}5$, то $\Delta\nu = 100$ ГГц. Меньшие разности $\Delta\nu$ требуют пропорционально больших разностей хода Δd.

Спектральная чувствительность маршрутизатора на ИМЦ может быть найдена с помощью матрицы межсоединения

$$\mathbf{T} = \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \begin{bmatrix} \exp\left(-j2\pi\dfrac{d_0 + \Delta d}{\lambda}\right) & 0 \\ 0 & \exp\left(\dfrac{-j2\pi d_0}{\lambda}\right) \end{bmatrix} \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}, \tag{23.13}$$

где d_0 и $d_0 + \Delta d$ — длины пути в плечах интерферометра, первая и третья матрицы в данном произведении представляют 3-дБ-делители. Для входного поля единичной мощности на входном порте 2 мощность, приходящая на порты 1 и 2, составляет $P_1 = |T_{21}|^2$ и $P_2 = |T_{22}|^2$ соответственно, так что

$$P_1 = \cos^2\left(\frac{\pi\Delta d}{\lambda}\right); \quad P_1 = \sin^2\left(\frac{\pi\Delta d}{\lambda}\right). \tag{23.14}$$

Графики зависимости этих мощностей от λ показаны на рис. 23.21 (см. цв. вклейку). Из них видно, что чем меньше отношение $\lambda/\Delta d$, тем более быстро меняются эти функции между 0 и 1, т. е. тем больше возможность разделения компонент с близко расположенными длинами волн.

Несколько ИМЦ можно расположить каскадом для разделения более чем двух длин волн. Например, четыре длины волны можно разделить в результате двухступенчатого процесса, как показано на рис. 23.22 (см. цв. вклейку). Первый ИМЦ разделяет длины волн с четными и нечетными номерами, а последующие ИМЦ осуществляют более тонкое разделение.

Маршрутизаторы на решетках волноводов

Для обеспечения более высокой избирательности по длине волны можно использовать другие интерферометрические схемы. Например, многолучевые интерферометры обладают высокой селективностью по длине волны, так как для них характерны очень узкие резонансы. Такие интерферометры можно создавать в планарных волноводах. Несколько интерферометров могут осуществлять многоволновую маршрутизацию в приборах с множеством входных и выходных портов. Принцип состоит в конфигурировании *каждого* соединения между входным и выходным портами как независимого многолучевого интерферометра, который пропускает только избранные длины волн. Поскольку многолучевой интерферометр аналогичен спектрометру с дифракционной

решеткой, такие маршрутизаторы называются **маршрутизаторами на решетках волноводов (waveguide grating routers, WGR)**. Эти приборы называются также **массивами планарных волноводов (arrayed waveguides, AWG)**.

Многолучевой интерферометр

Прежде чем рассматривать работу WGR, дадим обзор свойств многолучевого интерферометра (см. подразд. 2.5.2). L-лучевой интерферометр — соединение с L оптическими путями, длина которых постепенно увеличивается по линейному закону, так что соседние пути имеют одинаковую разность длин Δd. Волна, принимаемая на выходе, представляет собой сумму L волн с одинаковыми амплитудами и равноотстоящими на $\phi = 2\pi\Delta d/\lambda$ фазами при длине волны λ. Коэффициент пропускания по мощности

$$T = \frac{\sin^2(L\varphi/2)}{\sin^2(\varphi/2)} = \frac{\sin^2(L\pi\Delta d/\lambda)}{\sin^2(\pi\Delta d/\lambda)} \qquad (23.15)$$

является периодической функцией ϕ с острыми пиками при значениях ϕ, равных целому кратному 2π (см. рис. 2.28). Зависимость T от λ не является периодической, но содержит острые пики при $\lambda = \Delta d$ и целых частях этой величины, как видно на рис. 23.23. Чем больше число путей L, тем острее пики.

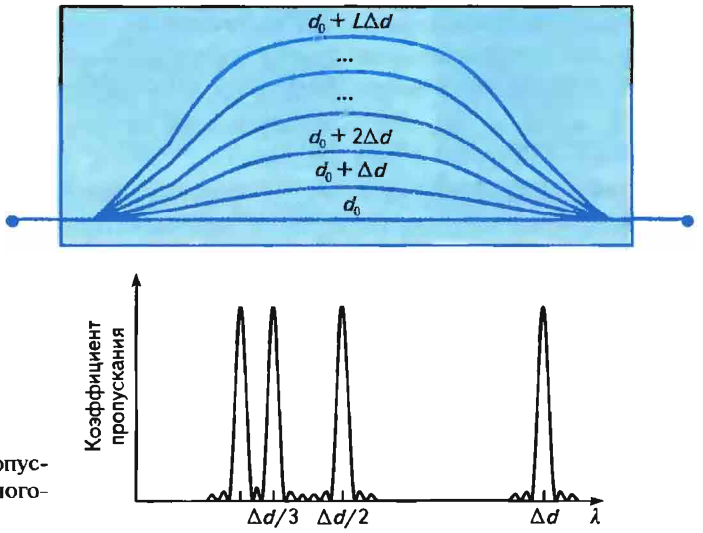


Рис. 23.23. Зависимость пропускания от длины волны в многолучевом интерферометре

WGR как демультинлексор с разделением по длине волны

Маршрутизатор на решетке волноводов (WGR) можно использовать как $1 \times N$-маршрутизатор с разделением по длине волны, который направляет каждую из N компонент с длинами волн λ_1, λ_2, ..., λ_N, приходящих на входной порт, на один из N выходных портов, как показано на рис. 23.24 (см. цв. вклейку).

Имеется N многолучевых интерферометров, по одному на каждый выходной порт. Каждый интерферометр имеет уникальную разность хода Δd, выбранную так, чтобы пропускалась только одна длина волны. Это достигается, если соединения, ведущие к выходному порту m, имеют разность хода Δd_m, равную целому числу λ_m, но не являющуюся целым числом других длин волн.

Конструкция проще, если длины волн $\lambda_1, \lambda_2, ..., \lambda_N$ распределены равномерно в виде убывающей последовательности $\lambda_m = \lambda_0 - m\Delta\lambda$, где $\Delta\lambda$ — расстояние между каналами по длине волны; $\lambda_0 = \lambda_1 - \Delta\lambda$. Необходимое условие работы демультиплексора имеет вид

$$\Delta d_m = \lambda_m = \lambda_0 - m\Delta\lambda, \quad m = 1, 2, \quad , N, \tag{23.16}$$

т. е. разность хода для соединений с m-м выходным портом убывает линейно с ростом m. Другое условие состоит в том, чтобы Δd_m не равнялось целому числу λ_ℓ для всех $\ell \neq m$. Это условие автоматически выполняется, если наименьшая длина волны λ_N превышает половину наибольшей длины волны λ_1, как показано на рис. 23.24 (см. цв. вклейку).

В схеме, показанной на рис. 23.24, длина каждого пути между входным и выходным портом складывается из длины волновода и расстояний, проходимых в звездообразных разветвителях. Длины волноводов можно выбрать постепенно увеличивающимися на постоянную величину Δd_w. Для звездообразного разветвителя с круговыми границами разность хода можно аппроксимировать линейно убывающей функцией m, так что

$$\Delta d_m = \Delta d_w + (\Delta d_a - m\Delta d_b), \tag{23.17}$$

где Δd_a и Δd_b — постоянные, зависящие от геометрии разветвителей. Тогда условие (23.16) может быть выполнено, если $\Delta d_w + \Delta d_a = \lambda_0$ и $\Delta d_b = \Delta\lambda$. Следовательно, разрешение демультиплексора по длине волны, т. е. минимальная разность длин волн $\Delta\lambda$, ограничено минимальным значением геометрического фактора Δd_b.

WGR как маршрутизатор $N \times N$

WGR можно использовать также в качестве более общего маршрутизатора $N \times N$ с разделением по длине волны. Соединения между ℓ-м входным и m-м выходным портами образуют многолучевой интерферометр с разностью хода $\Delta d_{\ell m} = \lambda_{00} - (\ell + m)\Delta\lambda$, которая убывает линейно и по ℓ, и по m (λ_{00} и $\Delta\lambda$ — постоянные, зависящие от геометрии WGR). Свет проходит между этими портами, если длина волны $\lambda_{\ell m}$ равна $\Delta d_{\ell m}$, т. е.

$$\boxed{\lambda_{\ell m} = \lambda_{00} - (\ell + m)\Delta\lambda, \quad \ell, m = 1, 2, ..., N.} \tag{23.18}$$

Уравнение WGR

Уравнение (23.18) является обобщением (23.17). Хотя WGR не обеспечивает маршрутизацию произвольных длин волн, он позволяет решить некоторые задачи маршрутизации, такие как одновременное выполнение операций мультиплексирования с разделением по длине волны.

23.2.2. Маршрутизаторы с разделением по поляризации, фазе и интенсивности

Маршрутизация с разделением по поляризации

Простейший пример пассивной оптической маршрутизации основан на поляризации. При демультиплексировании с разделением по поляризации компоненты с параллельной и перпендикулярной поляризацией разделяются при помощи поляризующего светоделителя (ПСД), как показано на рис. 23.25. Мультиплексирование осуществляется с использованием ПСД для совмещения пучков (свет идет справа налево, а не слева направо).

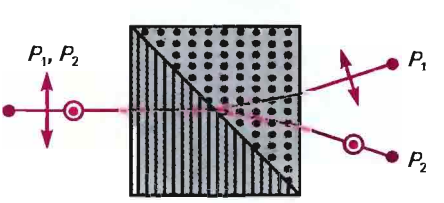


Рис. 23.25. Маршрутизация с разделением по поляризации при помощи поляризующего светоделителя (ПСД). Для пучков, идущих слева направо, призма является демультиплексором. Для пучков, идущих справа налево, это мультиплексор

Маршрутизация с разделением по фазе

Другой простой пример пассивной оптической маршрутизации основан на использовании фазы. Здесь последовательность оптических импульсов с фазами от 0 до π нужно рассортировать по фазе и направить в два выходных порта. Это можно осуществить с помощью простого интерферометра, как показано на рис. 23.26.

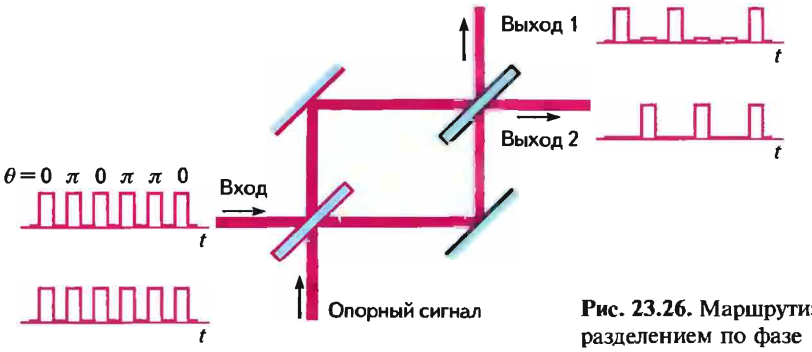


Рис. 23.26. Маршрутизация с разделением по фазе

Маршрутизация с разделением по интенсивности

Пучок света с изменяющейся во времени интенсивностью можно разделять на отдельные пучки по значению интенсивности. Например, пучок света, несущий последовательность импульсов с двумя интенсивностями, как показано на рис. 23.27, разделяется на два пучка: один с импульсами высокой

интенсивности, а другой — низкой. Такая операция демультиплексирования требует использования нелинейных оптических элементов. Часто для этого изменения интенсивности преобразуются в изменения фазы при помощи ячейки Керра (см. подразд. 21.3.1), как описано далее.

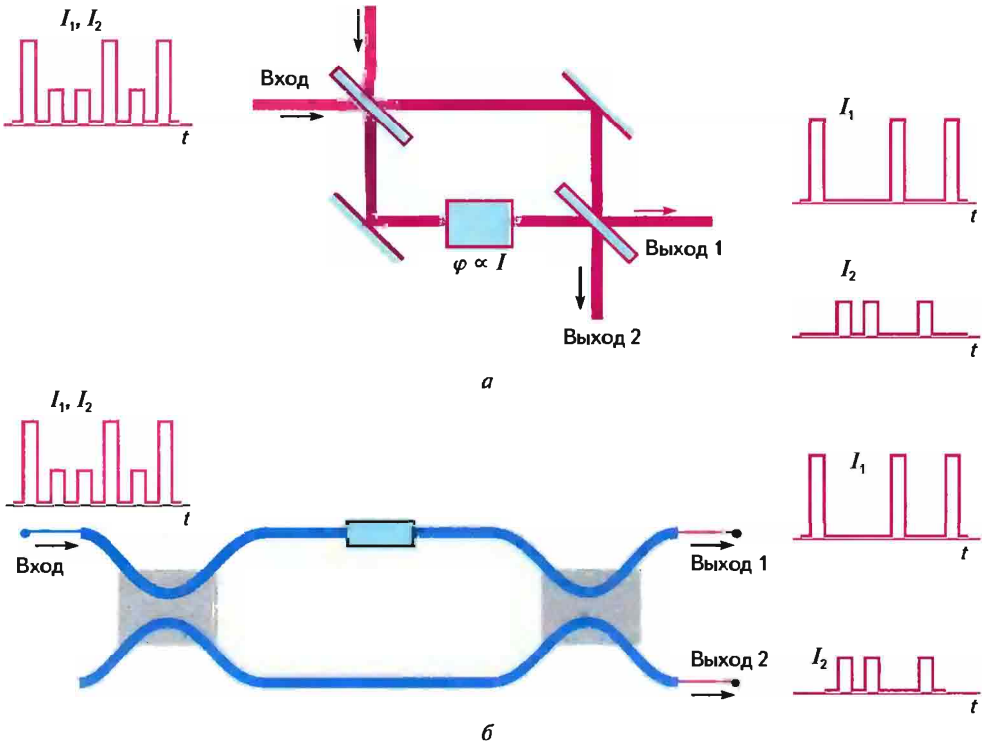


Рис. 23.27. 1 × 2-маршрутизатор с разделением по интенсивности на основе интерферометра Маха—Цендера с нелинейной ячейкой Керра в объемном (*а*) и волоконно-оптическом исполнении (*б*)

Нелинейный интерферометр Маха—Цендера (ИМЦ)

Нелинейный ИМЦ — это обычный ИМЦ с нелинейным оптическим элементом, таким как ячейка Керра, помещенным в одно из плеч интерферометра. Ячейка вносит фазовый сдвиг, пропорциональный интенсивности света. Система настроена так, что разность фаз между плечами интерферометра равна целому нечетному числу π для одной интенсивности и целому четному — для другой. Это разделяет поток импульсов на два порта, один из которых получает импульсы высокой, а другой — низкой интенсивности, как показано на рис. 23.27, *а*. Интерферометр может иметь волоконно-оптическое исполнение, как изображено на рис. 23.27, *б*.

Нелинейный асимметричный интерферометр Саньяка

1 × 2-маршрутизатор с разделением по интенсивности на основе нелинейного волоконного интерферометра Саньяка показан на рис. 23.28. В этом интерферометре свет входит через волокно 1 и разделяется на волны, распространяющиеся по часовой стрелке и против нее. Если оптические длины пути для этих волн одинаковы, происходит конструктивная интерференция и свет распространяется обратно в волокно 1 и направляется в выходной порт 1, так что устройство действует как зеркало. Так будет, если волокно линейное, либо волокно нелинейное, но интенсивности обеих волн одинаковы. Однако, если светоделитель, питающий петлю интерферометра, несимметричен, то интенсивности встречных волн неодинаковы, и за счет нелинейного эффекта Керра они в общем случае приобретают различный набег фазы. Если разность фаз равна π, то возникает деструктивная интерференция и свет направляется в волокно 2 и выходной порт 2. Поскольку разность фаз пропорциональна интенсивности падающей волны, система действует как самоуправляемый 1 × 2-маршрутизатор с разделением по интенсивности (демультиплексор).

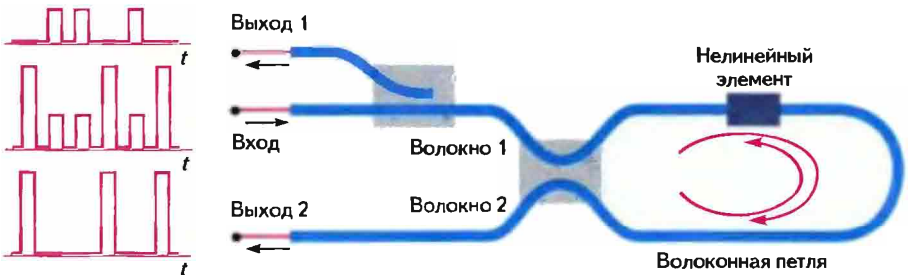


Рис. 23.28. 1 × 2 — маршрутизатор с разделением по интенсивности с использованием нелинейного интерферометра Саньяка, служащего нелинейным оптическим кольцевым зеркалом (НОКЗ)

Асимметрия между волнами, распространяющимися по часовой стрелке и против нее в интерферометре Саньяка, может вноситься также за счет асимметричного включения в петлю волоконного усилителя, легированного эрбием. Он усиливает одну из интерферирующих волн в течение первой половины ее прохождения по петле, так что при высокой интенсивности она проходит больше половины петли. Другая волна испытывает меньший нелинейный фазовый сдвиг. Система называется **нелинейным оптическим кольцевым зеркалом** (НОКЗ).

Нелинейный направленный разветвитель (ННР)

Волновод или волоконно-оптический направленный разветвитель из материала с керровской нелинейностью также может служить маршрутизатором с разделением по интенсивности, как показано на рис. 23.29. Если интенсивность входного импульса мала, то среда линейна и свет периодически переходит из одного волновода в другой по мере распространения (см. рис. 8.24). Если длина

разветвителя равна длине передачи L_0, свет полностью передается из входного волновода в другой волновод. Для импульсов высокой интенсивности постоянные распространения меняются из-за эффекта Керра, что создает зависящую от интенсивности расстройку синхронизма, которая меняется с расстоянием. Тогда распространение света подчиняется нелинейным связанным уравнениям

$$\frac{da_1}{dz} = -jC\exp(j\Delta\beta z)a_2(z) - j\gamma|a_1|^2 a_1; \tag{23.19}$$

$$\frac{da_2}{dz} = -jC\exp(-j\Delta\beta z)a_1(z) - j\gamma|a_2|^2 a_2, \tag{23.20}$$

которые являются обобщением линейных связанных уравнений (8.50) для линейного направленного разветвителя. Здесь $C = \pi/2L_0$ — коэффициент связи; γ пропорционально оптическому коэффициенту Керра n_2 [см. подразд. 21.3.1 и формулу (22.141)]. Система может быть устроена так, чтобы импульсы высокой интенсивности выходили из разветвителя по тому же волноводу, т. е. отделялись от импульсов низкой интенсивности.

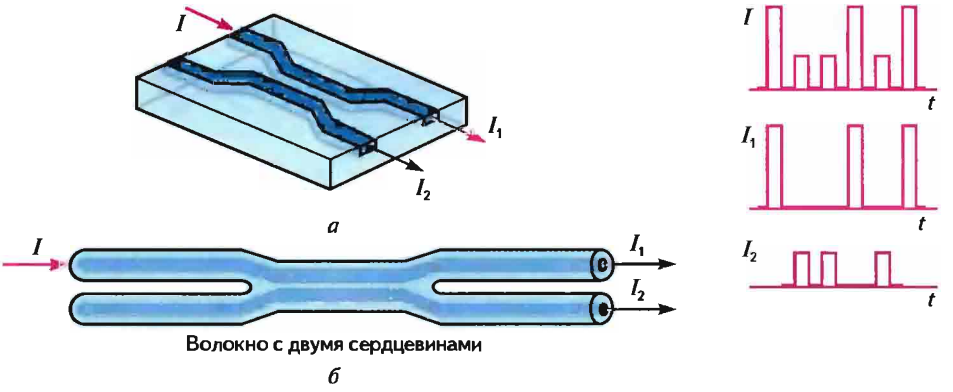


Рис. 23.29. 1 × 2-маршрутизатор с разделением по интенсивности с использованием направленного разветвителя, изготовленного из нелинейного материала. Устройство может быть реализовано с помощью интегрально-оптической (*а*) и волоконно-оптической технологии (*б*)

Солитонный направленный разветвитель

Маршрутизатор на ННР страдает недостатком, заключающимся в распаде импульса. Поскольку интенсивность оптического импульса меняется с течением времени, меняется и показатель преломления, а значит и соответствующая постоянная распространения в нелинейной среде. Различные части импульса, следовательно, кроссируют между двумя волокнами, что приводит к изменению формы и, возможно, распаду импульса. Этого не происходит в волоконно-оптическом ННР (см. рис. 23.29, *б*), если импульс является оптическим соли-

тоном (см. подразд. 22.5.2). Поскольку нелинейный сдвиг фазы оптического солитона постоянен по всей его огибающей, солитонный импульс остается неискаженным при передаче между связанными волокнами. Другим достоинством использования ННР в солитонном режиме является то, что переход между выходными портами намного резче зависит от мощности входного импульса.

23.3. ФОТОННЫЕ КОММУТАТОРЫ

23.3.1. Архитектуры пространственных коммутаторов

Коммутатор — устройство, устанавливающее и разъединяющее связи в каналах передачи коммуникационной системы или системы обработки данных. Управляющий блок обрабатывает команды соединения и посылает управляющий сигнал, заставляющий коммутатор срабатывать нужным образом. Если межсоединения всегда воздействуют на одинаковые приходящие сигналы одинаковым образом, то коммутаторы являются управляемыми, активными или реконфигурируемыми межсоединениями, которые модифицируются внешними командами. Примеры коммутаторов показаны на рис. 23.30.

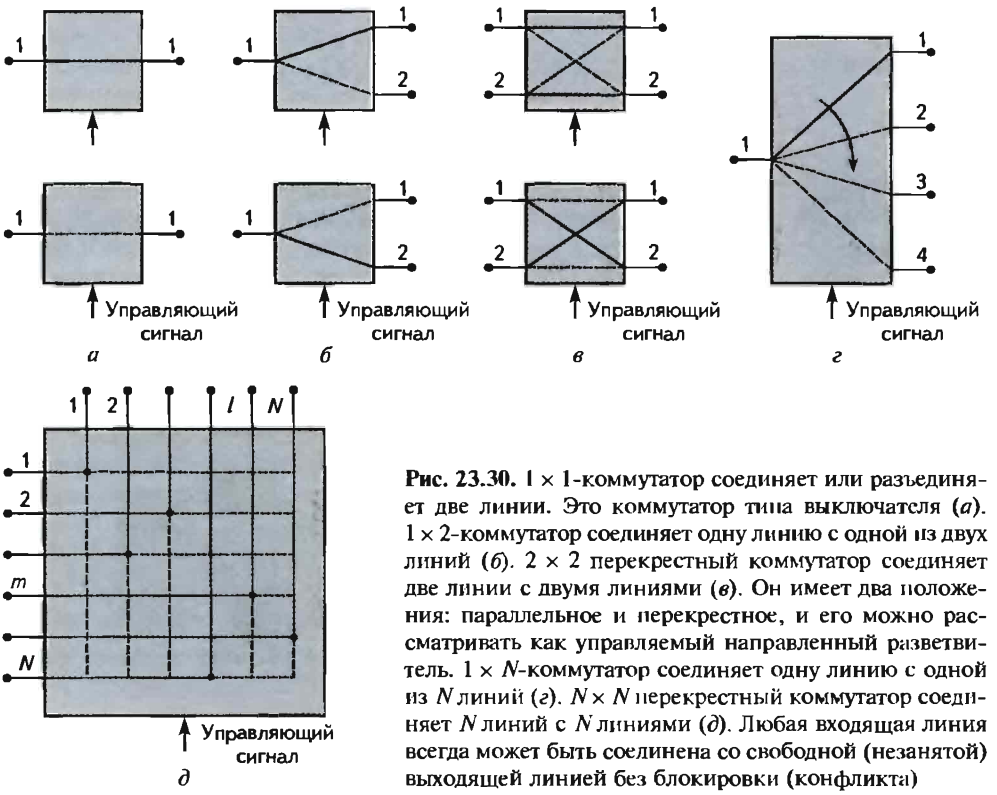
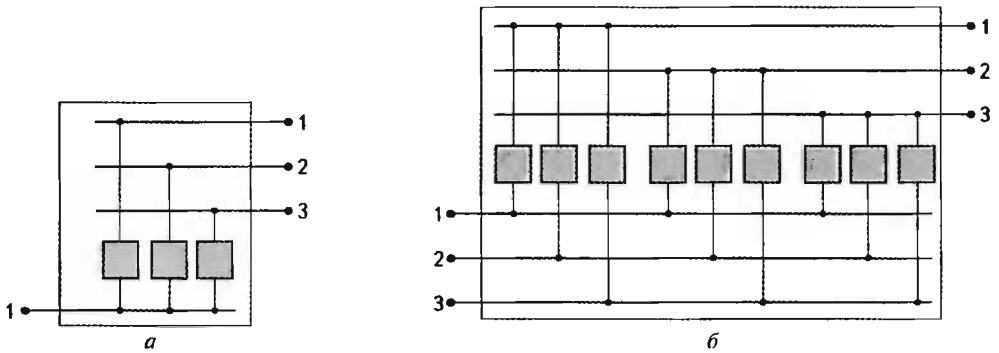


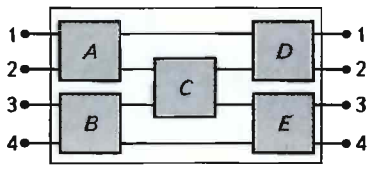
Рис. 23.30. 1 × 1-коммутатор соединяет или разъединяет две линии. Это коммутатор типа выключателя (*а*). 1 × 2-коммутатор соединяет одну линию с одной из двух линий (*б*). 2 × 2 перекрестный коммутатор соединяет две линии с двумя линиями (*в*). Он имеет два положения: параллельное и перекрестное, и его можно рассматривать как управляемый направленный разветвитель. 1 × *N*-коммутатор соединяет одну линию с одной из *N* линий (*г*). *N* × *N* перекрестный коммутатор соединяет *N* линий с *N* линиями (*д*). Любая входящая линия всегда может быть соединена со свободной (незанятой) выходящей линией без блокировки (конфликта)

1 × 1-коммутатор можно использовать как элементарное звено, из которого можно компоновать более сложные коммутаторы. Например $N \times N$ матричный перекрестный коммутатор можно построить с использованием массива из N^2 1 × 1-коммутаторов, организованных в матрицу $N \times N$, осуществляющую соединение или разъединение каждой из N входящих линий со свободной выходящей линией (рис. 23.30, *д*). Вход с номером m достигает всех элементарных переключателей m-й строки, а выход номер l соединен со всеми выходами элементарных переключателей l-го столбца. Соединение входа m с выходом l достигается переводом в положение «включено» элементарного 1 × 1-коммутатора (m, l). Примеры показаны на рис. 23.31.



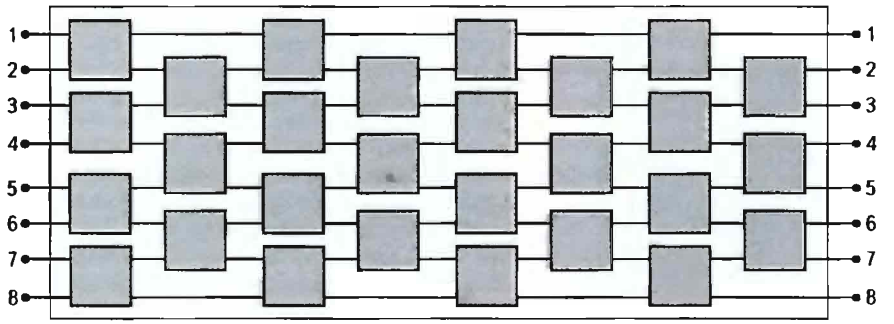
а *б*

Рис. 23.31. 1 × 3-коммутатор из трех 1 × 1-коммутаторов (*а*). 3 × 3-коммутатор из девяти 1 × 1-коммутаторов в конфигурации рассылка—выбор (**broadcast-and-select**) (*б*)



а

Рис. 23.32. 4 × 4-коммутатор из пяти 2 × 2-коммутаторов. Например, входящая линия 1 связывается с выходящей линией 3, если коммутаторы A и C находятся в перекрестном состоянии, а коммутатор E — в параллельном состоянии (*а*). 8 × 8-коммутатор из 28 2 × 2-коммутаторов (*б*)



б

$N \times N$-коммутатор можно также построить из 2 × 2-коммутаторов. Примеры показаны на рис. 23.32.

Характеристики коммутаторов

Коммутатор характеризуется следующими параметрами.

- *Размерность* (число входящих и выходящих линий) и *направление*, т. е. данные могут передаваться в одном или двух направлениях.
- *Время переключения* (время, необходимое для изменения конфигурации коммутатора).
- *Время задержки на распространение* (время, за которое сигнал проходит через коммутатор).
- *Пропускная способность* (максимальная скорость передачи данных через коммутатор).
- *Энергия переключения* (энергия, необходимая для включения и выключения коммутатора).
- *Диссипация мощности* (энергия, рассеиваемая в секунду в процессе переключения).
- *Вносимые потери* (спад мощности сигнала, вносимый соединением).
- *Перекрестные помехи* (нежелательная утечка мощности в другие линии).
- *Вероятность блокирования.* Вероятность того, что соединение не сможет быть установлено из-за конфликта с другим соединением.
- *Физические размеры.* Это важно при построении больших массивов коммутаторов.

23.3.2. Конструкции оптических пространственных коммутаторов

Оптоэлектронные коммутаторы

Электронные переключатели постоянно совершенствовались с начала появления телефонии, в целом следуя общему прогрессу микроэлектроники. В настоящее время наномасштабные основанные на КМОП (комплементарных металлооксидных полупроводниках) электронные затворы могут работать с временами переключения менее 0,1 нс и энергиями переключения менее 1 фДж. Усовершенствованные затворы на полевых транзисторах со структурой металл—оксид—полупроводник (МОП) могут переключаться за субпикосекундные времена. Легко доступны электронные микросхемы для перекрестного переключения с большим числом портов (например, 128 × 128). Поэтому естественно использовать эти устройства для оптической коммутации. Однако это требует преобразования оптического сигнала в электрический на входе коммутатора и обратного преобразования электрического сигнала в оптический на выходе, что схематически показано на рис. 23.33.

Поскольку оптоэлектронно-оптическое преобразование, необходимое для работы оптоэлектронных коммутаторов, вносит ненужную временную задержку и потери мощности, желательно разработать «прозрачные» фотонные коммутаторы, работающие непосредственно с оптическими сигналами с использованием механических, электрических, акустических, магнитных или тепловых эффектов, как описано далее в настоящем разделе.

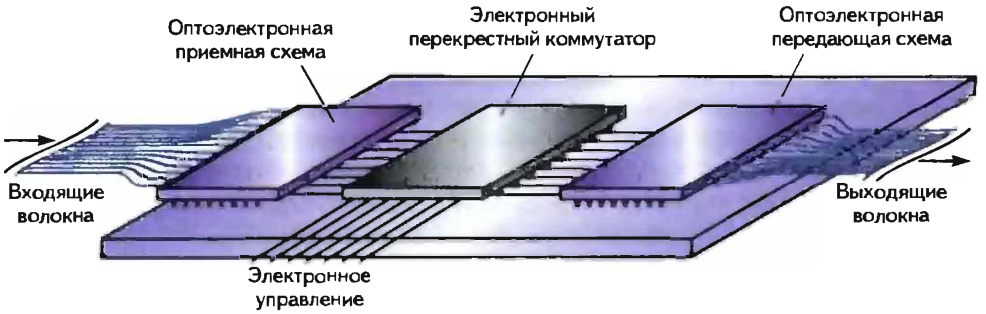


Рис. 23.33. Оптоэлектронный перекрестный коммутатор. Входящие по оптическим волокнам оптические сигналы детектируются массивом фотоприемников на оптоэлектронной схеме, переключаются электронным перекрестным коммутатором и восстанавливаются массивом источников света (например, поверхностно-излучающих лазеров с вертикальным резонатором, VCSEL), питающих выходящие оптические волокна

Некоторые основные конфигурации оптических коммутаторов

Наиболее элементарными оптическими коммутаторами являются оптический сканер и модулятор. Сканер, отклоняющий оптический пучок в N возможных направлениях, является $1 \times N$-коммутатором (рис. 23.34, *а*). Оптический модулятор, работающий в режиме включения—выключения, служит 1×1-коммутатором. Модуляция может быть *прямой*, опирающейся на некоторый физический эффект, благодаря которому свет проходит или блокируется, или *интерферометрической*, с использованием, например, оптического фазового модулятора, помещенного в одно из плеч интерферометра, который преобразует фазовую модуляцию в модуляцию интенсивности (рис. 23.34, *б*). Другим элементарным оптическим коммутатором является направленный разветвитель, действующий как 2×2-коммутатор. Такой коммутатор можно реализовать с использованием интерферометра с фазовым модулятором в одном или обоих плечах (рис. 23.34, *в*). Два плеча интерферометра могут также представлять две ортогональные поляризационные компоненты, и тогда фазовым модулятором является фазовая пластинка, которая вносит относительный фазовый сдвиг между двумя поляризациями.

Элементарные оптические коммутаторы можно комбинировать или располагать каскадами в свободном пространстве или в планарно-волноводном исполнении для создания коммутаторов большей размерности. Например, как показано на рис. 23.35, планарный массив из 16 оптических модуляторов, каждый из которых служит 1×1-коммутатором, можно организовать в оптическую систему, работающую как перекрестный 4×4-коммутатор в конфигурации broadcast-and-select (рассылка-выбор).

Модуляция и отклонение света может осуществляться с помощью механического, электромеханического, электрического, акустического, магнитного,

теплового или оптического управления. Коммутаторы в этих случаях называются оптомеханическими (или механооптическими), микроэлектромеханическими системами (МЭМС), электрооптическими, акустооптическими, магнитооптическими или термооптическими. В оставшейся части этого раздела дается краткий обзор соответствующих технологий. Полностью оптические или оптооптические коммутаторы описаны в подразд. 23.3.3. Времена переключения этих устройств сравниваются на следующей диаграмме:

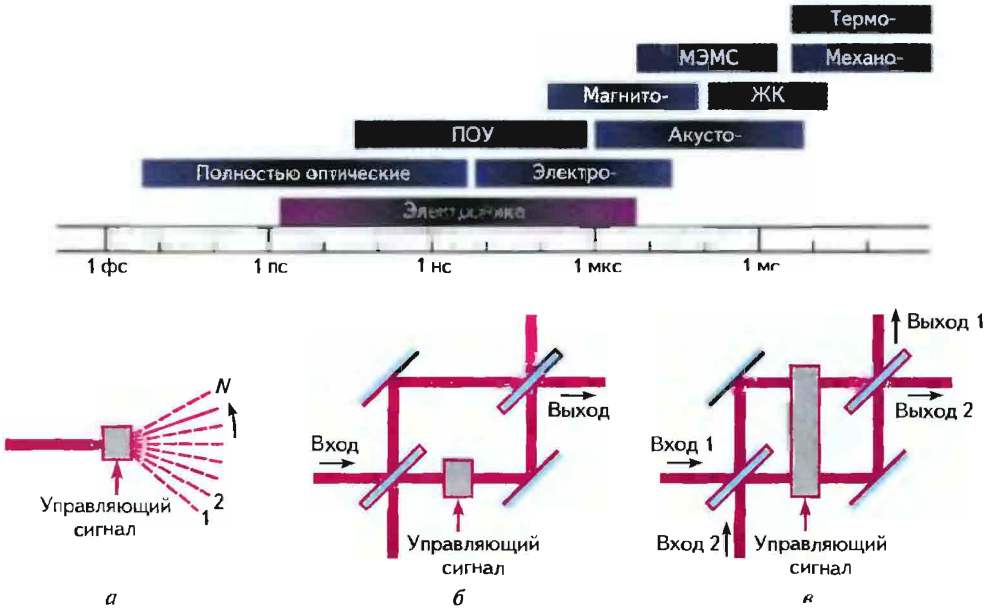
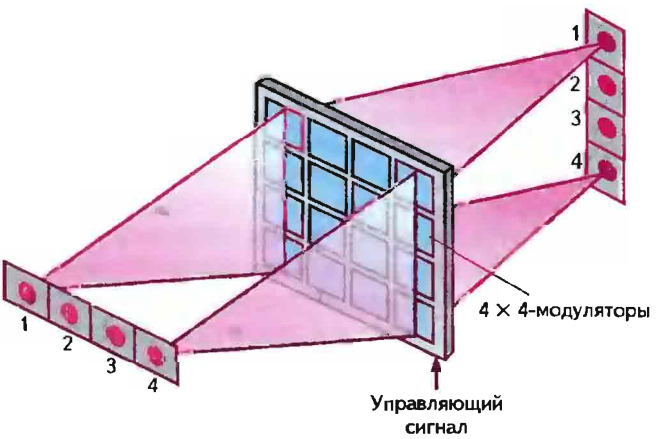


Рис. 23.34. Оптический сканер как $1 \times N$-коммутатор (*а*). Интерферометр с фазовым модулятором как 1×1-коммутатор (*б*). Интерферометр с фазовым модулятором как 2×2-коммутатор (*в*)

Рис. 23.35. 4×4 перекрестный коммутатор. Каждый из 16 элементов — 1×1-коммутатор, передающий или блокирующий свет в зависимости от управляющего сигнала. Свет из входной точки m посылается на все переключатели m-го столбца, $m = 1, 2, 3, 4$. Свет со всех переключателей l-й строки направляется на l-ю выходную точку, $l = 1, 2, 3, 4$. Система является конструктивным воплощением 4×4-коммутатора, изображенного на рис. 23.30, *д*



Механооптические коммутаторы

Механооптический (или оптомеханический) 1 × *N*-коммутатор (сканер) можно реализовать с помощью подвижного (вращающегося или изменяющегося) элемента — зеркала, призмы или голографической решетки, — который отклоняет пучок света в заданный набор направлений (рис. 23.36). Оптическое волокно можно соединить с любым из множества других оптических волокон путем механического смещения входящего волокна, обеспечивающего его юстировку с выбранным выходящим волокном и осуществляемого с помощью такого механизма, какой показан на рис. 23.36, *в*. Для более быстрого механического воздействия можно использовать пьезоэлектрические элементы.

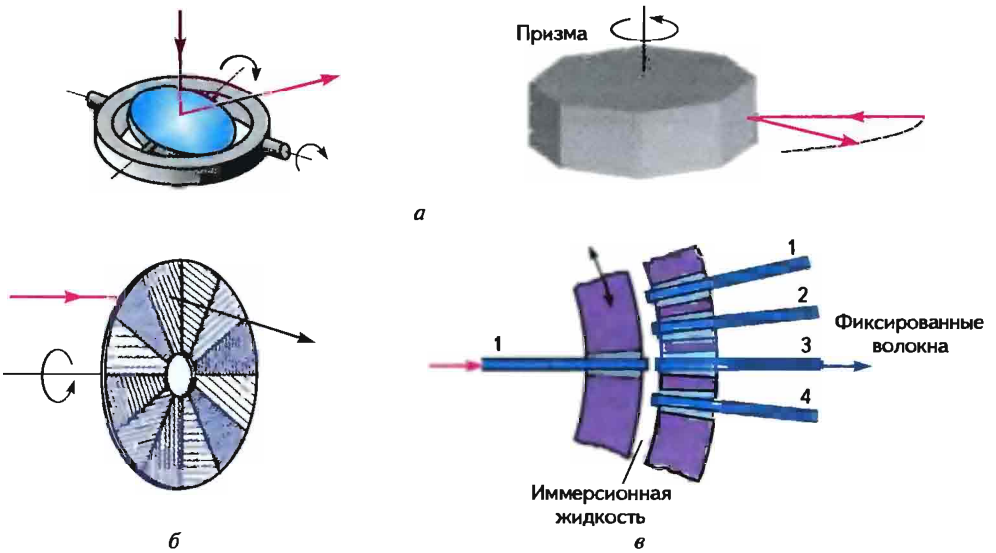


Рис. 23.36. Отклонение света в различных направлениях с помощью поворотного зеркала или призмы (*а*); поворотного голографического диска (*б*). Каждый сектор голографического диска содержит решетку, ориентация и период которой определяют плоскость и угол сканирования отклоняемого света. Оптическое волокно, присоединенное к поворачиваемому колесу, юстируется с одним из нескольких оптических волокон, закрепленных на неподвижном колесе (*в*). Волокна расположены в клиновидных канавках. Иммерсионная жидкость, обеспечивающая согласование показателей преломления, используется для улучшения качества оптической связи

Микроэлектромеханические системы (**МЭМС**) — миниатюризованные системы, приводимые в действие электростатическими приводами и изготавливаемые в виде больших массивов с использованием процессов, аналогичных применяемым в микроэлектронике. Скорости переключения лежат между 10 мс и 10 мкс. Например, перекрестный коммутатор можно создать с помощью набора микроэлектромеханических выпрыгивающих зеркал, как показано на рис. 23.37, *а*, либо набора подвижных зеркал, как показано на рис. 23.37, *б*.

Основным ограничением оптомеханических коммутаторов является их сравнительно медленный отклик (времена переключения лежат в миллисекундном диапазоне). Их основные достоинства — малость вносимых потерь и перекрестных помех.

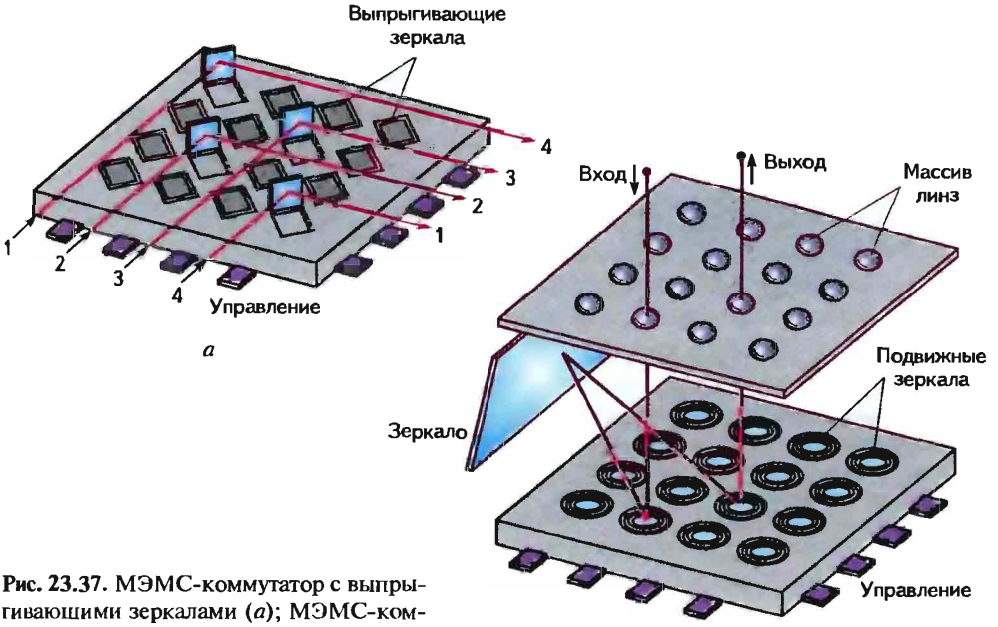


Рис. 23.37. МЭМС-коммутатор с выпрыгивающими зеркалами (*а*); МЭМС-коммутатор с подвижным зеркалом (*б*)

Электрооптические коммутаторы

Как обсуждалось в разд. 20.1, электрооптические материалы меняют свой показатель преломления в присутствии электрического поля. Их можно использовать как электрически управляемые фазовые модуляторы или фазовые пластинки. Будучи помещена в одно из плеч интерферометра или между скрещенными поляризаторами, электрооптическая ячейка служит электрически управляемым модулятором света или 1 × 1-коммутатором типа «включено—выключено» (см. подразд. 20.1.2).

В силу трудности изготовления больших массивов коммутаторов на основе объемных кристаллов наиболее перспективной технологией электрооптической коммутации является интегральная оптика (см. гл. 8 и разд. 20.1). Интегрально-оптические волноводы изготавливаются с использованием электрооптических диэлектрических подложек, таких как ниобат лития $LiNbO_3$, с полосками повышенного показателя преломления в местах расположения волноводов, создаваемыми путем диффузии титана в подложку.

Пример 1 × 1-коммутатора с использованием интегрально-оптического интерферометра Маха—Цендера (ИМЦ) описан в подразд. 20.1.2 и показан на рис. 23.38, *а*. Пример 2 × 2-коммутатора — направленный разветвитель, обсуж-

даемый в подразд. 20.1.4 и показанный на рис. 23.38, *б*. Два близко расположенные волновода оптически связаны; показатель преломления меняется путем наложения электрического поля, подобранного таким образом, чтобы оптическая мощность либо оставалась в том же волноводе, либо передавалась в другой волновод (рис. 23.38, *в*). Такие коммутаторы работают при нескольких вольтах со скоростями, которые могут превышать 20 ГГц.

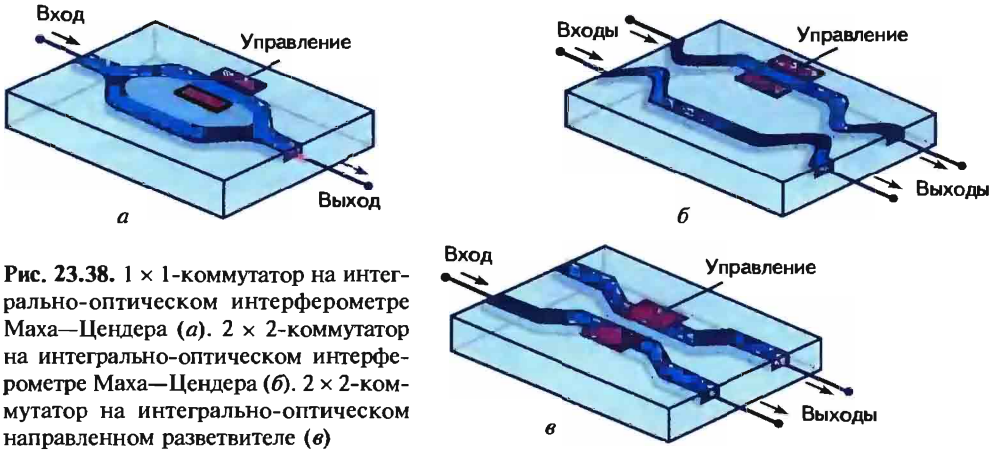


Рис. 23.38. 1 × 1-коммутатор на интегрально-оптическом интерферометре Маха—Цендера (*а*). 2 × 2-коммутатор на интегрально-оптическом интерферометре Маха—Цендера (*б*). 2 × 2-коммутатор на интегрально-оптическом направленном разветвителе (*в*)

Интегрально-оптический $N \times N$-коммутатор можно создать путем комбинации 2 × 2-коммутаторов. 4 × 4-коммутатор реализуется с помощью пяти 2 × 2-коммутаторов, соединенных, как показано на рис. 23.32, *а*. Такая конфигурация может быть выполнена на единой подложке в геометрии, показанной на рис. 23.39. На основе ниобата лития были изготовлены электрооптические коммутаторы размером до 32 × 32.

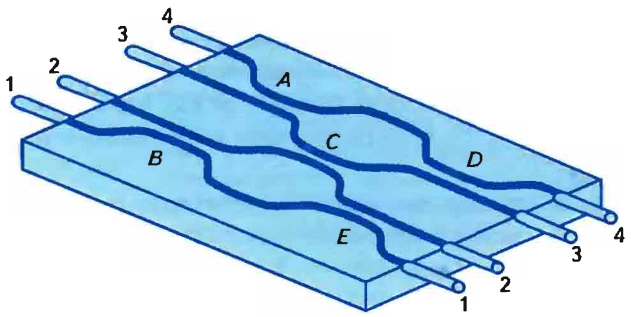


Рис. 23.39. Интегрально-оптический 4 × 4-коммутатор на пяти направленных разветвителях *A*, *B*, *C*, *D*, *E*, расположенных на единой подложке

Предельное число коммутаторов на единицу площади определяется сравнительно большими физическими размерами каждого направленного разветвителя и планарной природой межсоединений внутри микросхемы. Для уменьшения размеров и увеличения плотности упаковки коммутаторов вместо параллельных применяется пересекающиеся волноводы.

Из-за прямоугольной природы интегрально-оптической технологии трудно получить эффективную связь цилиндрических волноводов, например, оптических волокон. Возникают сравнительно высокие вносимые потери, особенно когда одномодовое волокно соединяется с интегрально-оптическим коммутатором. Поскольку коэффициент связи зависит от поляризации, поляризация направляемого света должна быть выбрана правильным образом. Это накладывает требование, чтобы входящее и выходящее волокна обладали свойством сохранения поляризации (см. подразд. 9.2.2). Для изготовления поляризационно-независимых коммутаторов требуются усовершенствованные схемы.

Полупроводниковые фотонные коммутаторы

Полупроводниковые приборы обладают рядом электронных и оптических свойств, которые можно использовать для быстрой оптической коммутации. Как описано в разд. 20.5, электропоглощение, основанное на эффекте Франца—Келдыша, и квантоворазмерный эффект Штарка (КРЭШ) в структурах с множественными квантовыми ямами (МКЯ) используются для управления светом в волноводах вблизи длины волны, соответствующей ширине запрещенной зоны, путем наложения электрического поля. Такие электрически управляемые оптические модуляторы используются как 1 × 1-коммутаторы, работающие на высоких скоростях с временами переключения короче 20 пс. Они могут быть изготовлены в виде больших массивов в поверхностно-нормальной конфигурации на кремниевых подложках, как схематически показано на рис. 23.40.

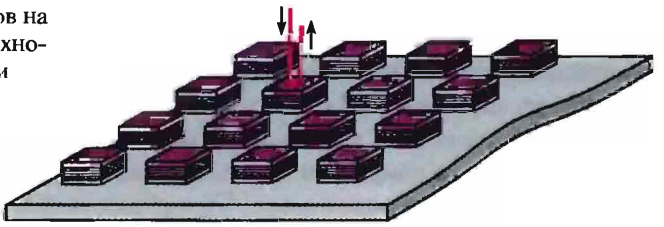


Рис. 23.40. Массив коммутаторов на МКЯ на основе КРЭШ в поверхностно-нормальной конфигурации

Другим важным устройством, используемым в оптической коммутации, является полупроводниковый оптический усилитель (ПОУ). Поскольку ПОУ можно быстро включать и выключать путем включения и выключения электрического тока инжекции (см. подразд. 17.2.1), его можно использовать как быстрый 1 × 1-коммутатор. Сообщалось о временах переключения в наносекундном режиме. В отсутствие усиления (т. е. когда устройство находится в положении «выключено») устройство действует как сильный поглотитель, а при усилении (в положении «включено») оно становится усилителем, так что получается очень высокий коэффициент контрастности (более 40 дБ). Коммутаторы с ПОУ на длинах волн 1,55 и 1,3 мкм реализованы на основе двойной гетероструктуры InGaAsP/InP, а также структур с МКЯ (см. подразд. 17.2.4).

Массивы ПОУ можно изготовить и соединить между собой через оптические волокна, что иллюстрируется примером на рис. 23.41. Поскольку ПОУ

обеспечивают усиление, их можно добавить в схему для компенсации больших потерь разделения. Гибридная и монолитная интеграция ПОУ-коммутаторов с планарными светопроводящими микросхемами (planar lightwave circuits, PLC) на основе кремния была получена для небольших схем, однако крупномасштабная интеграция остается нерешенной задачей.



Рис. 23.41. 2 × 2-коммутатор на основе четырех 1 × 1-коммутаторов на ПОУ в конфигурации рассылка-выбор (broadcast-and-select), показанной на рис. 23.31

Поскольку ПОУ могут работать как преобразователи длины волны, их можно использовать для коммутации с разделением по длинам волн (оптические данные, переносимые на одной длине волны «копируются» на другой длине волны). В силу своих нелинейных свойств ПОУ также используются как сверхбыстрые полностью оптические коммутаторы, как обсуждается в подразд. 23.3.3.

Жидкокристаллические коммутаторы

Жидкие кристаллы (ЖК) обеспечивают еще одну технологию, которую можно использовать для изготовления электрически управляемых оптических коммутаторов. Как описано в разд. 20.3, ЖК-ячейка может служить электрически управляемой фазовой пластинкой или вращателем поляризации. Изменение поляризации можно преобразовать в модуляцию интенсивности с помощью скрещенных поляризаторов. В другой конфигурации для переключения используется изменение показателя преломления ЖК под действием приложенного электрического поля. Свет входит в ЖК под углом из другой среды, показатель преломления которой подбирается так, чтобы полное внутреннее отражение возникало только при наложении электрического поля. Большой массив электродов, расположенных на одной жидкокристаллической панели, служит набором 1 × 1-коммутаторов (цифровым пространственным модулятором света), который можно использовать в конфигурации broadcast-and-select, показанной на рис. 23.35, в качестве перекрестного коммутатора $N \times N$.

Альтернативная конфигурация исполнения 2 × 2 перекрестного коммутатора на ЖК изображена на рис. 23.42. В этой схеме, представляющей собой поляризационную версию интерферометра Маха—Цендера, показанного на рис. 23.34, *в*, ЖК-ячейка поворачивает поляризацию пучков в плечах интерферометра на 90°, если включен управляющий сигнал. При этом происходит переключение из параллельного состояния в кросс-состояние. Этот коммутатор не зависит от поляризации, т. е. пучки направляются в нужные порты независимо от их поляризационного состояния.

Из-за относительно низкой скорости переключения ЖК-коммутаторы применяются в приложениях, где скорость — не главное, например защита от короткого замыкания и реконфигурируемые оптические мультиплексоры ввода-вывода в оптических волоконных сетях (см. 24.4.2).

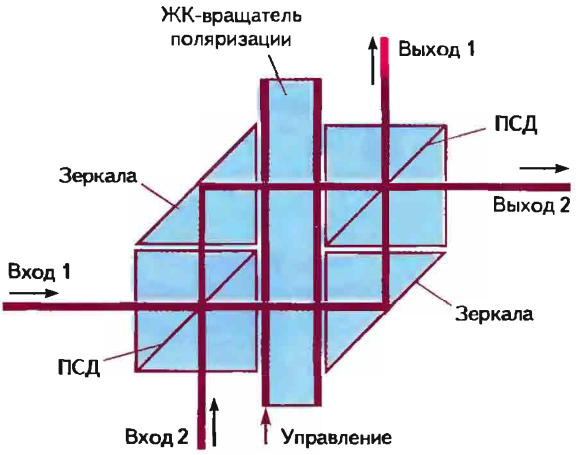


Рис. 23.42. Перекрестный 2 × 2 жидкокристаллический коммутатор. Две поляризационные компоненты входящего пучка разделяются левым поляризующим светоделителем (ПСД) и совмещаются правым ПСД после прохождения жидкокристаллической ячейки, которая служит вращателем поляризации на $\pi/2$ при включенном управляющем сигнале. Без поворота поляризации пучки, поступающие через входы 1 и 2, направляются на выходы 1 и 2 соответственно, т. е. коммутатор находится в параллельном состоянии. При повороте поляризации пучки направляются в противоположные выходы, что соответствует кросс-состоянию коммутатора

Акустооптические коммутаторы

В акустооптических коммутаторах используется свойство брэгговского отклонения света звуком (см. гл. 19). Мощность отклоняемого света управляется интенсивностью звука. Угол отклонения управляется частотой звука. Акустооптический модулятор является коммутатором 1 × 1. Акустооптический сканер (рис. 23.43) — это 1 × N-коммутатор, где N — число разрешаемых точек сканера (см. подразд. 19.2.2). Существуют акустооптические ячейки с $N = 2000$. Если разные части акустооптической ячейки несут акустические волны различной частоты, получается коммутатор или межсоединение $N \times M$. Ограничения на максимальное значение произведения NM, достижимое с акустооптическими ячейками, обсуждались в подразд. 19.2.2. Доступны также массивы электрооптических ячеек.

Магнитооптические коммутаторы

Магнитооптические материалы меняют свои оптические свойства под действием магнитного поля. Например, материалы с эффектом Фарадея в присутствии магнитного поля с индукцией B действуют как вращатели плоскости поляризации (см. подразд. 6.4.2); удельное вращение ρ (угол поворота на единицу длины) пропорционально составляющей вектора B в направлении распространения. Если такой материал поместить между скрещенными поляризаторами, то оптический коэффициент пропускания по мощности $\mathcal{T} = \sin^2 \theta$ за-

висит от угла поворота плоскости поляризации $\theta = \rho d$, где d — толщина ячейки. Прибор действует как коммутатор 1 × 1, управляемый магнитным полем.

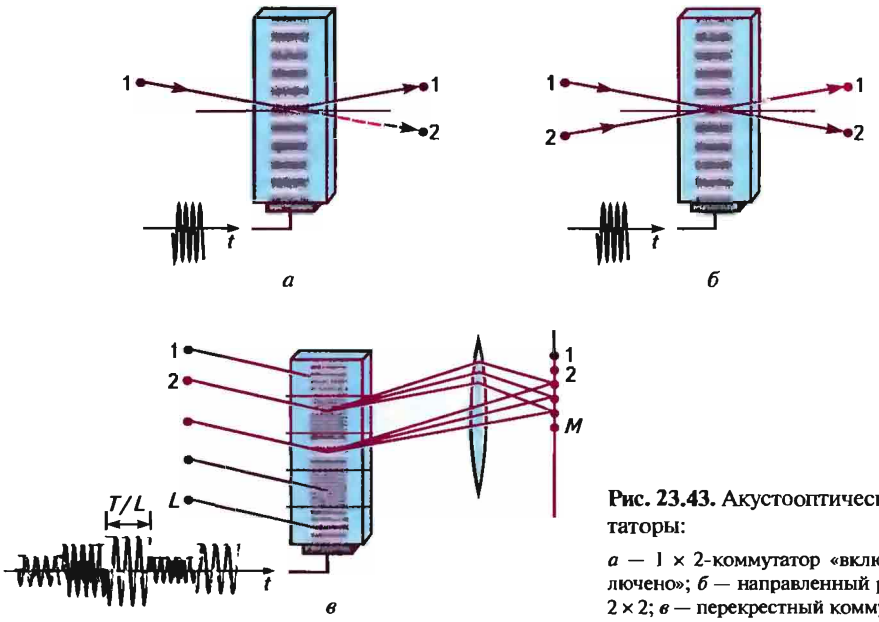


Рис. 23.43. Акустооптические коммутаторы:

а — 1 × 2-коммутатор «включено—выключено»; *б* — направленный разветвитель 2 × 2; *в* — перекрестный коммутатор $L \times M$

Магнитооптические материалы в последнее время привлекли повышенное внимание по причине их использования в системах записи на оптических дисках. В этих системах, однако, используется термомагнитный эффект, при котором намагничение меняется путем локального разогрева мощным сфокусированным лазером. Слабый линейно-поляризованный свет лазера используется для считывания.

Магнитооптические материалы обычно имеют вид пленки (например, из висмут-замещенного железистого граната), выращенной на немагнитной подложке. Магнитное поле накладывается с помощью двух пересекающихся проводников, несущих электрический ток. Система работает в двоичном режиме посредством переключения направления намагничивания.

Массивы магнитооптических коммутаторов можно построить путем вытравливания изолированных ячеек (размером 10 × 10 мкм каждая) на единой пленке. Проводники для управления посредством пропускания электрического тока затем наносятся с помощью фотолитографии. Большие массивы магнитооптических коммутаторов (1024 × 1024) стали теперь доступными, и данная технология продолжает быстро развиваться. Возможно время переключения 100 нс.

Термооптические коммутаторы

Термооптические коммутаторы в общем случае основаны на **термооптическом эффекте**, который заключается в изменении показателя преломления при изменении температуры материала. Например, термооптический

коэффициент кварцевого стекла составляет $dn/dT \approx 10^{-5}$ на один градус (Цельсия), а полимеры имеют более высокие коэффициенты. Поскольку указанное изменение очень мало, термооптический коммутатор часто выполняется в интерферометрической конфигурации.

Термооптические интегрально-оптические коммутаторы изготавливают в согласованных с волокном волноводах из кварца на кремнии (silica-on-Si, SOS). Примером является коммутатор на основе интерферометра Маха—Цендера, показанный на рис. 23.44. Нагревательный элемент в виде тонкой металлической пленки, нанесенной непосредственно на волновод, используется для управления температурой материала в одном из плеч интерферометра. Изменение температуры ΔT приводит к сдвигу фазы

$$\frac{2\pi L}{\lambda_0}\Delta n = \frac{2\pi L}{\lambda_0}\frac{dn}{dT}\Delta T,$$

где L — длина нагреваемого участка. Например, для коммутатора на основе кварца с $L/\lambda_0 = 2 \cdot 10^3$ изменение температуры, необходимое для сдвига фазы на π, составляет $\Delta T = 25$ °C.

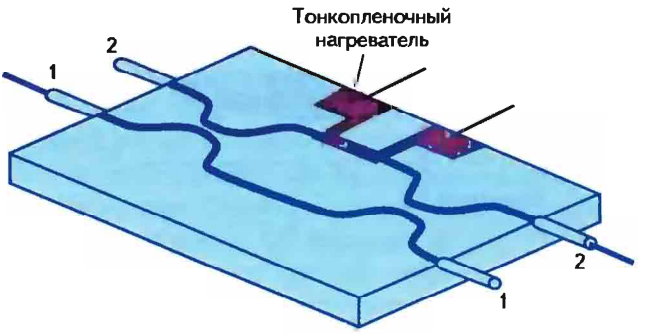


Рис. 23.44. Термооптический коммутатор на основе интерферометра Маха—Цендера

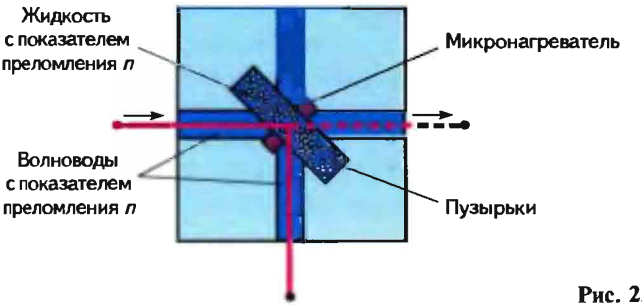
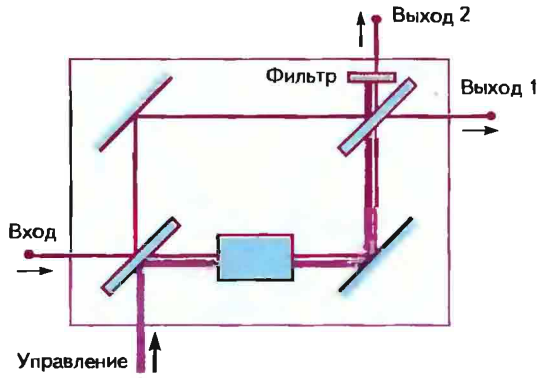


Рис. 23.45. Пузырьковый коммутатор

Еще одна интерферометрическая технология коммутации основана на изменении показателя преломления с использованием микронагревателей, создающих пузырьки в жидкости. Как показано на рис. 23.45, пузырьки превращают жидкость в зеркало, отражающее пучок света.

23.3.3. Полностью оптические пространственные коммутаторы

В полностью оптических (опто-оптических) коммутаторах свет управляет светом через посредство нелинейного оптического материала. Управляющий свет меняет какое-либо оптическое свойство нелинейного материала, которое влияет на какой-либо параметр управляемого света, переключая его с одного порта на другой. Например, в керровской среде показатель преломления меняется управляющим светом, что влияет на фазу управляемого света в интерферометре, переключающем свет с одного выходного порта на другой. Другие нелинейные взаимодействия, которые можно использовать для коммутации, включают светочувствительность задержки или коэффициента поглощения, а также столкновение оптических солитонов, сопровождающееся задержкой по времени или сдвигом по частоте.



а



б

Рис. 23.46. Полностью оптический 1 × 2-коммутатор с использованием интерферометра Маха—Цендера с оптической ячейкой Керра (*а*). Волоконно-оптический интерферометр Маха—Цендера (*б*)

Коммутатор с нелинейным интерферометром Маха—Цендера (ИМЦ)

Интерферометр Маха—Цендера (ИМЦ) с нелинейным оптическим элементом в одном из плеч (рис. 23.46, *а*) можно использовать как полностью оптический 1 × 2-коммутатор, направляющий оптический пучок, входящий в

один из двух входных портов, на любой из двух выходных портов. Коммутатор управляется оптическим пучком, освещающим нелинейный элемент. В отсутствие управляющего пучка интерферометр сбалансирован так, что входящий свет направляется в один из выходных портов. При включении управляющего света возникает изменение показателя преломления Δn, которое, в свою очередь, вносит фазовый сдвиг, равный π, и входящий свет направляется в другой порт. Такой интерферометр допускает объемное (рис. 23.46, *а*) или волоконно-оптическое (рис. 23.46, *б*) исполнение. Конечно, такой 1 × 2-коммутатор можно использовать и как 1 × 1-коммутатор для включения-выключения, если просто игнорировать один из выходных портов.

Сверхбыстрый коммутатор на нелинейном асимметричном интерферометре Саньяка

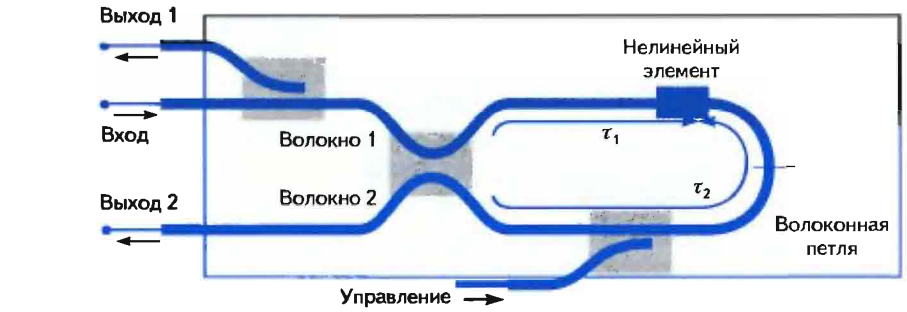
Скорость переключения полностью оптического коммутатора ограничена временем отклика нелинейно-оптического эффекта. Оно в типичном случае состоит из короткого времени подъема при включении управляющего оптического импульса и более длительного времени затухания после его выключения. Например, скорость переключения нелинейного ИМЦ с ячейкой Керра ограничена временем отклика эффекта Керра.

Намного большие скорости переключения, ограниченные малым временем нарастания нелинейного эффекта, могут быть достигнуты при использовании оригинальной интерферометрической схемы, у которой оба плеча интерферометра содержат один и тот же нелинейный элемент, а импульс света, который нужно переключить, проходит через этот элемент в разные моменты времени. Это легко осуществить в волоконном интерферометре Саньяка с нелинейным элементом, расположенным несимметрично в петле волокна, как показано на рис. 23.47.

Когда входной оптический импульс попадает в петлю через волокно 1, он расщепляется симметричным делителем на два импульса равной амплитуды, один из которых распространяется по часовой стрелке, а другой — против нее. Если оба импульса приобретают одинаковый сдвиг фазы за полный обход контура петли, то они рекомбинируют, возвращаются в то же волокно и покидают систему через выходной порт 1. Если они приобретают фазовые сдвиги, различающиеся на π, то после рекомбинации они попадают в другое волокно и покидают систему через выходной порт 2. Мы получаем два состояния коммутатора 1 × 2.

Нелинейный элемент управляется коротким оптическим импульсом, который меняет его показатель преломления на Δn. Это изменение появляется за короткое время τ_i и исчезает за гораздо более длительное время τ_r. Поскольку нелинейный элемент располагается несимметрично в петле волокна, два импульса пересекают его в разные моменты времени τ_1 и τ_2. Если оба импульса проходят через нелинейный элемент, когда он активен, т. е. при наличии полного изменения Δn, то они приобретают один и тот же фазовый сдвиг, и рекомбинирующий импульс направляется в волокно 1. То же самое происходит, если оба импульса проходят через нелинейный элемент в неактивном состоянии. Однако, если один импульс проходит нелинейный элемент в активном состоя-

нии, а другой — в неактивном, они приобретают различные фазовые сдвиги, и если их разность равна π, выходной импульс появляется в волокне 2 и выходит через порт 2. Переключающее действие, таким образом, зависит от разности времен $\tau_1 - \tau_2$, которая пропорциональна расстоянию от средней точки петли до нелинейного элемента. Если $\tau_1 - \tau_2$ немного превышает время нарастания τ_i, то переключением можно управлять с точностью, ограниченной малым временем нарастания, а не полным временем отклика τ_r. Сообщалось о фемтосекундных временах переключения, так что такой коммутатор может работать на терагерцевых скоростях. Данный коммутатор использовался для демультиплексирования с разделением по времени и известен как **терагерцевый оптический асимметричный демультиплексор (ТОАД)**[1].



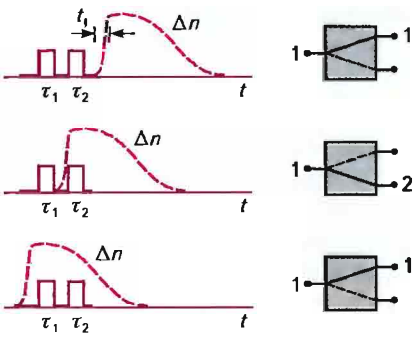


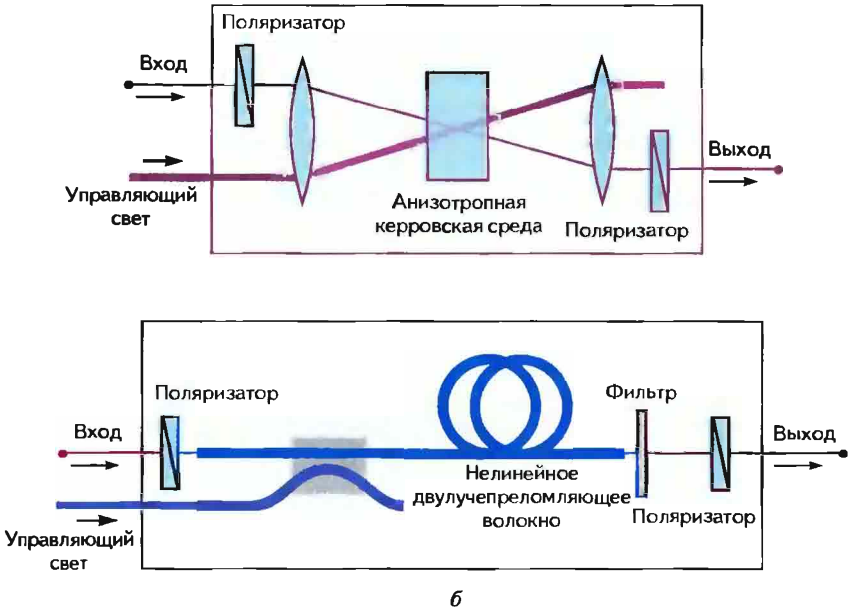
Рис. 23.47. Полностью оптический нелинейный асимметричный интерферометр Саньяка, используемый как коммутатор 1 × 2. Коммутатор управляется оптическим импульсом, взывающим изменение показателя преломления Δn в нелинейном элементе, расположенном несимметричным образом внутри петли интерферометра. Импульс, приходящий из волокна 1, расщепляется на два встречных импульса, проходящих через нелинейный элемент в разные моменты времени. Коммутатор осуществляет переключение с выходного порта 1 на порт 2, если один из встречных импульсов приходит непосредственно перед, а другой — сразу после установления Δn. В результате возникает разность фаз π и выходной импульс направляется в порт 2

Нелинейно-оптический коммутатор на фазовой пластинке

Полностью оптический коммутатор можно создать на основе нелинейного эффекта Керра в анизотропной среде. Действие управляющего оптического импульса создает разные изменения главных значений показателя преломления, так что среда может использоваться как оптически управляемая

[1] См. *Sokoloff J.P., Prucnal P.R., Glesk I., Kane M.* A Terahertz Optical Asymmertic Demultiplexer (TOAD). IEEE Photonics Technology Letters. Vol. 5, 1993. P. 787—790.

фазовая пластинка. Когда среда (например, кристалл или оптическое волокно) помещается между скрещенными поляризаторами, как показано на рис. 23.48, она работает как выключатель. Если запаздывание равно 0, свет блокируется и коммутатор находится в положении «выключено». При запаздывании по фазе, равном π, свет проходит и коммутатор находится в положении «включено».



б

Рис. 23.48. Анизотропная керровская среда в качестве полностью оптического коммутатора: кристалл (*а*) и анизотропное (*б*) (двулучепреломляющее) оптическое волокно. В присутствии управляющего света среда вносит запаздывание по фазе, равное π, так что плоскость поляризации линейно-поляризованного света поворачивается на угол 90° и свет проходит через выходной поляризатор. В отсутствие управляющего света фазовая задержка не вносится и свет блокируется. Фильтр используется для блокирования управляющего света, имеющего другую длину волны

Солитонные коммутаторы

Оптические солитоны — ультракороткие импульсы, распространяющиеся в нелинейной диспергирующей среде без расплывания (см. подразд. 22.5.2). Полностью оптический 1 × 2-коммутатор можно реализовать с помощью одного оптического солитона, управляющего переключением другого солитона между двумя выходными портами. Взаимодействие солитонов может иметь вид столкновения или рекомбинации в один векторный солитон. В любом случае некоторое оптическое свойство входящего солитона меняется в процессе взаимодействия, и изменяемое свойство используется для осуществления маршрутизации. Солитонная технология позволила осуществить субпикосекундные скорости переключения при энергиях переключения в десятки пикоджоулей.

Коммутация за счет столкновения солитонов

Если два солитона со слегка отличающимися частотами и, следовательно, слегка различными групповыми скоростями сталкиваются, т. е. проходят друг через друга, то время прихода и фаза каждого солитона меняются. Один из импульсов служит управляющим, а другой — сигнальным. Либо временная задержка, либо фазовый сдвиг, сопровождающий столкновение, используется для маршрутизации сигнального импульса. Маршрутизация с разделением по времени осуществляется с помощью оптического затвора, который открывается в течение предписанного временного окна. Маршрутизация с разделением по фазе реализуется с помощью интерферометра.

Коммутация с использованием векторных солитонов

Векторный солитон состоит из двух ортогонально поляризованных оптических импульсов, совместно распространяющихся в нелинейном двулучепреломляющем волокне. Поскольку оба импульса должны присутствовать для формирования векторного солитона, систему можно использовать как оптический коммутатор, в котором один импульс служит для управления другим.

Два импульса с ортогональными поляризациями распространяются в двулучепреломляющем волокне со слегка различными групповыми скоростями и, следовательно, разделяются во времени. Это явление известно как эффект сноса (см. подразд. 22.5.1). Если волокно еще и нелинейное, то фазовая кроссмодуляция (ФКМ) (см. подразд. 21.3.3) приводит к сдвигу частоты вверх у одного импульса и вниз — у другого. Из-за дисперсии групповой скорости (ДГС) эти сдвиги сопровождаются изменением групповых скоростей. Когда разность групповых скоростей из-за двулучепреломления в точности компенсируется ДГС (за счет ФКМ) оба импульса распространяются вместе как единый векторный солитон. Это явление называется также *захватом солитона.*

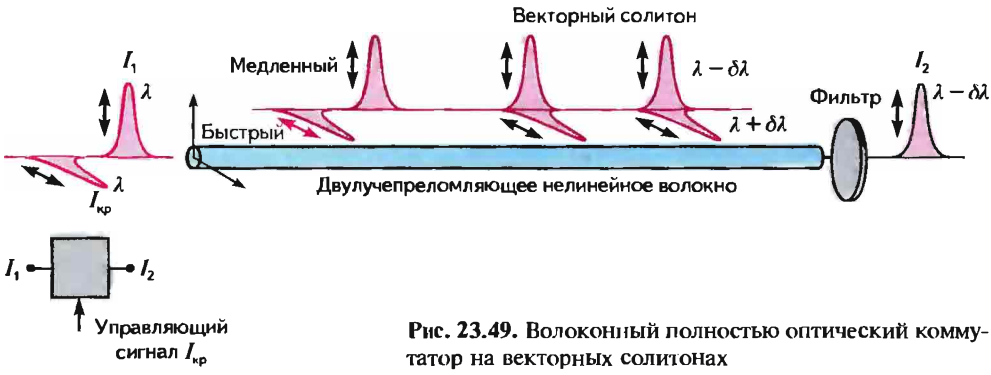


Рис. 23.49. Волоконный полностью оптический коммутатор на векторных солитонах

Как показано на рис. 23.49, солитонный коммутатор 1 × 1 реализуется при использовании одного из ортогонально поляризованных импульсов в качестве управляющего, а другого — в качестве сигнального, который пропускается или

блокируется. Если оба импульса имеют одинаковую длину волны λ, то при распространении через двулучепреломляющую среду они образуют векторный солитон, компоненты которого имеют сдвинутые длины волн $\lambda \pm \delta\lambda$. Одна из этих компонент выделяется фильтром и образует выходной сигнал коммутатора. В отсутствие управляющего импульса векторный солитон не образуется и длина волны не смещается, так что свет блокируется фильтром.

Фундаментальные ограничения полностью оптических коммутаторов

Минимальные значения энергии переключения E и времени переключения T полностью оптических коммутаторов подчиняются следующим фундаментальным физическим ограничениям.

Флуктуации числа фотонов

Минимальная энергия, необходимая для переключения, в принципе равна энергии одного фотона. Однако, поскольку существует неустранимая случайность числа фотонов, испускаемых лазером или светодиодом, требуется большее число фотонов, чтобы гарантировать переключающее действие почти всегда, когда это желательно. Для этих источников света и при определенных условиях (см. подразд. 12.2.3) число фотонов, приходящих в течение фиксированного интервала времени, представляет собой случайное число n с пуассоновским распределением вероятности

$$p(n) = \bar{n}^n \exp\frac{-\bar{n}}{n!},$$

где $\bar{n}$ — среднее число фотонов. Если $\bar{n} = 21$, то вероятность того, что не будет зарегистрировано ни одного фотона, равна $p(0) = e^{-21} \approx 10^{-9}$. Следовательно, 21 фотон в среднем — это минимальное число, гарантирующее доставку по крайней мере одного фотона со средним числом ошибок 1 из каждых 10^9 попыток. Соответствующая энергия равна $E = 21h\nu$. Для света с длиной волны $\lambda_0 = 1$ мкм $E = 21 \cdot 1{,}24 \approx 26$ эВ $= 4{,}2$ аДж. Это можно считать нижней границей энергии переключения. Следует отметить, что это скорее практическая граница, а не фундаментальный предел, поскольку в принципе можно использовать свет, сжатый по числу фотонов (см. подразд. 12.3.2). По менее оптимистичным оценкам, для опорного сигнала нужно минимум 100 фотонов. Это соответствует минимальной энергии переключения в 20 аДж при $\lambda_0 = 1$ мкм. Отметим, что на оптических частотах $h\nu$ намного больше тепловой энергии $k_B T$ при комнатной температуре ($k_B T = 0{,}026$ эВ при $T = 300$ °K).

Соотношение неопределенностей «энергия—время»

Другой фундаментальный квантовый принцип выражается соотношением неопределенностей «энергия—время» $\sigma_E \sigma_T \geq h/4\pi$ [см. (12.16)]. Произведение минимальной энергии переключения E и минимального времени переключения T та-

ким образом обязано быть больше $h/4\pi$ (т. е. $E \geq h/4\pi T = h\nu/4\pi\nu T$). Эта граница энергии меньше, чем энергия фотона, в $4\pi\nu T$ раз. Поскольку время переключения не меньше оптического периода $1/\nu$, величина $4\pi\nu T$ всегда больше единицы. Поскольку E выбирается больше энергии одного фотона $h\nu$, отсюда следует, что соотношение неопределенностей «энергия—время» всегда удовлетворяется.

Время переключения

Единственное фундаментальное ограничение, накладываемое на минимальное время переключения, возникает из-за соотношения неопределенностей «энергия—время». В действительности легко генерируются оптические импульсы длительностью в несколько фемтосекунд (несколько оптических периодов). Соответствующие скорости переключения не могут быть достигнуты в полупроводниковых электронных коммутаторах (а также лежат за пределами современных возможностей приборов на эффекте Джозефсона). Субпикосекундные скорости переключения были получены в ряде оптических коммутационных устройств. В принципе энергии переключения у этих устройств также могут быть намного меньше, чем в полупроводниковой электронике.

Габариты

Ограничения на размеры фотонных коммутаторов определяются дифракционными эффектами, которые затрудняют ввод-вывод оптической мощности для устройств, размеры которых меньше длины волны света

Практические ограничения

Главное ограничение для полностью оптической коммутации является результатом слабости нелинейных эффектов в ныне доступных материалах, что делает требуемую энергию переключения весьма высокой. Другое важное практическое ограничение связано с трудностью отвода тепла, выделяемого в процессе переключения. Это ограничение особенно существенно, когда переключение повторяется. Если минимальная энергия каждого переключения равна E, то каждую секунду используется полная энергия E/T. Для очень малых времен переключения эта мощность может быть весьма большой. Максимальная скорость, с которой выделяемая мощность должна отводиться, определяет предел, делающий непригодным сочетание очень малых времен и очень высоких энергий переключения. Заметим, однако, что тепловые эффекты накладывают менее строгие ограничения, если прибор работает при частотах повторения, которые ниже максимальной частоты, так что энергия, выделяемая при одном акте переключения, рассеивается за время, превышающее время прохождения одного бита.

23.3.4. Коммутаторы с разделением по длине волны

Коммутаторы, описанные до сих пор, осуществляют переключение сигналов в пространстве, т. е. устанавливают пути передачи, по которым оптические пучки направляются из одного физического положения в другое (из вход-

ного порта в выходной). Их логические аналоги в области длин волн называются **коммутаторами с разделением по длине волны.** Рассмотрим некоторые примеры.

Пример 23.2

Реконфигурируемый селектор длины волны

Примером оптического устройства, сочетающего пассивные маршрутизаторы с разделением по длине волны с пространственными коммутаторами, является селектор длины волны, показанный на рис. 23.50. Этот коммутатор выделяет одну или несколько длин волн из приходящего пучка с N длинами волн. В нем используются демультиплексор, разделяющий компоненты с N длинами волн, набор из N коммутаторов 1 × 1 для выбора желательных длин волн и мультиплексор, формирующий выходной пучок, как показано на рис. 23.50. В целом система является сочетанием пассивных маршрутизаторов с разделением по длине волны и пространственных коммутаторов.

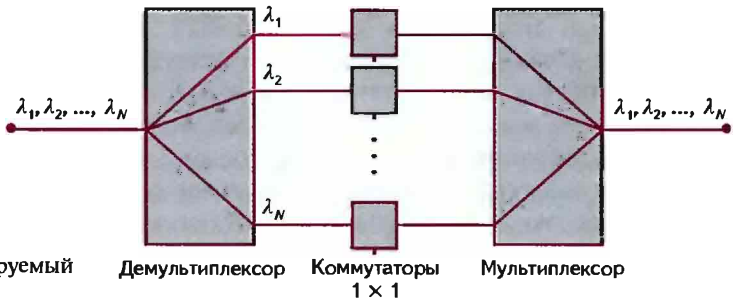


Рис. 23.50. Реконфигурируемый селектор длин волн

Пример 23.3

Реконфигурируемый оптический мультиплексор ввода-вывода (РОМВВ)

РОМВВ — это реконфигурируемый ОМВВ с возможностью добавления, удаления или пропускания, как показано на рис. 23.51. В нем используются демультиплексор, мультиплексор, также 1 × 2-коммутатор и 2 × 1-коммутатор на каждый канал ввода-вывода.

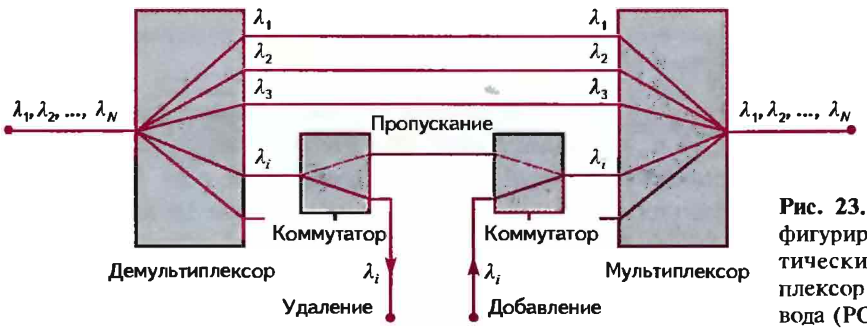


Рис. 23.51. Реконфигурируемый оптический мультиплексор ввода-вывода (РОМВВ)

Пример 23.4

Обмен данными между каналами с разными длинами волн (wavelength-channel interchange, WCI)

Коммутатор WCI, называемый также **лямбда-коммутатором**, осуществляет маршрутизацию данных между каналами с разной длиной волны внутри одного оптического пучка. Коммутатор WCI $N \times N$ можно реализовать путем отображения каналов с разными длинами волн на пространственную область с помощью демультиплексора, преобразования длин волн с помощью набора из N **преобразователей длины волны (ПДВ)** и совмещения каналов в один пучок с помощью $N \times 1$-соединителя, как показано на рис. 23.52 (см. цв. вклейку). Преобразователь длины волны изменяет длину волны пучка без изменения данных, т. е. копирует данные из одного канала в другой.

Многомерные коммутаторы с разделением по пространству и длине волны

Предыдущие примеры коммутаторов с разделением по длине волны обрабатывали один оптический пучок с множеством каналов, различающихся длиной волны. Коммутация может осуществляться с множеством многоканальных пучков. Рассмотрим, например, коммутацию N пучков, каждый из которых с одним из N каналов, отличающихся длиной волны. Коммутатор перераспределяет эти каналы между пучками. Две схемы реализации такого процесса показаны на рис. 23.53.

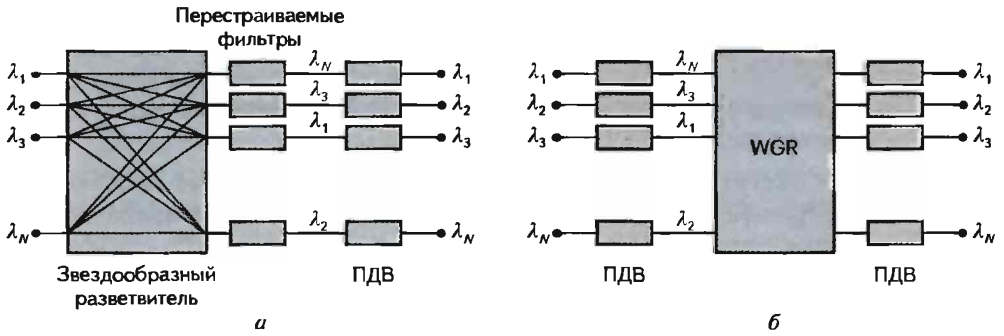


Рис. 23.53. Коммутаторы типа рассылка—выбор с разделением по пространству и длине волны (*а*); WGR (*б*)

В первой схеме маршрутизатор типа рассылка—выбор перенаправляет каналы с разными длинами волн в различные порты. Это достигается с использованием звездообразного разветвителя, который рассылает содержание всех N пучков на набор спектральных фильтров, каждый из которых настроен на длину волны одного канала (рис. 23.53, *а*). В конце концов для дальнейшей обработки длины волн переключаемых каналов преобразуются к исходным значениям (без изменения данных) с использованием набора преобразователей длины волны (ПДВ).

В другом исполнении используется два набора ПДВ, между которыми находится маршрутизатор на дифракционной решетке с разделением по длине волны (WGR), как показано на рис. 23.53, *б*. Первый ПДВ преобразует длины волн к значениям, удовлетворяющим уравнению WGR (23.18) с определенными адресами назначения.

WGR-коммутатор более эффективен, чем коммутатор типа рассылка—выбор, поскольку последний рассеивает значительную мощность на фильтрах. Однако преимуществом коммутатора типа рассылка—выбор является возможность изменения конфигурации.

Конструкции преобразователей длины волны

Преобразователь длины волны (ПДВ) переносит данные, содержащиеся в оптическом пучке на одной длине волны, на другую длину волны. Эти длины волн часто соответствуют каналам волоконной системы связи с разделением по длине волны (см. подразд. 24.3.3) и их разность невелика, так как они лежат в одном диапазоне спектра. Преобразование длины волны осуществляется с использованием нелинейных оптических устройств, параметрических или непараметрических.

В непараметрических ПДВ интенсивность одного пучка, промодулированная данными, вызывает изменение какого-либо оптического свойства среды, например коэффициента усиления, коэффициента поглощения или показателя преломления полупроводника, пропорциональное интенсивности, так что данные «записываются» в среде. При распространении второго пучка с другой длиной волны он модулируется за счет измененного свойства среды, так что данные «считываются» и передаются второму пучку.

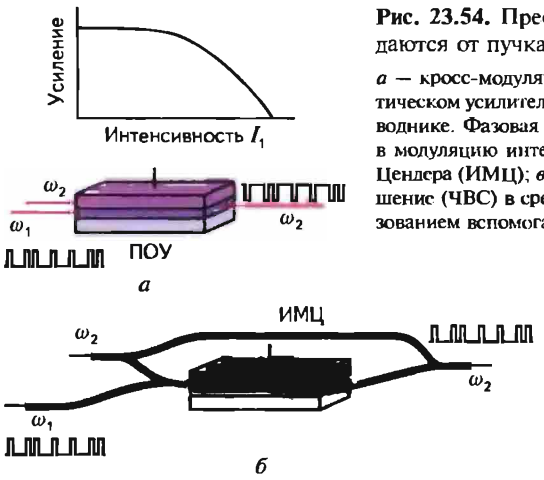


Рис. 23.54. Преобразование длины волны. Данные передаются от пучка с частотой ω_1 пучку с частотой ω_2:

а — кросс-модуляция усиления (КМУ) в полупроводниковом оптическом усилителе; *б* — кросс-модуляция фазы (КМФ) в полупроводнике. Фазовая модуляция преобразуемого пучка превращается в модуляцию интенсивности с помощью интерферометра Маха—Цендера (ИМЦ); *в* — частично вырожденное четырехволновое смешение (ЧВС) в среде с нелинейностью третьего порядка с использованием вспомогательной волны на частоте $\omega_0 = (\omega_1 + \omega_2)/2$

Как изображено на рис. 23.54, *а*, усиление насыщенного полупроводникового оптического усилителя (ПОУ) является убывающей функцией интен-

сивности. Когда исходный пучок, модулированный по интенсивности, проходит через такую среду, то усиление модулируется как обратная величина; то же самое происходит с интенсивностью считывающего пучка. Процесс называется **кросс-модуляцией усиления (КМУ)**.

В ненасыщенном ПОУ пишущим пучком модулируется показатель преломления, поскольку он зависит от плотности носителей. Следовательно, считывающий пучок модулируется по фазе. Процесс называется **кросс-модуляцией фазы (КМФ).** Для преобразования фазовой модуляции в модуляцию интенсивности требуется интерферометр, как показано на рис. 23.54, *б*.

В ПДВ на основе параметрических взаимодействий пучки с различными длинами волн связаны нелинейным эффектом (см. гл. 21). Например, в среде с нелинейностью второго порядка волна с частотой ω_1 может быть преобразована в волну с более низкой частотой $\omega_2 = \omega_3 - \omega_1$ с помощью вспомогательной волны с частотой ω_3. Амплитуда волны, преобразованной с понижением частоты, связана с амплитудой исходной волны, так что данные, заключенные в ее величине или фазе, переносятся в преобразованную волну. Основная трудность этого процесса трехволнового смешения заключается в том, что если частоты ω_1 и ω_2 близки, то частота ω_3 вспомогательной волны должна быть примерно вдвое больше. Если нужно использовать только волны с примерно одинаковыми частотами, могут осуществляться каскадные нелинейные параметрические процессы. Первым таким процессом является генерация второй гармоники (ГВГ), при котором ω_1 преобразуется в $2\omega_1$, а вторым — процесс трехволнового смешения с понижением частоты, при котором генерируется волна с частотой $\omega_2 = 2\omega_3 - \omega_1$. Все три волны теперь имеют приблизительно одинаковую частоту. Альтернативой является процесс четырехволнового смешения в среде с нелинейностью третьего порядка, например, в оптическом волокне (рис. 23.54, *в*). Как описано в разд. 21.3, этот процесс включает взаимодействие четырех волн, частоты которых удовлетворяют соотношению $\omega_1 + \omega_2 = \omega_3 + \omega_4$. В частично вырожденном случае $\omega_3 = \omega_4 = \omega_0$, так что $\omega_2 = 2\omega_0 - \omega_1$.

23.3.5. Коммутаторы с разделением по времени

Коммутатор с разделением по времени осуществляет маршрутизацию сигналов между временны́ми интервалами (рис. 23.55, см. цв. вклейку). В цифровых системах связи сигнал делится на последовательность временны́х кадров одинаковой длительности, каждый из которых разделен на N квантов времени, где находятся данные. Примером коммутации с разделением по времени является межинтервальный обмен (time-slot interchange, TSI), при котором данные из ℓ-го кванта времени каждого кадра передаются в m-й квант того же кадра. Это соответствует обмену между каналами с различными длинами волн (WCI), описанному в предыдущем разделе.

Двумерные пространственно-временны́е коммутаторы используют комбинацию временны́х и пространственных коммутаторов. Такой коммутатор соединяет набор входящих линий, каждая из которых несет цифровой сигнал, состоящий из

последовательности временны́х кадров, с аналогичным набором выходящих линий. Данные из каждого временно́го кванта в каждой входящей линии передаются в один или несколько временны́х квантов в одной или нескольких выходящих линиях в соответствии с некоторым правилом. Примером является коммутатор «время—пространство—время» (**time—space—time, TST**), состоящий из каскада TSI, пространственного коммутатора и другого TSI, как показано на рис. 23.56.

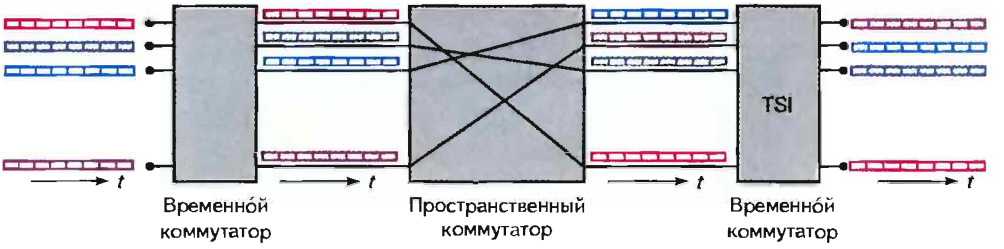


Рис. 23.56. Коммутатор «время—пространство—время» (TST)

Мультиплексирование и демультиплексирование с разделением по времени

Простым примером пространственно-временно́го коммутатора является демультиплексор с разделением по времени. Он имеет одну входящую линию и N выходящих линий, где N равно числу временны́х квантов в каждом кадре. Коммутатор направляет данные из ℓ-го временно́го кванта входящей линии в ℓ-й временно́й квант ℓ-й выходящей линии; $\ell = 1, 2, ..., N$. Процесс повторяется периодически во всех кадрах. Следовательно, такой коммутатор эквивалентен отображению временно́й области на пространственную.

Например, в демультиплексоре с разделением по времени, показанном на рис. 23.57, имеется $N = 4$ временны́х кванта в каждом кадре. Кванты содержат данные в виде импульсов различной высоты. Коммутатор направляет первый импульс в первый выходной порт, второй импульс — во второй выходной порт и так далее. Такой коммутатор можно построить с использованием пространственного коммутатора $1 \times N$, соединяющего входной порт последовательно с одним из четырех выходных портов.

Обратное по отношению к демультиплексору с разделением по времени устройство, называемое **мультиплексором с разделением по времени (time-division multiplexer, TDM)**, обеспечивает поочередное выстраивание импульсов из N отдельных портов в единую последовательность импульсов. Эта обратная операция наглядно представлена на рис. 23.57, где входные и выходные порты меняются ролями, т. е. импульсы проходят не слева направо, а наоборот, справа налево. $1 \times N$-демультиплексор с разделением по времени можно реализовать с помощью N коммутаторов 1×1 (типа «включено—выключено»), как показано на рис. 23.31, *а*, последовательно включаемых и выключаемых управляющими импульсами от задающего генератора.

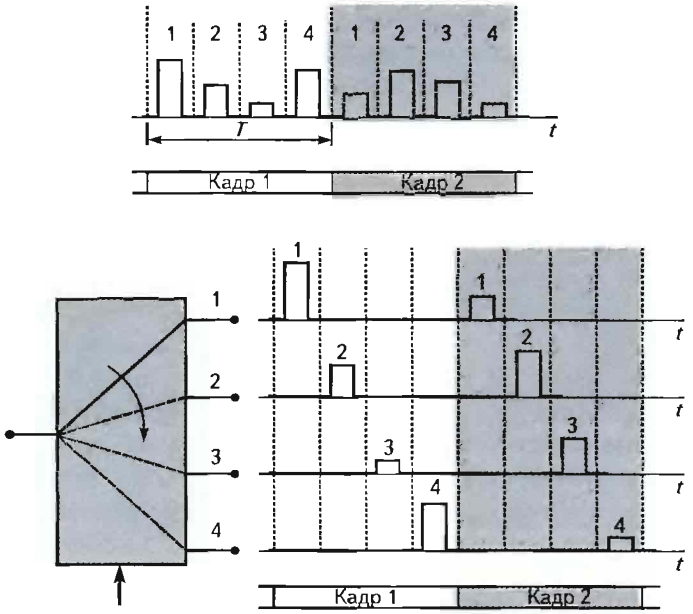


Рис. 23.57. Демультиплексирование с разделением по времени при $N = 4$

Оптическое мультиплексирование с разделением по времени (TDM)

Оптическое исполнение TDM показано на рис. 23.58, *а*. Копии входящего пучка пропускаются через набор из N оптических коммутаторов 1×1, управляемых набором оптических импульсов от задающего генератора, задержанных на целое число временны́х отрезков T/N, где T — длительность кадра.

В другом исполнении копии входящего пучка последовательно задерживаются на целое число T/N, так что N входящих импульсов синхронизируются во времени, но разделяются в пространстве; 1×1-коммутаторы управляются таким же сигналом задающего генератора (рис. 23.58, *б*). Система аналогична той, которая используется для регистрации временно́го профиля оптического импульса (см. рис. 22.51 в подразд. 22.6.1). Оптические задержки могут быть обеспечены с помощью оптических волокон (примерно 5 нс/м для волокна из кварцевого стекла). 1×1-коммутаторы могут иметь оптическое исполнение на основе полностью оптического нелинейного интерферометрического коммутатора. Пример — терагерцевый оптический асимметричный демультиплексор (TOAД), описанный в подразд. 23.3.3.

Оптический межинтервальный обмен (TSI)

TSI-коммутатор (см. рис. 23.55) — это временно́й коммутатор, осуществляющий обмен данными между временны́ми квантами внутри каждого

кадра. Его исполнение может быть основано на комбинации пространственных коммутаторов с пространственно-временны́ми. Например, в конфигурации, показанной на рис. 23.59 (см. цв. вклейку), используется демультиплексор с разделением по времени, направляющий временны́е кванты на отдельные строки (отображение «время—пространство»); перед вводом в пространственный перекрестный коммутатор $N \times N$ для синхронизации импульсов в одном кванте времени длительностью T/N вносятся временны́е задержки. Другой набор временны́х задержек затем вносится для восстановления расстановки импульсов по временны́м квантам, после чего мультиплексор с разделением по времени сводит эти временны́е кванты в единую строку (отображение «пространство—время»).

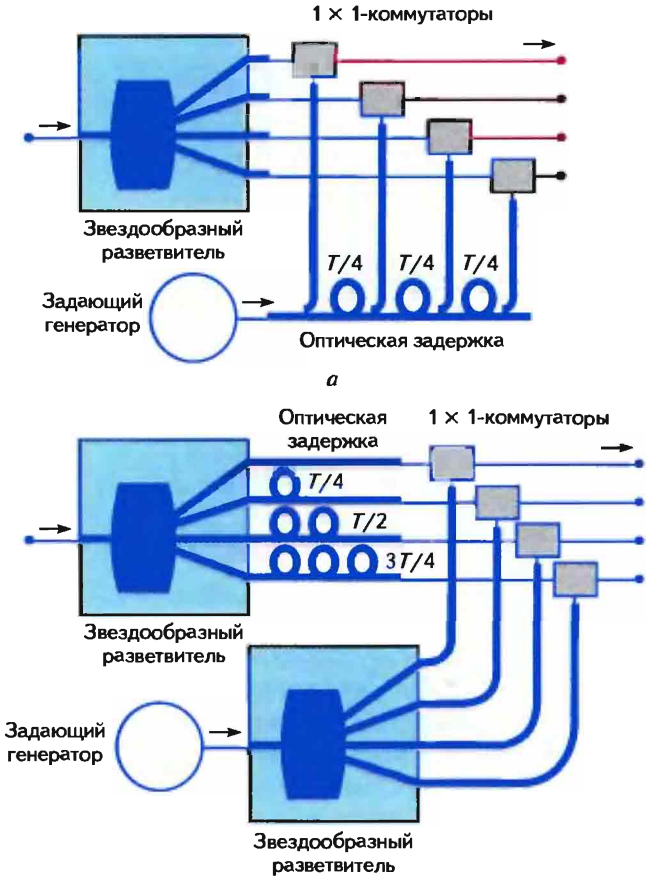


Рис. 23.58. Реализация демультиплексирования с разделением по времени с помощью набора элементов оптической задержки и 1×1 оптических коммутаторов. На этой иллюстрации $N = 4$

Оптические программируемые задержки и буферы

Управляемые временны́е задержки — существенные компоненты временно́й коммутации. Буферы — это элементы памяти для вре́менного хранения данных или компенсации разницы в скоростях потока данных. Как видно на рис. 23.58 и 23.59, такие задержки вносятся с помощью оптических волокон подходящей длины (≈5 нс/м). Программируемые задержки можно осуществить, давая возможность оптическим импульсам циркулировать программируемое число раз по волоконным петлям. Как показано на рис. 23.60, это достигается с использованием перекрестного коммутатора, который позволяет импульсу входить в петлю в нужное время и выпускает его после определенного числа обходов петли.

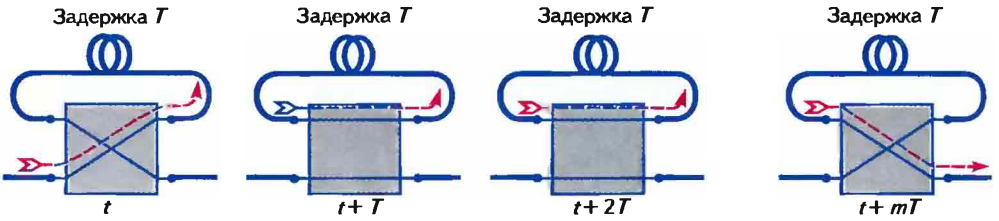


Рис. 23.60. Программируемая линия задержки с использованием волоконной петли и перекрестного коммутатора. В момент $t = 0$ коммутатор находится в перекрестном состоянии, так что оптический импульс пропускается в петлю. В момент $t = T$ импульс возвращается назад во входной порт коммутатора, который теперь находится в параллельном состоянии, так что импульс совершает еще один обход с дополнительной задержкой T. В момент $t = mT$ импульс выпускается путем перевода коммутатора в перекрестное состояние

23.3.6. Коммутаторы пакетов

Коммутаторы, представленные до сих пор в данном разделе, — это реляционные коммутаторы, которые устанавливают отображение между портами входа и выхода в зависимости от состояния коммутатора, управляемого внешними сигналами, не зависящими от данных, приходящих на входные порты. Такой тип коммутации называется **коммутацией цепей**. У другого типа коммутаторов, называемых коммутаторами пакетов, конфигурация устанавливается в соответствии с адресной информацией, содержащейся в самих входных данных. Данные организованы в пакеты, каждый из которых имеет заголовок, содержащий адрес назначения пакета, как показано на рис. 23.61. Коммутатор пакетов имеет устройство распознавания заголовка, который считывает адреса и посылает управляющий сигнал, устанавливающий коммутатор в нужное состояние.

Система распознавания адресов в заголовке может содержать банк корреляторов, которые сопоставляют битовую последовательность, представляющую адрес во входящем пакете, с битовыми последовательностями, представляю-

щими каждый из возможных адресов доставки в таблице соответствия, и идентифицируют адрес с наивысшей корреляцией. Например, если адрес входящего пакета есть битовая последовательность (a_1, a_2, ..., a_N), а один из адресов в таблице есть (b_1, b_2, ..., b_N), то корреляция определяется суммой $a_1b_1 + a_2b_2 +$ $+ a_Nb_N$. Поскольку биты поступившего заголовка приходят последовательно во времени, для осуществления операции корреляции требуются устройства задержки, умножитель и сумматор. В одном из оптических исполнений используется оптическое волокно с N отражателями в виде волоконных решеток Брэгга (ВРБ), расположенных на равных расстояниях, как показано на рис. 23.62. Отражатели имеют коэффициенты отражения (b_1, b_2, ..., b_N) и служат умножителями. Задержки на время полного обхода, вносимые участками волокна, обеспечивают синхронность битов приходящего адреса, так что они складываются и дают корреляционную сумму.

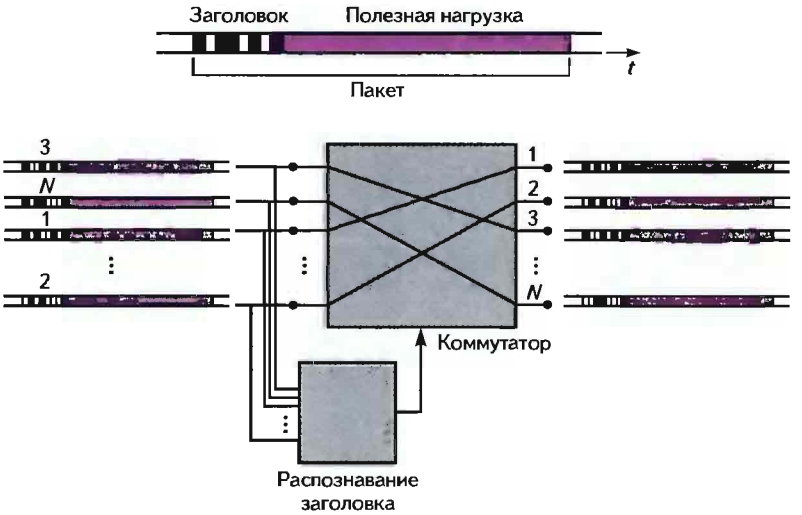


Рис. 23.61. Пакеты и коммутаторы пакетов

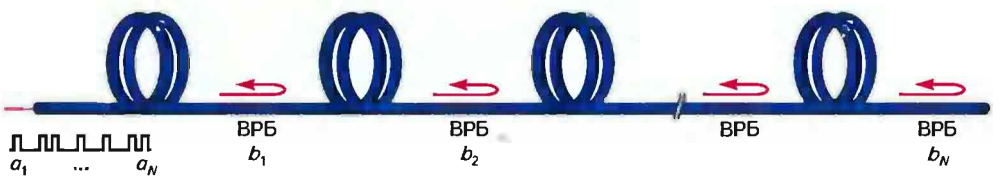


Рис. 23.62. Оптический коррелятор для распознавания адреса в заголовке

Коммутатор пакетов может иметь также последовательное исполнение с использованием набора элементарных 2 × 2-коммутаторов, каждый из которых направляет входящий пакет в свой верхний или нижний выходной порт в зависимости от одного бита в адресе заголовка. Например, если этот бит есть 1 или 0,

то коммутатор направляет данные в верхний или нижний порт соответственно. В других системах 2 × 2-коммутатор сортирует два входящих пакета и направляет пакет с большим номером адреса в нижний порт выхода, а другой пакет — в верхний.

Например, 8 × 8 трехступенчатый коммутатор, показанный на рис. 23.63 и называемый **коммутатором типа «баньян»**, включает двенадцать 2 × 2-коммутаторов с автоматической маршрутизацией. Адрес каждого пакета выражается двоичным числом (x_1, x_2, x_3). Маршрутизация на первой ступени основана на наиболее значимом бите x_1, а на ступенях 2 и 3 — на битах x_2 и x_3 соответственно. В каждом случае, если бит содержит 1, пакет направляется в нижний выходной порт, в противном случае — в верхний. Коммутатор устроен таким образом, что после трех ступеней пакет направляется по адресу назначения. Однако нетрудно показать, что между двумя пакетами может возникнуть конфликт, если оба они должны быть на один и тот же выходной порт 2 × 2-коммутатора. Для предотвращения такого внутреннего блокирования были разработаны более сложные конфигурации. Например, в сетях, использующих сочетание блоков сортировки и маршрутизации, внутреннего блокирования удается полностью избежать.

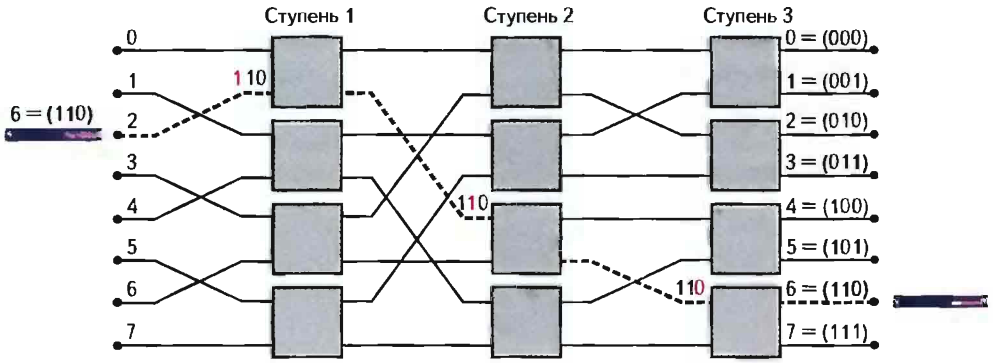


Рис. 23.63. 8 × 8 трехступенчатый коммутатор типа «баньян». Пакет, приходящий во входной порт номер 2 с адресом заголовка 6, направляется точно по назначению, в выходной порт номер 6, после прохождения через три 2 × 2-коммутатора с автоматической маршрутизацией. Поскольку адрес представляется двоичным числом 6 = (110), пакет направляется в нижний, нижний и верхний порты этих коммутаторов соответственно, следуя по пути, показанному штриховой линией, и в конечном счете достигает выходного порта 6 = (110)

Конфликт возникает, когда пакеты из различных входных портов одновременно назначаются в один и тот же выходной порт. Методы разрешении конфликтов включают направление конфликтующего пакета по другому пути или задержку его на другое время с помощью буфера. В оптической области пакет можно также преобразовать на другую длину волны и передать по другому каналу. Оптические буферы и преобразователи длины волны были описаны в подразд. 23.3.4.

23.4. ОПТИЧЕСКИЕ ЛОГИЧЕСКИЕ ЭЛЕМЕНТЫ

Высокоорганизованные цифровые электронные системы (например, цифровой компьютер) содержат большое число связанных между собой основных элементов: коммутаторов, логических элементов и элементов памяти (триггеров). В данном разделе вводятся бистабильные устройства, которые дают основу для оптических логических устройств и триггеров.

23.4.1. Бистабильные системы

Бистабильная (имеющая два устойчивых положения) система имеет выходной сигнал, принимающий только одно из двух различных значений при любом сигнале, поступающем на вход. Переключение между этими значениями может быть достигнуто путем временного изменения уровня входного сигнала. Например, в системе, показанной на рис 23.64, выходной сигнал принимает свое нижнее значение при низких уровнях входного сигнала и верхнее — при высоких. Когда рост сигнала на входе приводит к превышению некоторого критического уровня (порога) ϑ_2, выходной сигнал перескакивает с нижнего значения на верхнее. Последующее снижение входного сигнала вызывает скачкообразное изменение выходного сигнала на нижнее значение при другом критическом значении входного сигнала $\vartheta_1 < \vartheta_2$, так что кривая зависимости «вход—выход» образует петлю гистерезиса.

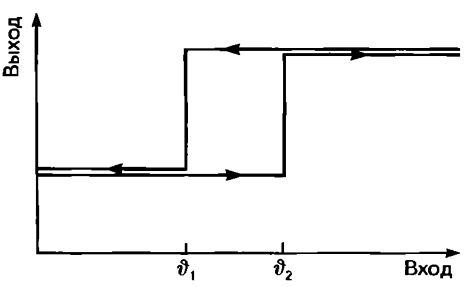


Рис. 23.64. Связь входного и выходного сигналов в бистабильной системе

Между ϑ_1 и ϑ_2 имеется область промежуточных значений входного сигнала, в которой возможно и высокое, и низкое значения выходного сигнала в зависимости от предыстории изменения сигнала на входе. В этой области система ведет себя как детские качели в виде доски, уравновешенной в центре. Если выходной сигнал имеет низшее значение, то большой положительный всплеск на входе переключает его на высшее значение. Большой отрицательный всплеск на входе переключает его обратно. Система имеет триггерное поведение; ее состояние зависит от предшествующей истории (был ли последний всплеск положительным или отрицательным, рис. 23.65).

Бистабильные устройства важны в цифровой электронике и являются основными строительными блоками компьютерных систем. Они используются в коммутаторах, логических элементах и элементах памяти. Параметры устройства могут быть подобраны так, чтобы два критических значения (пороги ϑ_1 и ϑ_2)

совместились в одно значение ϑ. В результате получается однопороговая крутая S-образная нелинейная связь между входным и выходным сигналами. При подходящем смещении устройство может иметь большое дифференциальное усиление и использоваться в качестве усилителя, как транзистор. Его можно использовать также как устанавливающий пороговый элемент, в котором выходной сигнал переключается между двумя значениями, когда входной сигнал превышает порог, для формирования импульсов или в качестве ограничителя (рис. 23.66). Для этих операций требуется стабильный порог и стабильное смещение.

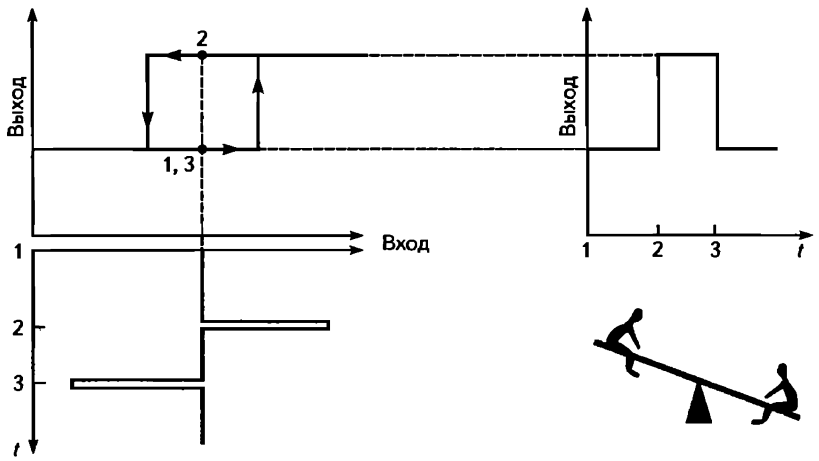


Рис. 23.65. Переключение в бистабильной системе. В момент времени 1 выходной сигнал низкий. Положительный импульс на входе в момент 2 перебрасывает систему в состояние с высоким выходным сигналом. Она остается в этом состоянии до тех пор, пока отрицательный импульс на входе не перебросит ее обратно. Система работает как выключатель с фиксацией положения или элемент памяти

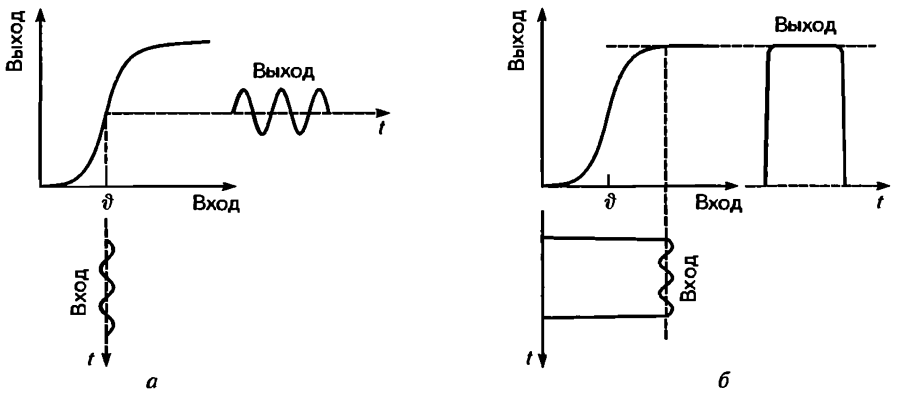


Рис. 23.66. Бистабильное устройство как усилитель (*а*) или как формирователь или ограничитель импульсов (*б*)

Бистабильные устройства используются также как логические элементы. Двоичные данные представляются импульсами, которые складываются, и их сумма используется в качестве входного сигнала бистабильного устройства. При должном выборе высот импульсов по отношению к порогу (рис. 23.67) устройство можно заставить переключаться вверх только при наличии обоих импульсов одновременно, так что оно работает как логический элемент «И». Это цифровое устройство с двумя двоичными входами и одним двоичным выходом. Для того чтобы на выходе было состояние «1», необходимо, чтобы «1» было на каждом из входов, в противном случае на выходе будет «0».

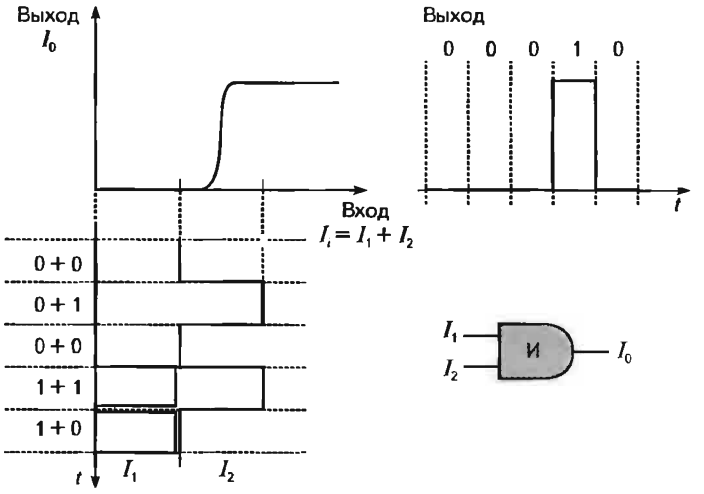


Рис. 23.67. Бистабильное устройство как логический элемент «И». Входной сигнал $I_i = I_1 + I_2$, где I_1 и I_2 — импульсы, представляющие двоичные данные. Выходной сигнал I_0 высокий тогда и только тогда, когда присутствуют оба импульса

Логические элементы можно использовать как коммутаторы. Например, при использовании одного из входов в схему «И» в качестве управляющего эта схема превращается в 1 × 1-коммутатор «включено—выключено».

23.4.2. Основы оптической бистабильности

Для создания бистабильного устройства требуются два свойства: *нелинейность* и *обратная связь*. Электронная бистабильная схема (триггер) получается, если подключить выход одного из двух транзисторов к входу другого (см. любой учебник по цифровой электронике). Оптическая бистабильная система реализуется на нелинейном оптическом элементе, выходной пучок которого используется в системе обратной связи для управления прохождением света через сам элемент.

Рассмотрим типичную оптическую систему, показанную на рис. 23.68. За счет обратной связи выходная интенсивность I_0 неким образом управляет ко-

эффициентом пропускания $\mathcal{T}$ системы, так что $\mathcal{T}$ становится некоторой нелинейной функцией $\mathcal{T} = \mathcal{T}(I_0)$. Поскольку $I_0 = \mathcal{T} I_i$,

$$I_i = \frac{I_0}{\mathcal{T}(I_0)}. \qquad (23.21)$$

Связь входа и выхода в бистабильной системе

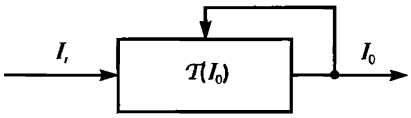


Рис. 23.68. Оптическая система, коэффициент пропускания которой $\mathcal{T}$ является функцией выходной интенсивности I_0

Если $\mathcal{T}(I_0)$ — немонотонная функция, например, колоколообразная, как на рис. 23.69, *а*, то I_i также будет немонотонной функцией I_0, как показано на рис. 23.69, *б*. Следовательно, I_0 будет многозначной функцией I_i с более чем одним соответствующим значением I_0, как видно на рис. 23.69, *в*.

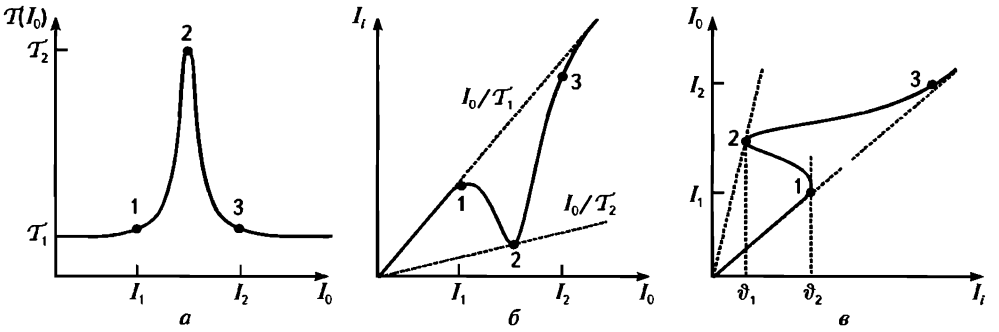


Рис. 23.69. Коэффициент пропускания $\mathcal{T}(I_0)$ в зависимости от выходного сигнала I_0 (*а*). Входной сигнал $I_i = I_0/\mathcal{T}(I_0)$ как функция выходного I_0 (*б*). Для $I_0 < a$ или $I_0 > b$, $\mathcal{T}(I_0) = \mathcal{T}_1$ и $I_i = I_0/\mathcal{T}_1$ представляет собой линейное соотношение с наклоном $1/\mathcal{T}_1$. При промежуточном значении I_0, для которого $\mathcal{T}$ имеет свое максимальное значение $\mathcal{T}_1$ (точка 2), I_i падает ниже линии $I_i = I_0/\mathcal{T}_1$ и касается нижней прямой $I_i = I_0/\mathcal{T}_2$ в точке 2. Зависимость выходного сигнала I_0 от входного I_i получается простым перестроением кривой (*б*) с заменой осей друг на друга (график поворачивается на 90° против часовой стрелки и зеркально отражается относительно вертикальной оси) (*в*)

Таким образом, система имеет бистабильное поведение. При малых ($I_i < \vartheta_1$) или больших ($I_i > \vartheta_2$) входных сигналах каждому значению входного сигнала соответствует единственное значение выходного сигнала. Однако в области промежуточных значений ($\vartheta_1 < I_i < \vartheta_2$) каждому значению сигнала на входе соответствует три возможных значения выходного сигнала. Верхнее и нижнее значения устойчивы, а среднее (линия, соединяющая точки 1 и 2 на рис. 23.69, *в*) неустойчиво. Любое слабое возмущение, добавленное на входе, заставляет выходной сигнал перейти на верхнюю или нижнюю ветвь. Если начиная в малых

значений увеличивать сигнал на входе, то по достижении порога ϑ_2 выходной сигнал изменит свое значение на верхнее без перехода через неустойчивое состояние. При дальнейшем уменьшении входного сигнала значение выходного следует верхней ветви до ϑ_1, после чего скачком перемещается на нижнюю ветвь, как показано на рис. 23.70.

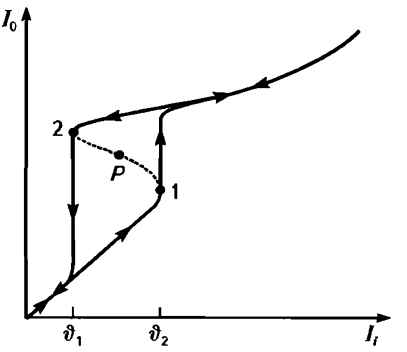


Рис. 23.70. Выходной сигнал в зависимости от входного для бистабильного устройства, показанного на рис. 23.68. Штриховая линия изображает неустойчивое состояние

В неустойчивости промежуточного состояния можно убедиться, если рассмотреть точку P на рис. 23.70. Небольшое увеличение выходного сигнала I_0 вызывает резкое увеличение коэффициента пропускания $\mathcal{T}(I_0)$, поскольку наклон $\mathcal{T}(I_0)$ положителен и велик (см. рис. 23.69, *а* и заметьте, что P лежит на линии, соединяющей точки 1 и 2). Это, в свою очередь, вызывает дальнейший рост $\mathcal{T}(I_0)$, который еще более увеличивает значение I_0. В результате происходит переход в верхнее устойчивое состояние. Аналогичным образом небольшое уменьшение I_0 вызывает переход в нижнее устойчивое состояние.

Нелинейная колоколообразная функция $\mathcal{T}(I_0)$ использовалась только для иллюстрации. Множество других нелинейных функций приводят к бистабильности (и, возможно, мультистабильности, когда более чем два устойчивых состояния выхода соответствуют одному и тому же состоянию входа).

Упражнение 23.3

Примеры нелинейных функций, демонстрирующих бистабильность

С помощью компьютера постройте график зависимости между I_0 и $I_i = I_0/\mathcal{T}(I_0)$ для каждой из следующих функций:

а) $\mathcal{T}(x) = \dfrac{1}{(x-1)^2 + a^2}$;

б) $\mathcal{T}(x) = \dfrac{1}{1 + a^2 \sin^2(x+\theta)}$;

в) $\mathcal{T}(x) = \dfrac{1}{2} + \dfrac{1}{2}\cos(x+\theta)$;

г) $\mathcal{T}(x) = \operatorname{sinc}^2 \sqrt{a^2 + x^2}$;

д) $\mathcal{T}(x) = \dfrac{(x+1)^2}{(x+a)^2}$.

Выберите подходящие значения постоянных a и θ для достижения бистабильности. Функции в примерах от (б) до (д) применяются к бистабильным системам, которые будут обсуждаться далее.

23.4.3. Бистабильные оптические устройства

Для реализации вышеописанных основных принципов можно использовать множество оптических схем. Возможно использование двух типов оптических элементов (рис. 23.71).

1. Диспергирующие нелинейные элементы, у которых показатель преломления n является функцией оптической интенсивности.

2. Диссипативные нелинейные элементы, у которых функцией интенсивности является коэффициент поглощения a.

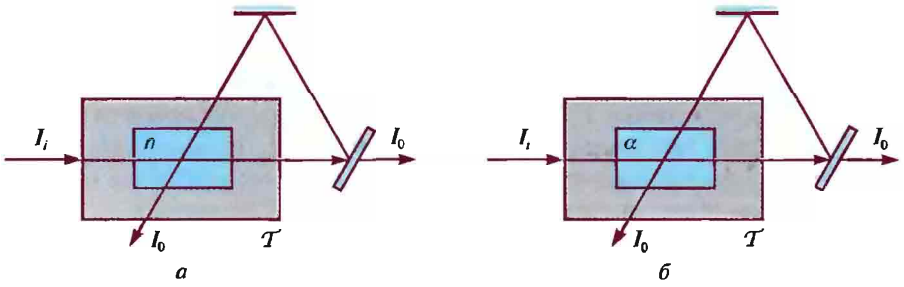


Рис. 23.71. Диспергирующая бистабильная оптическая система (*а*). Коэффициент пропускания $\mathcal{T}$ является функцией показателя преломления n, который управляется выходной интенсивностью I_0. Диссипативная бистабильная оптическая система (*б*). Коэффициент пропускания является функцией коэффициента поглощения α, который управляется выходной интенсивностью I_0

Оптический элемент помещается в оптическую систему, и выходная интенсивность I_0 управляет коэффициентом пропускания системы в соответствии с некоторым нелинейным законом $\mathcal{T}(I_0)$.

Диспергирующие нелинейные элементы

Можно предложить ряд оптических систем, у которых коэффициент пропускания $\mathcal{T}$ является немонотонной функцией зависящего от интенсивности показателя преломления $n = n(I_0)$. Примерами являются интерферометры, такие как интерферометр Маха—Цендера и эталон Фабри—Перо, у которых среда обладает оптическим эффектом Керра,

$$n = n_0 + n_2 I_0, \tag{23.22}$$

где n_0 и n_2 — постоянные.

В *интерферометре Маха—Цендера* нелинейная среда помещается в одном плече, как показано на рис. 23.72. Коэффициент пропускания системы по мощности дается выражением (см. подразд. 2.5.1)

$$\mathcal{T} = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{d}{\lambda_0}n + \varphi_0\right), \tag{23.23}$$

где d — длина активной среды; λ_0 — длина волны в свободном пространстве; φ_0 — постоянная. Подставляя (23.22), получаем

$$\boxed{\mathcal{T}(I_0) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{d}{\lambda_0}n_2 I_0 + \varphi\right),} \tag{23.24}$$

где φ — другая постоянная,

$$\varphi = \varphi_0 + \frac{2\pi d}{\lambda_0}n_0.$$

Как показано на рис. 23.72, это нелинейная функция, состоящая из периодических повторений типичной колоколообразной функции, использованной ранее для демонстрации бистабильности (см. рис. 23.69, *а*).

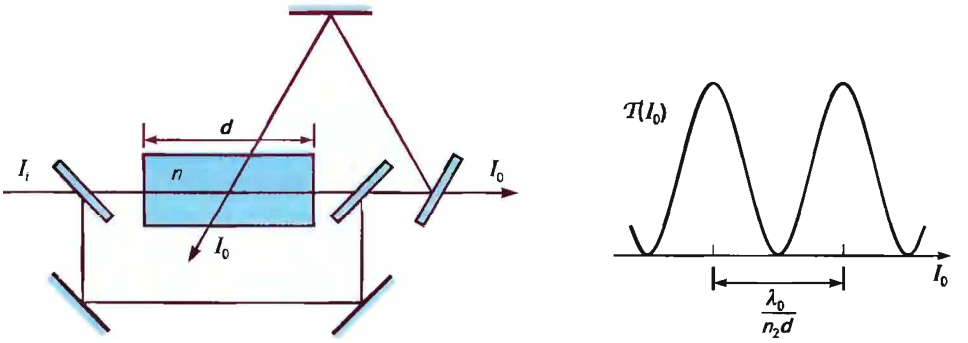


Рис. 23.72. Интерферометр Маха—Цендера с нелинейной средой, показатель преломления которой n управляется прошедшей интенсивностью I_0 за счет оптического эффекта Керра

В *эталоне Фабри—Перо* с расстоянием d между зеркалами коэффициент пропускания по интенсивности дается формулой (см. подразд. 2.5.2):

$$\mathcal{T} = \frac{\mathcal{T}_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2\left[(2\pi d/\lambda_0)n + \varphi_0\right]}, \tag{23.25}$$

где $\mathcal{T}_{\max}$, $\mathcal{F}$ и φ_0 — постоянные; λ_0 — длина волны в свободном пространстве. Подставляя n из (23.22), получаем

$$\boxed{\mathcal{T}(I_0) = \frac{\mathcal{T}_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2\left[(2\pi d/\lambda_0)n_2 I_0 + \varphi\right]},} \tag{23.26}$$

где φ — новая постоянная. Как показано на рис. 23.73, эта функция является периодической последовательностью острых пиков — узких колоколообразных функций. Следовательно, система также бистабильна.

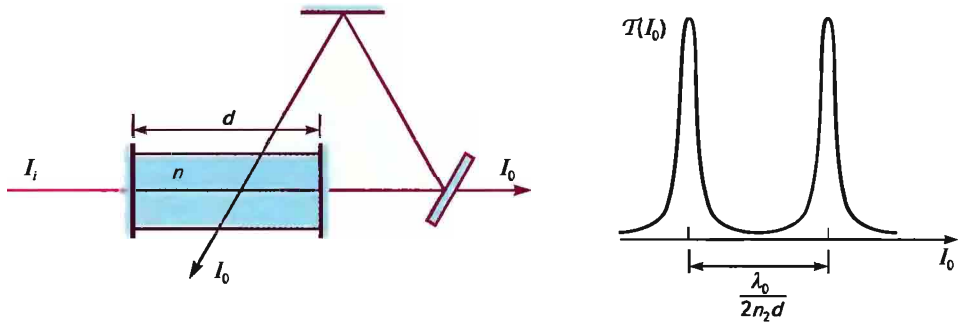


Рис. 23.73. Интерферометр Фабри—Перо, содержащий среду с показателем преломления n, управляемым интенсивностью пропускаемого света I_0

Внутренне бистабильные оптические устройства

Оптическая обратная связь, необходимая для бистабильности, может быть не внешней, а внутренней. Например, система, показанная на рис. 23.74, имеет резонатор с оптически нелинейной средой, показатель преломления которой n управляется интенсивностью света внутри резонатора I, а не выходной оптической интенсивностью I_0. Поскольку $I_0 = I\mathcal{T}_0$, где $\mathcal{T}_0$ — коэффициент пропускания выходного зеркала, действие внутренней интенсивности I имеет такой же эффект, как и действие выходной интенсивности I_0, с точностью до постоянного множителя. Если, например, среда обладает оптическим эффектор Керра, то показатель преломления — линейная функция оптической интенсивности $n = n_0 + n_2 I$, и коэффициент пропускания эталона Фабри—Перо

$$\mathcal{T}(I_0) = \frac{\mathcal{T}_{max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2\left[(2\pi d/\lambda_0) n_2 I_0/\mathcal{T}_0 + \varphi\right]}. \qquad (23.27)$$

Это устройство действует как самонастраивающаяся система.

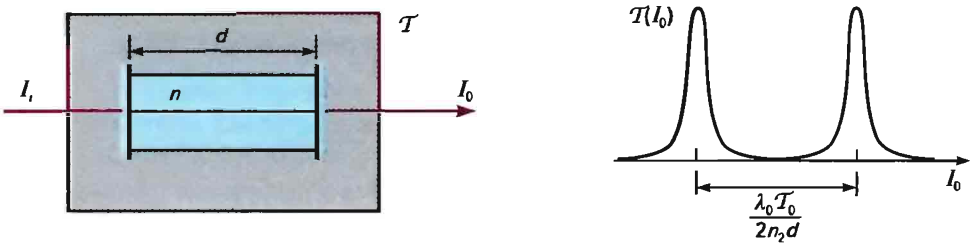


Рис. 23.74. Внутренне бистабильное устройство. Внутренняя интенсивность света I управляет активной средой и, следовательно, полным коэффициентом пропускания системы $\mathcal{T}$

Диссипативные нелинейные элементы

Диссипативный нелинейный материал имеет коэффициент поглощения, который зависит от оптической интенсивности I. Примером является насыщаемый поглотитель, обсуждавшийся в подразд. 14.4.1, у которого коэффициент поглощения — нелинейная функция I:

$$\alpha = \frac{\alpha_0}{1 + I/I_s}, \tag{23.28}$$

где α_0 — коэффициент поглощения слабого сигнала; I_s — интенсивность насыщения. Если поглотитель помещен внутри эталона Фабри—Перо длиной d, которая настроена на максимальное пропускание (рис. 23.75), то

$$\mathcal{T} = \frac{\mathcal{T}_1}{\left(1 - \mathcal{R}e^{-\alpha d}\right)^2}, \tag{23.29}$$

где $\mathcal{R} = \sqrt{\mathcal{R}_1\mathcal{R}_2}$; $\mathcal{R}_1$ и $\mathcal{R}_2$ — коэффициенты отражения зеркал; $\mathcal{T}_1$ — постоянная (см. подробности в подразд. 2.5.2 и 10.1.1). Если $\alpha d \ll 1$, т. е. среда оптически тонкая, то $e^{-\alpha d} \approx 1 - \alpha d$ и

$$\mathcal{T} \approx \frac{\mathcal{T}_1}{\left[1 - (1 - \alpha d)\mathcal{R}\right]^2}. \tag{23.30}$$

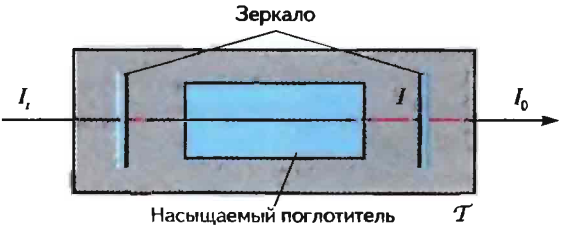


Рис. 23.75. Бистабильное устройство, состоящее из насыщаемого поглотителя в резонаторе

Поскольку α — нелинейная функция I, то $\mathcal{T}$ — также нелинейная функция I. С помощью соотношения $I = I_0/\mathcal{T}_0$ и формул (23.28) и (23.30), находим

$$\mathcal{T}(I_0) \approx \mathcal{T}_2\left[\frac{I_0 + I_{s_1}}{I_0 + (1 + a)I_{s_1}}\right]^2, \tag{23.31}$$

где

$$\mathcal{T}_2 = \frac{\mathcal{T}_1}{(1 - \mathcal{R})^2}; \quad a = \frac{\alpha_0 d\mathcal{R}}{1 - \mathcal{R}}; \quad I_{s_1} = I_s\mathcal{T}_0.$$

При некоторых значениях a система бистабильна [вспомните упражнение 23.3, пункт (д)].

Предположим теперь, что насыщаемый поглотитель заменили на усиливающую среду с насыщением усиления

$$\gamma = \frac{\gamma_0}{1 + I_i / I_s}. \tag{23.32}$$

Такая система — не что иное, как оптический усилитель с обратной связью, т. е. лазер. Если $\mathcal{R} \exp(\gamma_0 d) < 1$, лазер находится ниже порога, но если $\mathcal{R} \exp(\gamma_0 d) > 1$, система становится неустойчивой, и мы получаем лазерную генерацию. Итак, лазеры обладают бистабильным поведением. Однако теория подобных явлений выходит за рамки данной книги.

В некотором смысле *диспергирующие бистабильные оптические системы — это аналоги лазера, в которых вместо нелинейного усиления имеется нелинейный показатель преломления.*

Материалы

Оптическая бистабильность наблюдалась в множестве материалов с оптическим эффектом Керра (например, пары натрия, сероуглерод и нитробензол). Коэффициент нелинейности n_2 у этих материалов очень мал. Поэтому требуется большая длина d и, следовательно, велико время отклика (наносекундный режим). Требуемая мощность переключения также высока.

Полупроводники, такие как GaAs, InSb, InAs и CdS, обладают высокой оптической нелинейностью, благодаря экситонным эффектам на длинах волн вблизи краев запрещенной зоны. Бистабильное устройство можно сделать просто из слоя полупроводникового материала с двумя параллельными частично отражающими поверхностями, действующими как зеркала эталона Фабри—Перо (рис. 23.76). Из-за большой нелинейности такой слой может быть тонким, что позволяет уменьшить время отклика.

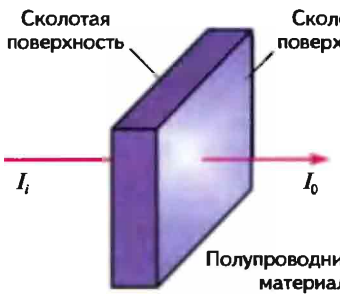


Рис. 23.76. Тонкий слой полупроводника с двумя параллельными отражающими поверхностями может служить бистабильным устройством

Коммутаторы из GaAs, основанные на этом эффекте, оказались наиболее удачными. Достигнуты времена включения в несколько пикосекунд, однако время выключения, определяемое главным образом сравнительно медленной рекомбинацией носителей, гораздо больше и составляет несколько наносекунд. Время выключения в 200 пс было достигнуто с использованием специально приготовленных образцов, в которых поверхностная рекомбинация усилена. Энергия переключения составляет от 1 до 10 пДж. В принципе возможно понизить энергию переключения до фемтоджоулей. InAs и InSb имеют более длитель-

ные времена выключения (до 200 нс). Однако их действие можно ускорить за счет увеличения энергии переключения. Показано, что полупроводниковые структуры с множественными квантовыми ямами (см. подразд. 16.1.7 и 17.2.5) также проявляют бистабильность, она присуща также органическим материалам.

Ключевым условием практического использования бистабильных оптических устройств является возможность их изготовления в виде больших массивов. Массивы бистабильных элементов могут размещаться на одном чипе с отдельными пикселями, определяемыми световыми пучками. Альтернативно для определения пикселей может использоваться реактивное ионное травление. Существующая технология позволяет изготовить массив 100 × 100 пикселей на пластине GaAs площадью 1 см^2. Основной трудностью является отвод тепла. Если энергия переключения $E = 1$ пДж, а время переключения $T = 100$ пс, то для $N = 10^4$ пикселей/см^2 тепловая нагрузка составляет $NE/T = 100$ Вт/см^2. Это приемлемо при хорошей технике теплоотвода. Устройство может выполнять 10^{14} битовых операций в секунду, что много в сравнении с электронными суперкомпьютерами (которые работают со скоростями около 10^{10} битовых операций в секунду).

Гибридные бистабильные оптические устройства

Бистабильные оптические системы, обсуждавшиеся до сих пор, являются полностью оптическими. Были предложены также гибридные частично электрические, частично оптические бистабильные системы, в которых задействованы электрические поля. Четыре примера таких систем показаны на рис. 23.77.

В первом примере (рис. 23.77, *а*) ячейка Поккельса помещена внутрь эталона Фабри—Перо. Выходящий свет детектируется фотоприемником и напряжение, пропорциональное принимаемой интенсивности, прикладывается к ячейке, так что изменение показателя преломления получается пропорциональным выходной интенсивности. Используя $LiNbO_3$ в качестве электрооптического материала, удалось достичь времен переключения 1 нс при мощности переключения ≈1 мкВт и энергии переключения ≈1 фДж. Был реализован также второй пример (рис. 23.77, *б*), представляющий собой интегрально-оптическую версию первого.

В третьем примере (рис. 23.77, *в*) электрооптический модулятор с использованием фазовой пластинки на основе ячейки Поккельса помещен между двумя скрещенными поляризаторами (см. подразд. 20.1.2). Снова регистрируется интенсивность выходящего света I_0 и пропорциональное ей напряжение прикладывается к ячейке. Коэффициент пропускания модулятора — нелинейная функция V

$$\mathcal{T} = \sin^2\left(\frac{\Gamma_0}{2} - \frac{\pi V}{2V_\pi}\right), \tag{23.33}$$

где Γ_0 и V_π — постоянные. Поскольку V пропорционально I_0, $\mathcal{T}(I_0)$ является немонотонной функцией и система проявляет бистабильность.

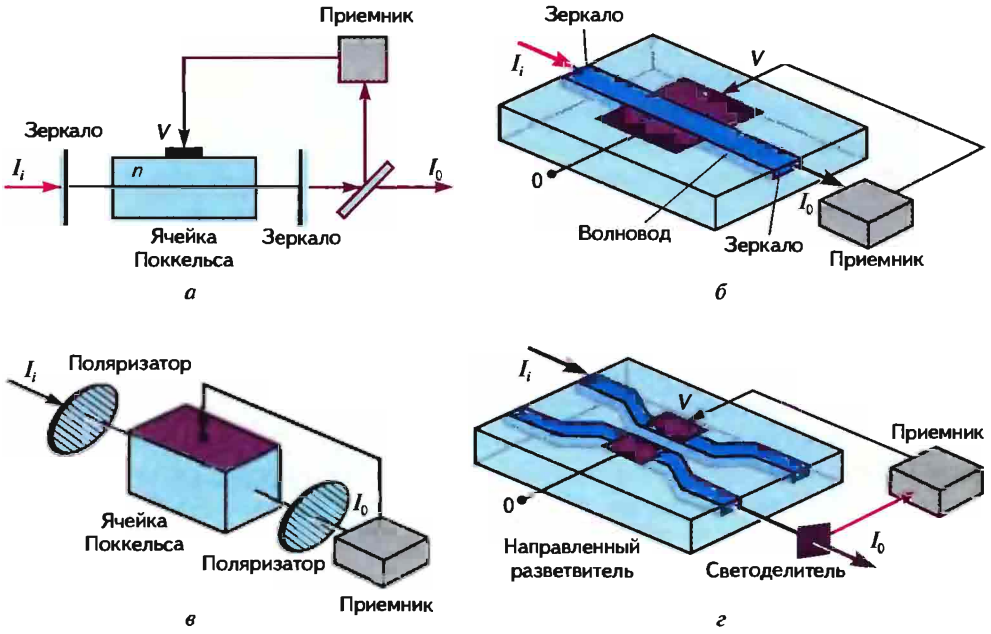


Рис. 23.77. Интерферометр Фабри—Перо, содержащий электрооптическую среду (ячейку Поккельса) (*а*). Выходная оптическая мощность детектируется и пропорциональное ей электрическое поле прикладывается к среде, меняя показатель преломления и, таким образом, меняя коэффициент пропускания интерферометра. Интегрально-оптическое исполнение (*б*). Бистабильная схема с ячейкой Поккельса в качестве управляемой фазовой пластинки (*в*). Бистабильная система на интегрально-оптическом направленном разветвителе, управляемом электрическим сигналом с выхода системы (*г*)

В четвертом примере (рис. 23.77, *г*) используется интегрально-оптический направленный разветвитель. Свет I_i входит через один волновод, а свет I_0 выходит через другой; отношение $\mathcal{T} = I_0/I_i$ — коэффициент передачи оптической мощности (см. подразд. 20.1.4). С помощью (20.23) получаем

$$\mathcal{T} = \left(\frac{\pi}{2}\right)^2 \operatorname{sinc}^2\left[\frac{1}{2}\sqrt{1+3\left(\frac{V}{V_0}\right)^2}\right], \qquad (23.34)$$

где V — приложенное напряжение; V_0 — постоянная. Бистабильность системы обеспечивается пропорциональностью V выходной интенсивности I_0 [см. упражнение 23.3, пункт (г)].

Пространственные модуляторы света (ПМС) можно использовать для построения массивов бистабильных элементов (рис. 23.78). Например, в жидкокристаллическом ПМС с оптической адресацией (см. подразд. 20.3.2) коэффициент отражения $\mathcal{R}$ каждого элемента является нелинейной функцией интен-

сивности света, освещающего его записывающую сторону. За счет обратных связей записанная интенсивность пропорциональна интенсивности I_0, отраженной от самого элемента, т. е. $\mathcal{R} = \mathcal{R}(I_0)$ и $I_0 = I_i\mathcal{R}(I_0)$, так что возникает бистабильное поведение. Разные точки поверхности прибора могут адресоваться раздельно, так что модулятор служит массивом бистабильных оптических элементов. Типичные времена переключения составляют десятки миллисекунд, а мощности переключения меньше 1 мкВт.

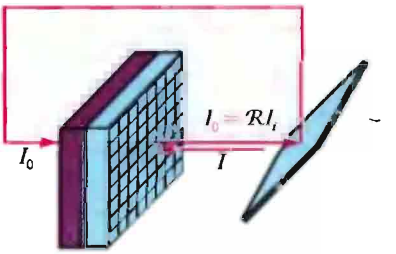


Рис. 23.78. Пространственный модулятор света с оптической адресацией действует как массив бистабильных оптических элементов. Коэффициент отражения оптического вентиля со стороны считывания (*справа*) в каждой точке поверхности является функцией $\mathcal{R} = \mathcal{R}(I_0)$ интенсивности I_0 на стороне записи (*слева*)

Электрооптические свойства полупроводников предоставляют много возможностей для создания бистабильных оптических устройств. Как упоминалось выше, важным примером является лазерный усилитель, в котором нелинейность неизбежно присутствует в виде насыщения усиления. Лазерные диодные усилители на InGaAsP использовались в качестве бистабильных переключателей с оптической энергией переключения менее 1 фДж и временем переключения менее 1 нс.

Устройство на самоиндуцированном электрооптическом эффекте

Еще одно полупроводниковое электрооптическое устройство, демонстрирующее бистабильность, основано на **самоиндуцированном электрооптическом эффекте** (self-electro-optic-effect device, **SEED**). Оно представляет собой p—i—n-фотодиод, в собственной области которого имеется гетероструктура с множественными квантовыми ямами (МКЯ) (рис. 23.79, *а*). Диод имеет обратное смещение, так что в МКЯ создается большое электрическое поле. За счет квантово-размерного эффекта Штарка (КРЭШ) (см. разд. 20.5) оптический коэффициент поглощения является нелинейной функцией $\alpha(V)$ напряжения V на МКЯ (рис. 23.79, *б*). Следовательно, оптический коэффициент пропускания $\mathcal{T}(V)$ — нелинейная функция V, зеркально отражающая $\alpha(V)$.

Бистабильное поведение SEED — результат механизма обратной связи, вносимой электрической цепью фотодиода, которая делает напряжение V зависящим от падающей оптической мощности P_i. Это происходит, поскольку поглощенный свет создает пропорциональный фототок $i_p = \Re(V)P_i$, который протекает по внешней цепи, вызывая спад напряжения V. Здесь $\Re(V)$ — токовая чувствительность, пропорциональная коэффициенту поглощения $\alpha(V)$. Например, если в цепи используется внешний источник напряжения V_0 с последова-

тельным сопротивлением нагрузки R_L, то $V = V_0 - i_p R_L = V_0 - R_L \Re(V) P_i$. Следовательно, прибор описывается двумя уравнениями

$$P_0 = \mathcal{T}(V) P_i; \tag{23.35}$$

$$P_i = \frac{1}{\Re(V)} \frac{V_0 - V}{R_L}, \tag{23.36}$$

где $\Re(V)$ пропорциональна $\alpha(V)$; $\mathcal{T}(V)$ является зеркальным отражением $\alpha(V)$; $\alpha(V)$ — нелинейная функция, показанная на рис. 23.79, *б*. Эти два уравнения определяют параметрическое соотношение между входной и выходной оптической мощностью, которое приводит к бистабильности, как схематически показано на рис. 23.79, *г*. Сопротивление в управляющей цепи можно заменить другим электронным устройством, таким как полевой транзистор (ПТ) или еще один SEED.

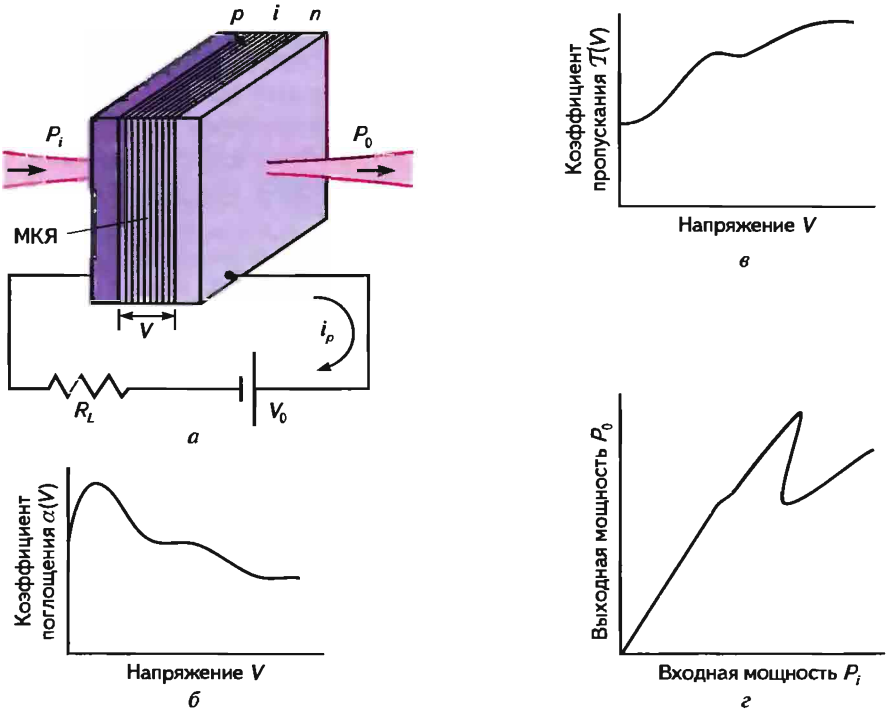


Рис. 23.79. Устройство с самоиндуцированным электрооптическим эффектом (SEED) — обратносмещенный МКЯ-фотодиод с оптически управляемым коэффициентом пропускания (*а*). Зависимость коэффициента поглощения от напряжения V за счет КРЭШ (*б*). Эта связь получается из рис. 20.36, *б*. Зависимость оптического коэффициента пропускания от напряжения V (*в*). Бистабильная связь между входной и выходной оптической мощностью (*г*)

Поскольку КРЭШ сильно зависит от длины волны (см. рис. 20.36, *б*), бистабильные характеристики SEED также имеют такую зависимость.

SEED работает без резонатора, поскольку обратная связь создается в электрической цепи оптически генерируемым фототоком. Но это не полностью оптическое устройство, поскольку оно включает электрические процессы в материале и внешней цепи и требует внешнего источника напряжения. SEED можно изготовлять в виде массивов, работающих на умеренно высоких скоростях (времена переключения — десятки наносекунд) и при очень низких мощностях.

Рекомендуемая литература

КНИГИ

El-Bawab T.S. Optical Switching. Springer-Verlag, 2006.

Pavesi L., Guillot G., eds. Optical Interconnects: The Silicon Approach. Springer-Verlag, 2006.

Ukita H. Micromechanical Photonics. Springer-Verlag, 2006.

Kabaciński W. Nonblocking Electronic and Photonic Switching Fabrics. Springer-Verlag, 2005.

Kawai S., ed. Handbook of Optical Interconnects. CRC Press/Taylor & Francis, 2005.

Gehani N. Bell Labs: Life in the Crown Jewel. Silicon Press, 2003.

Ramaswami R., Sivarajan K.N. Optical Networks: A Practical Perspective. Morgan Kaufmann, 2nd ed. 2002, Chapter 14.

Mouftah H.T., Elmirghani J.M.H., eds. Photonic Switching Technology: Systems and Networks. IEEE Press, 1999.

Tocci C.S., Caulfield H.J., eds. Optical Interconnection: Foundations and Applications. Artech, 1994.

Hinton H.S. An Introduction to Photonic Switching Fabrics. Plenum, 1993.

Midwinter J.E., ed. Photonics in Switching. Vol. 1: Background and Components. Academic Press, 1993.

Midwinter J.E., ed. Photonics in Switching. Vol. 2: Systems. Academic Press, 1993.

Islam M.N. Ultrafast Fiber Switching Devices and Systems. Cambridge University Press, 1992.

McAulay A.D. Optical Computer Architectures: the Application of Optical Concepts to Next Generation Computers. Wiley, 1991.

Arrathoon R., ed. Optical Computing: Digital and Symbolic. Marcel Dekker, 1989.

Gustafson T.K., Smith P.W., eds. Photonic Switching. Springer-Verlag, 1988.

Gibbs H.M. Optical Bistability: Controlling Light with Light. Academic, 1985.

Bowden C.M., Cifton M., Roble H.R., eds. Optical Bistability. Plenum, 1981.

СТАТЬИ

Holonyak, Jr. N., Feng M. The Transistor Laser. IEEE Spectrum. Vol. 43, № 2, 2006. P. 50—55.

Glesk I., Wang B.C., Xu L., Baby V., Prucnal P.R. Ultra-Fast All-Optical Switching in Optical Networks, in Progress in Optics. E. Wolf, ed. Elsevier. Vol. 45, 2003. P. 53—117.

Huang D., Sze T., Landin A., Lytel R., Davidson H.L. Optical Interconnects: Out of the Box Forever? IEEE Journal of Selected Topics in Quantum Electronics. Vol. 9, 2003. P. 614—623.

Keeler G.A., Nelson B.E., Agarwal D., Debaes C., Helman N.C., Bhatnagar A., Miller D.A.B. The Benefits of Ultrashort Optical Pulses in Optically Interconnected Systems. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 9, 2003. P. 477—485.

Potasek M.J. All-Optical Switching for High Bandwidth Optical Network. Optical Networks Magazine. Vol. 3, № 6, 2002. P. 30—43.

Lin L.Y., Goldstein E.L. Opportunities and Challenges for MEMS in Lightwave Communications. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, 2002. P. 163—172.

Special issue on arrayed grating routers/WDM MUX/DEMUXs and related applications/uses. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 8, № 6, 2002.

Forbes M., Gourlay J., Desmulliez M. Optically Interconnected Electronic Chips: A Tutorial and Review of the Technology. Electronics & Communication Engineering Journal. Vol. 13, 2001. P. 221—232.

Miller D.A.B. Rationale and Challenges for Optical Interconnects to Electronic Chips. Proceedings of the IEEE. Vol. 88, 2000. P. 728—749.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Special issue on optical interconnections for digital systems. Proceedings of the IEEE. Vol. 88, № 6, 2000.

Yao S., Mukherjee B., Dixit S. Advances in Photonic Packet Switching: An Overview. IEEE Communications Magazine. Vol. 38, № 2, 2000. P. 84—94.

Bishop D.J., Aksyuk V.A. Optical MEMS Answer High-Speed Networking Requirements. Electronic Design. April 5, 1999. P. 85—99.

Krishnamoorthy A.V., Chirovsky L.M.F., Hobson W.S., Leibenguth R.E., Hui S.P., Zydzik C.J., Goossen K.W., Wynn J.D., Tseng B.J., Lopata J., Walker J.A., Cunningham J.E., D'Asaro L.A. Vertical-Cavity Suiface-Emitting Lasers Flip-Chip Bonded to Gigabit-per-Second CMOS Circuits. IEEE Photonics Technology Letters. Vol. 11, № 1, 1999. P. 128—130.

Issue on smart photonic components, interconnects, and processing. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 5, № 2, 1999.

Miller D.A.B. Physical Reasons for Optical Interconnection. International Journal of Optoelectronics. Vol. 11, 1997. P. 155—168.

Krishnamoorthy A.V., Miller D.A.B. Scaling Optoelectronics-VLSI Circuits into the 21st Century: A Technology Roadmap. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 2, № 1, 1996. P. 55—76.

Marrakchi A., ed. Selected Papers on Photonic Switching. SPIE Optical Engineering Press (Milestone Series. Vol. 121), 1996.

Tsai C.S. Integrated Acoustooptic and Magnetooptic Devices for Optical Information Processing. Proceedings of the IEEE. Vol. 84, 1996. P. 853—869.

Reinisch R., Vitrant G. Optical Bistability. Progress in Quantum Electronics. Vol. 18, 1994. P. 1—38.

Kalman R.F., Kazovsky L.G., Goodman J.W. Space Division Switches Based on Semiconductor Optical Amplifiers. IEEE Photonics Technology Letters. Vol. 4, 1992. P. 1048—1051.

Miller D.A.B. Quantum-Well Self-Electro-Optic Effect Devices. Optical and Quantum Electronics. Vol. 22, 1990. P. S61—S98.

Stegeman G.I., Wright E.M. All-Optical Waveguide Switching. Optical and Quantum Electronics. Vol. 22, 1990. P. 95—122.

Issue on optical interconnects. Applied Optics: Information Processing. Vol. 29, № 8, 1990.

Issue on optical interconnections and networks. SPIE Proceedings. Vol. 1281, 1990.

Issue on nonlinear optical materials and devices for photonic switching. SPIE Proceedings. Vol. 1216, 1990.

Issue on optical interconnects in the computer environment. SPIE Proceedings. Vol. 1178, 1990.

Silberberg Y. Photonic Switching Devices. Optics News. Vol. 15, № 2, 1989. P. 7—12.

Issue on photonic switching. IEEE Journal of Selected Areas in Communicatiuns. Vol. 6, № 7, 1988.

Midwinter J.E. Digital Optics, Smart Interconnect or Optical Logic? Part 1. Physics in Technology. Vol. 19, 1988. P. 101—108.

Midwinter J.E. Digital Optics, Smart Interconnect or Optical Logic? Part 2. Physics in Technology. Vol. 19, 1988. P. 153—165.

Hartman D.H. Digital High Speed Interconnects: A Study of the Optical Alternative. Optical Engineering. Vol. 25, 1986. P. 1086—1102.

Haugen P.R., Rychnovsky S., Husain A., Hutcheson L.D. Optical Interconnects for High Speed Computing. Optical Engineering. Vol. 25, 1986. P. 1076—1085.

Sawchuk A.A., Jenkins B.K. Dynamic Optical Interconnections for Parallel Processors. SPIE Proceedings. Vol. 625, 1986. P. 143—153.

Su S.F., Jou L., Lenart J. A Review on Classification of Optical Switching Systems. IEEE Communications Magazine. Vol. 24, № 5, 1986. P. 50—55.

Miller D.A.B., Chemla D.S., Damen T.C., Wood T.H., Burrus, Jr. C.A., Gossard A.C., Wiegmann W. The Quantum Well Self-Electrooptic Effect Device: Optoelectronic Bistability and Oscillation and Self-Linearized Modulation. IEEE Journal of Quantum Electronics. Vol. 21, 1985. P. 1462—1476.

Goodman J.W., Leonberger F.I., Kung S.Y., Athale R.A. Optical Interconnections for VLSI Systems. Proceedings of the IEEE. Vol. 72, 1984. P. 850—866.

Lugiato L.A. Theory of Optical Bistability. In Progress in Optics. E. Wolf, ed. North-Holland. Vol. 21, 1984.

Smith P.W. Applications of All-Optical Switching and Logic. Philosophical Transactions of the Royal Society of London. Vol. A313, 1984. P. 349—355.

Smith P.W., Tomlinson W.J. Bistable Optical Devices Promise Subpicosecond Switching. IEEE Spectrum. Vol. 18, № 6, 1981. P. 26—33.

Cutrona L.J., Leith E.N., Palermo C.J., Porcello L.J. Optical Data Processing and Filtering Systems. IRE Transactions on Information Theory. Vol. IT-6, 1960. P. 386—400.

Задачи

К РАЗДЕЛУ 23.1

1. **Голограмма межсоединения для конформного отображения.** Спроектируйте голограмму, реализующую геометрическое преобразование

$$x' = \psi_x(x, y) = \ln\sqrt{x^2 + y^2};$$

$$y' = \psi_y(x, y) = \operatorname{arctg}\left(\frac{y}{x}\right).$$

Это преобразование от декартовых координат к полярным с последующим логарифмическим преобразованием радиальной координаты $r = (x^2 + y^2)^{1/2}$. Найдите выражение для фазовой функции $\varphi(x, y)$ требуемой голограммы.

К РАЗДЕЛУ 23.2

1. **Каскад «мультиплексор—демультиплексор» на ИМЦ.** Три интерферометра Маха—Цендера (ИМЦ) расположены каскадом, как показано на рис. 23.22, для мультиплексирования—демультиплексирования четырех каналов с разделением по длине волны $\Delta\lambda = 0,2$ нм и центральной длиной волны 1550 нм. Определите необходимые разности хода Δd в каждом интерферометре, если показатель преломления $n = 2,3$.

2. **WGR-демультиплексор.** Маршрутизатор на решетке волноводов (WGR) (см. рис. 23.24) используется для демультиплексирования четырех каналов с разделением по длине волны с разностью длин волн $\Delta\lambda = 0,2$ нм и центральной длиной волны 1550 нм. Определите параметр разности хода Δd_b, который должен вноситься звездообразным разветвителем, если его показатель преломления $n = 2,3$.

3. **WGR как 2 × 2 маршрутизатор с разделением по длине волны.** WGR имеет конфигурацию 2 × 2 маршрутизатора с разделением по длине волны. Входной порт 1 имеет два канала на длинах волн λ_1 и λ_2, входной порт 2 также имеет два канала на длинах волн λ_3 и λ_4. Спроектируйте маршрутизатор, который перераспределяет входные длины волн между двумя выходными каналами, а именно, направляет λ_1 и λ_3 в выходной порт 1, а λ_2 и λ_4 в выходной порт 2. Запишите условия маршрутизации через четыре оптические разности хода Δd_{11}, Δd_{12}, Δd_{21}, Δd_{22} многолучевых интерферометров, соединяющих каждый из входных портов с каждым из выходных.

К РАЗДЕЛУ 23.3

1. **Потери мощности и перекрестные помехи.** 4 × 4-коммутатор можно реализовать с помощью пяти 2 × 2-коммутаторов. Если каждый из них вносит потери мощности 0,5 дБ и перекрестные помехи −30 дБ, определите наихудшие значения потерь мощности и перекрестных помех для всего 4 × 4-коммутатора.

2. **Перекрестный коммутатор с ИМЦ.** Электрооптический интерферометр Маха—Цендера используется в качестве перекрестного коммутатора. Приложение напряжения $V = V_\pi$ к электрооптическому материалу вносит фазовый сдвиг π. Если коммутатор установлен в параллельное состояние при $V = 0$, какое напряжение нужно приложить для переключения коммутатора в перекрестное состояние? Определите перекрестные помехи (в дБ), связанные с 1-%-й ошибкой в величине приложенного напряжения.

3. **Коммутатор межинтервального обмена (TSI).** Как показано на рис. 23.59, коммутатор межинтервального обмена (TSI) можно реализовать посредством пятиступенчатого процесса: преобразование временно́го разделения в пространственное, внесение временны́х задержек, пространственная коммутация, внесение временны́х задержек, преобразование пространственного разделения во временно́е. Сконструируйте альтернативную схему с использованием программируемых линий задержки, показанных на рис. 23.60.

К РАЗДЕЛУ 23.4

1. **Оптическая логика**. На рис. 23.67 показано, как можно использовать нелинейный порогообразующий оптический элемент для построения логической ячейки «И». Покажите, как аналогичная система может быть использована в качестве логической ячейки «НЕ И», «ИЛИ» и «НЕ ИЛИ». Можно ли так реализовать логическую операцию «ИСКЛЮЧИТЕЛЬНОЕ ИЛИ»? Можно ли использовать такую же систему для получения «ИЛИ» от N двоичных входов?

2. **Бистабильный интерферометр**. Кристалл с оптическим эффектом Керра помещен в одно из плеч интерферометра Маха—Цендера. Прошедшая интенсивность I_0 направляется назад и освещает кристалл. Покажите, что коэффициент пропускания по интенсивности системы равен

$$\frac{I_0}{I_i} = \mathcal{T}(I_0) = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{\pi I_0}{I_\pi} + \varphi\right),$$

где I_π и φ — постоянные. Предполагая, что $\varphi = 0$, изобразите графически зависимость I_0 от I_i и выведите выражение для максимального дифференциального усиления $\mathrm{d}I_0/\mathrm{d}I_i$.

ГЛАВА 24 ВОЛОКОННО-ОПТИЧЕСКИЕ СИСТЕМЫ СВЯЗИ

До середины 1970-х гг. практически все системы связи опирались на передачу информации по электрическим кабелям или использовали радиочастотное и микроволновое излучение, распространяющееся в свободном пространстве. Казалось бы, использование света было бы более естественным выбором для передачи сообщений, поскольку, в отличие от электричества и радиоволн, его не нужно было открывать. Причин для задержки в развитии такой технологии можно назвать две:

1) трудность изготовления источника света, который можно было бы быстро включать и выключать, тем самым кодируя информацию с высокой скоростью;

2) тот факт, что свет легко задерживается непрозрачными объектами, таким как дымка, облака, туман, дым.

В отличие от радиоволн и микроволнового излучения, свет редко подходит для связи через свободное пространство.

Тем не менее световая связь нашла свое место и в нынешней реальности является предпочтительной технологией во многих приложениях, включая передачу данных, голосовую и видеосвязь, телеметрию, связь на коротких расстояниях и локальные сети, а также дальнюю связь и Интернет-траффик. Оптические технологии предоставляют пользователю огромную пропускную способность, большое расстояние между повторителями, защиту от электромагнитного вмешательства, сравнительную легкость установки. Оптическая передача — единственная технология, отвечающая обширным и экспоненциально растущим запросам глобальной связи, и сейчас она достигает индивидуальных жилых домов через широкополосные системы «волокно в дом» (fiber-to-the-home, FTTH).

Корни впечатляющих успехов оптической связи — в двух критических изобретениях фотоники: это светоизлучающий диод (СИД) и оптическое волокно с низкими потерями как переносчик света. Довольно рано стали доступными подходящие приемники света, хотя в течение последнего десятилетия их функциональные характеристики были разительно улучшены. Интерес к оптической связи первоначально возник после изобретения лазера в 1960-х гг. Однако в первом поколении оптических волоконных систем связи в качестве источников использовались СИД, да и сейчас они применяются во многих действующих локальных системах связи. Тем не менее большинство оптических систем связи (таких как системы дальней связи на одномодовых оптических волокнах

и системы ближней связи через свободное пространство) существенно выигрывает от использования высокой оптической мощности, малой спектральной ширины и высокой направленности, обеспечиваемых лазером.

О данной главе

Эта глава содержит введение в оптические системы связи и сети. Двухточечное коммуникационное соединение состоит из трех основных элементов, как показано на рис. 24.1: компактного источника света, модулируемого электрическим сигналом, оптического волокна с низкими потерями и малой дисперсией и фотоприемника, преобразующего оптический сигнал обратно в электрический. Эти оптические элементы подробно обсуждались в гл. 17, 9 и 18 соответственно. Оптические усилители также доказали свою необходимость в волоконных системах; они обсуждались в гл. 14. Чтобы сделать данную главу самодостаточной, в разд. 24.1 приводится краткая сводка необходимых свойств волокон, источников, приемников и усилителей в общем контексте разработки, действия и рабочих характеристик оптической связи. Другие оптические аксессуары, такие как сращивания оптических волокон, коннекторы, разветвители, коммутаторы и устройства мультиплексирования, также важны для успешной работы оптической связи и сетей. Принципы построения некоторых из них описаны в гл. 23 и других частях настоящей книги.



Рис. 24.1. Оптическая волоконная система связи. Электрический сигнал преобразуется в оптический сигнал (Э/О) посредством модуляции источника света. Оптический сигнал передается по волокну к приемнику. Приемник преобразует оптический сигнал обратно в электрический (О/Э). Для длинных волокон могут использоваться оптические усилители, поддерживающие ослабляемый оптический сигнал. В других схемах несколько оптических соединений располагаются каскадом, образуя более длинное соединение, с использованием промежуточного электрического усиления и восстановления сигнала между смежными соединениями. Такие блоки называют **регенераторами** или **ретрансляторами**

В разд. 24.2 вводятся основные принципы проектирования дальних цифровых и аналоговых оптических систем связи с использованием модуляции интенсивности. Определена максимальная длина волокна, которое можно использовать для передачи данных с заданной скоростью и уровнем качества. Качество падает, если скорость передачи данных превышает ширину полосы волокна или если принимаемая мощность меньше, чем обнаружительная способность приемника (так что сигнал невозможно распознать на фоне шумов). Затем в разд. 24.3 дается введение в системы модуляции и мультиплексирования, используемые в волоконно-оптической связи.

Волоконно-оптические сети — коммуникационные соединения, связывающие множество пользователей, разбросанных по некоторой географической области, и управляемые набором маршрутизаторов и коммутаторов. В разд. 24.4 дается введение в такие сети, включая те, в которых используется **мультиплексирование с разделением по длине волны (WDM)**.

В *когерентных* оптических системах связи, которые вводятся в разд. 24.5, свет используется не как носитель управляемой мощности, а как электромагнитная волна с управляемой амплитудой, фазой или частотой. Эти системы представляют собой естественное обобщение радиочастотных и микроволновых систем связи на оптический диапазон. Они дают существенное увеличение чувствительности приема, позволяют увеличить расстояние между ретрансляторами и повысить скорость передачи данных.

24.1. ВОЛОКОННО-ОПТИЧЕСКИЕ КОМПОНЕНТЫ

24.1.1. Оптические волокна

Оптическое волокно — цилиндрический диэлектрический волновод, сделанный из материала с низкими потерями, обычно из плавленого кварцевого стекла высокой химической чистоты. У простейшего оптического волокна со ступенчатым профилем показателя преломления сердцевина имеет постоянный показатель преломления, немного больший, чем у оболочки (внешней среды), так что свет направляется вдоль оси волокна благодаря полному внутреннему отражению.

Как описано в гл. 9, прохождение света через волокно можно исследовать, рассматривая траектории лучей внутри сердцевины. В соответствии с более полным анализом, основанным на электромагнитной теории, свет распространяется по волокну в виде мод, каждая из которых представляет собой волну с определенным пространственным распределением, поляризацией, постоянной распространения, групповой скоростью и коэффициентом ослабления. Существует соответствие между модами и лучами, бегущими внутри сердцевины по определенным траекториям.

Волокно со ступенчатым профилем показателя преломления характеризуется радиусом сердцевины a, показателями преломления сердцевины и оболочки n_1 и n_2, а также относительным перепадом показателя преломления $\Delta = (n_1 - n_2)/n_1$, который обычно очень мал ($\Delta = 0{,}001$—$0{,}02$). Лучи света, образующие с осью волокна углы, меньшие дополнительного к критическому

$$\bar{\theta}_{\text{кр}} = \arccos\left(\frac{n_2}{n_1}\right),$$

направляются по сердцевине в результате многократного полного внутреннего отражения на границе сердцевины и оболочки. Угол $\bar{\theta}_{\text{кр}}$ внутри волокна соот-

ветствует входной угловой апертуре $\theta_a = \arcsin(\mathrm{NA})$ для лучей, входящих в сердцевину из воздуха, где NA — числовая апертура,

$$\boxed{\mathrm{NA} = \sin\theta_a = \sqrt{n_1^2 - n_2^2} \approx n_1\sqrt{2\Delta}.} \qquad (24.1)$$

Числовая апертура

Многомодовые волокна (MMB)

Число направляемых мод M определяется параметром

$$V = 2\pi\frac{a}{\lambda_0}\mathrm{NA},$$

где a/λ_0 — отношение радиуса сердцевины a к длине волны λ_0. В волокне с $V \gg 1$ существует большое число мод $M \approx V^2/2$. Поскольку моды распространяются с различными групповыми скоростями, это приводит к расплыванию импульса, которое нарастает линейно по длине волокна. Этот эффект называется **модовой дисперсией**. Когда импульс света проходит по волокну расстояние L, он превращается в последовательность импульсов, центры которых отстоят друг от друга на время модовой задержки, как показано на рис. 24.2. Составной импульс имеет приближенную среднеквадратичную ширину

$$\boxed{\sigma_{\text{мод}} \approx \frac{\Delta}{2c_1}L,} \qquad (24.2)$$

Время отклика (ступенчатое MMB)

где $c_1 = c_0/n_1$. Следовательно, желательно использовать волокна с малым Δ. Например, если $n_1 = 1{,}46$ и $\Delta = 0{,}01$, то время отклика на километр примерно равно $\Delta/2c_1 \approx 24$ нс/км. Для волокна длиной 100 км импульс расплывается до ширины 2,4 мкс.

Модовую дисперсию можно уменьшить также с использованием **градиентных волокон** (graded-index, GRIN). В таких волокнах показатель преломления сердцевины плавно меняется от максимального значения n_1 в центре до минимального n_2 на границе сердцевины и оболочки. Лучи следуют искривленным траекториям, длина которых меньше, чем в ступенчатом волокне. Кратчайшее расстояние проходит осевой луч с наименьшей фазовой скоростью (при наибольшем показателе преломления), а наклонные лучи проходят большее расстояние с большими фазовыми скоростями (при меньшем показателе преломления), так что времена задержки приблизительно выравниваются. Если профиль показателя преломления, имеющий приблизительно параболическую форму, подобран оптимально, то скорость расплывания импульса (пс/км) равна скорости в эквивалентном ступенчатом волокне, умноженной на $\Delta/2$. Например, при $\Delta = 0{,}01$ расплывание импульса уменьшается в 500 раз. На практике, однако, этот показатель редко достигается из-за трудности изготовления волокон с идеальным профилем показателя преломления.

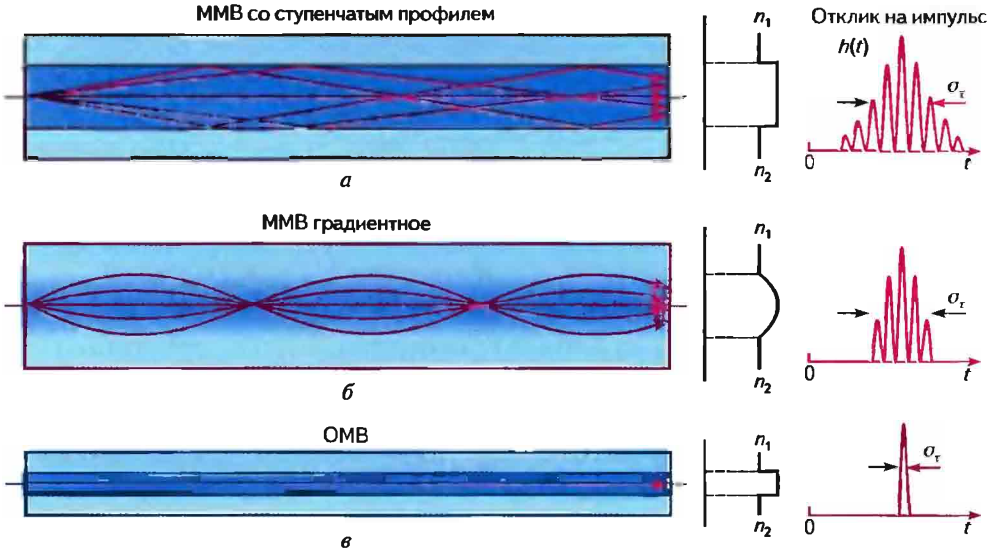


Рис. 24.2. Многомодовое волокно (ММВ) со ступенчатым профилем показателя преломления (*а*): сравнительно большой радиус сердцевины; однородный показатель преломления в сердцевине и оболочка; большое расплывание импульса из-за модовой дисперсии. ММВ с градиентным показателем преломления (GRIN): показатель преломления в оболочке меняется плавно; число мод меньше; расплывание импульса из-за модовой дисперсии понижено (*б*). Одномодовое волокно (ОМВ): малый радиус сердцевины; нет модовой дисперсии; уширение импульса происходит только из-за материальной и волноводной дисперсии (*в*)

Одномодовые волокна (ОМВ)

Если радиус сердцевины и числовая апертура волокна со ступенчатым профилем показателя преломления достаточно малы, так что $V < 2{,}405$, возможна только одна мода и волокно называется **одномодовым** (ОМВ). Одним из преимуществ одномодовых волокон является исключение расплывания импульсов, обусловленного модовой дисперсией. Тем не менее расплывание существует, так как начальный импульс имеет конечную спектральную ширину, а групповая скорость (и, следовательно, время задержки) зависит от длины волны. Этот эффект называется **хроматической дисперсией**.

Хроматическая дисперсия имеет два источника: **материальную дисперсию**, возникающую из-за зависимости показателя преломления от длины волны, и волноводную дисперсию, являющуюся следствием зависимости групповой скорости моды от отношения радиуса сердцевины к длине волны. Материальная дисперсия обычно больше, чем волноводная.

Короткий оптический импульс со спектральной шириной σ_λ расплывается до временной ширины

$$\boxed{\sigma_{\text{вол}} = |D|\sigma_\lambda L,} \qquad (24.3)$$

Время отклика (ОМВ)

пропорциональной пройденному расстоянию L (км) и ширине линии источника σ_λ (нм), где D — коэффициент дисперсии [пс/(км · нм)]. Параметр D учитывает сочетание волноводной и материальной дисперсии. Для слабонаправляющих волокон ($\Delta \ll 1$) D можно представить как сумму $D_\lambda + D_w$ вкладов материальной и волноводной дисперсии.

Например, для ОМВ с источником света со спектральной шириной $\sigma_\lambda = 1$ нм (типичной для одномодовых лазеров) и коэффициентом дисперсии $D = 1$ пс/(км · нм) (при рабочей длине волны вблизи $\lambda_0 = 1300$ нм с минимальной волноводной дисперсией) время отклика, даваемое формулой (24.3), составляет $\sigma_\tau / L = 1$ пс/км. Волокно длиной 100 км имеет время отклика 100 пс.

Геометрия, профиль показателя преломления и уширение импульса в многомодовом волокне со ступенчатым и градиентным профилем показателя преломления и в одномодовом волокне схематически сравниваются на рис. 24.2.

Материальная дисперсия и ослабление

Зависимость коэффициентов ослабления от длины волны для волокон из плавленого кварцевого стекла показана на рис. 24.3. По мере роста длины волны с выходом из видимого диапазона ослабление снижается до минимума примерно в 0,3 дБ/км при $\lambda_0 = 1300$ нм, слегка растет при 1,4 мкм из-за поглощения ионами ОН и снова падает до своего абсолютного минимума ≈0,16 дБ/км при $\lambda_0 = 1550$ нм, после чего быстро растет. Недавно были разработаны волокна с подавлением поглощения, обусловленного ионами ОН.

Зависимость коэффициента дисперсии D_λ от длины волны для плавленого кварцевого стекла также показана на рис. 24.3. Он меняется от отрицательных значений на коротких волнах к положительным значениям на длинных волнах и проходит через ноль при $\lambda_0 \approx 1312$ нм. В среде с отрицательной дисперсией коротковолновые компоненты импульса распространяются медленнее, чем длинноволновые. Такая дисперсия называется **нормальной**. Противоположный случай (**аномальная дисперсия**) имеет место в среде с положительным коэффициентом дисперсии (см. разд. 5.6). Хотя знак коэффициента дисперсии не влияет на скорость уширения импульса, он может играть важную роль в распространении импульса через среды, состоящие из каскада материалов с разными знаками, как описано в подразд. 24.2.4 (см. также разд. 22.3).

Волокна с модифицированной дисперсией (DSF)

Как описано в подразд. 9.3.2, в усовершенствованных вариантах одномодовых волокон используются градиентные сердцевины со специальным профилем показателя преломления, подобранным так, чтобы полный коэффициент хроматической дисперсии D имел желательные величины на заданных длинах волн либо зависимость от длины волны, удобную для волоконных систем связи, как в следующих примерах.

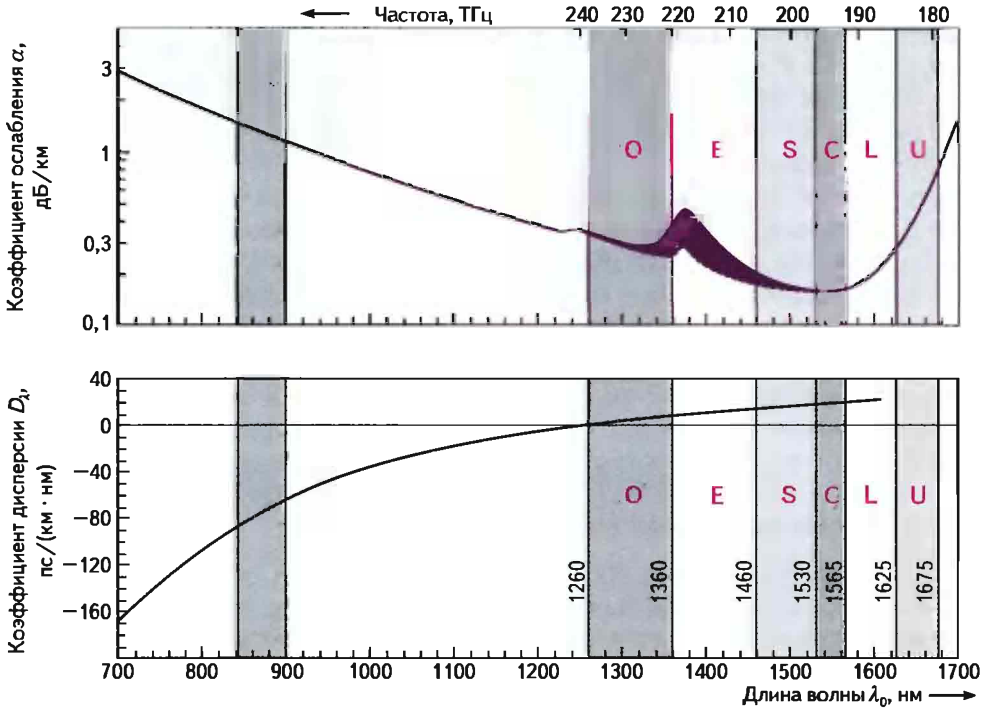


Рис. 24.3. Зависимость коэффициента ослабления α и коэффициента материальной дисперсии D_λ от длины волны для волокон из плавленого кварцевого стекла. Штриховой линией показано ослабление для волокон с подавленным поглощением OH. Отмечены три спектральные области: область с центром на 870 нм, которая использовалась в ранних системах, имеет $\alpha = 1{,}5$ дБ/км и $D_\lambda = -80$ пс/(км · нм); O (оригинальная) полоса с центром на 1310 нм, для которой $\alpha = 0{,}3$ дБ/км, а дисперсия минимальна C (стандартная) полоса, для которой ослабление минимально ($\alpha = 0{,}16$ дБ/км), а $D_\lambda = +80$ пс/(км · нм). Три дополнительные полосы используются в системах с мультиплексированием по длине волны (WDM): E = extended (расширенная), S = short (коротковолновая), L = long (длинноволновая), U = ultra-long (ультрадлинноволновая)

• В **волокнах со смещенной дисперсией (DSF)** D исчезает при $\lambda_0 = 1550$ нм, где ослабление минимально, а не при 1312 нм. В волокнах с **ненулевой смещенной дисперсией (NZ-DSF)** коэффициент D существенно понижен в области 1500—1600 нм, но не равен нулю. Небольшая дисперсия может быть полезна для снижения нелинейных искажений, испытываемых короткими импульсами высокой интенсивности. Зависимость D от длины волны в DSF и NZ-DSF показана на рис. 24.4 (см. также рис. 9.28, *а*).

• В **волокнах со сглаженной дисперсией** D исчезает на двух длинах волн и понижено в промежуточной области (см. рис. 9.28, *б*).

• В **волокнах с компенсирующей дисперсией (DCF)** коэффициент D пропорционален тому, который имеется у обычных волокон со ступенчатым профилем показателя преломления, в широкой области длин волн, но имеет проти-

воположный знак. Короткое волокно с большим коэффициентом дисперсии обратного знака можно использовать для компенсации расплывания импульса, вносимого обычным волокном большой длины (см. рис. 9.28, *в*).

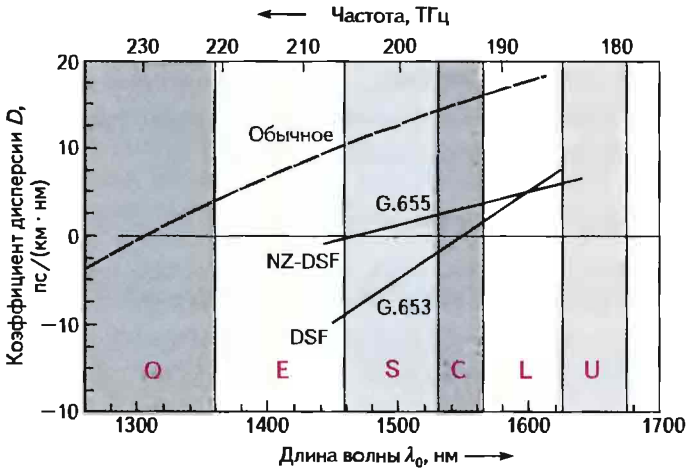


Рис. 24.4. Зависимость коэффициента хроматической дисперсии D от длины волны для обычного волокна и примеров волокна со смещенной дисперсией (DSF) и ненулевой смещенной дисперсией (NZ-DSF). Маркировка G.653 и G.655 соответствует спецификации Международного Союза Телекоммуникаций ITU

Другие типы волокон с измененной дисперсией включают **дырчатые волокна** (holey fibers) и **фотонно-кристаллические волокна** (ФКВ), описанные в разд. 9.4. В этих волокнах над хроматической дисперсией преобладает волноводная дисперсия, сильно зависящая от геометрии отверстий. Можно достичь сглаживания дисперсии в широких областях длин волн, а также смещения дисперсии на длины волн ниже, чем длина волны нулевой материальной дисперсии. Дырчатое волокно можно спроектировать так, чтобы оно действовало как одномодовый волновод в широком диапазоне длин волн (бесконечно одномодовое волокно). В волокнах с полой сердцевиной и оболочкой с отверстиями, образующими периодическую структуру, свет направляется вдоль сердцевины благодаря отражению от окружающей фотонно-кристаллической среды. Поскольку свет проходит по полой сердцевине, он испытывает меньшие потери и пониженное влияние нелинейных эффектов.

Поляризационно-модовая дисперсия

Другой механизм расплывания импульсов, известный как **поляризационно-модовая дисперсия (ПМД)**, обусловлен случайными изменениями анизотропии волокна, вносимыми за счет изменения окружающих условий по его длине. Случайные изменения величины и ориентации двулучепреломления вносят дифференциальные задержки между двумя поляризационными модами

и, как описано в подразд. 9.3.2, среднеквадратичное уширение импульса за счет ПМД пропорционально квадратному корню из длины волокна:

$$\sigma_{\text{ПМД}} = D_{\text{ПМД}}\sqrt{L}, \tag{24.4}$$

Поляризационно-модовая дисперсия

где $D_{\text{ПМД}}$ — дисперсионный параметр, обычно лежащий в интервале от 0,1 до 1 пс/$\sqrt{\text{км}}$. ПМД становится важной при высоких скоростях передачи данных, когда другие виды дисперсии скомпенсированы.

Нелинейные оптические эффекты

В волокнах из плавленого кварцевого стекла наблюдается два вида нелинейных эффектов: нелинейность третьего порядка, лежащая в основе оптического эффекта Керра, и нелинейное неупругое рассеяние, включающее вынужденное комбинационное (рамановское) рассеяние и рассеяние Мандельштама—Бриллюэна. При передаче оптических импульсов высокой мощности по одномодовому волокну с малой площадью поперечного сечения оптическая интенсивность может быть достаточно большой для того, чтобы указанные нелинейные взаимодействия могли вызвать множество вредных эффектов, искажающих сигнал в системах связи:

- *Самомодуляция фазы* (СМФ) — форма нелинейной дисперсии, обусловленная оптическим эффектом Керра (слабой зависимостью показателя преломления, а значит фазовой скорости, от оптической интенсивности, описанной в подразд. 21.3.1). Поскольку различные части оптического импульса распространяются с различными скоростями, возникает расплывание импульса (см. подразд. 22.3.2). Результатом оптического эффекта Керра могут быть также перекрестные помехи между встречными волнами в двухсторонних системах связи.
- *Кросс-модуляция фазы* (КМФ) возникает за счет нелинейного смешения волн, в результате которого фазовая скорость волны с одной частотой начинает зависеть от интенсивностей волн на других частотах, одновременно распространяющихся в том же волокне (см. подразд. 21.3.3). В системах мультиплексирования с разделением по длине волны (WDM) КМФ может вызывать серьезные перекрестные помехи между различными каналами.
- *Четырехволновое смешение* (ЧВС) также связано с нелинейными эффектами третьего порядка (см. подразд. 21.3.4). Оно вызывает перекрестные помехи между четырьмя волнами с различной длиной волны, которые одновременно распространяются в одном волокне, за счет обмена энергией между волнами. ЧВС вносит зависящее от интенсивности усиление/ослабление в каналы системы WDM.
- *Вынужденное комбинационное рассеяние* (ВКР) *и вынужденное рассеяние Мандельштама—Бриллюэна* (ВРМБ) — неупругие процессы рассеяния при взаимодействии света с молекулярными колебаниями или акустическими колебаниями среды. В этих процессах две оптические волны разных частот взаимодей-

ствуют через посредство молекулярной колебательной моды (ВКР) или акустической волны (ВРМБ) (см. подразд. 13.5.3, 14.3.4 и 15.3.1). Такие взаимодействия также приводят к нежелательным перекрестным помехам между каналами системы WDM.

Нелинейные свойства волокна можно использовать для полезных приложений в системах связи. Нелинейная дисперсия за счет СМФ может быть подобрана так, чтобы компенсировать хроматическую дисперсию волокна. В результате получаются нерасплывающиеся импульсы, известные как оптические солитоны (подразд. 22.5.2). Нелинейные взаимодействия также можно использовать для обеспечения полезного усиления за счет ЧВС или ВКР. Оптические ВКР-усилители описаны в подразд. 14.3.4.

24.1.2. Источники для оптических передатчиков

Основные требования к источникам света для оптических систем связи зависят от типа проектируемого приложения (дальняя связь, локальная сеть и т. д.). Ниже перечислены основные показатели.

- *Мощность.* Мощность источника должна быть достаточной для того, чтобы после передачи по волокну сигнал можно было детектировать с требуемой точностью.
- *Скорость.* Должна быть возможность модулировать мощность источника с частотой, желательной для сообщения информации.
- *Ширина линии.* Источник должен иметь узкую спектральную линию, чтобы минимизировать влияние хроматической дисперсии в волокне.
- *Шум.* Случайные флуктуации мощности источника должны быть исключены, особенно для когерентных систем связи.
- *Другие показатели.* Другими важными показателями являются прочность, нечувствительность к изменениям окружающей среды, например температуры, надежность, низкая стоимость и длительный срок службы.

В качестве источников для волоконно-оптических систем связи используются как светоизлучающие диоды (СИД), так и лазерные диоды (ЛД). Эти приборы обсуждались в гл. 17.

Светоизлучающие диоды (СИД) изготовляются в двух базовых конструкциях: с поверхностным и торцевым излучателем. Преимущества *СИД с поверхностным излучателем* состоят в прочности, надежности, низкой стоимости, большом сроке службы и простоте конструкции. Основные ограничения их использования связаны со сравнительно большой шириной линии излучения, которая может превышать 100 нм в диапазоне 1300—1600 нм (см. рис. 17.60). При работе на максимальной мощности возможны частоты модуляции до 100 МБ/с, однако при понижении мощности доступны и более высокие скорости (до 500 МБ/с). *Диод с торцевым излучателем* имеет структуру, аналогичную лазерному диоду без механизма обратной связи. Он дает большую мощность на выходе при сравнительно более узкой линии излучения за счет усложнения конструкции.

Лазерные диоды обладают преимуществом высокой мощности (десятки Вт), высокой скорости (многие ГБ/с), узкой линии излучения (десятки МГц) и простоты согласования с одномодовыми волокнами. Однако они чувствительны к изменениям температуры. Многомодовые лазерные диоды страдают от шума токоразделения, представляющего собой случайное распределение лазерной мощности между модами. В результате хроматической дисперсии в волокне это приводит к случайным флуктуациям интенсивности и искажению формы передаваемых импульсов. Лазерные диоды подвержены также частотному чирпу — изменению частоты при модуляции оптической мощности. Чирп возникает из-за изменения показателя преломления, сопровождающего изменение концентрации носителей заряда при модуляции тока инжекции.

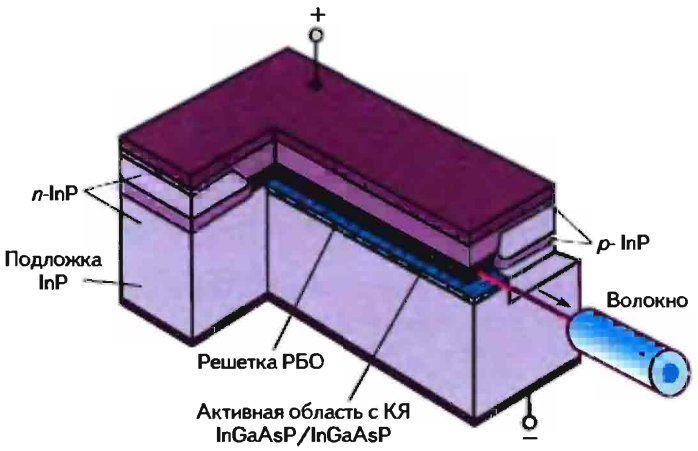


Рис. 24.5. Лазер с распределенной обратной связью на квантовых ямах со скрытой гетероструктурой, используемый в системах волоконной оптической связи. Диэлектрическая пленка обеспечивает направленность усиления, а чередующиеся *n*- и *p*-слои обеспечивают протекание тока только в окрестности активной области. Такие лазеры обеспечивают достаточное усиление при умеренных уровнях тока и выдают мощность 1 Вт и более в одной пространственной моде. Типичные значения порогового тока и дифференциальной токовой чувствительности составляют $i_{пор} < 10$ мА и $\mathcal{R}_d \approx 0{,}4$ Вт/А соответственно

Как обсуждалось в гл. 17, СИД на AlInGaP/InGaP часто используются в недорогих системах связи на пластиковых волокнах, работающих в диапазоне 600—650 нм (см. рис. 17.20). Более распространенным, однако, является $In_{1-x}Ga_xAs_{1-y}P_y$ — сплав переменного состава, широко применяемый при изготовлении как СИД, так и ЛД в ближнем инфракрасном диапазоне спектра. Он является прямозонным полупроводником, который за счет состава может перестраиваться в широком диапазоне длин волн, сохраняя при этом согласование решеток с подложкой из InP. InGaAsP применяется для изготовления СИД для систем связи на малых расстояниях с умеренной скоростью передачи данных, работающих на длине волны $\lambda_0 = 1{,}3$ мкм (см. рис. 17.19). Системы дальней связи

с высокими скоростями передачи данных обычно работают на λ_0 = 1,55 мкм с использованием лазерных диодов, а не СИД, поскольку коллимированное до высокой степени излучение ЛД намного легче ввести в одномодовое волокно.

Наиболее распространенной конструкцией лазерного диода является лазер с распределенной обратной связью (РОС) (рис. 24.5). Как обсуждалось в подразд. 17.4.3, в этом приборе применяется решетка в виде гофрированного слоя, прилегающего к активной области, который действует как распределенный отражатель, заменяющий зеркала резонатора Фабри—Перо. Такая конструкция совместима с интеграцией в микросхеме. Такие лазеры с торцевым излучателем дают узкие спектральные линии, что критично для эффективной работы оптических систем связи с мультиплексированием по длине волны (WDM) на 1,3 и 1,55 мкм.

24.1.3. Оптические усилители

Оптические усилители — обязательная составная часть современных оптических волоконных систем дальней связи. Они применяются как пост-усилители (усилители мощности), усилители на линии и предусилители. Как показано на рис. 24.6, усилители мощности повышают оптическую мощность перед ее закачкой в оптическое волокно, усилители на линии действуют как повторители, поддерживающие уровень сигнала в процессе его передачи по волокну (см. рис. 24.1), а предусилители обеспечивают усиление перед фотоприемником.

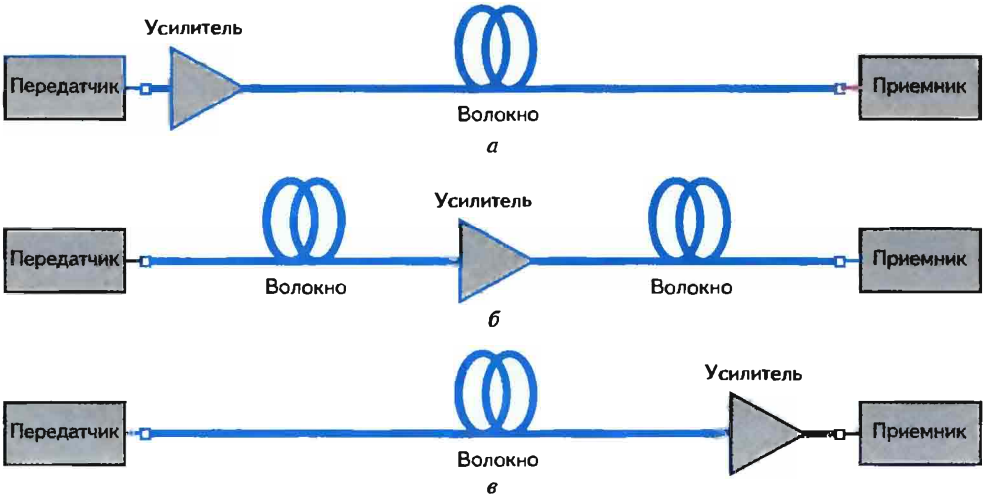


Рис. 24.6. Оптические волоконные усилители (ОВУ) используются в системе волоконной оптической связи как пост-усилители (*а*); усилители на линии (*б*); предусилители (*в*)

Мы рассмотрим три вида оптических усилителей.

• Оптические волоконные усилители (ОВУ). К ним относятся волоконные усилители, легированные эрбием (EDFA) (см. подразд. 14.3.3), волоконные

усилители, легированные редкоземельными элементами (rare-earth-doped fiber amplifiers, REFA) (см. подразд. 14.3.3), и волоконные ВКР-усилители (Raman fiber amplifiers, RFA).

- Полупроводниковые оптические усилители (ПОУ) (см. подразд. 14.3.4).
- Оптические параметрические усилители (ОПУ) (см. подразд. 21.4.3).

За исключением ОПУ, все оптические усилители, перечисленные выше, являются непараметрическими устройствами, поскольку их действие опирается на обмен энергией между полем и усиливающей средой (см. введение к гл. 21). EDFA и RFA оказались самыми подходящими усилителями для оптической волоконной связи, как видно из последующего обсуждения.

Оптические волоконные усилители (ОВУ)

ОВУ включают три разновидности: EDFA, REFA и RFA.

Волоконные усилители, легированные эрбием (EDFA)

Эти исторически первые ОВУ широко используются в системах волоконно-оптической связи. Как обсуждалось в подразд. 14.3.3, они имеют высокое поляризационно-независимое усиление, высокую выходную мощность, низкие вносимые потери, низкий уровень шума и широкую линию перехода вблизи $\lambda = 1{,}55$ мкм (что соответствует длине волны минимальных потерь для оптических волокон из кварца, как показано на рис. 24.3). Накачка осуществляется путем продольного ввода света в усиливающую среду, обычно от лазерных диодов на InGaAs с напряженными квантовыми ямами, работающих на длине волны $\lambda_0 = 980$ нм. Излучение накачки может инжектироваться в прямом или обратном направлении либо в обоих направлениях одновременно.

Усиление, превышающее 50 дБ, может быть достигнуто в EDFA при мощности накачки в десятки милливатт, при этом легко достигаются мощности выходного сигнала, превышающие 100 Вт. Достижимая ширина полосы составляет $\Delta\lambda \approx 40$ нм, что соответствует $\Delta\nu \approx 5{,}3$ ТГц и подходит для C-полосы. Легко покрывается и L-полоса, хотя параметры оптимизации EDFA в двух полосах не совпадают. Высокое усиление и ширина полосы этих усилителей делают их весьма подходящими для использования в системах мультиплексирования с разделением по длине волны (WDM) (см. подразд. 24.3.3).

Волоконные усилители, легированные редкоземельными элементами (REFA)

Ряд других, чем Er^{3+}, ионов (например, Pr^{3+}, Tm^{3+}, Nd^{3+}) также используется для создания волоконных усилителей, легированных редкоземельными элементами (REFA), которые покрывают полосы O/E/S/U (см. рис. 24.10). REFA можно использовать для расширения полосы усиления далеко за 60 нм (7,5 ТГц), доступные при использовании индивидуально оптимизированных EDFA в полосах C и L. К сожалению, волоконные усилители, легированные не Er^{3+}, а другими ионами, намного лучше работают с фторидным и теллуритным стек-

лами, чем с кварцевым стеклом. Трудности, связанные с такой заменой материала, нелегко преодолеть по ряду причин:

1) волокно из кварцевого стекла широко используется и технология его изготовления является устоявшейся;

2) изготовление каждого типа REFA требует своей собственной матрицы;

3) сращивание волокон из разных стекол не является простой задачей;

4) каждый тип REFA требует отдельной накачки лазерным диодом с соответствующей длиной волны.

Тем не менее путем смешивания и согласования REFA с Er^{3+} и Tm^{3+} доступная ширина полосы $\Delta\lambda$ может быть расширена от 60 нм примерно до 150 нм, что соответствует $\Delta\nu \approx 18{,}8$ ТГц на 1550 нм.

Волоконные ВКР-усилители (RFA)

ВКР-усилители работают на основе вынужденного комбинационного (рамановского) рассеяния (см. подразд. 13.5.3). Как обсуждалось в подразд. 14.3.4, существует две стандартные конфигурации ВКР-усилителей:

1) распределенные ВКР-усилители, где сигнал и накачка вместе посылаются по передающему волокну, которое служит усиливающей средой;

2) сосредоточенные усилители, в которых усиление обеспечивает короткий отрезок сильно нелинейного волокна.

Как и в EDFA, излучение накачки может распространяться в прямом или обратном направлении или быть двунаправленным.

Ширина полосы ВКР-усиления для кварцевого волокна составляет около 100 нм (что соответствует примерно 12,5 ТГц на 1550 нм), так что эти усилители обычно обеспечивают большую ширину полосы, чем EDFA. Более того, многочастотная накачка позволяет обеспечить еще большую ширину полосы, так что в принципе ВКР-усиление можно использовать во всей области прозрачности волокна. Усиление RFA, составляющее около 20 дБ, существенно ниже, чем у EDFA, как и эффективность, однако их можно повысить при использовании волокна, компенсирующего дисперсию. Требуется также накачка с другой поляризацией. Сравнительные свойства EDFA и RFA рассматривались в подразд. 14.3.4. Несмотря на недостатки RFA по сравнению с EDFA, их большая ширина полосы (которую можно еще увеличить за счет многочастотной накачки), произвольная рабочая длина волны и совместимость с существующими системами все больше повышают конкурентоспособность среди оптических волоконных усилителей.

Полупроводниковые оптические усилители (ПОУ)

Полупроводниковые оптические усилители (ПОУ) (см. разд. 17.3) могут работать в любой области оптического спектра, если правильно выбрать полупроводниковый материал. Они компактны и совместимы с интегральными оптоэлектронными схемами, в частности, как пост- или предусилители на интегральной оптоэлектронной схеме, и допускают электрическую накачку.

ПОУ, применяемые в системах оптической передачи данных в ближнем инфракрасном диапазоне обычно изготавливают из InGaAsP, InGaAs или InP. В полосе связи 1,3—1,6 мкм ширина полосы для ПОУ на квантовых ямах достигает $\Delta\lambda \approx 50$ нм, что соответствует $\Delta\nu \approx 6{,}5$ ТГц на $\lambda_0 = 1550$ нм, а ширина полосы ПОУ на квантовых точках — примерно 200 нм. Однако из-за низкого усиления (≈15 дБ) их применения принципиально ограничены оптическими сетями внутри крупных городов, где небольшое усиление достаточно для компенсации потерь, связанных с многочисленными оптическими узлами ввода-вывода.

Как обсуждалось в подразд. 17.2.4, ПОУ имеют ряд недостатков по сравнению с ОВУ: они несовместимы с геометрией волокна, обладают значительной межканальной и межсимвольной интерференцией, высоким уровнем шума, чувствительностью к температуре, а также остаточной чувствительностью к поляризации сигнала. В результате они более привлекательны для таких приложений, как полностью оптическая коммутация в оптических сетях (см. гл. 23) и преобразование длины волны, чем в качестве оптических усилителей на линии.

Оптические параметрические усилители (ОПУ)

Оптические параметрические усилители (ОПУ), обсуждавшиеся в подразд. 21.2.3 и 21.4.3, обеспечивают высокое усиление и обладают широкой полосой перестройки, простирающейся от инфракрасного до видимого диапазона. Однако они имеют ряд свойств, которые ограничивают возможности их эксплуатации в системах с мультиплексированием по длине волны (WDM):

- Сигналы WDM должны удовлетворять условиям фазового синхронизма с накачкой, что требует сглаживания дисперсии.
- Крупномасштабная реализация WDM с равноотстоящими по частоте каналами затруднена присутствием четырехволнового смешения.
- Такой усилитель чувствителен к поляризации сигнала, так что требуется накачка с поляризационным мультиплексированием.

Мы приходим к заключению, что ПОУ и ОПУ уступают ОВУ в применении к системам волоконной оптической связи.

24.1.4. Детекторы для оптических приемников

Обстоятельное обсуждение полупроводниковых детекторов фотонов содержится в гл. 18. В оптических системах связи обычно используются два типа детекторов: *p—i—n*-фотодиод и лавинный фотодиод (ЛФД). Преимуществом ЛФД является то, что он обеспечивает усиление до первой ступени электронного усиления в приемнике, снижая таким образом пагубное влияние шумов электрической схемы. Однако его механизм усиления сам вносит шум и имеет конечное время отклика, что может уменьшить ширину полосы приемника. Более того, ЛФД требует более высокого напряжения и более сложной схемы для компенсации своей чувствительности к температурным флуктуациям. Отношение сигнал—шум и обнаружительная способность приемников, в которых используются *p—i—n*-фотодиоды, обсуждались в разд. 18.6.

Детекторы в области 870 нм

На этой длине волны используются кремниевые p—i—n-фотодиоды и ЛФД. В современных предусилителях кремниевые ЛФД имеют преимущество в 10–15 дБ по обнаружительной способности по сравнению с p—i—n-фотодиодами, так как их внутреннее усиление делает шум предусилителя относительно менее важным.

Детекторы в области 1300—1600 нм

В этой области длин волн нельзя использовать кремний, поскольку он прозрачен (см. рис. 5.14); это происходит потому, что длина волны, соответствующая ширине запрещенной зоны, лежит ниже длины волны света ($\lambda_g < \lambda_0$). Вместо этого используются p—i—n-фотодиоды из InGaAs и Ge. Из них предпочтение отдается InGaAs, имеющему меньший темновой шум и большую температурную устойчивость. Типичные InGaAs p—i—n-фотодиоды имеют квантовую эффективность $\eta \approx 0{,}8$, токовую чувствительность $\mathfrak{R} \approx 1$ А/Вт (см. рис. 18.19), ширину полосы ≈10 ГГц на 50 Ом и темновой ток ≈0,1 нА. Волноводные структуры обладают большей шириной полосы.

Широкое применение находят ЛФД на InGaAs. Однако, как и все материалы с узкой запрещенной зоной, InGaAs страдает от больших туннельных токов утечки при наложении сильных электрических полей. Эта проблема решается путем использования гетероструктуры, в которой материал с узкой запрещенной зоной используется в области поглощения, а материал с широкой запрещенной зоной — в области умножения (separate absorption-multiplication avalanche photodiode, SAM APD). На рис. 24.7 показана вариация на эту тему: ЛФД с раздельными слоями поглощения и умножения, между которыми находится переходный (grading) слой (SAGM APD). Поглощение происходит в InGaAs, умножение в InP, а слой InGaAs обеспечивает плавный переход между ними.

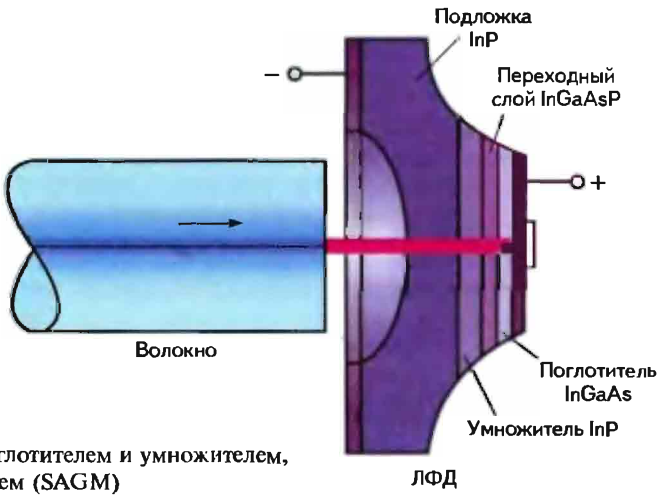


Рис. 24.7. Структура ЛФД с поглотителем и умножителем, разделенными переходным слоем (SAGM)

Поскольку в этом приборе размножаются дырки, ионизационное отношение есть $1/k$ (см. рис. 18.31). Для InP $1/k \approx 0{,}5$, когда среднее усиление $\bar{G} = 10$, т. е. шум усиления значительно больше, чем для Si. Тем не менее эти приборы хорошо работают; в типичном случае они имеют квантовую эффективность $\eta \approx 0{,}8$, токовую чувствительность $\mathfrak{R} \approx 10$ А/Вт, ширину полосы ≈10 ГГц, произведение усиления на ширину полосы ≈100 ГГц и темновой ток ≈0,1 нА.

24.2. ВОЛОКОННО-ОПТИЧЕСКИЕ СИСТЕМЫ СВЯЗИ

Простейшей системой связи является линия связи между двумя точками. Информация переносится сигналом — физической переменной (электрической, электромагнитной, оптической и т. д.), которая *модулируется* в одной точке и регистрируется в другой. Для одновременной передачи более чем одного сигнала по одной и той же линии сигналы должны быть отмечены некоторым различающимся признаком (время, частота, длина волны) или идентифицироваться по некоторому индивидуальному коду. Такая схема называется *мультиплексированием*.

Одной из рабочих характеристик аналоговой системы связи является ширина полосы B (Гц). Это максимальная частота, на которой модулированная оптическая мощность может быть передана по линии так, чтобы принимаемый сигнал детектировался с заданным уровнем отношения сигнал—шум. Ширина полосы определяется временем отклика всего канала связи, ослаблением и уровнем шума на приемнике.

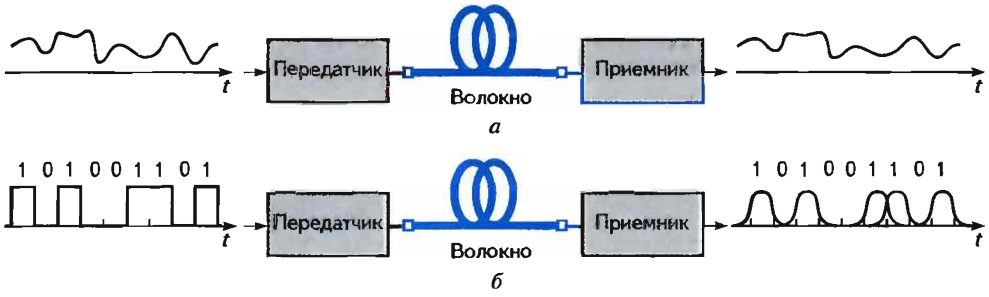


Рис. 24.8. Системы оптоволоконной связи с модуляцией интенсивности:

а — аналоговая система: мощность источника света пропорциональна сигналу, являющемуся непрерывной функцией времени; примером является аудио- или видеосигнал; *б* — цифровая двоичная система «ДА—НЕТ»: биты 1 и 0 представлены присутствием или отсутствием оптического импульса соответственно

В системах оптической волоконной связи линией является оптическое волокно, по которому пропускается световая волна, модулированная сигналом. Модулируемая физическая величина, несущая информацию, может представлять собой оптическую интенсивность, амплитуду, частоту, фазу или поляризацию. Простейший пример — это система связи с модуляцией интенсивности,

показанная на рис. 24.8. Простейший пример оптического мультиплексирования — это мультиплексирование с разделением по длине волны (WDM), при котором множество сигналов передается через одно волокно на разных длинах волны, как показано на рис. 24.9.

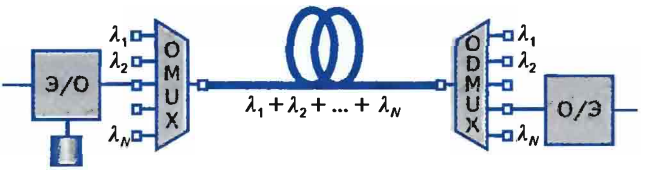


Рис. 24.9. Мультиплексирование с разделением по длине волны (WDM)

Аналогичной рабочей характеристикой цифровой системы связи является максимальная **битовая скорость передачи** B_0 (бит в секунду, б/с), с которой биты принимаемого сигнала можно различить при частоте ошибок, не превышающей заданного уровня. Битовая скорость передачи определяется ослаблением и расплыванием импульсов, вносимыми системой, а также уровнем шумов приемника. Битовые скорости передачи, установленные для оптических носителей стандартом Синхронной оптической сети (SONET), принятым в технологии оптической связи, приведены в табл. 24.1.

Таблица 24.1. **Примерные битовые скорости передачи стандарта SONET**

OC-1	OC-3	Oc-12	OC-24	OC-48	OC-192	OC-768
52 Мб/с	156 Мб/с	622 Мб/с	1,25 Гб/с	2,5 Гб/с	10 Гб/с	40 Гб/с

Этот раздел начинается с обзора развития оптических волоконных систем связи, затем проводится количественный анализ пределов работоспособности простых цифровых и аналоговых систем с модуляцией интенсивности.

24.2.1. Эволюция волоконно-оптических систем связи

Как видно на рис. 24.3, минимальное ослабление в кварцевом стекле происходит на ≈1550 нм, тогда как минимальная материальная дисперсия имеет место при ≈1312 нм. Выбор между этими двумя длинами волн зависит от относительной важности потери мощности и расплывания импульса, как объясняется в подразд. 24.2.2. Однако доступность подходящего источника света — тоже фактор. Системы оптоволоконной связи первого поколения работали на ≈870 нм (длина волны светоизлучающих и лазерных диодов на AlGaAs), где и ослабление сигнала, и материальная дисперсия сравнительно велики. Более продвинутые системы работают на 1300 и 1550 нм.

Различные рабочие длины волн, материалы и типы волокон, источников света, детекторов и усилителей, которые можно использовать для построения оптической линии, образуют ряд возможных комбинаций, часть которых приведена на рис. 24.10. Прогресс оптических волоконных сетей исторически сле-

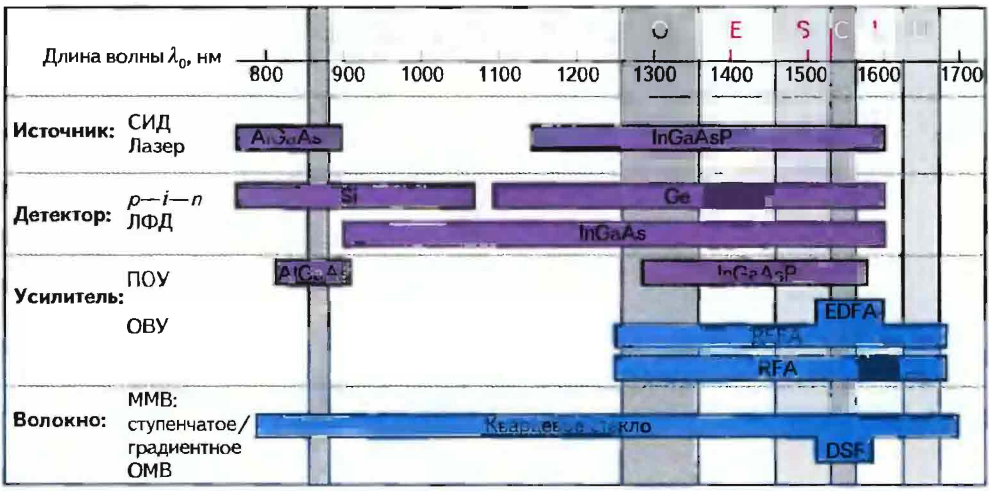


Рис. 24.10. Типы и материалы оптических источников, детекторов, усилителей и волокон, использованные на различных длинах волн. Первые поколения оптических линий связи работали на длинах волн вблизи 870, 1310 и 1550 нм

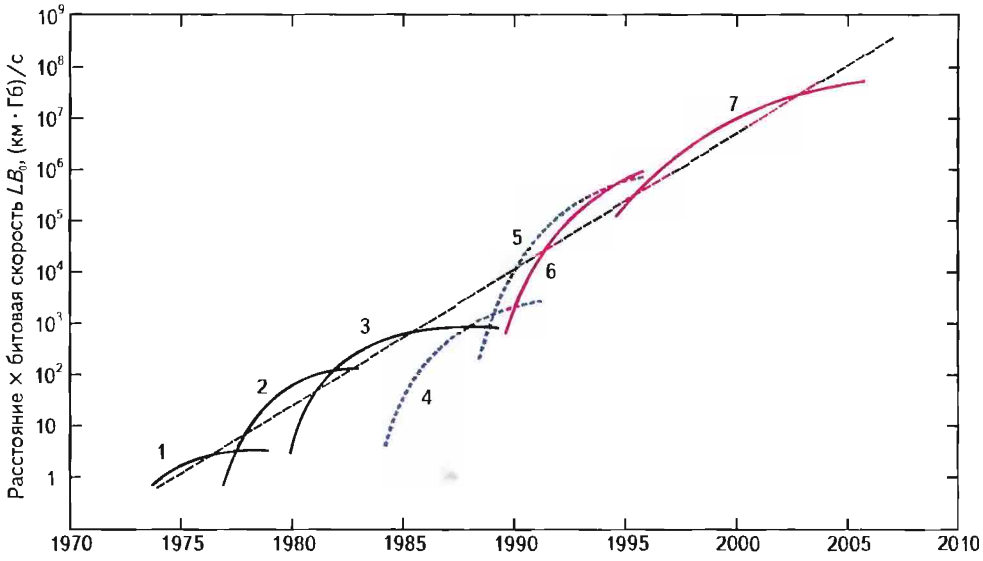


Рис. 24.11. История систем волоконной оптической связи представляет собой историю непрерывного повышения произведения расстояние на битовую скорость LB_0:

1 — 870 нм ММВ; 2 — 1310 нм ОМВ; 3 — 1550 нм, ОМВ (DSF), одночастотный лазер; 4 — когерентные; 5 — солитон; 6 — ОВУ; 7 — WDM

довал по пути увеличения длин волн, от многомодовых к одномодовым волокнам, от светоизлучающих диодов к диодным лазерам, от p—i—n-фотодиодов к лавинным фотодиодам и от полупроводниковых оптических усилителей к волоконным. Чтобы сделать этот прогресс возможным, потребовалась разработка подходящих материалов для более длинных волн (например, четырехкомпонентных полупроводников для источников и детекторов).

Эволюция волоконных компонентов и систем мотивировалась желанием повысить битовую скорость передачи B_0 (б/с) и увеличить длину L (км) линии связи (расстояния между ретрансляторами); произведение LB_0 (км б/с) использовалось как количественная мера продвижения вперед. Следующие семь систем характеризуют эту эволюцию, а рис. 24.11 отражает рост LB_0 год от года. Первые три системы, часто называемые первыми тремя поколениями оптоволоконных систем, позволили достичь тысячекратного повышения LB_0 за период с 1974 по 1990 г. Эти технологии используются как примеры при обсуждении рабочих характеристик системы в подразд. 24.2.2. Последующий прогресс привел к расширению этих основных систем во многих направлениях, что привело к росту LB_0 еще на пять порядков величины за период с 1990 по 2005 г. Такой десятикратный рост в течение каждых четырех лет был назван «оптическим законом Мура».

Система 1. Многомодовое волокно (MMB) на 870 нм

Это ранняя технология 1970-х гг. Волокна либо со ступенчатым, либо с градиентным профилем показателя преломления. Источник света — либо СИД, либо лазер (сначала GaAs, а затем AlGaAs). Фододиоды кремниевые, как p—i—n, так и лавинные. Эксплуатационные характеристики системы ограничены высоким ослаблением в волокне и модовой дисперсией. Типичная внутригородская линия связи этого периода работала при B_0 = 100 Мб/с с расстоянием между ретрансляторами L = 10 км, т. е. LB_0 = 1 км Гб/с.

Система 2. Одномодовое волокно (OMB) на 1310 нм

Переход к одномодовым волокнам и области длин волн, где материальная дисперсия минимальна, привел к существенному улучшению эксплуатационных характеристик, ограниченных теперь только ослаблением в волокне. используются лазеры на InGaAsP в сочетании с p—i—n- или лавинными фотодетекторами на InGaAs (иногда используются также германиевые ЛФД). Типичная дальняя линия связи этого класса работает по OC-12 (622 Мб/с) с расстоянием между ретрансляторами L = 40 км и $LB_0 \approx$ 25 км Гб/с.

Система 3. Одномодовое волокно (OMB) на 1550 нм

На этой длине волны волокно имеет наименьший возможный коэффициент ослабления. Эксплуатационные характеристики ограничены материальной дисперсией, которая уменьшена за счет использования одночастотных лазеров на InGaAsP с распределенной обратной связью (РОС) и малым чирпом. Последующее использование волокон со смещенной дисперсией (DSF) смягчило

проблему дисперсии и повысило эксплуатационные характеристики. Примером такой системы является дальняя наземная или проходящая по морскому дну линия связи, работающая при 2,5 Гб/с (OC-48) на расстоянии $L = 100$ км, для которой $LB_0 \approx 250$ км Гб/с. Улучшение передатчиков и приемников продвинуло эту систему до 10 Гб/с (OC-192) и довело LB_0 до одного км Тб/с.

Система 4. Когерентная система

Как описано в разд. 24.5, такая система не измеряет интенсивность светового сигнала непосредственно фотодетектором, а производит **когерентное детектирование**, при котором свет от опорного источника (называемого гетеродином) смешивается с сигнальным светом на детекторе. Использование когерентного детектирования повышает обнаружительную способность приемника, что дает возможность увеличить длину линии связи. Однако это достигается ценой увеличения сложности системы. Из-за этого промышленное внедрение когерентных систем запоздало по сравнению с внедрением систем прямого детектирования, особенно в связи с появлением оптических волоконных усилителей.

Система 5. Линия связи с оптическими усилителями

Появление полупроводниковых и оптических волоконных усилителей (см. подразд. 24.1.3) оказало мощное воздействие на эксплуатационные характеристики волоконных оптических систем связи. Помещенные периодически вдоль волокна, эти усилители компенсируют ослабление сигнала и, следовательно, увеличивают расстояние между электронными ретрансляторами. Примером является пересекающая Тихий океан система TPC-5, работающая на битовых скоростях до $B_0 = 10$ Гб/с на расстояниях до $L = 20\,000$ км, т. е. $LB_0 = 200$ (км · Тб)/с.

Система 6. Оптическая солитонная система

Солитоны — короткие (обычно от 1 до 50 пс) оптические импульсы, которые могут распространяться по длинному оптическому волокну без изменения формы своей огибающей. Как обсуждалось в подразд. 22.5.2, влияние дисперсии волокна в этом случае точно компенсирует влияние самомодуляции фазы (например, из-за оптического эффекта Керра), так что импульсы ведут себя так, как в линейной недиспергирующей среде. Эрбиевые волоконные усилители эффективно компенсируют потери мощности из-за поглощения и рассеяния. Экспериментальные системы работали при 10 Гб/с при длинах волокна более 12 000 км ($LB_0 = 1200$ (км · Тб)/с).

Система 7. Мультиплексирование с разделением по длине волны (WDM)

Введение WDM обеспечило существенный рост пропускной способности системы за счет одновременной передачи многих каналов на разных длинах волн по одному волокну. Для одновременного усиления всех каналов используются широкополосные усилители. Примером является система TPC-6, для которой $B_0 = 100$ Гб/с, $L = 9000$ км и $LB_0 = 900$ (км · Тб)/с. В сочетании с

дисперсионно-управляемой передачей и системой коррекции ошибок с избыточным кодированием теперь можно достичь скоростей 5—10 Тб/с на одно волокно на расстояниях в 10 000 км.

24.2.2. Эксплуатационные показатели оптических волоконных систем

Первым шагом в оценке эксплуатационных показателей волоконной системы связи является создание математической модели, описывающей влияние различных компонентов системы, главным образом оптического волокна, на модулированный сигнал. Это позволяет оценить форму принимаемого искаженного сигнала и, следовательно, определить отношение сигнал—шум в аналоговых системах и ожидаемую частоту появления ошибочных битов в цифровых системах.

В большинстве приложений волокно можно рассматривать как линейную систему, описываемую функцией отклика на импульсное воздействие $h(t)$ или ее Фурье-образом — передаточной функцией $H(f)$, где f — частота модуляции. Эти функции характеризуются тремя важными параметрами.

• *Коэффициент пропускания по мощности.* Это отношение выходной мощности к входной для стационарного (немодулированного) оптического излучения. Он равен передаточной функции $H(f)$ при $f = 0$. Поскольку $H(f)$ — Фурье-образ функции $h(t)$,

$$H(0) = \int h(t)\,dt$$

является площадью под графиком $h(t)$. Для волокна длиной L с коэффициентом ослабления α (дБ/км) $H(0) = \exp[-\alpha L]$, где $\alpha = 0{,}23\alpha$ — коэффициент ослабления в единицах км$^{-1}$. Локализованные потери мощности на соединениях также можно включить в α и выразить в распределенных единицах дБ/км.

• *Время отклика* σ_τ — это ширина функции $h(t)$. Оно определяет временное расплывание оптических импульсов и, следовательно, устанавливает максимальную скорость передачи данных, которую можно использовать в цифровых системах. Время отклика пропорционально длине волокна. Например, в одномодовом волокне $\sigma_\tau = |D|\sigma_\lambda L$, где σ_λ (нм) — ширина линии источника; D [пс/(км · нм)] — коэффициент дисперсии.

• *Ширина полосы* σ_f — ширина передаточной функции $|H(f)|$. В аналоговой системе ширина полосы определяет максимальную частоту, с которой входной сигнал может быть промодулирован и успешно зарегистрирован приемником. Поскольку $H(f)$ и $h(t)$ связаны преобразованием Фурье, ширина полосы σ_f обратно пропорциональна времени отклика σ_τ. Коэффициент пропорциональности зависит от конкретного профиля $h(t)$ (см. приложение А, разд. А.2). С целью иллюстрации здесь мы будем использовать соотношение $\sigma_f = 1/2\pi\sigma_\tau$.

Максимальная длина волокна, которую можно использовать для передачи сигнала с заданным уровнем качества, устанавливается следующими основными факторами ухудшения сигнала, вносимыми системой.

• *Ослабление* приводит к экспоненциальному спаду оптической мощности как функции расстояния (рис. 24.12, *а*). На расстоянии, для которого принимаемая мощность становится меньше обнаружительной способности приемника (минимальной мощности, требуемой для приема), эксплуатационные характеристики системы становятся неприемлемыми.

• *Дисперсия* приводит к увеличению ширины оптических импульсов, представляющих биты данных в цифровой системе, с увеличением расстояния (рис. 24.12, *б*). Когда ширина превышает расстояние между битами, соседние импульсы перекрываются, что приводит к межсимвольной интерференции, вносящей ошибки. В аналоговой системе дисперсия смазывает высокочастотные компоненты модулированного сигнала и уменьшает ширину полосы системы.

• *Шум*, добавляемый оптическими компонентами, такими как оптические усилители, а также случайными эффектами распространения, такими как дисперсия поляризационных мод, вносит дополнительные ошибки.

• *Нелинейное искажение*, связанное с интенсивными оптическими импульсами, приводит к перекрестному смешению спектральных компонент и интерференции каналов в системах мультиплексирования с разделением по длине волны (WDM).

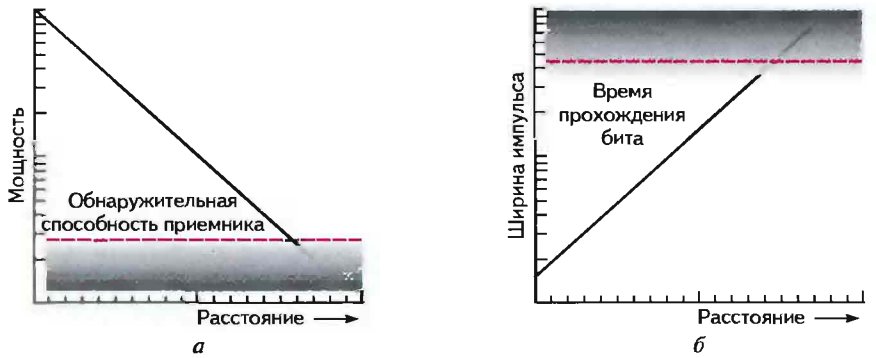


Рис. 24.12. Зависимость оптической мощности от расстояния (*а*). Зависимость ширины импульса от расстояния (*б*). Максимальная длина оптической линии определяется либо ослаблением (*а*), когда принимаемая мощность опускается ниже обнаружительной способности приемника, либо дисперсией (*б*), когда ширина импульса превышает время прохождения бита

Системы связи более чувствительны к искажениям передачи на высоких битовых скоростях (или высоких частотах модуляции) из-за следующих эффектов.

• При фиксированной средней мощности более высокая битовая скорость передачи данных соответствует меньшему числу фотонов на один бит и, следовательно, большему фотонному шуму. Другие источники шума в приемнике также становятся более важными при высоких скоростях передачи данных. Следовательно, обнаружительная способность приемника является возрастающей функцией битовой скорости передачи данных (рис. 24.13, *а*).

• Более высокая битовая скорость передачи данных соответствует более коротким импульсам (рис. 24.13, *б*) с более широким спектром и большей дисперсией. Такие импульсы претерпевают большее уширение, что приводит к большей межсимвольной интерференции (МСИ).

• При фиксированной оптической энергии на один бит более высокая битовая скорость передачи данных (более короткое время прохождения бита) требует большей оптической мощности (рис. 24.13, *в*), что вызывает нелинейные взаимодействия, ведущие к нелинейной межсимвольной интерференции.

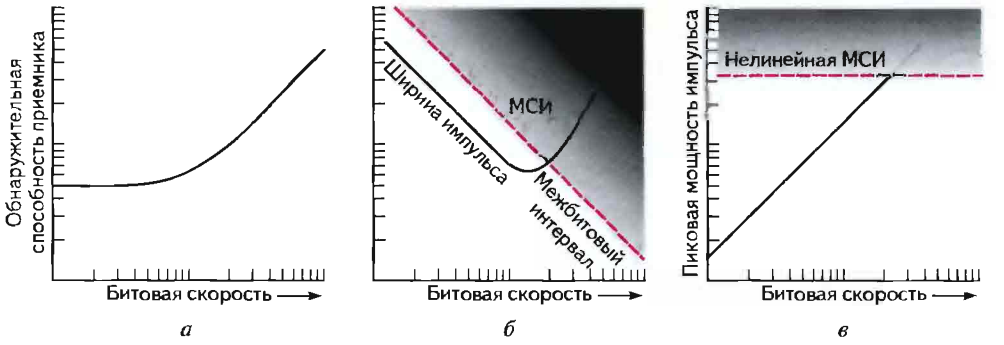


Рис. 24.13. Влияние битовой скорости передачи данных на обнаружительную способность приемника (*а*), ширину импульса на приемнике (*б*) и пиковую мощность (*в*). На более высоких частотах система связи более чувствительна к ослаблению, дисперсии и нелинейным эффектам

Как мы увидим в оставшейся части данного раздела, проектирование волоконно-оптической линии дальней связи включает выбор волокон с наименьшим ослаблением и/или дисперсией, а также тщательный выбор мощности и длительности импульсов для защиты от нелинейных эффектов, связанных с импульсами высокой интенсивности.

Частота ошибок по битам

Качество работы *цифровой* системы связи измеряется **вероятностью ошибки** на один бит, которая называется **частотой ошибок по битам** (bit error rate, BER). Для системы с амплитудной манипуляцией, пример которой показан на рис. 24.8, биты «1» и «0» представляются соответственно присутствием и отсутствием оптического импульса. Если p_1 — вероятность ошибки «1» для «0», а p_0 — вероятность ошибки «0» для «1» и если оба бита имеют одинаковую вероятность пройти через систему, то $\text{BER} = p_1/2 + p_0/2$. Типичное приемлемое значение BER составляет 10^{-9} (т. е. в среднем одна ошибка на каждые 10^9 битов).

Ошибки возникают из-за шума в принимаемом сигнале или из-за расплывания импульса на соседние биты, что приводит к межсимвольной интерференции. На рис. 24.14 показан пример случайных реализаций импульса, соответствующего биту «1», наложенных на случайные реализации сигнала, полу-

ченного от возможных соседних импульсов, когда бит равен «0». Эта диаграмма называется **глазковой (индикаторной) диаграммой**. Чем больше раскрыт «глазок», тем более различимы биты «1» и «0» и меньше вероятность ошибки.

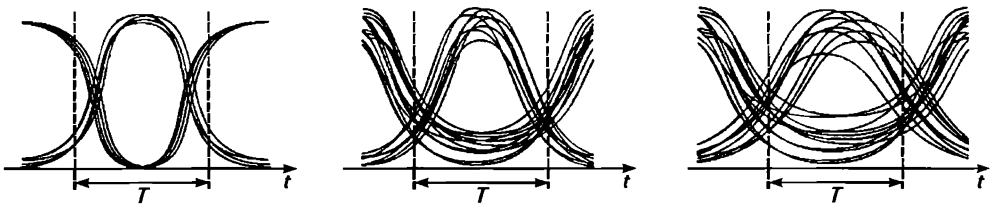


Рис. 24.14. Закрытие глазковой диаграммы (слева направо) с ростом шума и уширения импульса

Обнаружительная способность приемника

Обнаружительная способность цифрового оптического приемника определяется как минимальное число фотонов (или соответствующая оптическая энергия) на один бит, необходимое для гарантии того, что частота ошибок по битам (BER) меньше предписанного значения (например, 10^{-9}). Ошибки возникают из-за случайности числа фотоэлектронов, детектируемых в течение каждого бита, а также из-за шума схемы самого приемника. Обнаружительная способность приемников на базе различных фотодетекторов обсуждается в подразд. 18.6.5.

Например, если источником света служит стабилизированный лазер, детектор имеет квантовый выход единица, а схема приемника свободна от шумов, то в среднем требуется по крайней мере $\bar{n}_0 = 10$ фотонов, чтобы гарантировать BER $\leq 10^{-9}$. Следовательно, *обнаружительная способность идеального приемника составляет 10 фотонов/бит*. Это означает, что бит «1» должен нести, по меньшей мере, 20 фотонов, поскольку бит «0» фотонов не несет. При наличии других видов шума требуется большее число фотонов.

Обнаружительная способность $\bar{n}_0$ фотонов соответствует оптической энергии $h\nu\bar{n}_0$ на один бит и оптической мощности

$$P_r = \frac{h\nu\bar{n}_0}{1/B_0},$$

$$P_r = h\nu\bar{n}_0 B_0, \tag{24.5}$$

которая пропорциональна битовой скорости передачи данных B_0. С ростом этой величины требуется все большая оптическая мощность для поддержания постоянного числа фотонов на один бит (и, следовательно, BER).

В подразд. 18.6.5 показано, что когда шум схемы является важным, обнаружительная способность приемника $\bar{n}_0$ зависит от ширины полосы (т. е. от скорости передачи данных B_0). Такое поведение усложняет задачу проектирования. Для простоты мы предположим в последующем анализе, что обнаружительная способность приемника (число фотонов на бит) не зависит от B_0.

24.2.3. Системы, ограниченные по ослаблению и дисперсии

В данном разделе мы изучим ограничения, накладываемые *ослаблением* и *дисперсией* на рабочие характеристики системы с амплитудной манипуляцией. Нелинейные эффекты игнорируются, а сама проводящая волоконная система считается не вносящей шума.

Рассмотрим оптическую линию связи, работающую как цифровая коммуникационная система со скоростью передачи данных B_0 (бит/с) на расстояние L (км). Источник имеет мощность P_s (мВт), длину волны λ (нм) и спектральную ширину σ_λ (нм). Волокно имеет коэффициент ослабления α (дБ/км) и коэффициент хроматической дисперсии D_λ [пс/(км · нм)]. Приемник имеет обнаружительную способность $\bar{n}_0$ (фотонов/бит), что соответствует минимальной мощности

$$P_r = \frac{hc}{\lambda}\bar{n}_0 B_0 \quad (\text{мВт}),$$

которую должен иметь получаемый системой сигнал, чтобы она работала с приемлемой частотой ошибок.

Пределы работоспособности устанавливаются путем определения максимального расстояния L, на которое линия может передать B_0 (бит/с), не превышая предписанной частоты ошибок по битам. Очевидно, что L уменьшается с ростом B_0. Наоборот, мы можем определить максимальную битовую частоту B_0, которую линия длиной L способна передавать не превышая допустимого предела ошибок. Максимальное произведение битовой скорости передачи на длину LB_0 служит единым числом, которое описывает возможности системы. Мы определим типичную зависимость L от B_0 и выведем выражения для максимального произведения LB_0 для различных типов волокна.

Для приемлемой работы линии требуется выполнение двух условий.

1. Принимаемая мощность должна быть, по крайней мере, равна обнаружительной способности по мощности P_r. Это условие выполняется за счет выбора баланса мощности, из которого определяется максимальная длина волокна. Мощность обычно задается с запасом в 6 дБ над значением P_r.

2. Ширина принимаемых импульсов не должна заметно превышать битовый временной интервал $1/B_0$, иначе соседние импульсы перекрываются и вызывают *межсимвольную интерференцию*, увеличивающую частоту ошибок. Выполнение этого условия достигается подбором баланса расплывания импульса, вносимого передатчиком, приемником и различными видами дисперсии в волокне.

Если битовая скорость передачи B_0 фиксирована, а длина линии L увеличивается, может возникнуть две ситуации, ведущие к ухудшению характеристик: принимаемая мощность становится меньше обнаружительной способности приемника P_r или принимаемые импульсы становятся шире, чем время прохождения бита $1/B_0$. Если сначала возникает первая из названных ситуаций, то линия связи называется *ограниченной по ослаблению*. Если сначала возникает вторая ситуация, то говорят, что линия *ограничена по дисперсии*.

Ограничение по ослаблению: баланс мощности

При ограничении по ослаблению характеристики оцениваются путем составления баланса мощности. Так как ослабление волокна измеряется в децибелах, удобно измерять мощность также в децибелах. Если взять 1 мВт за опорное значение, то мощность в децибел-милливаттах (дБм) определяется формулой

$$\boxed{\mathcal{P} = 10 \lg P,} \quad P \text{ — в мВт; } \mathcal{P} \text{ — в дБм.} \tag{24.6}$$

Например, $P = 0$, 1 мВт, 1 мВт и 10 мВт соответствует $\mathcal{P} = -10$, 0 и +10 дБм. В этих логарифмических единицах потери мощности аддитивны.

Если $\mathcal{P}_s$ — мощность источника (дБм), α — потери в волокне (дБ/км), $\mathcal{P}_c$ — потери на стыках и соединениях (дБ), а L — максимальная длина волокна, такая что мощность, подаваемая на приемник, равна его обнаружительной способности $\mathcal{P}_r$ (дБм), то

$$\mathcal{P}_s - \mathcal{P}_c - \mathcal{P}_m - \alpha L = \mathcal{P}_r \text{ (в дБ)}, \tag{24.7}$$

где $\mathcal{P}_m$ — запас для безопасности. Схематический график оптической мощности как функции расстояния от передатчика приведен на рис. 24.15.

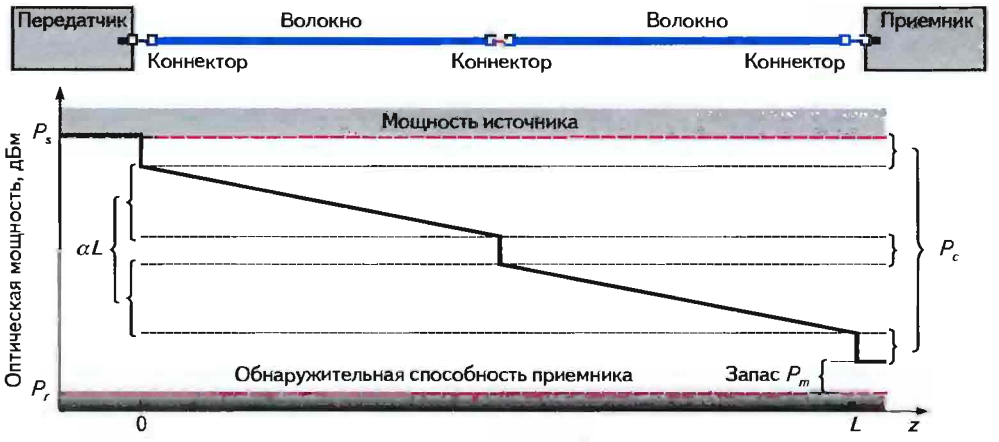


Рис. 24.15. Баланс мощностей оптической линии связи

Обнаружительная способность приемника по мощности $\mathcal{P}_r = 10 \lg_{10} P_r$ (дБм) получается из (24.5):

$$\mathcal{P}_r = 10 \lg \frac{\bar{n}_0 h\nu B_0}{10^{-3}}, \text{ дБм.} \tag{24.8}$$

Таким образом, $\mathcal{P}_r$ с ростом B_0 увеличивается по логарифмическому закону и баланс мощностей нужно приспосабливать к каждому значению B_0, как показано на рис. 24.16.

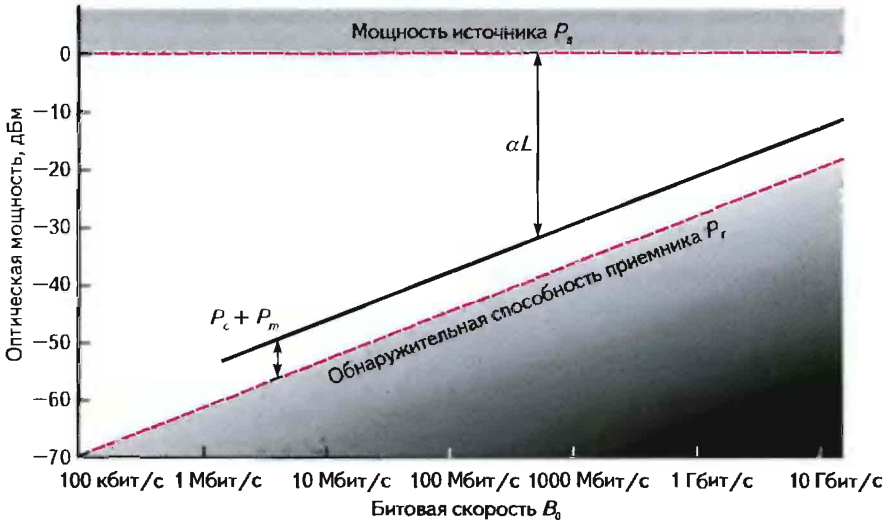


Рис. 24.16. Баланс мощности как функция битой скорости передачи данных B_0. С ростом B_0 требуемая для приема мощность $\mathcal{P}_r$ растет (так что энергия на один бит остается постоянной), а максимальная длина L уменьшается

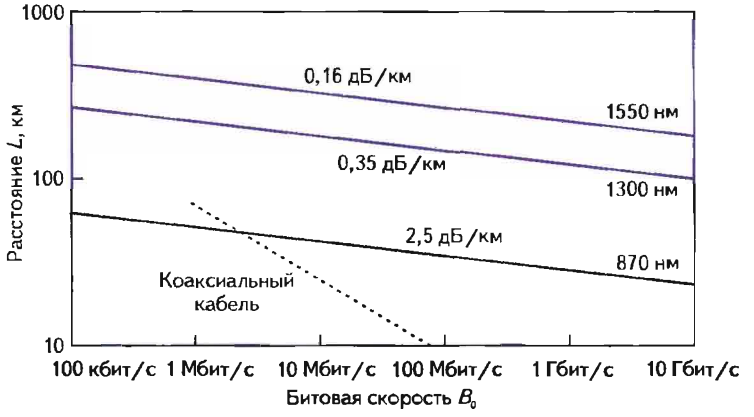


Рис. 24.17. Максимальная длина волокна L как функция битовой скорости B_0 в условиях ограничения по ослаблению для волокна из плавленого кварцевого стекла на рабочих длинах волн $\lambda_0 = 870$, 1300 и 1550 нм при коэффициентах ослабления волокна α = 2,5; 0,35 и 0,16 дБ/км соответственно. Мощность источника $P_s = 1$ мВт ($\mathcal{P}_s = 0$ дБм); обнаружительная способность приемника $\bar{n}_0 = 300$ фотонов/бит для приемников, работающих на 870 и 1300 нм, и $\bar{n}_0 = 1000$ для приемника, работающего на 1550 нм; $\mathcal{P}_c = \mathcal{P}_m = 0$. Для сравнения показано соотношение L—B_0 для типичного коаксиального кабеля

Максимальная длина связи получается путем подстановки (24.8) в (24.7)

$$L = \frac{1}{\alpha}\left(\mathcal{P}_s - \mathcal{P}_c - \mathcal{P}_m - 10\lg\frac{\bar{n}_0 h\nu B_0}{10^{-3}}\right) \text{ (дБ)}, \qquad (24.9)$$

откуда

$$\boxed{L = L_0 - \frac{10}{\alpha}\lg B_0,} \qquad (24.10)$$

Волокно с ограничением по ослаблению

где

$$L = \frac{\mathcal{P}_s - \mathcal{P}_c - \mathcal{P}_m - 30 - 10\lg(\bar{n}_0 h\nu)}{\alpha}.$$

Длина падает с ростом B_0 по логарифмическому закону с наклоном $10/\alpha$. На рис. 24.17 это соотношение проиллюстрировано графически для рабочих длин волн 870, 1300 и 1550 нм.

Ограничение по дисперсии: баланс времени

Когда импульс, представляющих бит данных, генерируется передатчиком, распространяется по волокну и детектируется приемником, он теряет мощность и приобретает ширину. Финальная ширина импульса σ_r зависит от исходной ширины σ_s, времени отклика передатчика $\sigma_{\text{пер}}$, времени отклика волокна σ_τ, возникающего из-за различных видов дисперсии, и времени отклика приемника $\sigma_{\text{пр}}$. Фактическая форма принимаемого импульса может быть определена путем свертки исходного импульса с функциями отклика на импульсное воздействие передатчика, волокна и приемника (в предположении, что все три системы линейны). Если все функции гауссовы, то квадрат ширины финального импульса равен сумме квадратов ширин всех составляющих функций, так что

$$\sigma_0^2 = \sigma_s^2 + \sigma_{\text{сист}}^2, \qquad (24.11)$$

где

$$\sigma_{\text{сист}}^2 = \sigma_{\text{пер}}^2 + \sigma_{\text{пр}}^2 + \sigma_\tau^2; \qquad (24.12)$$

$\sigma_{\text{сист}}$ — ширина функции отклика всей коммуникационной системы (передатчик + волокно + приемник). Эти соотношения используются при практическом проектировании, даже если функции отклика не гауссовы.

Принципиальным условием при проектировании линии связи является гарантия того, что ширина принимаемого импульса не превышает установленной доли битового периода T/B_0, чтобы избежать межсимвольной интерференции (МСИ). Необходимо составить баланс времени (рис. 24.18), гарантирую-

щий выполнение этого условия. Выбор указанной доли произволен, и на практике используется несколько значений, подобранных для разных случаев. Например, некоторые разработчики требуют, чтобы время отклика системы не превышало 70 % битового периода для импульсов без возвращения к нулю и 35 % для импульсов с возвращением к нулю (см. определения соответствующих форматов модуляции на рис. 24.27 в подразд. 24.3.1).

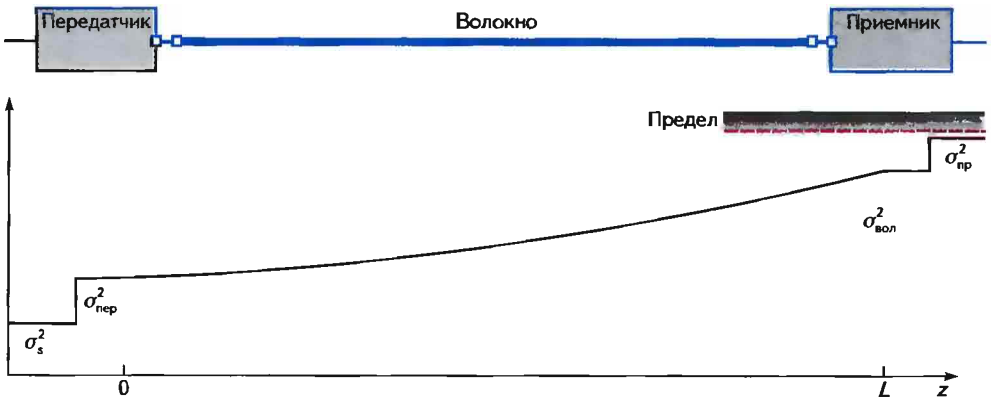


Рис. 24.18. Баланс временной ширины импульса

Для данных передатчика и приемника центральной задачей проектирования линии связи становится определения максимальной длины волокна L. Поскольку единственный зависящий от длины вклад в $\sigma_{сист}$ вносится волокном (σ_τ), при последующем анализе мы примем проектировочное условие, что максимальное допустимое значение σ_τ должно составлять 25 % от битового интервала T, т. е.

$$\sigma_\tau = \frac{1}{4}T = \frac{1}{4B_0}. \tag{24.13}$$

Выбор множителя 1/4, очевидно, произволен и служит только для сравнения различных типов волокна. Рассмотрим теперь соотношения между расстоянием и битовой скоростью, вытекающие из этого условия для различных случаев ограничения по дисперсии, упомянутых в подразд. 24.1.1. Результаты изображены графически на рис. 24.19.

• *Многомодовое волокно* (*ММВ*). Для многомодового волокна ширина принимаемого импульса после прохождения расстояния определяется прежде всего модовой дисперсией. Для волокна со ступенчатым профилем показателя преломления формулы (24.2) и (24.13) приводят к соотношению

$$\boxed{LB_0 = \frac{c_1}{2\Delta},} \tag{24.14}$$

ММВ со ступенчатым профилем

где $c_1 = c_0/n_1$ — скорость света в материале сердцевины; $\Delta = (n_1 - n_2)/n_1$ — относительный перепад показателя преломления волокна. В градиентных волокнах с оптимальным профилем показателя преломления, близким к параболическому, ширина импульса меньше в $2/\Delta$ раз, а LB_0 во столько же раз больше. При $n_1 = 1{,}46$ и $\Delta = 0{,}01$ для градиентного волокна произведение $LB_0 \approx 10$ (км · Мб)/с.

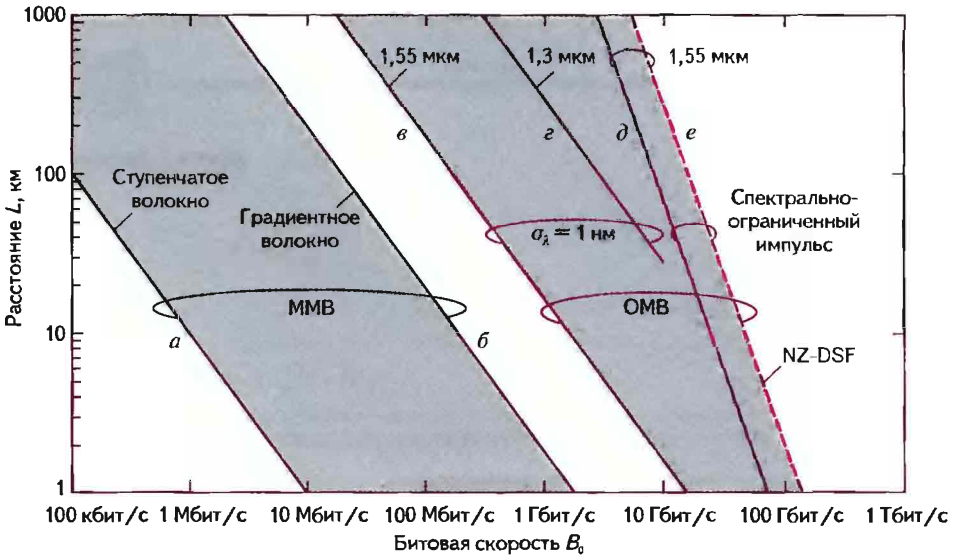


Рис. 24.19. Ограниченная дисперсией максимальная длина волокна как функция битовой скорости для многомодовых волокон (ММВ) и одномодовых волокон (ОМВ). Показаны шесть линий (слева направо):

а — ММВ со ступенчатым профилем ($n_1 = 1{,}46$, $\Delta = 0{,}01$), $LB_0 = 10$ (км · Мб)/с; *б* — ММВ с градиентным параболическим профилем ($n_1 = 1{,}46$, $\Delta = 0{,}01$), $LB_0 = 2$ (км · Гб)/с; *в* — ОМВ, ограниченное по материальной дисперсии, работающее на 1550 нм с $D_\lambda = 17$ пс/(км · нм) и $\sigma_\lambda = 1$ нм, $LB_0 \approx 15$ (км · Гб)/с; *г* — ОМВ, ограниченное по материальной дисперсии, работающее на 1300 нм с $|D_\lambda| = 1$ пс/(км · нм) и $\sigma_\lambda = 1$ нм, $LB_0 = 250$ (км · Гб)/с; *д* — ОМВ со спектрально-ограниченными импульсами, работающее на 1550 нм с $|D_\lambda| = 17$ пс/(км · нм); *е* — то же, что в случае (*д*), но для волокна с ненулевым смещением дисперсии (NZ-DSF) с коэффициентом хроматической дисперсии $D_\lambda = 4$ пс/(км · нм)

• *Одномодовое волокно* (*ОМВ*). Если предположить, что уширение импульса в одномодовом волокне является результатом только материальной дисперсии (т. е. пренебречь волноводной дисперсией), то для источника с шириной линии σ_λ длительность принимаемого импульса будет определяться формулой (24.3), т. е.

$$LB_0 = \frac{1}{4|D_\lambda|\sigma_\lambda}, \quad \text{ОМВ} \tag{24.15}$$

где D_λ — коэффициент дисперсии материала волокна. Вблизи рабочей длины волны $\lambda_0 = 1300$ нм коэффициент $|D_\lambda|$ может быть мал и достигать 1 пс/(км · нм). Считая, что $\sigma_\lambda = 1$ нм (ширина линии одномодового лазера), находим для

произведения битовой скорости на расстояние $LB_0 \approx 250$ (км · Гб)/с. При работе вблизи $\lambda_0 = 1550$ нм $D_\lambda = 17$ пс/км нм, и для той же ширины линии источника $\sigma_\lambda = 1$ нм имеем $LB_0 \approx 15$ (км · Гб)/с.

• *Одномодовое волокно со спектрально-ограниченными импульсами.* Для уменьшения хроматической дисперсии спектральная ширина источника σ_λ должна быть малой. Спектральные ширины, составляющие малую долю 1 нм, получаются с помощью одномодовых лазеров с внешними модуляторами. Однако предельно малая спектральная ширина несовместима с предельно коротким импульсом, поскольку спектральное и временное распределения связаны преобразованием Фурье. Как описано в разд. А.2 приложения А, импульсы с минимальным произведением длительности и спектральной ширины имеют гауссов профиль. Следовательно, такие спектрально-ограниченные импульсы испытывают наименьшее влияние дисперсии.

Спектрально-ограниченный импульс шириной τ_0 с комплексной огибающей $\exp(-t^2/\tau_0^2)$ имеет гауссову спектральную интенсивность с шириной по половине высоты максимума $\sigma_\nu = 0{,}375/\tau_0$ (см. подразд. 22.1.2). Это соответствует

$$\sigma_\lambda = \left|\frac{\partial \lambda_0}{\partial \nu}\right| \sigma_\nu = \left(\frac{\lambda_0^2}{c}\right)\sigma_\nu = 0{,}375\frac{\lambda_0^2}{c\tau_0}.$$

Если импульс имеет ширину, равную половине битового периода, т. е. $\tau_0 = T/2 = 1/2B_0$, то

$$\sigma_\lambda = 0{,}75\frac{\lambda_0^2}{c_0} B_0, \tag{24.16}$$

что прямо пропорционально битовой скорости B_0. Например, для $\lambda_0 = 1550$ нм и $B_0 = 10$ Гб/с $\sigma_\lambda = 0{,}06$ нм. Как описано в подразд. 22.3.2, при распространении спектрально-ограниченного импульса длительностью τ_0 через диспергирующую среду с коэффициентом дисперсии D_ν он уширяется в $\sqrt{2}$ раз на характеристической длине $z_0 = \pi\tau_0^2/D_\nu$. На таком расстоянии импульс с первоначальной длительностью $\tau_0 = T/2$ растягивается на время

$$\frac{(\sqrt{2}-1)T}{2} \approx 0{,}21T$$

Таким образом, мы можем принять z_0 за максимальную приемлемую длину L линии связи. Пользуясь соотношениями

$$L = z_0 = \frac{p\tau_0^2}{D_\nu}; \quad \tau_0 = \frac{T}{2} = \frac{1}{2}B_0; \quad D_\nu = \frac{D_\lambda \lambda_0^2}{c_0^2},$$

получаем соотношение между расстоянием и битовой скоростью

$$\boxed{LB_0^2 = \frac{\pi}{4}\frac{c_0}{|D_\lambda|\lambda_0^2}.} \tag{24.17}$$

ОМВ

Спектрально-ограниченный импульс

Таким образом, максимальное расстояние обратно пропорционально B_0^2, т. е. спадает быстрее с ростом скорости передачи данных, чем в предыдущих случаях. Также произведение LB_0 обратно пропорционально скорости передачи данных. На рис. 24.19 показано соотношением для $\lambda_0 = 1550$ нм и $D_\lambda = 17$ пс/(км·нм). Например, при $B_0 = 10$ Гб/с, $L = 64$ км, но при $B_0 = 40$ Гб/с L падает до 4 км. Следовательно, использование спектрально-ограниченных импульсов существенно расширяет границы битовой скорости, устанавливаемые дисперсией, хотя при дальнейшем увеличении битовой скорости полученный выигрыш быстро убывает.

• *Одномодовое волокно с компенсацией дисперсии и спектрально-ограниченными импульсами.* В одномодовом волокне при использовании спектрально-ограниченных импульсов максимальная длина волокна при заданной битовой скорости передачи достигает своего максимального значения, ограниченного только коэффициентом дисперсии. Этот коэффициент можно уменьшить с использованием волокон со смещенной дисперсией (DSF). Как показано на рис. 24.19, уменьшение с 17 до 4 пс/(км·нм) при использовании DSF увеличивает максимальное расстояние с 64 до 272 км при 10 Гб/с. Однако волокна со смещенной дисперсией имеют несколько больший коэффициент ослабления.

Комбинация ограничений по ослаблению и дисперсии

Соотношения между длиной связи и битовой скоростью передачи данных в случае одновременного присутствия ограничений по ослаблению и дисперсии показаны на рис. 24.20, который получен путем наложения рис. 24.17 и 24.19 и выбора наименьшего из расстояний, ограниченных ослаблением или дисперсией. Эти соотношения описывают характеристики поколений оптических волокон, работающих на $\lambda_0 = 870$ (многомодовых), 1300 и 1550 нм (одномодовых). Для построения этой карты было сделано несколько упрощающих допущений и произвольных выборов, так что полученные величины следует рассматривать только как индикаторы порядков величины относительных характеристик волокон различных типов. Тем не менее можно сделать ряд важных заключений:

• При низких битовых скоростях волоконная линия в общем случае ограничена по ослаблению; L спадает с ростом B_0 по логарифмическому закону. При высоких битовых скоростях основным является ограничение по дисперсии; L обратно пропорционально B_0 для оптических импульсов, ограниченных шириной линии источника, и обратно пропорционально B_0^2 для спектрально-ограниченных импульсов.

• Для дальних линий связи с высокими скоростями передачи данных существенно использование одномодовых волокон. Выбор между 1300 и 1550 нм не очевиден, так как для обычных волокон хроматическая дисперсия наименьшая на 1300 нм, а ослабление — на 1550 нм. Это объясняет пересечение линий L—B_0 на этих длинах волн.

• С применением волокон со смещенной дисперсией (DSF) можно уменьшить полный коэффициент хроматической дисперсии на 1550 нм, что делает работу на этой длине волны более выгодной, чем на 1300 нм.

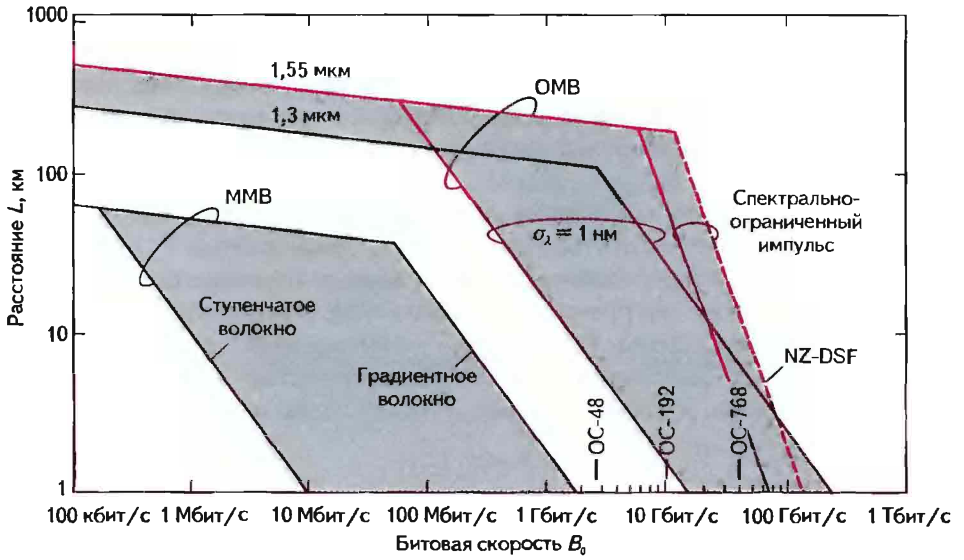


Рис. 24.20. Зависимость максимальной дальности L от битовой скорости B_0 для шести типов волоконной связи. График получен путем наложения рис. 24.17 и 24.19. Каждая линия представляет максимальную дальность L связи при данной битовой скорости B_0, которая удовлетворяет ограничениям как по ослаблению, так и по дисперсии, т. е. одновременно гарантирует требуемую мощность и ширину принимаемого импульса

Характеристики аналоговой системы связи

Как и у цифровых систем связи, рабочие характеристики аналоговой линии связи ограничены ослаблением и/или дисперсией в волокне. Из-за ослабления принимаемый сигнал может оказаться неразличимым на фоне шума. Из-за дисперсии ширина полосы передачи $\sigma_f = 1/2\pi\sigma_\tau$ ограничена, так что высокочастотные компоненты сигнала ослабляются сильнее низкочастотных, что приводит к искажению сигнала. Вредное влияние обоих эффектов увеличивается с ростом длины волокна L. Принимаемая оптическая мощность экспоненциально убывает с ростом L, а полоса пропускания волокна обратно пропорциональна L. Нелинейные эффекты не играют роли в аналоговых системах, так как мощность распределена во времени и не сконцентрирована в коротких импульсах.

Максимальная длина аналоговой линии связи определяется выполнением двух условий:

- Ослабление в волокне должно быть достаточно мало, чтобы принимаемая мощность была больше, чем обнаружительная способность приемника по мощности $\mathcal{P}_r$.
- Ширина полосы пропускания волокна σ_f должна быть больше, чем спектральная ширина B передаваемого сигнала.

Как говорилось в разд. 18.6, обнаружительная способность аналогового оптического приемника — это наименьшая оптическая мощность, необходимая для того, чтобы отношение сигнал—шум (ОСШ) для фототока превышало наперед заданное значение ОСШ_0. Для идеального приемника (квантовый выход равен единице, шумы схемы отсутствуют)

$$\text{ОСШ} = \bar{n} = \frac{P/h\nu}{2B},$$

где B — ширина полосы приемника; P — оптическая мощность (Вт); $\bar{n}$ — среднее число фотонов, принимаемых за время $1/2B$, которое характеризует временное разрешение системы. Если ОСШ_0 — минимальное допустимое отношение сигнал—шум, то обнаружительная способность становится равной $\bar{n}_0 = \text{ОСШ}_0$ фотонов за время разрешения системы, а соответствующая мощность

$$P_r = h\nu\bar{n}_0(2B). \tag{24.18}$$

Это совпадает с выражением (24.4) для обнаружительной способности по мощности цифрового приемника, если время разрешения $1/2B$ аналоговой системы приравнять к битовому времени $1/B_0$ цифровой системы.

Ввиду эквивалентности (24.18) и (24.4) и в силу применимости уравнения баланса мощности (24.7) к аналоговым системам соотношения между L и B_0 полученные ранее для двоичной цифровой системы можно применять к аналоговой системе с заменой B_0 на $2B$ при условии, что приемлемым эксплуатационным показателем для аналоговой системы является $\text{ОСШ}_0 = 10$. Например, волоконная линия длиной 1 км, способная передавать цифровые данные со скоростью 2 Гб/с с BER не более 10^{-9}, может также использоваться для передачи аналоговых данных с шириной полосы 1 ГГц и отношением сигнал—шум не менее 10.

Однако у аналоговых систем требуемое отношение сигнал—шум обычно намного больше 10, так что обнаружительная способность приемника должна быть намного больше 10 фотонов на время разрешения. Для высококачественных аудио- и видеосигналов, например, часто требуется отношение сигнал—шум, соответствующее 60 дБ. Это соответствует $\text{ОСШ}_0 = 10^6$ или $\bar{n}_0 = 10^6$ фотонов за время разрешения. Для аналоговых систем особенно важны дополнительные проектировочные соображения. Например, нелинейные отклики источника света и фотодетектора вызывают дополнительную деградацию сигнала и накладывают ограничения на динамический диапазон передаваемых форм волны.

24.2.4. Компенсация ослабления и дисперсии и управление ими

Компенсация ослабления

Эксплуатационные характеристики волокна с ограничением по ослаблению можно значительно улучшить с помощью оптических волоконных усилителей, расположенных на соответствующем расстоянии друг от друга вдоль

линии волоконной связи, как показано на рис. 24.21. Усилители поднимают убывающую оптическую мощность, так что принимаемая мощность остается выше обнаружительной способности приемника при больших длинах связи. Этот процесс в конечном счете ограничен шумом, вносимым самими усилителями, поскольку, в отличие от ретрансляторов, оптические усилители не регенерируют точный цифровой сигнал. Однако, прежде чем этот предел будет достигнут, в игру вступает дисперсия и система становится дисперсионно-ограниченной. Следовательно, компенсация дисперсии обязательна в системах дальней оптоволоконной связи с использованием оптических усилителей.

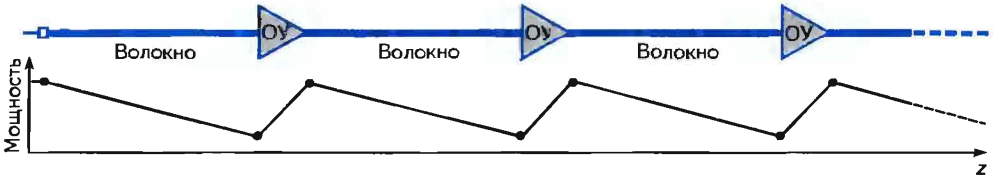


Рис. 24.21. Компенсация ослабления с использованием оптических волоконных усилителей

Компенсация дисперсии

Расплывание импульса, вносимое распространением по оптическому волокну длины L с коэффициентом дисперсии D_λ, можно обратить с использованием другого волокна, называемого волокном, компенсирующим дисперсию, с коэффициентом дисперсии D'_λ противоположного знака и длиной L', подобранной так, чтобы величины дисперсии, вносимой обоими волокнами, были равны, т. е.

$$D'_\lambda L' = -D_\lambda L. \tag{24.19}$$

Расплывание и сжатие импульса, вносимое такими чередующимися волокнами, показано на рис. 24.22. Компенсирующее волокно часто сравнительно короткое, поэтому его коэффициент дисперсии должен быть большим. Поскольку дисперсия в обычных волокнах положительна на длинах волн выше 1310 нм, компенсирующее волокно должно иметь отрицательную дисперсию в этой области. Это можно осуществить с помощью волокон со смещенной дисперсией (DSF).

Вместо DSF можно использовать другие оптические компоненты. Как описано в разд. 22.2, распространение оптического импульса через диспергирующую среду эквивалентно действию фильтра с квадратичным чирпом, у которого коэффициент чирпа равен $b = D_\lambda L$. Действие такого фильтра можно исключить с помощью обратного компенсирующего фильтра — другого фильтра с квадратичным чирпом, у которого коэффициент чирпа имеет такую же величину, но противоположный знак, $b' = -b$. Компенсирующее дисперсию волокно играет именно эту роль, однако можно использовать и другие оптические элементы, такие как решетки и интерферометры (см. разд. 22.2).

Компенсирующий фильтр может быть помещен на передающем конце линии связи для упреждающей компенсации дисперсии, которая затем вносится передающим волокном. В качестве альтернативы фильтр может быть помещен на приемном конце для компенсации уже возникшего дисперсионного расплывания импульсов непосредственно перед их детектированием. Чаще же несколько компенсирующих фильтров периодически располагаются внутри волокна, обеспечивая распределенную компенсацию. В условиях линейного распространения фактическое местоположение фильтров не важно. Однако во избежание вредного влияния нелинейных эффектов компенсирующие фильтры располагаются так, чтобы избежать формирования коротких импульсов на больших расстояниях внутри волокна.

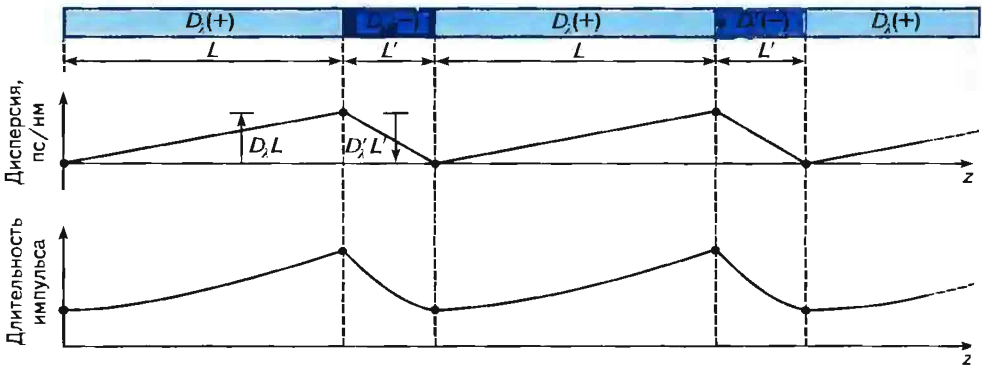


Рис. 24.22. Компенсация дисперсии с помощью участков волокна с противоположной дисперсией

Широкополосная компенсация дисперсии: управление дисперсией

Для широкополосных систем связи, таких как WDM, условие компенсации дисперсии (24.22) должно удовлетворяться на всех длинах волн в пределах спектральной полосы, т. е. ошибка $e_\lambda = D'_\lambda L' - D_\lambda L$ должна всюду равняться нулю. Поскольку коэффициенты дисперсии зависят от длины волны, это условие трудновыполнимо. На рис. 24.23 (см. цв. вклейку) показана ситуация, при которой $e_\lambda = 0$ на длине волны λ_1 в середине полосы, где компенсация полная, e_λ положительна на длине волны λ_2, что соответствует в целом положительной дисперсии, и e_λ отрицательна на длине волны λ_3, что соответствует в целом отрицательной дисперсии

Если D_λ и D'_λ — приблизительно линейные функции λ с одинаковым наклоном и если $e_\lambda = 0$ на центральной длине волны λ_1, то $e_\lambda \approx 0$ везде. Проектирование компенсирующего дисперсию фильтра с подходящей величиной коэффициента дисперсии и наклона его зависимости от длины волны известно как **управление дисперсией.**

24.2.5. Солитонная оптическая связь

Максимальная компенсация дисперсии естественно происходит в оптических солитонах. Эти нерасплывающиеся импульсы имеют интенсивность, достаточно высокую для того, чтобы в их формировании принципиальную роль играли нелинейные оптические свойства волокна. Как описано в подразд. 22.5.2, оптические солитоны — это импульсы, для которых нелинейная дисперсия (зависимость фазовой скорости от интенсивности за счет оптического эффекта Керра) полностью компенсирует линейную хроматическую дисперсию, в результате чего импульс распространяется без изменения своей ширины и формы. Усиление, обеспечиваемое волоконным усилителем, можно использовать для компенсации ослабления в волокне, так что импульс будет сохранять свою пиковую интенсивность и продолжать распространяться как солитон.

Согласно формуле (22.139), для импульса длительностью τ_0 с пиковой интенсивностью I_0 и длиной волны в свободном пространстве λ_0 на центральной частоте условие формирования солитона выражается как

$$\frac{2\pi}{\lambda_0} n_2 I_0 = \frac{-\beta''}{\tau_0^2}, \qquad (24.20)$$

Условие солитона

где n_2 — коэффициент Керра;

$$-\beta'' = \frac{\lambda_0^2}{2\pi c_0} D_\lambda$$

пропорционально коэффициенту дисперсии D_λ. Профиль интенсивности солитона описывается выражением

$$I(t) = I_0 \operatorname{sech}^2\left(\frac{t}{\tau_0}\right)$$

и имеет колоколообразную форму с шириной на высоте половины максимума, равной $1{,}76\tau_0$.

В цифровых коммуникационных системах солитон с шириной τ_0, много меньшей битового интервала T, представляет бит «1», в то время как бит «0» представляется отсутствием солитона. Это обязательно формат модуляции с возвратом к нулю.

Солитонные системы связи не ограничены ни по ослаблению, ни по дисперсии. Вместо этого основным ограничением служит нелинейная межсимвольная интерференция, возникающая из-за нелинейного взаимодействия между хвостами солитонов, представляющих соседние биты. Например, когда два идентичных солитона, разделенных битовым интервалом T, проходят достаточно большое расстояние по одному и тому же волокну, они в конце концов сливаются в один импульс, который затем снова разделяется на два исходных

импульса. Как отмечалось в подразд. 22.5.2, этот процесс повторяется периодически с периодом [см. (22.158)]

$$L_p = \pi e^{r/2} z_0, \tag{24.21}$$

где $r = T/\tau_0$ — отношение интервала между солитонами к их ширине; $2z_0$ — дисперсионное расстояние волокна,

$$2z_0 = -\frac{\tau_0^2}{\beta''} = 2\pi c_0 \frac{\tau_0^2}{\lambda_0^2} D_\lambda.$$

Период L_p экспоненциально растет с увеличением отношения r. Если $r \gg 1$, т. е. битовый интервал T много больше ширины солитона τ_0, то L_p можно сделать гораздо большим, чем z_0. Если длина волокна L много меньше L_p, то взаимодействие между соседними битами минимально. Для фиксированного отношения r условие $L \ll L_p$ можно выразить через битовую скорость $B_0 = 1/T$ как

$$LB_0^2 \ll \frac{\pi^2 c_0}{\lambda_0^2 D_\lambda} \frac{e^{r/2}}{r^2}. \tag{24.22}$$

Это устанавливает границу максимального расстояния и скорости передачи.

Пример 24.1

Солитонная система связи

Солитонная система связи передает данные со скоростью 10 Гб/с по одномодовому волокну со смещенной дисперсией на $\lambda_0 = 1550$ нм с помощью солитонных импульсов с шириной на уровне половины максимума 10 пс. На этой длине волны коэффициент дисперсии $D_\lambda = 1$ пс/(нм · км), а коэффициент нелинейности $n_2 = 2{,}6 \cdot 10^{-20}$ м^2/Вт. Эффективное сечение волокна $A_{эфф} = 60$ мкм^2. Перейдем к определению оптической мощности источника и максимальной длины линии связи.

Ширина импульса на уровне половины максимума 10 пс соответствует постоянной времени $\tau_0 = 10/1{,}76 = 5{,}7$ пс. Чтобы удовлетворить условию формирования солитона (24.20), пиковая интенсивность должна быть $I_0 = 3{,}75 \cdot 10^8$ Вт/м^2, что соответствует пиковой мощности $I_0 A_{эфф} = 22{,}5$ мВт, которую должен обеспечить источник. Дисперсионная длина волокна

$$2z_0 = \frac{2\pi c_0}{\lambda_0^2} \frac{\tau_0^2}{D_\lambda} \approx 25 \text{ км}.$$

Поскольку интервал между битами $T = 1/B_0 = 100$ пс, отношение $r = T/\tau_0 = 17{,}6$, а период взаимодействия, определяемый (24.21), равен $L_p \approx 2{,}1 \ \ 10^4 z_0$. Длина волокна должна быть намного меньше этой длины. В данном примере (24.22) представляет $LB_0^2 \ll 26$ (Тб/с)^2 км.

24.3. МОДУЛЯЦИЯ И МУЛЬТИПЛЕКСИРОВАНИЕ

24.3.1. Модуляция

Системы оптической модуляции классифицируются по оптической переменной, которая подвергается модуляции. Используется два основных типа модуляции: модуляция поля и модуляция интенсивности.

Модуляция поля

Поле монохроматической оптической волны служит синусоидальным носителем очень высокой частоты (например, 200 ТГц при $\lambda_0 = 1500$ нм). В системах амплитудной модуляции (АМ), фазовой модуляции (ФМ) и частотной модуляции (ЧМ) амплитуда, фаза или частота меняются пропорционально сигналу (рис. 24.24). Из-за предельно высокой частоты оптической несущей возможна большая ширина полосы модуляции и в принципе можно передавать очень большие объемы информации. Хотя модуляция оптического поля является очевидным обобщением обычных радиочастотных и микроволновых систем связи на оптический диапазон, ее достаточно сложно осуществить по следующим причинам:

- Для нее требуется источник, у которого амплитуда, частота и фаза стабильны и свободны от флуктуаций, т. е. лазер с высокой степенью когерентности.
- Прямую модуляцию фазы или частоты лазера обычно трудно осуществить. Например, может потребоваться внешний модулятор на основе электрооптического эффекта.
- Из-за предполагаемой высокой степени когерентности источника многомодовые волокна проявляют большой модовый шум, следовательно, нужно одномодовое волокно.
- Если не используется волокно, сохраняющее поляризацию, требуется механизм отслеживания поляризации и управления ею.
- Приемник должен быть способен измерять величину и фазу оптического поля. Это обычно достигается с помощью системы гетеродинного приема.

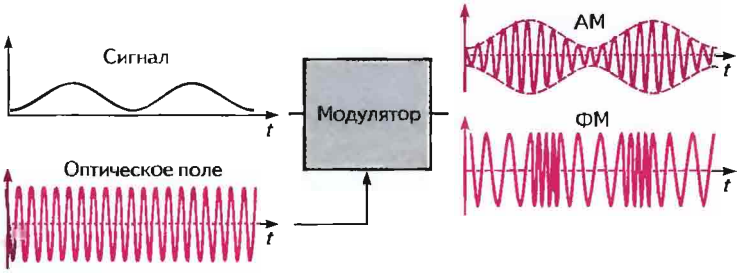


Рис. 24.24. Амплитудная и частотная модуляция оптического поля

В силу требования когерентности оптические системы связи, использующие модуляцию поля, называются **когерентными системами связи**. Эти системы обсуждаются в разд. 24.5.

Модуляция интенсивности

В системе с модуляцией интенсивности оптическая интенсивность (или мощность) пропорциональна сигналу или его кодированной версии, как показано на рис. 24.25. В большинстве серийных волоконно-оптических систем связи в настоящее время используется модуляция интенсивности. Мощность источника модулируется путем изменения тока инжекции в СИД или лазерном диоде. Волокно может быть многомодовым или одномодовым, а принимаемая оптическая мощность измеряется **приемником прямого детектирования**. Высокочастотные колебания оптического поля не играют роли в модуляции и демодуляции, только *мощность* меняется на передатчике и детектируется на приемнике. Однако длина волны света может использоваться для идентификации различных сигналов, передаваемых по одной линии, — процесс, известный как мультиплексирование с разделением по длине волны (WDM).

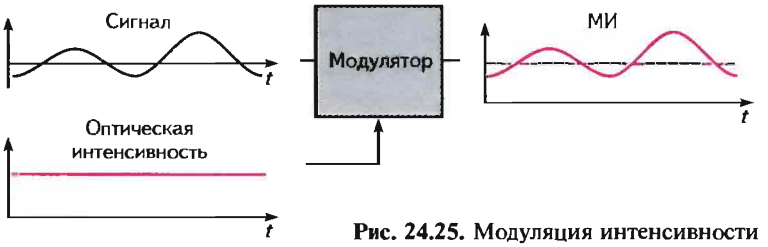


Рис. 24.25. Модуляция интенсивности

Виды модуляции

После того как выбрана модулируемая величина (интенсивность, частота или фаза), можно выбрать любой из общепринятых видов модуляции (аналоговый, импульсный или цифровой). Важным примером является **импульсно-кодовая модуляция** (ИКМ) При ИКМ производится периодическая выборка значений аналогового сигнала в равноотстоящие моменты времени с определенной частотой, каждое выбранное значение представляется одним из дискретных уровней конечного набора. Каждому уровню сопоставляется двоичный код, который передается в виде последовательности битов «1» и «0», представленных импульсами, в течение интервала времени между двумя соседними актами выборки (рис. 24.26).

При модуляции интенсивности каждый бит представляется наличием или отсутствием импульса света. Такой тип модуляции называется **амплитудной манипуляцией** (АМн). При частотной или фазовой модуляции биты представляются двумя значениями частоты или фазы. Такую модуляцию называют **частотной манипуляцией** (ЧМн) или **фазовой манипуляцией** (ФМн). Эти схемы модуляции иллюстрируются рис. 24.27. Можно также модулировать интенсивность света гармонической функцией (поднесущей), амплитуда, частота или фаза которой модулируются сигналом (в формате АМ, ЧМ, ФМ, ЧМн или ФМн).

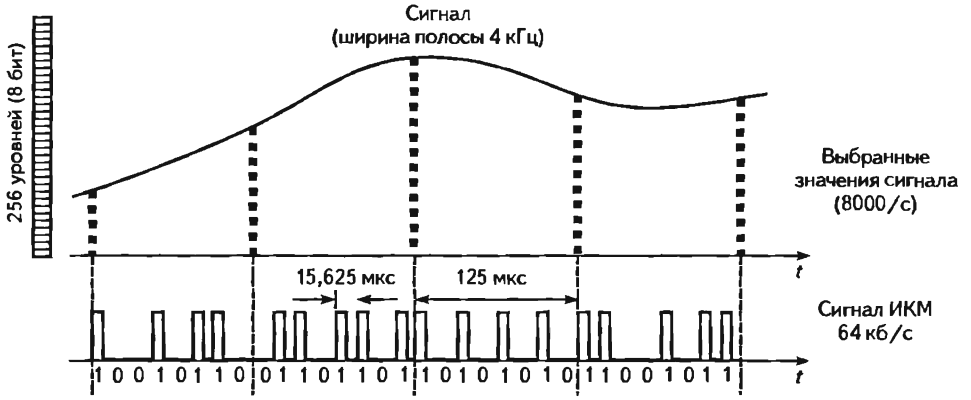


Рис. 24.26. Пример ИКМ. Голосовой сигнал с частотой 4 кГц подвергается дискретной выборке с частотой 8×10^3 значений в секунду. Дискретное представление выбранных значений осуществляется набором из $2^8 = 256$ дискретных уровней, каждому из которых сопоставляется двоичный код из 8 бит; в результате сигнал передается в виде последовательности битов с частотой 64 кб/с

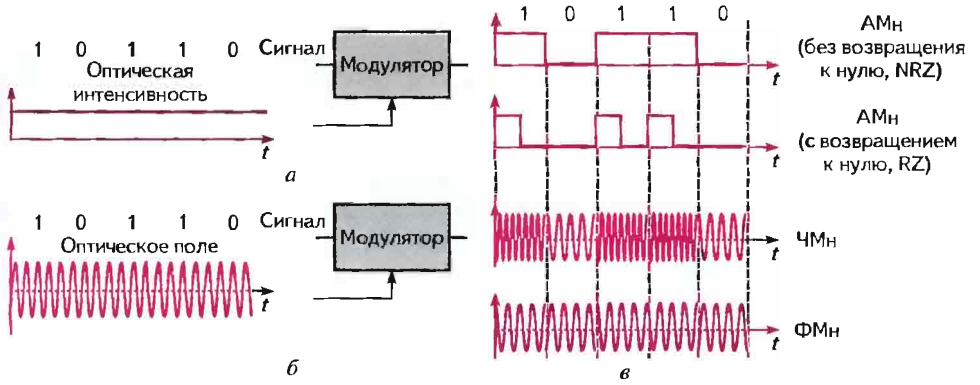


Рис. 24.27. Примеры двоичной модуляции света:

а — амплитудная манипуляция — модуляция интенсивности (АМн-МИ); *б* — частотная манипуляция (ЧМн); *в* — фазовая манипуляция (Мн) — модуляция поля

24.3.2. Мультиплексирование

Мультиплексирование позволяет передавать и принимать более одного сигнала по одной линии связи, как показано на рис. 24.28. Это достигается маркировкой каждого сигнала посредством дискретной физической метки или кода, распознаваемого приемником.

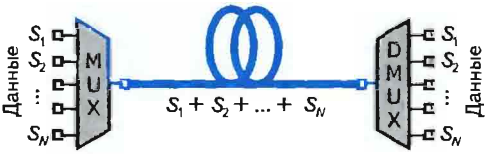


Рис. 24.28. Передача сигналов по одному каналу с использованием мультиплексора (MUX) и демультиплексора (DMUX)

Существует три стандартные системы мультиплексирования с **разделением по частоте** (FDM), **времени** (TDM) и **коду** (CDM).

FDM

При FDM несущие разных частот модулируются разными сигналами. На приемнике сигналы идентифицируются с использованием фильтров, настроенных на несущие частоты, как показано на рис. 24.29, *а*.

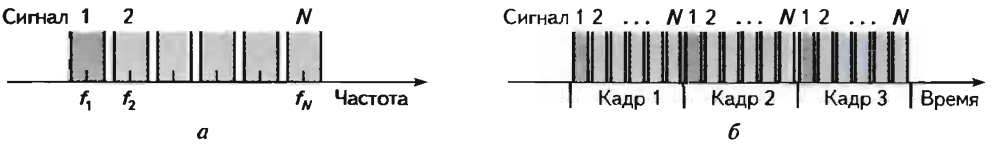


Рис. 24.29. Мультиплексирование с разделением по частоте (FDM), каждому сигналу приписана частотная полоса, центры полос образуют дискретный ряд (*а*). При мультиплексировании с разделением по времени (TDM) каждому сигналу приписана последовательность квантов времени (*б*). Кванты времени разных сигналов перемежаются

TDM

При TDM данные передаются в виде последовательности временных кадров, каждый из которых имеет набор квантов времени, приписанных битам или байтам различных сигналов, как показано на рис. 24.29, *б*. Эти биты должны синхронизироваться одним задающим генератором. На приемнике каждый сигнал идентифицируется по своему положению внутри кадра. Пример иерархической системы TDM—Т-система, показанная на рис. 24.30.

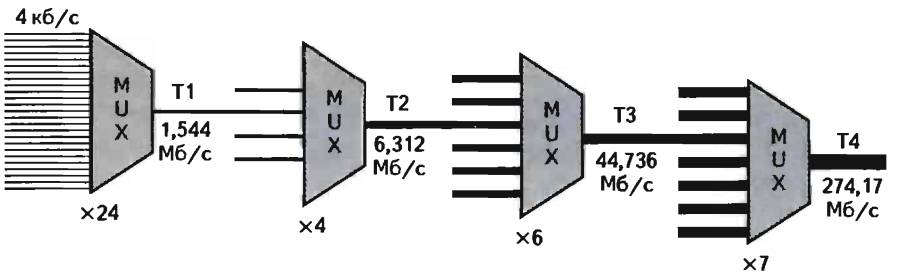


Рис. 24.30. Т-система. Набор из 24 сигналов с битовой скоростью 4 кб/с подается в мультиплексор с разделением по времени (TDM), который генерирует составной сигнал T1 на 1,544 Мб/с. Мультиплексирование четырех таких сигналом дает составной сигнал T2 и так далее

CDM

При CDM каждый сигнал сопровождается кодом (или ключом) в виде уникальной функции времени, определенной внутри интервала побитовой передачи. Коды различных сигналов должны быть некоррелированы (ортогональны),

так что их можно разделить на приемнике с помощью коррелятора. В одной из схем кодирования, показанной на рис. 24.31, каждый бит «1» исходных данных замещен кодовой последовательностью. Каждый приемник коррелирует свой собственный код с принимаемым сигналом. Принимаются только биты, записанные таким же кодом, что и у приемника, а остальные игнорируются.

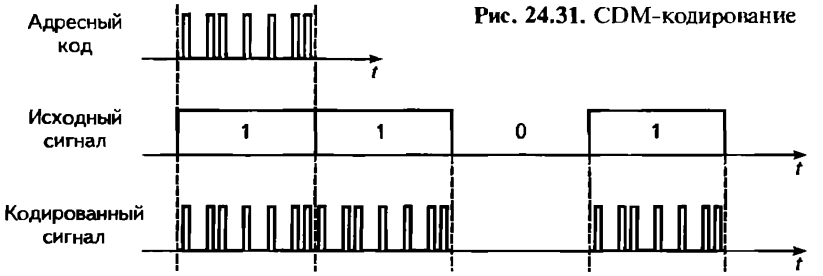


Рис. 24.31. CDM-кодирование

Мультиплексирование может быть электронным или оптическим. При **электронном** методе сигналы мультиплексируются (FDM или TDM) с образованием составного электронного сигнала, который используется для модуляции источника света по любой из схем оптической модуляции, обсуждавшихся в подразд. 24.3.1. Например, электронный FDM-сигнал можно генерировать с помощью набора несущих частот, называемых «поднесущими», которые осуществляют модуляцию интенсивности (МИ) источника света. На приемнике свет детектируется, а демультиплексирование осуществляется с помощью электронных фильтров. Другой пример — электронный TDM-сигнал, такой как T4-сигнал на рис. 24.30, используется для модуляции интенсивности источника света; демультиплексирование детектируемого сигнала осуществляется электронным способом.

При **оптическом мультиплексировании** метки, различающие уплотняемые сигналы являются оптическими по своей природе. Например, при оптическом FDM разные оптические частоты являются несущими для разных сигналов. Эти частоты разделяются на приемнике с помощью оптических фильтров.

Когда при оптическом FDM несущие частоты далеко отстоят друг от друга (например, более чем на 20 ГГц), этот вид оптического FDM называют **мультиплексированием с разделением по длине волны** (WDM). Системы WDM популярны, так как их можно использовать для увеличения пропускной способности существующей волоконной сети без прокладки дополнительного волокна.

24.3.3. Мультиплексирование с разделением по длине волны (WDM)

В системе WDM используются источники света с различными длинами волн, каждый из которых модулируется своим электрическим сигналом. Модулированные пучки света смешиваются и вводятся в волокно с помощью оптического мультиплексора (OMUX). Демультиплексирование осуществляет-

ся на приемном конце с помощью оптического демультиплексора (ODMUX), который разделяет различные длины волн и направляет их на разные детекторы. Оптические мультиплексоры и демультиплексоры описаны в подразд. 23.2.1. Электронный сигнал, ассоциированный с каждой длиной волны, часто представляет собой набор других сигналов, мультиплексированных электронным образом; в таком случае на приемном конце необходимо электронное демультиплексирование. Полная система показана на рис. 24.32.

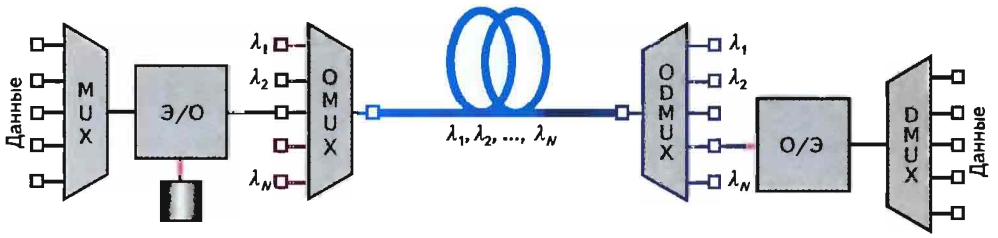


Рис. 24.32. Мультиплексирование с разделением по длине волны (WDM)

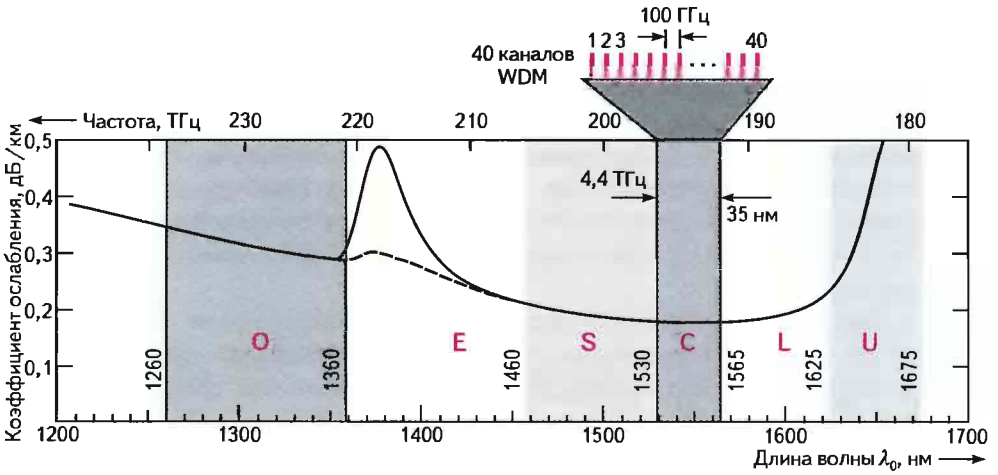


Рис. 24.33. 40-канальная система WDM в спектральной полосе C, где ослабление в волокне минимально. Расстояние между каналами 100 ГГц

Спектральные полосы, используемые в современных сетях оптоволоконной связи, показаны на рис. 24.33 вместе с ослаблением в волокнах из кварцевого стекла. Системы WDM используют любые комбинации длин волн внутри этих полос. Промежуток между длинами волн различных каналов должен достаточно превосходить спектральную ширину модулированного света в каждом канале, которая определяется шириной линии источников света, а также спектральной шириной данных, передаваемых по каналу. Расстояние между каналами должно быть достаточно большим, чтобы обеспечить возможность оптического мультиплексирования и демультиплексирования с минимальными перекрестными помехами.

Системы WDM подразделяются на две категории — грубые и плотные — в зависимости от числа каналов и расстояния между ними.

Грубые WDM- (coarse WDM, CWDM) системы имеют небольшое число каналов с далеко отстоящими длинами волн (20 нм и более). Примером является система с двумя длинами волн: одна 1310 нм, а другая 1550 нм. CWDM применяется в сетях кабельного телевидения, где разные длины волн используются для принимаемых и передаваемых сигналов. Стандарт физического уровня Ethernet LX-4 является другим примером, где используются четыре длины волны вблизи 1310 нм, каждая из которых несет поток данных 3,125 Гб/с. В сетях больших городов используются системы CWDM с длинами волн, отстоящими друг от друга на 20 нм.

Плотные WDM- (dense WDM, DWDM) системы имеют большое число каналов (обычно, более 8) с близко расположенными длинами волн. На длине волны 1550 нм в полосе С частотный интервал $\Delta\nu = 200$ соответствует интервалу по длине волны

$$\Delta\lambda = \left(\frac{\lambda_0^2}{c_0}\right)\Delta\nu = 1{,}6 \text{ нм}.$$

В системах DWDM используются такие малые частотные расстояния между каналами, как 50 ГГц или даже 25 ГГц, что соответствует интервалам длин волн 0,4 и 0,2 нм. Как показано на рис. 24.33, ширина С-полосы равна 35 нм или примерно 4,4 ТГц. В нее может поместиться 40 каналов с шагом 100 ГГц (0,8 нм). DWDM имеет тенденцию к применению на более высоком уровне (и при больших скоростях передачи данных) в иерархии систем связи, например в магистральных линиях Интернет. Проектирование DWDM-систем значительно более сложно, чем CWDM, так как лазеры должны быть значительно более стабильными, а для предотвращения теплового дрейфа длины волны часто требуется прецизионный температурный контроль.

24.4. ВОЛОКОННО-ОПТИЧЕСКИЕ СЕТИ

Сеть связи включает набор линий связи, соединяющих множество пользователей (терминалов), распределенных по некоторой географической области. Сообщения или данные могут передаваться от одного терминала к другому через одну или несколько линий связи по пути, контролируемому маршрутизаторами и коммутаторами. **Локальная сеть** (local-area network, LAN), например, соединяет терминалы (компьютеры, принтеры, видеомониторы, факсы и копировальные машины) в пределах здания, университетского городка или промышленного предприятия. К большим сетям относятся телефонные сети, глобальная сеть Телекс и Интернет. В сети могут использоваться электрические кабели, оптические волокна или спутниковая связь. В состав оптоволоконных сетей входят волоконно-оптические линии связи вместе с электронными или оптическими маршрутизаторами и коммутаторами (см. гл. 23).

24.4.1. Топологии сетей и коллективный доступ

Сеть с N узлами можно построить с помощью выделенных линий, соединяющих каждый узел со всеми остальными. Для этого требуется $N(N-1)$ дуплексных (действующих в обоих направлениях) двухточечных линий связи, $2N(N-1)$ передатчиков и $2N(N-1)$ приемников.

Топологии с меньшим числом двухточечных линий связи, приемников и передатчиков включают шинную, звездообразную и кольцевую топологии, показанные на рис. 24.34, а также систему доступа к линии связи коллективного пользования. В этих сетях необходимо иметь только N передатчиков и N приемников. В звездообразной сети каждый узел связан со всеми другими узлами через звездообразный разветвитель в центре сети; мощность, передаваемая одним узлом, распределяется поровну между всеми другими узлами. В кольцевой и шинной сетях волокно проходит через узлы и данные могут быть выделены из оптического сигнала или добавлены в него каждым узлом. Поскольку свет, переданный каждым узлом, проходит различное расстояние до различных других узлов, приемники должны быть способны обрабатывать полученную мощность на различных уровнях, т. е. иметь большой динамический диапазон. Более общей конфигурацией является ячеистая сеть.

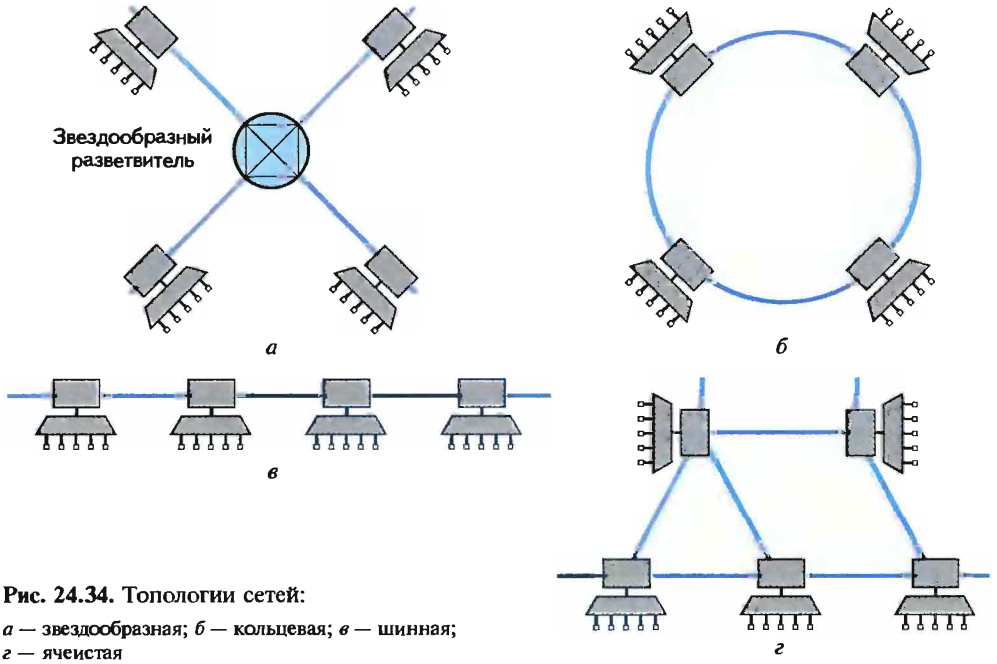


Рис. 24.34. Топологии сетей:
а — звездообразная; *б* — кольцевая; *в* — шинная; *г* — ячеистая

Несколько сетей одинаковой или различной топологии часто соединяют в большую сеть, как в примере, изображенном на рис. 24.35.

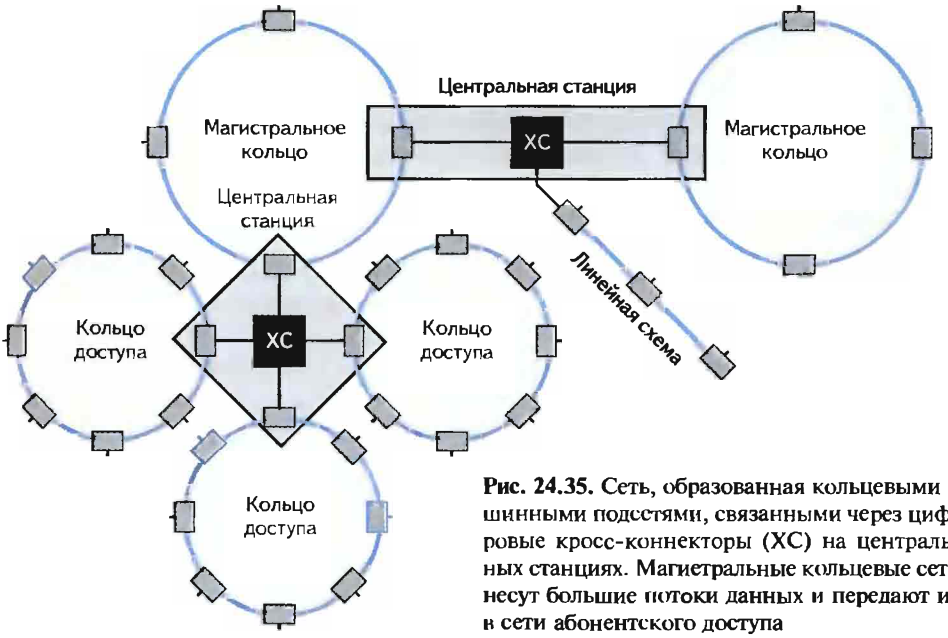


Рис. 24.35. Сеть, образованная кольцевыми и шинными подсетями, связанными через цифровые кросс-коннекторы (ХС) на центральных станциях. Магистральные кольцевые сети несут большие потоки данных и передают их в сети абонентского доступа

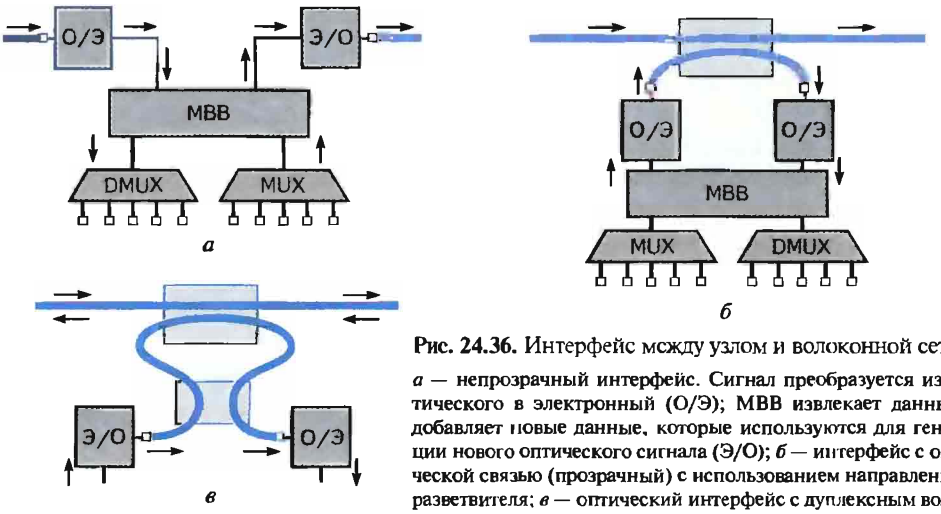


Рис. 24.36. Интерфейс между узлом и волоконной сетью:
а — непрозрачный интерфейс. Сигнал преобразуется из оптического в электронный (О/Э); МВВ извлекает данные и добавляет новые данные, которые используются для генерации нового оптического сигнала (Э/О); *б* — интерфейс с оптической связью (прозрачный) с использованием направленного разветвителя; *в* — оптический интерфейс с дуплексным волокном с использованием двух направленных разветвителей

Интерфейс

Интерфейс между терминалом и волоконной сетью на каждом узле включает приемник, передатчик и электронный мультиплексор ввода-вывода (МВВ), как показано на рис. 24.36, *а*. Приемник детектирует оптический сигнал, а МВВ извлекает данные и добавляет новые данные, которые модулируют

источник света, после чего новый оптический сигнал передается по другому волокну. Такой интерфейс называется *непрозрачным*, поскольку свет детектируется и регенерируется в узле.

Прозрачный интерфейс связан с волоконной сетью оптически, как показано на рис. 24.36, *б*. Оптические направленные разветвители описаны в разд. 23.1 и 23.3. Оптический интерфейс с двунаправленным (дуплексным) волокном включает два направленных разветвителя для передачи и приема в любом направлении, как показано на рис. 24.36, *в*.

Коллективный доступ

Сигналы, передаваемые узлами сети, совместно используют одно волокно (**среду**). Во избежание путаницы необходима схема **коллективного доступа** или **доступа к среде**. Используются системы временно́го, частотного и кодового коллективного доступа

• **Коллективный доступ с разделением по времени (time-division multiple access,** TDMA) аналогичен мультиплексированию с разделением по времени, которое используется в обычных двухточечных системах связи (см. подразд. 24.3.2). Узлы посылают свои данные через общую среду в течение перемежающихся временных промежутков (слотов). Для хранения данных в ожидании соответствующего времени могут использоваться буферы. Поскольку невозможно синхронизировать тактовое время всех узлов, необходимы защитные интервалы, отделяющие последовательные слоты друг от друга.

• **Коллективный доступ с разделением по частоте (frequency-division multiple access, FDMA)** аналогичен мультиплексированию с разделением по частоте (FDM) (см. подразд. 24.3.2). Здесь узлы посылают свои данные через общую среду в предписанных частотных полосах, и нет необходимости синхронизировать биты поступающих сигналов. В оптических сетях FDMA называется **коллективным доступом с разделением по длине волны (WDMA)**, аналогично мультиплексированию с разделением по длине волны (WDM).

• **Коллективный доступ с разделением по коду (CDMA)** аналогичен мультиплексированию с разделением по коду (CDM) (см. подразд. 24.3.2). При CDMA каждому узлу приписывается уникальный адресный код. Данные, передаваемые узлом, включают адресный код получателя. Каждый узел коррелирует свой собственный адресный код с кодом входящего сигнала. Таким образом выделяются только биты, адресованные данному узлу, а остальные биты игнорируются. Данные приходят в виде последовательности пакетов, каждый из которых имеет адрес назначения (см. подразд. 23.3.6).

Синхронная оптическая сеть (SONET).

SONET [и ее международная версия Синхронная цифровая иерархия (SDH)] — это TDM-стандарт для передачи по оптическим волокнам. Она решает проблему мультиплексирования с разделением по времени сигналов со слегка различающимися тактовыми частотами путем встраивания этих сигналов в кадры

большей длительности. Полезная нагрузка (сигнальные биты) может плавать внутри кадров, но сами кадры абсолютно синхронны. SONET обеспечивает иерархию мультиплексированных сигналов, в которой основной элемент, называемый сигналом STS-1, или оптической несущей 1 (OC-1), переносит данные со скоростью 51,84 Мб/с. Комбинация N таких сигналов образует сигнал OC-N, который имеет в N раз большую скорость передачи, как указывается в табл. 24.2. Например, OC-192 и OC-768 работают примерно на 10 и 40 Гб/с соответственно.

Таблица 24.2. **Скорости передачи данных (Мб/с) в STS иерархии, используемые в сети SONET**

OC-1	OC-3	OC-12	OC-24	OC-48	OC-192	OC-768
51,84	155,52	622,08	1244,16	2488,32	9995,33	39813,12

Пример 24.2

Кольцевая сеть

Пример оптоволоконной сети с четырьмя узлами, работающими с разными скоростями передачи данных, показан на рис. 24.37. Каждый из четырех узлов передает данные другим трем узлам с использованием либо OC-12 (≈622 Мб/с), либо OC-24 (≈1,24 Гб/с). Участок волокна, соединяющий узлы 1 и 2, пропускает наибольший поток данных с комбинированной скоростью OC-12 + OC-12 + OC-24 = OC-48 (≈2,5 Гб/с). Участки 2—3 и 3—4 пропускают меньшие потоки данных со скоростью OC-24.

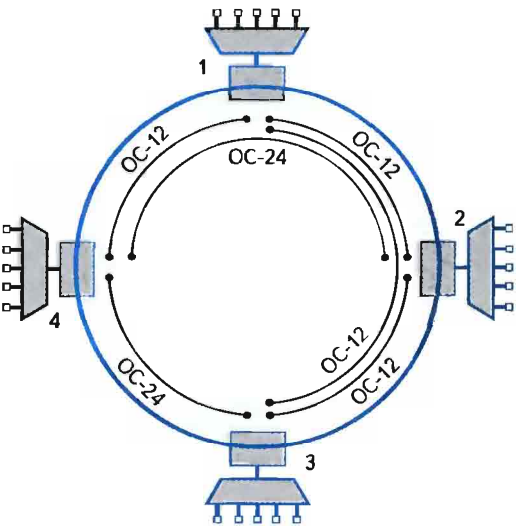


Рис. 24.37. Четырехузловая кольцевая сеть

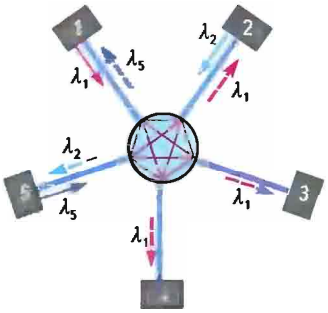
24.4.2. Сети, использующие мультиплексирование с разделением по длине волны (WDM)

При мультиплексировании с разделением по длине волны (WDM) в волоконно-оптической сети используются грубое или плотное WDM для связи по волокнам и WDMA для доступа в среду. Узлы соединены в некоторой топологии (звездообразной, кольцевой, шинной или ячеистой), и каждый узел

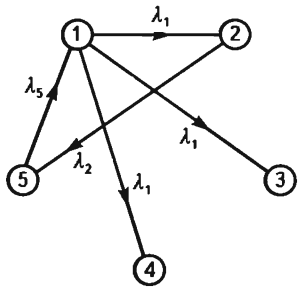
передает данные в один или несколько каналов с заданной длиной волны и получает данные из одного или нескольких таких каналов. Существование множества каналов с разными длинами волн для каждого физического соединения добавляет новое измерение в сеть и обеспечивает некоторую гибкость за счет известного усложнения.

WDM-сеть «рассылка—выбор»

Простейшей сетью WDM является сеть типа «рассылка—выбор». Каждый узел передает на уникальной фиксированной волне и *рассылает* передачу по всем другим узлам через пассивные оптические разветвители. Приемник каждого узла выбирает одну длину волны, адресованную ему, с помощью перестраиваемого фильтра. Например, в сети с пятью узлами, показанной на рис. 24.38, *а*, узлы 1, 2, ..., 5 передают на длинах волн λ_1, λ_2, ..., λ_5 соответственно. Оптический звездообразный разветвитель рассылает каждую передачу по всем другим узлам. В показанном для примера состоянии узел 1 настроен на канал λ_5, узлы 2, 3 и 4 — на канал λ_1, а узел 5 — на канал λ_2. Как показано на эквивалентной схеме соединений на рис. 24.38, *б*, узел 2 передает данные узлу 5, узел 5 — узлу 1, а узел 1 — сразу трем узлам 2, 3 и 4.

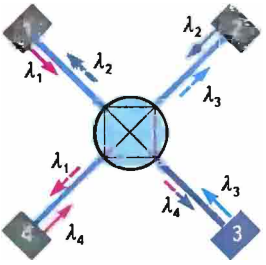


а

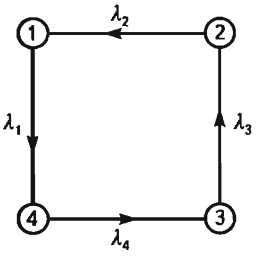


б

Рис. 24.38. Сеть WDM типа «рассылка—выбор» (*а*) и эквивалентная ей схема соединений (*б*)



а



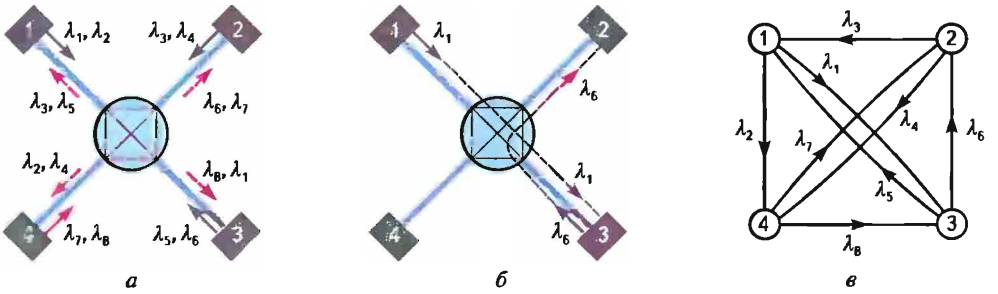
б

Рис. 24.39. Сеть WDM звездообразной физической топологией (*а*) эквивалентна кольцевой логической топологии (*б*)

В другом примере, показанном на рис. 24.39, *а*, приемник каждого узла настроен на длину волны передачи своего соседа. В результате сеть, имеющая физическую топологию «*звезда*», эквивалентна *кольцевой* сети, что видно из рис. 24.39, *б*.

Многопересылочные WDM-сети типа «рассылка—выбор»

Требование, чтобы каждый узел сети типа «рассылка—выбор» мог принимать селективно любую из длин волн, передаваемых всеми другими узлами, может быть трудно выполнить. Это требование смягчается в многопересылочной сети, у которой каждому узлу сопоставляется два разных канала для передачи и всего два разных канала для приема. В любой момент времени узел может передавать данные на одной из двух длин волн, назначенных ему для передачи, и принимать данные на одной из двух длин волн, назначенных для приема. Каналы приписаны узлам так, чтобы узел мог иметь доступ к любому другому узлу либо напрямую (за одну пересылку), либо через промежуточный узел (за две пересылки). Например, в сети, показанной на рис. 24.40, *а*, узел 2 может передавать данные на узел 1 напрямую по каналу λ_3. Хотя узел 1 не может передавать данные на узел 2 непосредственно, такая передача может произойти в два приема: 1—3 по каналу λ_1, а затем 3—2 по каналу λ_6. Поэтому такая конфигурация называется **многопересылочной сетью «рассылка—выбор»**.



а *б* *в*

Рис. 24.40. Многопересылочная WDM-сеть типа «рассылка—выбор» (*а*). Двухпересылочное соединение узла 1 с узлом 2 через узел 3 (*б*). Логическая топология сети (*в*)

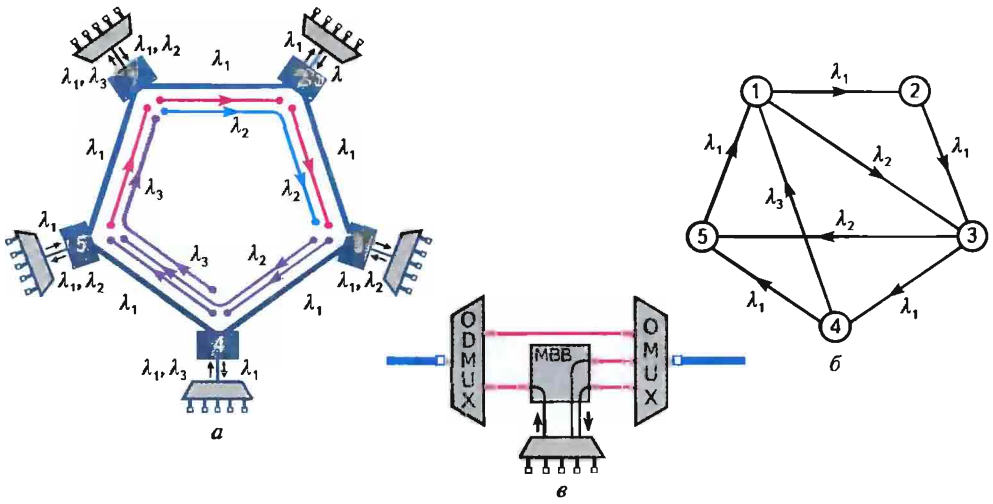
Конфигурация «рассылка—выбор» с одной или несколькими пересылками не годится для сетей с большим числом узлов. Поскольку мощность, передаваемая каждым узлом, должна достигать всех остальных узлов, система становится неэффективной для большого числа узлов. Число используемых каналов, которое должно равняться числу узлов или превышать его, также является препятствием для больших сетей.

Сети с маршрутизацией по длине волны

В **сетях с маршрутизацией по длине волны** пара узлов связывается с помощью одного из каналов, следуя некоторому маршруту соединения. Другая пара узлов может использовать канал с той же длиной волны, если их маршрут соединения не использует общих линий связи с маршрутом первой пары. Например, в сети, показанной на рис. 24.41, *а*, узлы 1 и 2 связываются на длине волны λ_1. Ту же длину волны используют для связи узлы 2 и 3. Однако узлы 1 и 3 должны использовать другую длину волны λ_2, если они используют марш-

рут, проходящий через узел 2. Аналогично, узлы 4 и 1 должны использовать для связи третий канал λ_3, поскольку их маршрут связи содержит линии, использующие каналы λ_1 и λ_2.

В такой сети каждая линия связи передает одну или несколько длин волн (но не обязательно все длины волн, как в случае сети типа «рассылка—выбор»). Например, линия связи между узлами 4 и 5 передает потоки данных на трех каналах, каждый со своей длиной волны, однако каждая из четырех других линий связи передает данные только на двух длинах волн. Каждый узел также передает и принимает данные на одной или нескольких длинах волн. Например, узел 5 принимает данные с узла 4 на λ_1, а с узла 3 на λ_2; он же передает данные на узел 1 на λ_1; данные, передаваемые по каналу λ_2, проходят через этот узел без детектирования. Логическая схема соединений в этой сети показана на рис. 24.41, *б*.



а *б* *в*

Рис. 24.41. Трехканальная пятиузловая кольцевая сеть с маршрутизацией по длине волны (*а*). Ее логическая топология (*б*). Оптический мультиплексор ввода-вывода (OMBB), используемый на узле 5 (*в*)

Ключевым элементом WDM-сети с маршрутизацией по длине волны является оптический мультиплексор ввода-вывода (OMBB) (см. подразд. 23.2.1). Каждый узел имеет OMBB, который извлекает данные из некоторых каналов входящего волокна, добавляет данные в некоторые каналы выходящего волокна и позволяет данным на некоторых каналах входящего волокна проходить без изменения в выходящее волокно. OMBB состоит из оптического демультиплексора, мультиплексора ввода-вывода (MBB) и оптического мультиплексора. В качестве примера OMBB, используемый в узле 5 сети рис. 24.41, *а*, показан на рис. 24.41, *в*. В быстродействующих сетях используются реконфигурируемые OMBB (POMMB).

Сети с маршрутизацией по длине волны с конфигурацией, отличной от кольцевой, имеют узлы с несколькими входящими и выходящими волокнами. В таких узлах нужны более сложные маршрутизаторы. Например, в узле с двумя входящими и двумя выходящими волокнами, показанном на рис. 24.42, используется **оптический центральный коммутационный узел (ОЦКУ)**, который принимает данные от выбранных входящих волокон/каналов, добавляет данные в выбранные выходящие волокна/каналы и осуществляет маршрутизацию данных между избранными входящими и входящими каналами. В ОЦКУ используются многомерные пространственно-временные волновые коммутаторы и МВВ (см. подразд. 23.3.4). В сети с маршрутизацией по длине волны используется также узел-концентратор с сервером для обработки данных на всех волновых каналах.

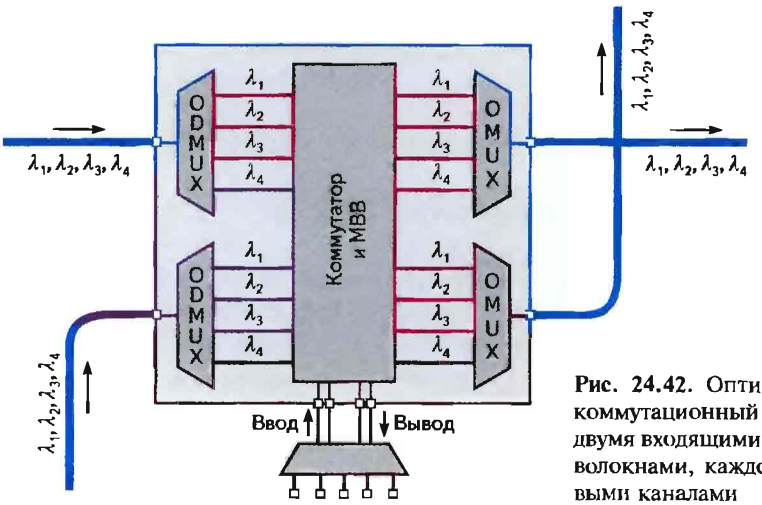


Рис. 24.42. Оптический центральный коммутационный узел (ОЦКУ) в узле с двумя входящими и двумя выходящими волокнами, каждое с четырьмя волновыми каналами

Пример 24.3

WDM-усовершенствование кольцевой сети

Кольцевая WDM-сеть с маршрутизацией по длине волны, имеющая 4 узла, работает на 3 каналах с длинами волн λ_1, λ_2 и λ_3 со скоростями передачи данных, указанными на рис. 24.43. Эта сеть — усовершенствованная версия сети, изображенной на рис. 24.37. Узлы 1 и 3 имеют доступ к каналам λ_1 и λ_2; узел 4 имеет доступ к λ_1 и λ_3; узел 2 имеет доступ ко всем трем длинам волн λ_1, λ_2 и λ_3. В усовершенствованной сети узлы связываются со скоростью вдвое больше, чем у исходной сети, но наибольшая скорость в каждом из WDM-каналов не превышает скорости в исходной сети. Участок волокна между узлами 1 и 2 несет наибольшую нагрузку, комбинированная скорость передачи здесь соответствует OC-96 (≈5 Гб/с), однако наивысшая скорость передачи в каждом на каждой из длин волн соответствует OC-48 (≈2,5 Гб/с).

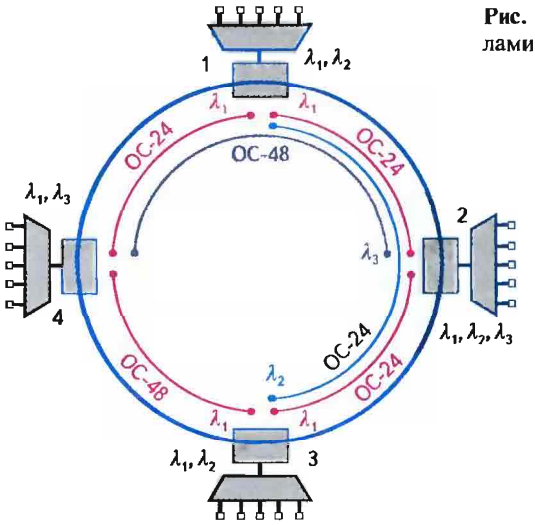


Рис. 24.43. Схема кольцевой WDM-сети с 4 узлами и 3 каналами

24.5. КОГЕРЕНТНАЯ ОПТИЧЕСКАЯ СВЯЗЬ

В когерентных системах оптической связи используется модуляция поля (амплитуды, фазы или частоты) вместо модуляции интенсивности. В них применяются когерентные источники света, одномодовые волокна и гетеродинные приемники. В данном разделе мы изучим принципы действия таких систем, определим преимущества их рабочих характеристик и кратко обсудим требования к компонентам системы.

Гетеродинные и гомодинные приемники

Фотодетекторы дают отклик на поток фотонов и, как таковые, не чувствительны к оптической фазе. Однако можно измерить комплексную амплитуду (величину и фазу) сигнального оптического поля путем смешивания его с когерентным опорным оптическим полем со стабильной фазой от дополнительного источника — **гетеродина** — и детектирования суперпозиции фотоприемником, как показано на рис. 24.44. В результате интерференции (биений) между двумя полями регистрируемый фототок содержит информацию как об амплитуде, так и о фазе сигнального поля.

Такой метод детектирования называется **оптическим гетеродинированием, оптическим смешением, фотосмешением, световыми биениями** (см. подразд. 2.6.2) или **когерентным оптическим детектированием** (в противопоставление **прямому детектированию**). Когерентный оптический приемник является оптическим аналогом супергетеродинного радиоприемника. Волны сигнала и гетеродина обычно имеют разные частоты (ν_s и ν_L). Когда $\nu_s = \nu_L$, приемник называется **гомодинным**.

Пусть $\mathcal{E}_s = \mathrm{Re}\{A_s \exp(j2\pi\nu_s t)\}$ — сигнальное оптическое поле с комплексной амплитудой $A_s = |A_s|\exp(j\varphi_s)$ и частотой ν_s. Модуль $|A_s|$ или фаза φ_s модулируется сигналом с частотой, намного меньшей ν_s. Поле гетеродина описывается аналогичным выражением с $\mathcal{E}_L$, A_L, ν_L, и φ_L. Оба поля смешиваются при помощи светоделителя или оптического разветвителя, как показано на рис. 24.44. Если падающие поля представляют собой идеально параллельные плоские волны, имеющие строго одинаковую поляризацию, то пространственная зависимость может быть подавлена и полное поле представляет собой сумму двух составляющих полей $\mathcal{E} = \mathcal{E}_s + \mathcal{E}_L$. Для квадрата модуля суммы комплексных волн получаем

$$\begin{aligned} &\left|A_s \exp(j2\pi\nu_s t) + A_L \exp(j2\pi\nu_L t)\right|^2 = \\ &= |A_s|^2 + |A_L|^2 + 2|A_s||A_L|\cos\left[2\pi(\nu_s - \nu_L)t + (\varphi_s - \varphi_L)\right]. \end{aligned} \tag{24.23}$$

Поскольку интенсивности I_s, I_L и I пропорциональны квадратам модулей комплексных амплитуд, имеем

$$I = I_s + I_L + 2\sqrt{I_s I_L}\cos\left[2\pi\nu_I t + (\varphi_s - \varphi_L)\right], \tag{24.24}$$

где $\nu_I = \nu_s - \nu_L$ — разностная частота (называемая также **промежуточной частотой**).

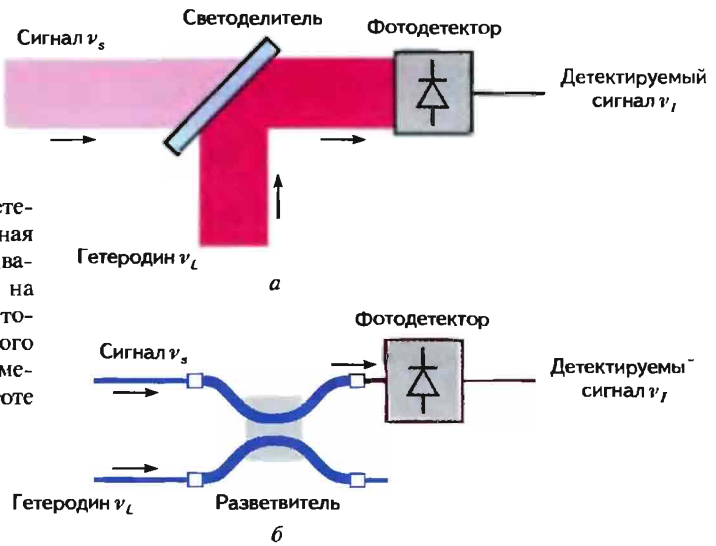


Рис. 24.44. Оптический гетеродинный прием. Сигнальная волна с частотой ν_s смешивается с волной гетеродина на частоте ν_L с помощью светоделителя (*а*) и оптического разветвителя (*б*). Фототок меняется на разностной частоте $\nu_I = \nu_s - \nu_L$

Оптическая мощность P, собираемая фотоприемником, равна произведению интенсивности и площади приемника, так что

$$P = P_s + P_I + 2\sqrt{P_s P_L}\cos\left[2\pi\nu_I t + (\varphi_s - \varphi_L)\right], \tag{24.25}$$

где P_s и P_L — мощности пучков сигнала и гетеродина соответственно. Третий член в (24.25) меняется со временем на разностной частоте ν_I с фазой $\varphi_s - \varphi_L$.

Если пучки сигнала и гетеродина близки по частоте, то разность ν_I может быть на много порядков меньше, чем каждая из частот.

Разъюстировка направлений двух волн уменьшает или размывает интерференционный вклад [третий член в (24.25)], поскольку фаза $\varphi_s - \varphi_L$ начинает зависеть от координат по синусоидальному закону в пределах площади приемника. Как легко понять из рис. 2.25, этого можно избежать, если поддерживать угол θ между волновыми фронтами достаточно малым, так чтобы выполнялось условие $\theta \ll \lambda/a$, где a — размер апертуры фотодетектора.

Фототок i, генерируемый полупроводниковым детектором фотонов, пропорционален падающему потоку фотонов Ф (см. подразд. 18.1.2). Когда ν_I намного меньше ν_s и ν_L, свет суперпозиции является квазимонохроматическим и полный поток фотонов

$$\Phi = \frac{P}{h\bar{\nu}}$$

пропорционален оптической мощности, где

$$\bar{\nu} = \frac{1}{2}(\nu_s + \nu_L).$$

Средний фототок тогда

$$\bar{i} = \eta e \Phi = \frac{\eta e}{h\bar{\nu}} P,$$

где e — заряд электрона; η — квантовый выход приемника, так что

$$\bar{i} = \bar{i}_s + \bar{i}_L + 2\sqrt{\bar{i}_s \bar{i}_L}\cos\left[2\pi\nu_I t + (\varphi_s - \varphi_L)\right], \tag{24.26}$$

где $\bar{i}_s$ и $\bar{i}_L$ — фототоки, создаваемые сигналом и гетеродином по отдельности,

$$\bar{i}_s = \frac{\eta e P_s}{h\bar{\nu}}; \quad \bar{i}_L = \frac{\eta e P_L}{h\bar{\nu}}.$$

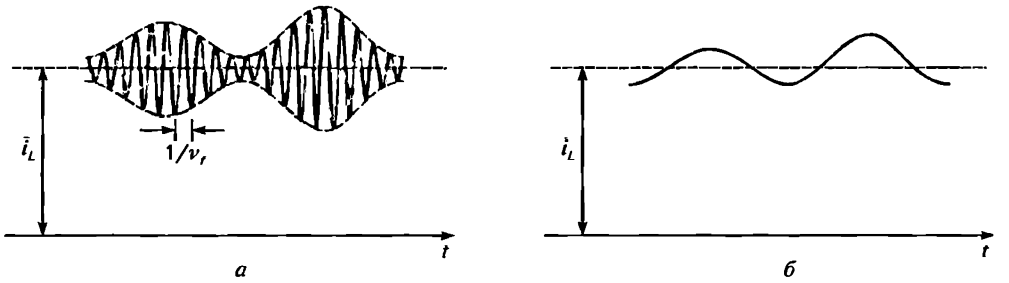
Гетеродин часто намного мощнее сигнала, так что первым слагаемым в (24.26) можно пренебречь и

$$\boxed{\bar{i} \approx \bar{i}_L + 2\sqrt{\bar{i}_s \bar{i}_L}\cos\left[2\pi\nu_I t + (\varphi_s - \varphi_L)\right].} \tag{24.27}$$

Фототок при гетеродинировании

Временна́я зависимость регистрируемого тока $\bar{i}$ схематически изображена на рис. 24.45, *а*. Второй член в (24.27), осциллирующий на разностной частоте ν_I, несет полезную информацию. По известным $\bar{i}_L$ и φ_L можно определить амплитуду и фазу этого члена и найти $\bar{i}_s$ и φ_s, откуда можно получить интенсивность и фазу (и, следовательно, комплексную амплитуду) измеряемого оптического сигнала. Информативные сигнальные переменные $\bar{i}_s$ и φ_s обычно меняются медленно по сравнению с разностной частотой ν_I и ведут себя как медленные

модуляции амплитуды и фазы гармонической функции $2\sqrt{\bar{i}_L}\cos(2\pi\nu_I t - \varphi_L)$. Этот ток, модулированный по амплитуде и фазе, можно демодулировать с привлечением обычных методов, используемых в АМ- и ЧМ-радиоприемниках.



а *б*

Рис. 24.45. Фототок, генерируемый гетеродинным приемником (*а*). Огибающая и фаза зависящей от времени компоненты несет полную информацию о комплексной амплитуде оптического поля, представляющего сигнал. Фототок, генерируемый гомодинным приемником (*б*)

С точки зрения оптики фотонов процесс можно понять как детектирование полихроматических (двухчастотных) фотонов (см. задачу 7 к разд. 12.1).

Гомодинная система — частный случай гетеродинной, у которой $\nu_s = \nu_L$ и $\nu_I = 0$. Процесс демодуляции в этом случае происходит по-другому. С помощью цепи фазовой синхронизации фаза гетеродина синхронизируется так, что $\varphi_L = 0$, после чего (24.27) дает

$$\bar{i} = \bar{i}_L + 2\sqrt{\bar{i}_s \bar{i}_L}\cos\varphi_s. \qquad (24.28)$$

Амплитудная и фазовая модуляция достигается путем изменения $\bar{i}_s$ и φ_s соответственно.

Преимущества гетеродинных и гомодинных приемников

По сравнению с приемником прямого детектирования гетеродинный приемник имеет следующие преимущества:

- Он способен измерять оптическую фазу и частоту.
- Он позволяет использовать мультиплексирование с разделением по длине волны (WDM) с меньшим расстоянием между каналами (≈100 МГц). В обычных системах с прямым детектированием расстояние между каналами порядка 100 ГГц.
- Он позволяет использовать электронное выравнивание для компенсации уширения импульсов в волокне. Уширение импульсов происходит благодаря дефазировке компонент с различными длинами волн (частотами) из-за различия их групповых скоростей. Поскольку приемник отслеживает фазу, эту дефазировку можно убрать путем соответствующей электронной фильтрации.
- При использовании сильного опорного поля гетеродинный приемник имеет внутренний коэффициент свободного от шумов усиливающего преобразования, которое эффективно увеличивает сигнал выше уровня шумов схемы.

• Он обеспечивает преимущество в 3-дБ ОСШ даже по сравнению с нешумящим приемником прямого детектирования, как показано в разд. 24.5.

• Он не чувствителен к нежелательному фоновому свету, с которым свет гетеродина не смешивается. Гетеродинирование — один из немногих способов достижения приема, ограниченного фотонным шумом, в инфракрасной области, где фоновые шумы преобладают.

Ценой этих преимуществ является усложнение системы, поскольку для гетеродинирования нужны стабильный источник опорной волны, оптический смеситель, в котором обе волны направляются строго параллельно, а также сложные схемы фазовой синхронизации.

Когерентные системы

Существенным условием правильного смешения полей гетеродина и принимаемой сигнальной волны является то, что они должны быть синхронизированы по фазе, параллельны и одинаково поляризованы, что делает возможной их интерференцию. Это предъявляет строгие требования к обоим лазерам и волокну. Лазеры должны быть одночастотными и иметь минимальные флуктуации фазы и интенсивности. Гетеродин синхронизируется по фазе с принимаемым оптическим полем с помощью управляющей системы, которая адаптивно устанавливает фазу и частоту гетеродина с помощью цепи фазовой синхронизации. Волокно должно быть одномодовым (во избежание модового шума). Оно должно сохранять поляризацию либо приемник должен содержать адаптивную систему поляризационной компенсации.

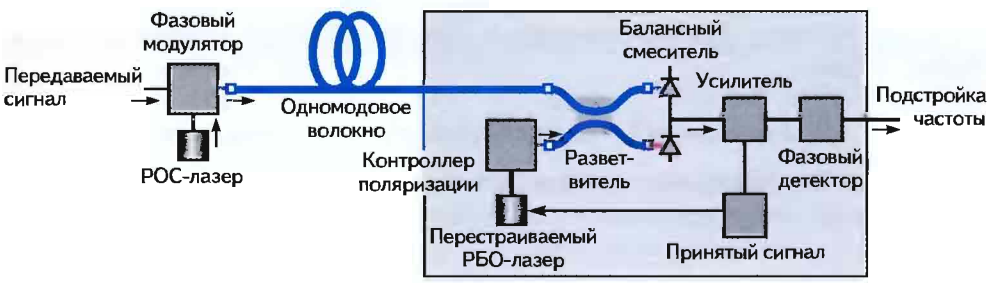


Рис. 24.46. Когерентная система оптоволоконной связи

Схема когерентной оптоволоконной системы связи с двумя лазерами и фазовой модуляцией показана на рис. 24.46. Поле гетеродина смешивается с принимаемым оптическим полем с помощью оптического направленного разветвителя. Одно плечо его выхода содержит сумму двух оптических полей, а другое — их разность. Определяемые формулой (24.26) токи детекторов

$$\bar{i}_{\pm} = \bar{i}_s + \bar{i}_L \pm 2\sqrt{\bar{i}_s \bar{i}_L}\cos\left[2\pi\nu_I t + (\varphi_s - \varphi_L)\right] \qquad (24.29)$$

вычитаются электронным образом, что дает сигнал $4\sqrt{\bar{i}_s\bar{i}_L}\cos\left[2\pi\nu_I t+(\varphi_s-\varphi_L)\right]$, который демодулируется для восстановления сообщения. Такой тип когерентного приемника называется **балансным смесителем**. Его преимущество состоит в исключении флуктуаций интенсивности гетеродина.

Ряд когерентных оптоволоконных систем связи был реализован на $\lambda_0 = 1550$ нм (где ослабление минимально) с произведениями длины на битовую скорость, соответствующими теоретическим ожиданиям. Один из примеров — система, работающая с битовой скоростью ≈4 Гб/с. РОС-лазер с шириной линии 15 МГц в непрерывном режиме подвергался прямой модуляции в формате частотной манипуляции (ЧМн). Гетеродином служил перестраиваемый РБО-лазер (см. подразд. 17.3.3). Эта система имеет чувствительность приема ≈190 фотонов/бит и использовалась для передачи по волокну длиной более 160 км.

*Рабочие характеристики аналоговой когерентной системы связи

Гетеродинный прием необходим всегда, когда нужно измерить фазу оптического поля. Однако гетеродинный прием может быть полезен также при измерении оптической интенсивности, поскольку он обеспечивает усиление благодаря присутствию сильного опорного поля. Как таковой, он представляет собой альтернативу как оптическому усилению (см. гл. 14 и разд. 17.2), так и усилению в лавинных фотодиодах (см. разд. 18.4). Это дает выигрыш в отношении сигнал—шум по сравнению с прямым детектированием, как мы покажем в настоящем разделе.

Средний фототок $\bar{i}$, генерируемый фотодиодом, сопровождается шумом с дисперсией

$$\sigma_i^2 = 2e\bar{i}B + \sigma_r^2, \tag{24.30}$$

где B — ширина полосы приемника; первый член обусловлен фотонным шумом, а второй — шумом схемы (см. разд. 18.6). Интенсивность гетеродина можно сделать достаточно большой, так что даже если сигнал слаб, полный ток $\bar{i}$ будет таким, что шум схемы σ_r^2 пренебрежимо мал по сравнению с фотонным шумом $2e\bar{i}B$.

Предполагая, что $\bar{i}_L \gg \bar{i}_s$ и $2e\bar{i}\,LB \gg \sigma_r^2$, используем (24.27) и (24.30)

$$\bar{i} \approx \bar{i}_L + 2\sqrt{\bar{i}_s\bar{i}_L}\cos\left[2\pi\nu_I t+(\varphi_s-\varphi_L)\right]; \tag{24.31а}$$

$$\sigma_i^2 \approx 2e\bar{i}_L B. \tag{24.31б}$$

В случае амплитудной модуляции сигнал представляется среднеквадратичным значением синусоидального профиля (24.31а), причем фаза игнорируется. Тогда мощность электрического сигнала

$$\frac{1}{2}\left[2\sqrt{\bar{i}_s\bar{i}_L}\right]^2 = 2\bar{i}_s\bar{i}_L,$$

а мощность шума

$$\sigma_i^2 = 2e\bar{i}_L B,$$

так что отношение сигнал—шум по мощности равно

$$\text{ОСШ} = \frac{2\bar{i}_s\,\bar{i}_L}{2e\bar{i}_L B} = \frac{\bar{i}_s}{eB}. \tag{24.32}$$

Если среднее число фотоэлектронов, дающих отсчеты за время разрешения $T = 1/2B$,

$$\bar{m} = \frac{\bar{i}}{2Be},$$

то

$$\boxed{\text{ОСШ} = 2\bar{m}.} \tag{24.33}$$

Отношение сигнал—шум
Гетеродинный приемник

Для сравнения ОСШ приемника прямого детектирования на фотодиоде, измеряющего тот же сигнальный ток $\bar{i}_s$ без использования преимуществ гетеродинирования, равно

$$\text{ОСШ} = \frac{\bar{i}_s^2}{2e\bar{i}_s B + \sigma_r^2} = \frac{\bar{m}^2}{\bar{m} + \sigma_q^2}, \tag{24.34}$$

где σ_q^2 — параметр шума схемы, обсуждавшийся в подразд. 18.6.3,

$$\sigma_q^2 = \left(\frac{\sigma_r}{2Be}\right)^2$$

Теперь очевидно принципиальное преимущество гетеродинной системы. Для сильного света или низкого уровня шума схемы ($\bar{m} \gg \sigma_q^2$) для прямого детектирования получаем ОСШ = $\bar{m}$. Гетеродинный приемник, для которого ОСШ = $2\bar{m}$, дает преимущество в два раза (3 дБ). Однако для слабого света (или большого шума схемы) преимущество еще больше, поскольку гетеродинный приемник имеет ОСШ = $2\bar{m}$, а для прямого детектирования ОСШ дополнительно уменьшается шумами схемы до значения

$$\text{ОСШ} = \frac{\bar{m}}{1 + \sigma_q^2/\bar{m}}.$$

Рабочие характеристики приемника прямого детектирования на лавинном фотодиоде также уступают характеристикам гетеродинного приемника на фотодиоде. В соответствии с (18.66), ОСШ для ЛФД с высоким усилением, призванным подавить шумы схемы, равно

$$\text{ОСШ} = \frac{\bar{m}}{F}, \tag{24.35}$$

где F — коэффициент избыточно шума ЛФД ($F > 1$). Следовательно, даже свободный от шума приемник на ЛФД ($F = 1$) вдвое уступает гетеродинному приемнику.

*Рабочие характеристики цифровых когерентных систем связи

В этом разделе определены рабочие характеристики и чувствительность цифровой коммуникационной системы в случае амплитудной и фазовой модуляции.

Гомодинная система с амплитудной манипуляцией (АМн)

Рассмотрим АМн-систему, передающую данные со скоростью B_0 бит/с, имеющую гомодинный приемник. Биты «1» и «0» представлены наличием или отсутствием сигнала $\bar{i}_s$ на протяжении битового интервала $T = 1/B_0$. Считая, что $\varphi_s = \varphi_L = 0$ и $\nu_I = \nu_s - \nu_L = 0$, для измеряемого тока получаем из (24.31а) и (24.31б) следующие средние значения и дисперсии:

$$\begin{aligned} &\text{среднее } \mu_1 \approx \bar{i}_L + 2\sqrt{\bar{i}_L \bar{i}_s}, \quad \text{дисперсия } \sigma_1^2 \approx 2\bar{i}_L eB \quad \text{для бита «1»;} \\ &\text{среднее } \mu 0 \approx \bar{i}_L, \qquad\quad \text{дисперсия } \sigma_0^2 \approx 2\bar{i}_L eB \quad \text{для бита «0».} \end{aligned} \tag{24.36}$$

Ширина полосы приемника $B = B_0/2$, поскольку битовое время $T = 1/B_0$ является шагом квантования по времени $1/2B$ для сигнала с шириной полосы $1/B$.

Рабочие характеристики двоичной системы связи в гауссовом приближении обсуждались в разд. 18.6. Частота появления ошибочных битов дается формулой (18.85), где

$$Q = \frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} = \sqrt{\frac{\bar{i}_s}{2eB}} = \sqrt{\bar{m}}; \tag{24.37}$$

$\bar{m}$ — среднее число детектируемых фотонов в бите «1»,

$$\bar{m} = \frac{\bar{i}_s}{2eB}.$$

При частоте появления ошибочных битов BER $= 10^{-9}$ $Q \approx 6$ и, следовательно, $\bar{m} = 36$, что соответствует различительной способности приемника

$$\bar{m}_a = \frac{1}{2}\bar{m} = 18$$

фотоэлектронов на бит (усредненной по обоим битам).

Таблица 24.3. **Различительная способность разных приемников и систем модуляции при идеальных условиях (в фотонах на бит)**

Система модуляции	Прямое детектирование	Гомодинный	Гетеродинный
АМн	10	18	36
ФМн	—	9	18
ЧМн	—	—	36

Гомодинная система фазовой манипуляции (ФМн)

Здесь биты «1» и «0» представляются фазовым сдвигом $\varphi_s = 0$ и π соответственно. Предполагая, что $\varphi_L = 0$, получаем из (24.31) средние значения и дисперсии фототоков для битов «1» и «0»:

$$\text{среднее } \mu_1 = \bar{i}_L + 2\sqrt{\bar{i}_L \bar{i}_s}, \quad \text{дисперсия } \sigma_1^2 = 2\bar{i}_L eB \quad \text{для бита «1»;}$$
$$\text{среднее } \mu_0 = \bar{i}_L - 2\sqrt{\bar{i}_L \bar{i}_s}, \quad \text{дисперсия } \sigma_0^2 = 2\bar{i}_L eB \quad \text{для бита «0»}$$

и, следовательно,

$$Q = \frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0} = 2\sqrt{\frac{\bar{i}_s}{2eB}} = 2\sqrt{\bar{m}}. \tag{24.38}$$

При BER = 10^{-9} $Q = 6$, откуда $\bar{m} = 9$. Поскольку в этом случае *каждый* из двух битов должен нести в среднем 9 фотоэлектронов, среднее число фотоэлектронов на бит равно $\bar{m}_a = \bar{m} = 9$. Следовательно, различительная способность приемника составляет 9 фотоэлектронов на бит. Гомодинный ФМн приемник в два раза чувствительнее гомодинного АМн приемника, потому что ему требуется в два раза меньше фотоэлектронов.

Сравнение

Различительная способность гетеродинного цифрового приемника можно определить с помощью аналогичного анализа. В табл. 24.3 перечислены различительные способности нескольких цифровых систем модуляции в предположении, что $\eta = 1$. Хотя кажется, что АМн-система прямого детектирования имеет почти такую же различительную способность, как и наилучшая когерентная система (гомодинная ФМн), на практике это не так. В гомодинной системе шум схемы подавляется, в то время как в системе прямого детектирования его нельзя игнорировать, если только не используется ЛФД. При использовании ЛФД в приемнике прямого детектирования шум схемы преодолевается, но шум усиления ухудшает различительную способность приемника, поднимая ее от 10 до по крайней мере $10F$, где F — коэффициент избыточного шума. Систе-

мы прямого детектирования могли бы иметь рабочие характеристики, сравнимые с таковыми для систем когерентного детектирования, если бы существовали идеальные ЛФД с $F = 1$ (без избыточного шума).

Рекомендуемая литература

КНИГИ ПО ВОЛОКОННО-ОПТИЧЕСКОЙ СВЯЗИ

См. также литературу к гл. 8, 9, 17, 18, 22 и 23.

Agrawal G.P. Nonlinear Fiber Optics. Academic Press, 1991; 4th ed. 2006.

Ramachandran S. Fiber-Based Dispersion Compensation. Springer-Verlag, 2006.

Damask J.N. Polarization Optics in Telecommunications. Springer-Verlag, 2005.

Ho K.-P. Phase-Modulated Optical Communication Systems. Springer-Verlag, 2005.

Palais J.C. Fiber Optic Communications. Prentice Hall, 5th ed. 2005.

Galtarossa A., Menyuk C.R., eds. Polarization Mode Dispersion, Springer-Verlag, 2005.

Lin C.-F. Optical Components for Communications: Principles and Applications. Springer-Verlag, 2004.

Kolimbiris H. Fiber Optics Communications. Prentice Hall, 2004.

Schneider T. Nonlinear Optics in Telecommunications. Springer-Verlag, 2004.

Cvijetic M. Optical Transmission Systems Engineering. Artech, 2004.

Agrawal G.P. Fiber-Optic Communication Systems. Wiley, 3rd ed. 2002.

Kaminow I.P., Li T., eds. Optical Fiber Telecommunications IVA: Components. Academic Press, 2002.

Kaminow I.P., Li T., eds. Optical Fiber Telecommunications IVB: Systems and Impairments. Academic Press, 2002.

DeCusatis C., ed. Fiber Optic Data Communication: Technological Trends and Advances. Academic Press, 2002.

DeCusatis C. Handbook of Fiber Optic Data Communication. Academic Press, 2nd ed. 2002.

Freeman R.L. Fiber-Optics Systems for Telecommunications. Wiley, 2002.

Van Stryland E.W., Bass M., eds. Fiber Optics Handbook: Fiber, Devices, and Systems for Optical Communications. McGraw-Hill, 2002.

Grote N., Venghaus H., eds. Fibre Optic Communication Devices. Springer-Verlag, 2001.

Mynbaev D.K., Scheiner L.L. Fiber-Optic Communications Technology. Prentice Hall, 2001.

Mahlke G., Gössing P. Fiber Optic Cables: Fundamentals, Cable Design, System Planning. Wiley-VCH, 4th revised and enlarged ed. 2001.

Keiser G. Optical Fiber Communications. McGraw-Hill, 3rd ed. 2000.

Sabella R., Lugli P. High Speed Optical Communications. Kluwer, 1999.

Guekos G., ed. Photonic Devices for Telecommunications: How to Model and Measure. Springer-Verlag, 1999.

Dutton H.J.R. Understanding Optical Communications. Prentice Hall, 1999.

Lachs G. Fiber Optic Communications: Systems, Analysis, and Enhancements. McGraw-Hill, 1998.

Kaminow I.P., Koch T.L., eds. Optical Fiber Telecommunications IIIA: Components. Academic Press, 1997.

Kaminow I.P., Koch T.L., eds. Optical Fiber Telecommunications IIIB: Systems and Impairments. Academic Press, 1997.

Alexander S.B. Optical Communication Receiver Design. SPIE Optical Engineering Press, 1997.

Kazovsky L., Benedetto S., Willner A.E. Optical Fiber Communication Systems. Artech, 1996.

Gagliardi R.M., Karp S. Optical Communications. Wiley, 1976; 2nd ed. 1995.

Ryu S. Coherent Lightwave Communication Systems. Artech, 1995.

Shimada S., ed. Coherent Lightwave Communications Technology. Chapman & Hall, 1995.

Senior J.M. Optical Fiber Communications: Principles and Practice. Prentice Hall, 2nd ed. 1992.

Midwinter J.E. Optical Fibers for Transmission. Wiley, 1979; Krieger, reissued 1992.

Miller S.E., Kaminow I.P., eds. Optical Fiber Telecommunications II. Academic Press, 1988.

Kao C.K. Optical Fibre. Institution of Electrical Engineers, 1988.

Personick S.D. Fiber Optics: Technology and Applications. Plenum, 1985.

Kao C.K. Optical Fiber Systems. McGraw-Hill, 1982.

Miller S.E., Chynoweth A.G., eds. Optical Fiber Telecommunications. Academic Press, 1979.

Saleh B.E.A. Photoelectron Statistics with Applications to Spectroscopy and Optical Communication. Springer-Verlag, 1978.

КНИГИ ПО ОПТОВОЛОКОННЫМ СЕТЯМ

Prucnal P.R., ed. Optical Code Division Multiple Access: Fundamentals and Applications. Taylor & Francis/CRC Press, 2006.

Mukherjee B. Optical WDM Networks. Springer-Verlag, 2006.

Desurvire E. Wiley Survival Guide in Global Telecommunications: Broadband Access, Optical Components and Networks, and Cryptography. Wiley, 2004.

Desurvire E. Wiley Survival Guide in Global Telecommunications: Signaling Principles, Protocols, and Wireless Systems. Wiley, 2004.

Bernstein G., Rajagopalan B., Saha D. Optical Network Control: Architecture, Protocols, and Standards. Addison-Wesley, 2004.

Greenfield D. The Essential Guide to Optical Networks. Prentice Hall, 2002.

Bates R.J. Optical Switching and Networking Handbook. McGraw-Hill, 2001.

Jukan A. QoS-Based Wavelength Routing in Multi-Service WDM Networks. Springer-Verlag, 2001.

Ramaswami R., Sivarajan K.N. Optical Networks: A Practical Perspective. Morgan Kaufmann, 2nd. ed. 2002.

Fatehi M.T., Wilson M. Optical Networking with WDM. McGraw-Hill, 2002.

Chaffee C.D. Building the Global Fiber Optics Superhighway. Plenum, 2001.

Stem T.E., Bala K. Multiwavelength Optical Networks: A Layered Approach. Prentice Hall, 1999.

Bononi A. Optical Networking. Springer-Verlag, 1999.

Green P.E. Fiber Optic Networks. Prentice Hall, 1993.

СТАТЬИ

Issue on optical communications. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 4, 2006.

Issue on optical communications. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 10, № 2, 2004.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Begley D.L., ed. Selected Papers on Free-Space Laser Communications II. SPIE Optical Engineering Press (Milestone Series. Vol. 100), 1994.

Rawson E.G., ed. Selected Papers on Fiber Optic Local Area Networks. SPIE Optical Engineering Press (Milestone Series. Vol. 91), 1994.

Hutcheson L.D., *Mettler S.C.*, eds. Selected Papers on Fiber Optic Communications. SPIE Optical Engineering Press (Milestone Series. Vol. 88), 1993.

Jacobs S.F. Optical Heterodyne (Coherent) Detection. American Journal of Physics. Vol. 56, 1988. P. 235—245.

Prucnal P.R., *Santoro M.A.*, *Fan T.R.* Spread Spectrum Fiber Optic Local Area Network Using Optical Processing. IEEE Journal of Lightwave Technology. Vol. LT-4, 1986. P. 547—554.

Teich M.C. Laser Heterodyning, Optica Acta. Journal of Modern Optics. Vol. 32, 1985. P. 1015—1021.

Задачи

К РАЗДЕЛУ 24.1

1. **Оптические волоконные системы.** Обсудите справедливость каждого из следующих утверждений и сформулируйте условия, при которых ваше заключение применимо.

а. Длина волны $\lambda_0 = 1300$ нм предпочтительна по сравнению с $\lambda_0 = 870$ нм для всех оптических волоконных систем связи.

б. Длина волны $\lambda_0 = 1550$ нм предпочтительна по сравнению с $\lambda_0 = 1300$ нм для всех оптических волоконных систем связи.

в. Одномодовые волокна лучше многомодовых, так как они имеют меньший коэффициент ослабления.

г. Расплывание импульса не происходит на $\lambda_0 \approx 1312$ нм в волокнах из кварцевого стекла.

д. Для оптоволоконных систем связи требуются приборы на сложных полупроводниках.

е. ЛФД сильнее шумят, чем p—i—n-фотодиоды, и потому не подходят для оптических волоконных систем.

2. **Компоненты для оптических волоконных систем.** Проектирование оптической волоконной системы связи включает множество актов выбора источников, волокон, усилителей и детекторов, некоторые из которых показаны на рис. 24.10. Сделайте необходимый выбор для каждого из приложений, перечисленных ниже. Возможно более одного правильного ответа. Некоторые сочетания, однако, могут быть несовместимыми.

а. Трансокеанский кабель, передающий данные со скоростью 2,5 Гб/с с расстоянием 100 км между ретрансляторами.

б. Кабель длиной 1 м, передающий аналоговые данные от датчика на 1 кГц.

в. Линия связи для компьютерной локальной сети, работающей при 500 Мб/с.

г. Линия связи длиной 1 км, работающая при 100 Мб/с в условиях изменения температуры ± 50 °C.

К РАЗДЕЛУ 24.2

1. **Характеристики пластиковой волоконной линии связи.** В системе связи ближнего действия с малой скоростью передачи данных используется волокно из пластика с коэффициентом ослабления 0,5 дБ/м, СИД, генерирующий 1 мВт на длине волны 870 нм, и фотодиод с обнаружительной способностью −20 дБм. Считая потери мощности на входном и выходном соединении равными по 3 дБ на каждом, определите максимальную длину связи. Считайте, что скорость передачи достаточно мала, чтобы не учитывать влияние дисперсии.

2. **Максимальная длина системы, ограниченной по ослаблению.** Линия оптоволоконной связи разрабатывается для работы при скорости 10 Мб/с. Источником служит 100-мкВт СИД, работающий на 870 нм, а волокно имеет коэффициент ослабления 3,5 дБ/км. Волокно состоит из отрезков длиной 1 км, коннекторы сегментов вносят потери 1 дБ каждый. На входном и выходном соединении теряется по 2 дБ. Запас надежности равен 6 дБ. Используется два приемника: Si p—i—n-фотодиод с обнаружительной способностью 5000 фотонов на бит и Si ЛФД с обнаружительной способностью 125 фотонов на бит. Определите обнаружительную способность $\mathcal{D}_r$ (в единицах дБм) и максимальную длину связи для каждого приемника.

3. **Максимальная скорость передачи данных системы с ограничением по ослаблению.** Линия связи длиной 50 км работает на длине волны 1550 нм. Источником является InGaAsP-лазер мощностью 2 мВт, а волокно имеет коэффициент ослабления 0,2 дБ/км. Коннекторы и разветвители вносят полные потери 8 дБ, а запас надежности составляет 6 дБ. Приемником является ЛФД из InGaAs с обнаружительной способностью 1000 фотонов на бит при частоте появления ошибочных битов 10^{-9}. Определите максимальную скорость передачи данных, считая систему ограниченной по ослаблению. Если требуемая частота появления ошибочных битов 10^{-11}, чему равна максимальная скорость передачи данных?

4. **Максимальная длина аналоговой связи.** Оптическая волоконная линия связи использует модуляцию интенсивности для передачи данных в полосе $B = 10$ МГц с отношением сигнал—шум 40 дБ. Источником служит светоизлучающий диод на $\lambda_0 = 870$ нм, дающий среднюю мощность в 100 мкВт с максимальной глубиной модуляции 0,5. Волокно со ступенчатым профилем показателя преломления имеет коэффициент ослабления 2,5 дБ/км. Приемником служит лавинный фотодиод со средним усилением $\bar{G} = 100$, коэффициентом избыточного шума $F = 5$ и токовой чувствительностью 0,5 А/Вт (исключая усиление). Предполагая, что шумом схемы можно пренебречь, используйте теорию из подразд. 18.6.4 для расчета обнаружительной способности приемника по оптической мощности и ограниченной ослаблением максимальной длины волокна L.

5. **Баланс времени для системы с ограничением по дисперсии.** Одномодовая волоконная линия связи длиной 100 км работает на длине волны 1550 нм. Источником является InGaAsP лазерный диод со спектральной шириной 0,2 нм и временем отклика 20 пс. Волокно имеет коэффициент дисперсии 17 пс/(км·нм). Приемник использует InGaAs ЛФД и имеет время отклика 0,1 нс. Определите максимальную скорость передачи данных на основании критерия, что время

отклика волокна не должно превышать 25 % времени передачи одного бита. Определите также максимальную скорость передачи данных на основании критерия, что время отклика всей системы не больше 70 % времени передачи бита. Если заменить волокно на таковое со смещенной дисперсией, чтобы коэффициент дисперсии понизился до 1 пс/(км нм), то какова будет максимальная скорость передачи, удовлетворяющая обоим критериям?

К РАЗДЕЛУ 24.3

1. **Число WDM-каналов**. Сколько WDM-каналов укладывается в С-полосе (1530—1565 нм) и в О-полосе (1260—1360 нм), если частотное расстояние между каналами 75 ГГц?

2. **Число узлов WDM-сети типа «рассылка—выбор»**. Максимальное число узлов N для WDM сети типа «рассылка—выбор» часто ограничено доступной оптической мощностью. Определите для локальной сети с использованием звездообразного разветвителя, соединенного с каждым узлом посредством волокна длиной 2 км с коэффициентом ослабления 0,3 дБ/км и потерями на коннекторах 1 дБ. Звездообразный разветвитель распределяет мощность равномерно между выходами и вносит дополнительные потери 3 дБ. Каждый узел имеет источник мощностью 1 мВт и приемник с различительной способностью −35 дБ. Предполагается, что запас надежности должен составлять 5 дБ.

3. **Кольцевая WDM-сеть с маршрутизацией по длине волны**. Рассмотрим WDM-сеть с 4 узлами и 6 каналами. Каждый узел имеет мультиплексор ввода-вывода для пропускания или приема на каждой из трех приписанных ему разных длин волн. Остальные длины волн свободно проходят через мультиплексор. Например, узел 1 может извлекать или добавлять данные в каналах λ_1, λ_2 и λ_3, но пропускает данные на λ_4, λ_5 и λ_6. Назначьте наборы из трех каналов ввода—вывода каждому узлу 2, 3 и 4 так, чтобы каждый узел на кольце мог связываться с каждым из других узлов. Идея в том, чтобы каждый узел имел один общий канал ввода-вывода с каждым из трех других узлов, но этот канал не был общим с промежуточными узлами.

ПРИЛОЖЕНИЕ А ПРЕОБРАЗОВАНИЕ ФУРЬЕ

В данном приложении дается краткий обзор преобразования Фурье и его свойств для функций одной и двух переменных.

А.1. ОДНОМЕРНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Гармоническая функция $F\exp(j2\pi\nu t)$ играет важную роль в науке и технике. Она имеет частоту ν и комплексную амплитуду F. Ее действительная часть $|F|\cos(2\pi\nu t + \arg\{F\})$ — функция косинус с амплитудой $|F|$ и фазой $\arg\{F\}$. Переменная t обычно представляет время, частота ν измеряется числом колебаний в секунду (Гц). Гармонические функции являются строительными блоками, из которых другие функции можно получить простой суперпозицией.

В соответствии с теоремой Фурье комплекснозначная функция $f(t)$, удовлетворяющая некоторым достаточно мягким условиям, может быть разложена в интеграл, имеющий вид суперпозиции гармонических функций с различными частотами и комплексными амплитудами:

$$f(t) = \int_{-\infty}^{\infty} F(\nu)\exp(j2\pi\nu t)\,d\nu. \qquad \text{(A.1)}$$

Обратное преобразование Фурье

Компонента на частоте ν имеет комплексную амплитуду $F(\nu)$, даваемую формулой

$$F(\nu) = \int_{-\infty}^{\infty} f(t)\exp(-j2\pi\nu t)\,dt. \qquad \text{(A.2)}$$

Преобразование Фурье

$F(\nu)$ называется **Фурье-образом** $f(t)$ и получается **прямым преобразованием Фурье**, а $f(t)$ — обратным преобразованием $F(\nu)$. Функции $f(t)$ и $F(\nu)$ образуют Фурье-пару: если одна из них известна, другая может быть определена.

В этой книге мы придерживаемся соглашения, что $\exp(j2\pi\nu t)$ — гармоническая функция с положительной частотой, а $\exp(-j2\pi\nu t)$ — с отрицательной. Противоположное соглашение используется многими другими авторами, кото-

рые определяют преобразование Фурье (А.2) со знаком «плюс», а обратное преобразование Фурье (А.1) со знаком «минус» в экспоненте.

В теории связи функции $f(t)$ и $F(\nu)$ представляют сигнал, причем $f(t)$ — его временное, а $F(\nu)$ — частотное представление. Квадрат модуля $|f(t)|^2$ называется мощностью сигнала, а $|F(\nu)|^2$ — спектральной плотностью энергии. Если $|F(\nu)|^2$ занимает широкую область частот, сигнал называется широкополосным.

Свойства преобразования Фурье

Ниже приведены некоторые важные свойства преобразования Фурье. Эти свойства можно доказать, непосредственно применяя определения (А.1) и (А.2) (см. любую книгу из списка рекомендуемой литературы).

- *Линейность.* Фурье-образ суммы двух функций равен сумме их Фурье-образов.
- *Масштабирование.* Если $f(t)$ имеет Фурье-образ $F(\nu)$ и τ — действительный масштабный множитель, то $f(t/\tau)$ имеет Фурье-образ $|\tau|F(\tau\nu)$. Это значит, что если $f(t)$ масштабируется множителем τ, то ее Фурье-образ масштабируется множителем $1/\tau$. Например, если $\tau > 1$, то $f(t/\tau)$ по сравнению с $f(t)$ растягивается, а $F(\tau\nu)$ по сравнению с $F(\nu)$ сжимается. Фурье-образ $f(-t)$ есть $F(-\nu)$.
- *Сдвиг по времени.* Если $f(t)$ имеет Фурье-образ $F(\nu)$, то Фурье-образ $f(t-\tau)$ равен $\exp(-j2\pi\nu\tau)F(\nu)$. Каким образом, временна́я задержка τ эквивалентна умножению Фурье-образа на фазовый множитель $\exp(-j2\pi\nu\tau)$.
- *Сдвиг по частоте.* Если $F(\nu)$ — Фурье-образ $f(t)$, то Фурье-образом $f(t)\exp(j2\pi\nu_0 t)$ является $F(\nu-\nu_0)$. Таким образом, умножение на гармоническую функцию частоты ν_0 эквивалентно сдвигу Фурье-образа по частоте вверх на величину ν_0.
- *Симметрия.* Если функция $f(t)$ действительна, то $F(\nu)$ имеет эрмитову симметрию, т. е. $F(-\nu) = F^*(\nu)$. Если $f(t)$ действительна и симметрична, то $F(\nu)$ также действительна и симметрична.
- *Теорема свертки.* Если $F_1(\nu)$ и $F_2(\nu)$ — Фурье-образы $f_1(t)$ и $f_2(t)$ соответственно и обратное преобразование Фурье от произведения

$$F(\nu) = F_1(\nu)F_2(\nu) \tag{А.3}$$

то

$$f(t) = \int_{-\infty}^{\infty} f_1(\tau)f_2(t-\tau)\,d\tau. \tag{А.4}$$

Свертка

Операция, определяемая выражением (А.4), называется сверткой $f_1(t)$ с $f_2(t)$. Свертка во временно́й области эквивалентна умножению в частотной области.

- *Теорема корреляции.* Функцией корреляции между двумя комплекснозначными функциями называется

$$f(t) = \int_{-\infty}^{\infty} f_1^*(\tau)f_2(t+\tau)\,d\tau. \tag{А.5}$$

Корреляционная функция

Фурье-образы $f_1(t)$, $f_2(t)$ и $f(t)$ связаны соотношением

$$F(\nu) = F_1^*(\nu) F_2(\nu). \tag{A.6}$$

• *Теорема Парсеваля.* Энергия сигнала, равная интегралу от мощности сигнала $|f(t)|^2$, равна интегралу от спектральной плотности энергии $|F(\nu)|^2$:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \int_{-\infty}^{\infty} |F(\nu)|^2 \, d\nu. \tag{A.7}$$

Теорема Парсеваля

Примеры

Фурье-образы некоторых важных функций, используемых в этой книге, приведены в табл. А.1. С помощью свойств линейности, масштабирования, задержки и частотного сдвига можно легко получить преобразование Фурье других функций. В указанной таблице:

• rect $(t) \equiv 1$ при $|t| \le 1/2$ и 0 во всех остальных точках, т. е. это импульс единичной высоты и единичной ширины с центром в точке $t = 0$;

• $\delta(t)$ — импульсная функция (дельта-функция Дирака), определяемая формулой

$$\delta(t) = \lim_{\alpha \to \infty} \alpha \, \text{rect}(\alpha t).$$

Это предел прямоугольного импульса единичной площади при ширине, стремящейся к нулю (при этом высота стремится к бесконечности);

• sinc $(t) = \sin(\pi t)/(\pi t)$ — симметричная функция с пиковым значением 1 в точке $t = 0$ и нулями при $t = \pm 1, \pm 2,$

А.2. ДЛИТЕЛЬНОСТЬ И СПЕКТРАЛЬНАЯ ШИРИНА

Часто необходимо иметь количественную меру ширины функции. Ширина функции времени $f(t)$ — это продолжительность импульсного сигнала, а ширина Фурье-образа $F(\nu)$ — это спектральная ширина, или ширина полосы сигнала. Поскольку не существует единственного определения ширины, используется множество разных определений. *Всем им, тем не менее, присуще общее свойство: спектральная ширина обратно пропорциональна временно́й ширине в соответствии со свойством масштабирования Фурье-преобразования.* В разных местах этой книги используются следующие определения.

Среднеквадратичная ширина

Среднеквадратичная ширина σ_τ неотрицательной функции $f(t)$ определяется соотношением

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} (t-\bar{t})^2 f(t)\,dt}{\int_{-\infty}^{\infty} f(t)\,dt}, \tag{A.8}$$

где

$$\bar{t} = \frac{\int_{-\infty}^{\infty} t f(t)\,dt}{\int_{-\infty}^{\infty} f(t)\,dt}.$$

Если $f(t)$ описывает распределение массы (t указывает положение точки), то $\bar{t}$ представляет собой центр масс, а σ_t — радиус инерции. Если $f(t)$ представляет собой функцию распределения вероятности (плотность вероятности), то указанные величины определяют среднее значение и стандартное отклонение соответственно. Например, гауссова функция

$$f(t) = \exp\left(-\frac{t^2}{2\sigma_t^2}\right)$$

имеет среднеквадратичную ширину σ_t. Ее Фурье-образ дается выражением

$$F(\nu) = \left(\frac{1}{\sqrt{2\pi}}\sigma_\nu\right)\exp\left(-\frac{\nu^2}{2\sigma_\nu^2}\right),$$

где σ_ν — среднеквадратичная спектральная ширина,

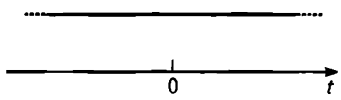
$$\sigma_\nu = \frac{1}{2\pi\sigma_t}. \tag{A.9}$$

Это определение не подходит для функций с отрицательными или комплексными значениями. Для таких функций используется среднеквадратичная ширина квадрата модуля функции $|f(t)|^2$

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} (t-\bar{t})^2 |f(t)|^2\,dt}{\int_{-\infty}^{\infty} |f(t)|^2\,dt},$$

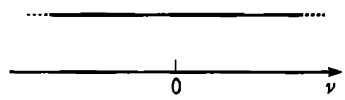
Таблица А. 1. **Некоторые функции и их Фурье-образы**

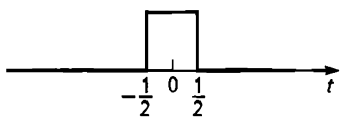
Функция	$f(t)$	$F(\nu)$
Постоянная	1	$\delta(\nu)$
Импульс	$\delta(t)$	1
Прямоугольная	$\mathrm{rect}\,(t)$	$\mathrm{sinc}\,(\nu)$
Экспоненциальная*	$\exp(-\lvert t\rvert)$	$\dfrac{2}{1+(2\pi\nu)^2}$
Гауссова	$\exp(-\pi t^2)$	$\exp(-\pi\nu^2)$

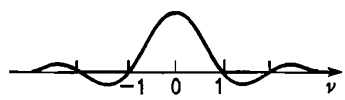


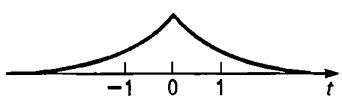


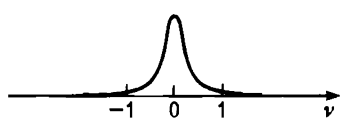


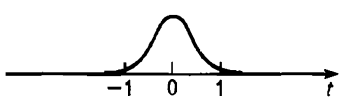


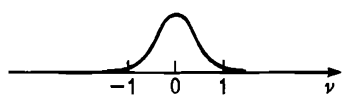


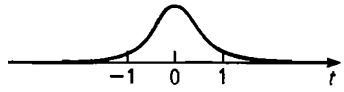
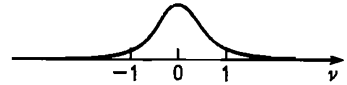
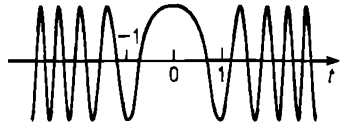
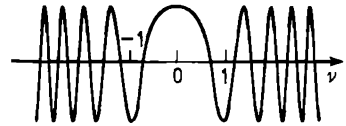
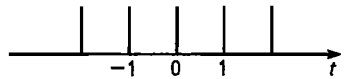
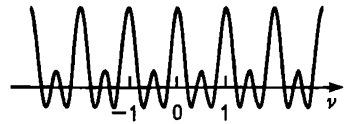
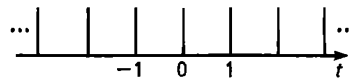
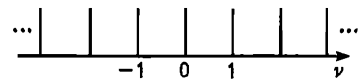










Гиперболический секанс		$\operatorname{sech}(\pi t)$	$\operatorname{sech}(\pi\nu)$ 
Чирп**		$\exp(j\pi t^2)$	$e^{j\pi/4}\exp(j\pi\nu^2)$ 
Импульсы $M = 2S + 1$		$\sum_{m=-S}^{S} \delta(t-m)$	$\dfrac{\sin(M\pi\nu)}{\sin(\pi\nu)}$ 
Гребенка		$\sum_{m=-\infty}^{\infty} \delta(t-m)$	$\sum_{m=-\infty}^{\infty} \delta(\nu-m)$ 

* Показана двухсторонняя экспоненциальная функция. Фурье-образ односторонней экспоненты $f(t) = \exp(-t)$ при $t \geq 0$ равен $F(\nu) = \dfrac{1}{1 + j2\pi\nu}$. Его модуль $\dfrac{1}{\sqrt{1 + (2\pi\nu)^2}}$.

** Показаны функции $\cos(\pi t^2)$ и $\cos(\pi\nu^2)$. Функция $\sin(\pi t^2)$ показана на рис. 4.27.

где

$$\bar{t} = \frac{\int_{-\infty}^{\infty} t|f(t)|^2 \, dt}{\int_{-\infty}^{\infty} |f(t)|^2 \, dt}.$$

Мы называем такой вариант σ_t *среднеквадратичной шириной по мощности.*

С помощью неравенства Шварца можно показать, что произведение среднеквадратичных ширин по мощности для любой функции $f(t)$ и ее Фурье-образа $F(\nu)$ должно быть не меньше $1/4\pi$

$$\boxed{\sigma_t \sigma_\nu \geq \frac{1}{4\pi}}, \qquad \text{(A.10)}$$

Обратное соотношение длительности и ширины полосы

где спектральная ширина σ_ν определяется как

$$\sigma_\nu^2 = \frac{\int_{-\infty}^{\infty} (\nu - \bar{\nu})^2 |F(\nu)|^2 \, d\nu}{\int_{-\infty}^{\infty} |F(\nu)|^2 \, d\nu},$$

где

$$\bar{\nu} = \frac{\int_{-\infty}^{\infty} \nu |F(\nu)|^2 \, d\nu}{\int_{-\infty}^{\infty} |F(\nu)|^2 \, d\nu}.$$

Таким образом, спектральная ширина и длительность импульса не могут быть одновременно как угодно малыми. Например, *гауссова функция*

$$f(t) = \exp\left(-\frac{t^2}{4\sigma_t^2}\right)$$

имеет среднеквадратичную ширину по мощности σ_t. Ее Фурье-образ $F(\nu)$ — также гауссова функция:

$$F(\nu) = \left(\frac{1}{2\sqrt{\pi}\sigma_\nu}\right)\exp\left(\frac{-\nu^2}{4\sigma_\nu^2}\right)$$

с среднеквадратичной шириной по мощности, равной

$$\sigma_\nu = \frac{1}{4\pi\sigma_t}. \qquad \text{(A.11)}$$

Поскольку $\sigma_t \sigma_\nu = 1/4\pi$, гауссова функция имеет минимальное допустимое произведение длительности на спектральную ширину $\omega = 2\pi\nu$. В угловых частотах

$$\sigma_t \sigma_\omega \geq \frac{1}{2}. \tag{A.12}$$

Если переменные t и ω, которые обычно описывают время и угловую частоту (рад/с), заменить координатной переменной x и пространственной угловой частотой k (рад/м) соответственно, то (А.12) превращается в

$$\sigma_x \sigma_k \geq \frac{1}{2}. \tag{A.13}$$

В квантовой механике положение частицы x описывается волновой функцией $\psi(x)$, а волновое число k описывается функцией $\phi(k)$, которая представляет собой Фурье-образ функции $\psi(x)$. Неопределенности x и k — среднеквадратичные ширины плотностей вероятности $|\psi(x)|^2$ и $|\phi(k)|^2$ соответственно, так что σ_x и σ_k интерпретируются как неопределенности координаты и волнового числа. Так как импульс частицы равен $p = \hbar k$, где $\hbar = h/2\pi$; h — постоянная Планка, неопределенность координаты-имиульса удовлетворяет неравенству

$$\boxed{\sigma_x \sigma_p \geq \frac{\hbar}{2},} \tag{A.14}$$

Соотношение неопределенностей Гайзенберга

известному как **соотношение неопределенностей Гайзенберга.**

Энергоэквивалентная ширина

Энергоэквивалентная ширина сигнала $f(t)$ — это энергия сигнала, деленная на его пиковую мощность. Например, если $f(t)$ имеет пиковое значение при $t = 0$, то ее энергоэквивалентная ширина есть

$$\tau = \int_{-\infty}^{\infty} \frac{|f(t)|^2}{|f(0)|^2} \mathrm{d}t. \tag{A.15}$$

Например, двухсторонняя экспоненциальная функция

$$f(t) = \exp\left(-\frac{|t|}{\tau}\right)$$

имеет энергоэквивалентную ширину τ, как и гауссова функция

$$f(t) = \exp\left(-\frac{\pi t^2}{2\tau^2}\right).$$

Это определение использовано в разд. 11.1, где время когерентности света определено как энергоэквивалентная ширина комплексной степени временно́й когерентности.

Аналогично, энергоэквивалентная спектральная ширина определяется как

$$\mathcal{B} = \int_{-\infty}^{\infty} \frac{|F(\nu)|^2}{|F(0)|^2} \mathrm{d}\nu. \tag{A.16}$$

Если $f(t)$ действительная, так что $|F(\nu)|^2$ симметрична, и если ее пик приходится на $\nu = 0$, то энергоэквивалентная ширина обычно определяется как положительно-частотная ширина

$$B = \int_{0}^{\infty} \frac{|F(\nu)|^2}{|F(0)|^2} \mathrm{d}\nu. \tag{A.17}$$

Например, в случае

$$F(\nu) = \frac{\tau}{1 + j2\pi\nu\tau},$$

$$B = \frac{1}{4\tau}. \tag{A.18}$$

Это определение используется в подразд. 18.6.1 для описания ширины полосы фотоприемников, чувствительных к фотонным и схемным шумам (см. также задачу 5 к разд. 18.5).

Используя теорему Парсеваля (A.7) и соотношение

$$F(0) = \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t,$$

формулу (A.17) можно записать в виде

$$B = \frac{1}{2T}, \tag{A.19}$$

где T — еще одно определение длительности [квадрат площади под графиком $f(t)$, деленный на площадь под графиком $f(t)^2$],

$$T = \frac{\left[\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t\right]^2}{\int_{-\infty}^{\infty} f^2(t)\,\mathrm{d}t}. \tag{A.20}$$

В этом случае произведение длительности на ширину полосы: $BT = 1/2$.

Ширина на уровне 1/e-, половины максимума и 3-дБ

Другой тип меры для ширины функции — это длительность на уровне заданной части ее максимального значения (например, $1/\sqrt{2}$, 1/2, $1/e$ или $1/e^2$). Используется либо полуширина, либо полная ширина по обе стороны пика. Две общепринятые меры — это полная ширина на уровне половины максимума (full-width-at-half-maximum, FWHM) и полуширина на уровне $1/\sqrt{2}$-максимума, называемая 3-дБ-шириной. Вот три важных примера.

- *Экспоненциальная функция* $f(t) = \exp(-t/\tau)$ при $t \geq 0$ и $f(t) = 0$ при $t < 0$, описывающая отклик множества электрических и оптических систем, имеет ширину на уровне $1/e$ максимума $\Delta t_{1/e} = \tau$. Модуль ее Фурье-образа $F(\nu) = \tau/(1 + j2\pi\nu\tau)$ имеет 3-дБ-ширину (полуширину на уровне $1/\sqrt{2}$-максимума)

$$\Delta\nu_{\text{3-дБ}} = \frac{1}{2\pi\tau}. \tag{A.21}$$

- *Двухсторонняя экспоненциальная функция*

$$f(t) = \exp\left(-\frac{|t|}{\tau}\right)$$

имеет полуширину на уровне $1/e$ максимума $\Delta t_{1/e} = \tau$. Ее Фурье-образ

$$F(\nu) = \frac{2\tau}{1 + (2\pi\nu\tau)^2},$$

известный как распределение Лоренца, имеет полную ширину на уровне половины максимума

$$\Delta\nu_{\text{ПШПМ}} = \frac{1}{\pi\tau} \tag{A.22}$$

и обычно записывается в виде

$$F(\nu) = \frac{\Delta\nu/2\pi}{\nu^2 + (\Delta\nu/2)^2},$$

где $\Delta\nu = \Delta\nu_{\text{ПШПМ}}$. Распределение Лоренца в ряде случаев описывает спектр испускания света (см. подразд. 13.3.4).

- *Функция Гаусса*

$$f(t) = \exp\left(-\frac{t^2}{2\tau^2}\right),$$

имеет полную ширину на уровне $1/e$ максимума $\Delta t_{1/e} = 2\sqrt{2}\,\tau$. Ее Фурье-образ

$$F(\nu) = \sqrt{2\pi}\tau\exp\left(-2\pi^2\tau^2\nu^2\right)$$

имеет полную ширину на уровне $1/e$ максимума

$$\Delta\nu_{1/e} = \frac{\sqrt{2}}{\pi\tau} \tag{A.23}$$

и полную ширину на уровне половины максимума

$$\Delta\nu_{\text{ПШПМ}} = \frac{\sqrt{2\ln 2}}{\pi\tau}, \tag{A.24}$$

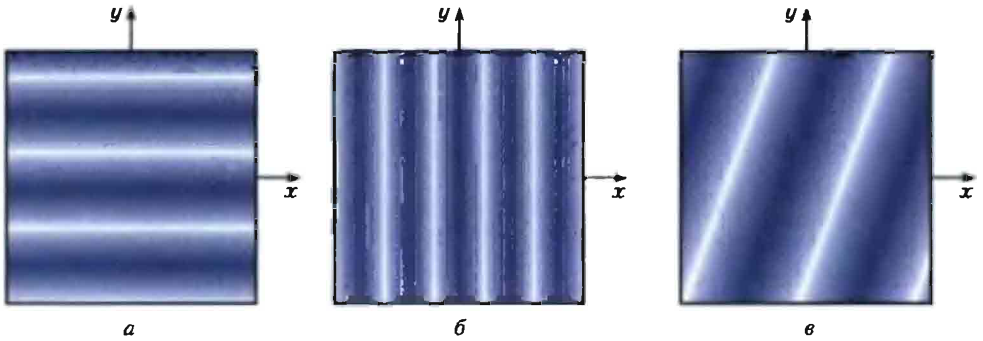
так что

$$\Delta\nu_{\text{ПШПМ}} = \sqrt{\ln 2}\Delta\nu_{1/e} = 0{,}833\Delta\nu_{1/e}. \tag{A.25}$$

Функция Гаусса также используется для описания спектра некоторых видов излучения света (см. подразд. 13.3.4), а также простанственного распределения световых пучков (см. разд. 3.1).

А.3. ДВУМЕРНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Рассмотрим теперь функцию двух переменных $f(x, y)$. Если x и y представляют собой координаты точки в двумерном пространстве, то $f(x, y)$ описывает пространственное распределение (например, оптическое поле в заданной плоскости). Гармонические функции $F\exp[-j2\pi(\nu_x x + \nu_y y)]$ являются строительными блоками, из которых другие функции можно построить в виде суперпозиции. Переменные ν_x и ν_y представляют собой пространственные частоты в направлениях x и y соответственно. Поскольку x и y измеряются в единицах длины (мм), ν_x и ν_y имеют размерность число колебаний/мм или число линий/мм. Примеры двумерных гармонических функций показаны на рис. А.1. Теорему Фурье можно обобщить на функции двух переменных.



а *б* *в*

Рис. А.1. Действительная часть $|F|\cos[2\pi\nu_x x + 2\pi\nu_y y + \arg\{F\}]$ двумерной гармонической функции:

а — $\nu_x = 0$; *б* — $\nu_y = 0$; *в* — произвольный случай. Здесь мы предполагали, что $\arg\{F\} = 0$, так что темные и светлые точки представляют положительные и отрицательные значения функции соответственно

Функция $f(x, y)$ может быть разложена в интеграл суперпозиции гармонических функций x и y

$$f(x, y) = \int_{-\infty}^{\infty}\int F(\nu_x, \nu_y)\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]d\nu_x d\nu_y, \qquad \text{(A.26)}$$

Обратное преобразование Фурье

где коэффициенты $F(\nu_x, \nu_y)$ определяются с помощью двумерного преобразования Фурье

$$F(\nu_x, \nu_y) = \int_{-\infty}^{\infty}\int f(x, y)\exp\left[j2\pi(\nu_x x + \nu_y y)\right]dx dy. \qquad \text{(A.27)}$$

Преобразование Фурье

Наши определения дву- и одномерного преобразования Фурье (A.27) и (A.2) различаются знаком в экспоненте. Выбор знака, конечно, произволен, так как в прямом и обратном преобразовании Фурье используются противоположные знаки. В этой книге мы приняли соглашение, что $\exp(j2\pi\nu t)$ имеет положительную временную частоту ν, а $\exp[-j2\pi(\nu_x x + \nu_y y)]$ — положительные пространственные частоты ν_x и ν_y. Выбор противоположных знаков в пространственном (двумерном) и временном одномерном) случаях сделан для упрощения обозначений, использованных в гл. 4 (Фурье-оптика), где бегущая волна $\exp(+j2\pi\nu t)\exp[-j(k_x x + k_y y + k_z z)]$ имеет пространственную и временну́ю зависимость с противоположными знаками.

Свойства

Многие свойства двумерного преобразования Фурье являются очевидными обобщениями свойств одномерного преобразования Фурье, в то время как другие являются уникальными для двумерного случая.

• *Теорема свертки.* Если $f(x, y)$ представляет собой двумерную свертку двух функций $f_1(x, y)$ и $f_2(x, y)$ с Фурье-образами $F_1(\nu_x, \nu_y)$ и $F_2(\nu_x, \nu_y)$ соответственно, так что

$$f(x, y) = \int_{-\infty}^{\infty}\int f_1(x', y')f_2(x - x', y - y')dx'dy', \qquad \text{(A.28)}$$

то Фурье-образ $f(x, y)$

$$F(\nu_x, \nu_y) = F_1(\nu_x, \nu_y)F_2(\nu_x, \nu_y). \qquad \text{(A.29)}$$

Таким образом, как и в одномерном случае, свертка в пространстве эквивалентна умножению в Фурье-области.

• *Разделение переменных.* Если функция $f(x, y) = f_x(x)f_y(y)$ — произведение двух функций, одна из которых зависит от x, а другая от y, то ее двумерный

Фурье-образ равен произведению одной функции от ν_x на другую от ν_y. Тогда двумерное преобразование Фурье функции $f(x, y)$ выражается через произведение одномерных преобразований $f_x(x)$ и $f_y(y)$ формулой $F(\nu_x, \nu_y) = F_x(-\nu_x)F_y(-\nu_y)$. Например, Фурье-образ функции $\delta(x - x_0)\delta(y - y_0)$, описывающей импульс, локализованный в точке (x_0, y_0), — гармоническая функция $\exp[j2\pi(\nu_x x_0 + \nu_y y_0)]$, а Фурье-образ гауссовой функции $\exp[-\pi(x^2 + y^2)]$ — гауссова функция $\exp\left[-\pi\left(\nu_x^2 + \nu_y^2\right)\right]$ и так далее.

• *Функции с круговой симметрией.* Фурье-образ функции, обладающей круговой симметрией, также обладает круговой симметрией. Например, Фурье-образ функции

$$f(x, y) = \begin{cases} 1, & \sqrt{x^2 + y^2} \le 1; \\ 0, & \sqrt{x^2 + y^2} > 1, \end{cases} \tag{A.30}$$

которая обозначается символом circ (x, y), равен

$$F\left(\nu_x, \nu_y\right) = \frac{J_1\left(2\pi\nu_\rho\right)}{\nu_\rho}; \quad \nu_\rho = \sqrt{\nu_x^2 + \nu_y^2}, \tag{A.31}$$

где J_1 — функция Бесселя первого порядка. Эти функции показаны на рис. А.2.

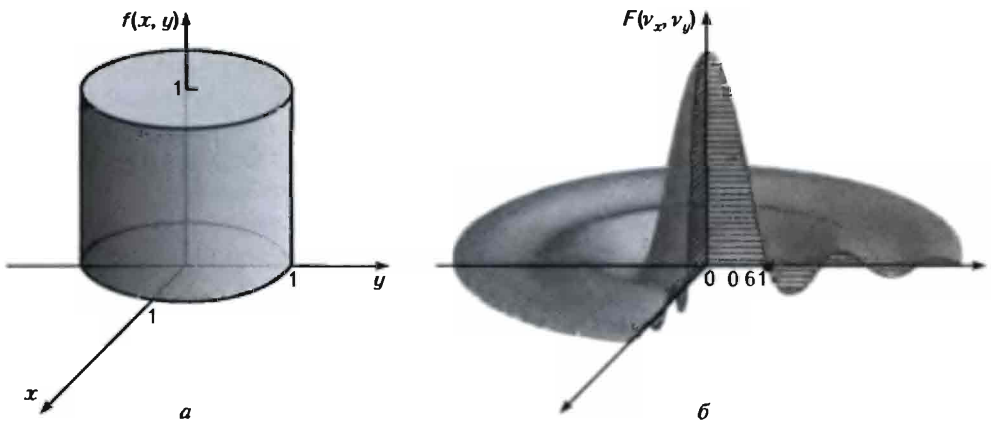


Рис. А.2. Функция circ (*а*) и ее двумерный Фурье-образ (*б*)

Рекомендуемая литература

Kamen E. Introduction to Signals and Systems. Macmillan, 1987; 2nd ed. 1990.

Gabel A., Roberts R.A. Signals and Linear Systems. Wiley, 1973; 3rd ed. 1987.

McGillem C.D., Cooper G.R. Continuous and Discrete Signal and System Analysis. Oxford University Press, 3rd ed. 1991.

Oppenheim A.V., Willsky A.S., Nawab S.H. Signals and Systems. Prentice Hall, 1983; 2nd ed. 1997.

Bracewell R.N. The Fourier Transform and Its Applications. McGraw-Hill, 3rd ed. 2000.

Gaskill J.D. Linear Systems, Fourier Transforms, and Optics. Wiley, 1978.

Franks L.E. Signal Theory. Prentice Hall, 1969; revised ed. 1981.

Papoulis A. Systems and Transforms with Applications in Optics. McGraw-Hill, 1968; Krieger, reissued 1986.

Papoulis A. The Fourier Integral and Its Applications. McGraw-Hill, 1962; reprinted 1987.

ПРИЛОЖЕНИЕ Б ЛИНЕЙНЫЕ СИСТЕМЫ

В этом приложении дается обзор основных характеристик одномерных и двумерных линейных систем.

Б.1. ОДНОМЕРНЫЕ ЛИНЕЙНЫЕ СИСТЕМЫ

Рассмотрим систему, у которой входным и выходным сигналами являются функции $f_1(t)$ и $f_2(t)$ соответственно. Примером является гармонический осциллятор под действием вынуждающей переменной силы $f_1(t)$, откликом на действие которой является смещение $f_2(t)$. Система характеризуется правилом, связывающим входной сигнал с выходным. В общем случае оно может иметь вид дифференциального уравнения, интегрального преобразования или простых математических соотношений, таких как $f_2(t) = \lg f_1(t)$.

Линейные системы

Система называется линейной, если она удовлетворяет принципу суперпозиции, т. е. отклик на сумму двух любых входных сигналов равен сумме откликов на каждый из них в отдельности. В общем случае выходной сигнал в момент времени t представляет собой взвешенную суперпозицию входных сигналов в различные моменты времени τ

$$f_2(t) = \int_{-\infty}^{\infty} h(t;\ \tau) f_1(\tau) \mathrm{d}\tau, \qquad \text{(Б.1)}$$

где $h(t;\ \tau)$ — весовая функция, представляющая вклад входного сигнала в момент τ в выходной сигнал в момент t. Если входной сигнал представляет собой импульс в момент τ, так что $f_1(t) = \delta(t - \tau)$, то (Б.1) дает $f_2(t) = h(t;\ \tau)$. Таким образом, $h(t;\ \tau)$ представляет собой **функцию отклика системы на импульсное воздействие** (известную также как **функция Грина**).

Линейные системы, инвариантные относительно сдвига

Линейная систем называется **инвариантной относительно сдвига (стационарной)**, если при сдвиге входного сигнала во времени выходной сигнал претерпевает такой же сдвиг, в остальном не меняясь. Тогда функция отклика

на импульсное воздействие зависит от разности времен $h(t; \tau) = h(t - \tau)$. При таких условиях (Б.1) приобретает вид

$$f_2(t) = \int_{-\infty}^{\infty} h(t-\tau) f_1(\tau) \mathrm{d}\tau. \qquad \text{(Б.2)}$$

Таким образом, выходной сигнал $f_2(t)$ представляет собой свертку входного сигнала $f_1(t)$ с функцией отклика на импульсное воздействие $h(t)$ [см. (А.4)]. Если $f_1(t) = \delta(t)$, то $f_2(t) = h(t)$, а если $f_1(t) = \delta(t - \tau)$, то $f_2(t) = h(t - \tau)$, как показано на рис. Б.1.

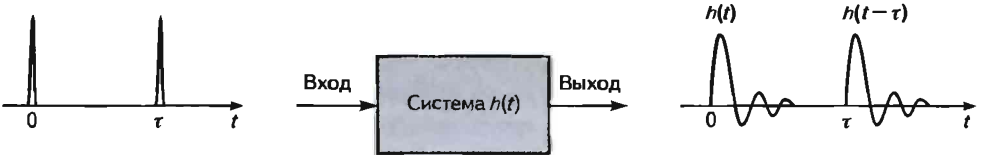


Рис. Б.1. Отклик линейной стационарной системы на импульсы

Передаточная функция

В соответствии с теоремой свертки, обсуждающейся в приложении А, Фурье-образы $F_1(\nu)$, $F_2(\nu)$ и $H(\nu)$ функций $f_1(t)$, $f_2(t)$ и $h(t)$, соответственно, связаны соотношением

$$F_2(\nu) = H(\nu) F_1(\nu). \qquad \text{(Б.3)}$$

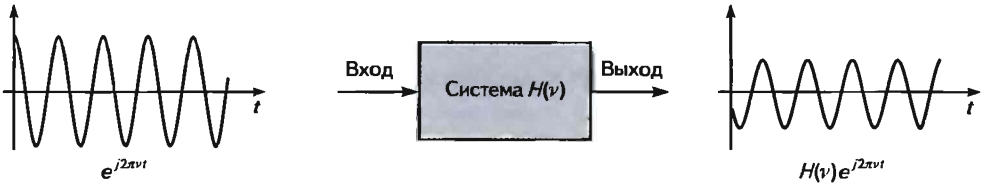


Рис. Б.2. Отклик линейной стационарной системы на гармоническую функцию

Если на входе $f_1(t)$ — гармоническая функция $F_1(\nu) \exp(j2\pi\nu t)$, то на выходе $H(\nu)F_1(\nu) \exp(j2\pi\nu t)$ — также гармоническая функция той же частоты, но с измененной комплексной амплитудой $F_2(\nu) = H(\nu)F_1(\nu)$, как показано на рис. Б.2. Множитель $H(\nu)$ называется **передаточной функцией** системы. Передаточная функция равна Фурье-образу функции отклика на импульсное воздействие. Уравнение (Б.3) является ключевым в использовании методов Фурье-анализа линейных стационарных систем. Для определения отклика системы на произвольный входной сигнал нужно просто разложить его на Фурье-компоненты, умножить их комплексные амплитуды на передаточную функцию на соответ-

ствующей частоте, а затем образовать суперпозицию получившихся гармонических функций.

Примеры

• *Идеальная система*: $H(\nu) = 1$ и $h(t) = \delta(t)$; выходной сигнал — точная копия входного.

• *Идеальная система с задержкой*: $H(\nu) = \exp(-j2\pi\nu\tau)$ и $h(t) = \delta(t - \tau)$; выходной сигнал копирует входной с задержкой на время τ.

• *Система с экспоненциальным откликом*:

$$H(\nu) = \frac{\tau}{1 + j2\pi\nu\tau}; \quad h(t) = e^{-t/\tau}$$

при $t \geq 0$ и $h(t) = 0$ при $t < 0$ представляют систему, описываемую линейным дифференциальным уравнением первого порядка, например *RC*-цепочку с постоянной времени τ. Импульс на входе вызывает экспоненциально затухающий отклик.

• *Система с чирпом*:

$$H(\nu) = \exp(-j\pi\nu^2); \quad h(t) = e^{-j\pi/4}\exp(j\pi t^2);$$

система искажает входной сигнал путем внесения фазового сдвига, пропорционального ν^2. Входной импульс генерирует выходной сигнал с чирпом, т. е. гармоническую функцию, мгновенная частота которой (производная от фазы) линейно растет со временем. Эта система описывает распространение оптических импульсов через среды с зависящей от частоты фазовой скоростью (см. разд. 5.6). Она описывает также изменения в пространственном распределении световых волн при их распространении в свободном пространстве (см. подразд. 4.1.3).

Линейные стационарные причинные системы

Функция отклика на импульсное воздействие $h(t)$ линейной стационарной *причинной* системы должна исчезать при $t < 0$, поскольку отклик системы (следствие) не может начаться до воздействия на входе (причины). Следовательно, функция $h(t)$ не симметрична, а ее Фурье-образ — передаточная функция $H(\nu)$ — должен быть комплексным. Можно показать[1], что если $h(t)$ при $t < 0$, то действительная и мнимая части $H(\nu)$, обозначаемые $H'(\nu)$ и $H''(\nu)$ соответственно, связаны соотношением

$$H'(\nu) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{H''(s)}{s - \nu}ds; \tag{Б.4}$$

$$H''(\nu) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{H'(s)}{\nu - s}ds, \tag{Б.5}$$

Преобразование Гильберта

[1] См., например, *Franks L.E.* Signal Theory. Prentice Hall, 1969; revised ed. 1981.

содержащие главные значения интегралов в смысле Коши, т. е.

$$\int_{-\infty}^{\infty} \equiv \lim_{\Delta \to 0}\left(\int_{-\infty}^{\nu-\Delta} + \int_{\nu+\Delta}^{\infty} \right), \quad \Delta > 0.$$

О функциях, удовлетворяющих (Б.4) и (Б.5), говорят, что они образуют пару преобразования Гильберта, при этом $H''(\nu)$ является **преобразованием Гильберта** функции $H'(\nu)$.

Если функция отклика на импульсное воздействие $h(t)$ действительна, то ее Фурье-образ должен быть симметричным, $H(-\nu) = H^*(\nu)$. Действительная часть $H'(\nu)$ тогда обладает четной, а мнимая — нечетной симметрией. Интегралы в формулах (Б.4) и (Б.5) тогда можно переписать как интегралы по интервалу $(0, \infty)$. Получающиеся равенства известны как **соотношения Крамерса—Кронига**

$$H'(\nu) = \frac{2}{\pi}\int_0^{\infty} \frac{sH''(s)}{s^2 - \nu^2}\,\mathrm{d}s; \tag{Б.6}$$

$$H''(\nu) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{\nu H'(s)}{\nu^2 - s^2}\,\mathrm{d}s. \tag{Б.7}$$

Соотношения Крамерса—Кронига

Итак, преобразования Гильберта или соотношения Крамерса—Кронига связывают действительную и мнимую части передаточной функции причинной системы, инвариантной относительно сдвига, так что если одна из них известна для всех частот, то другая может быть однозначно определена.

Пример: гармонический осциллятор

Линейная система, описываемая дифференциальным уравнением

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \sigma\frac{\mathrm{d}}{\mathrm{d}t} + \omega_0^2\right) f_2(t) = f_1(t), \tag{Б.8}$$

описывает гармонический осциллятор со смещением $f_2(t)$ под действием приложенной силы $f_1(t)$, где ω_0 — резонансная угловая частота; σ — коэффициент, представляющий эффекты затухания. Передаточную функция $H(\nu)$ этой системы можно получить путем подстановки $f_1(t) = \exp(j2\pi\nu t)$ и $f_2(t) = H(\nu)\exp(j2\pi\nu t)$ в (Б.8), что дает

$$H(\nu) = \frac{1}{(2\pi)^2}\frac{1}{\nu_0^2 - \nu^2 + j\nu\Delta\nu}, \tag{Б.9}$$

где $\nu_0 = \omega_0/2\pi$ — резонансная частота; $\Delta\nu = \sigma/2\pi$. Следовательно, действительная и мнимая части $H(\nu)$ равны

$$H'(\nu) = \frac{1}{(2\pi)^2} \frac{\nu_0^2 - \nu^2}{\left(\nu_0^2 - \nu^2\right)^2 + (\nu\Delta\nu)^2}; \tag{Б.10}$$

$$H''(\nu) = -\frac{1}{(2\pi)^2} \frac{\nu\Delta\nu}{\left(\nu_0^2 - \nu^2\right)^2 + (\nu\Delta\nu)^2}. \tag{Б.11}$$

В силу причинности системы, $H'(\nu)$ и $H''(\nu)$ удовлетворяют соотношениям Крамерса—Кронига. Когда $\nu_0 \gg \Delta\nu$, $H'(\nu)$ и $H''(\nu)$ существенно отличны от нуля в узкой области с центром в ν_0. При $\nu_0 \approx \nu$ имеем

$$\left(\nu_0^2 - \nu^2\right) \approx 2\nu_0(\nu_0 - \nu),$$

так что (Б.10) и (Б.11) можно приближенно записать как

$$H''(\nu) = -\frac{1}{(2\pi)^2} \frac{\Delta\nu/4\nu_0}{(\nu_0 - \nu)^2 + (\Delta\nu/2)^2}; \tag{Б.12}$$

$$H'(\nu) = 2\frac{\nu - \nu_0}{\Delta\nu} H''(\nu). \tag{Б.13}$$

Передаточная функция гармонического осциллятора используется в гл. 5 и 14 для описания диэлектрических и атомных систем. Функция (Б.12) имеет лоренцеву форму.

Б.2. ДВУМЕРНЫЕ ЛИНЕЙНЫЕ СИСТЕМЫ

Двумерная система связывает две двумерные функции $f_1(x, y)$ и $f_2(x, y)$, называемые функциями входа и выхода. Эти функции могут, например, представлять оптические поля в двух параллельных плоскостях, при этом (x, y) представляют собой координаты на плоскости. Система тогда включает участки свободного пространства и оптические элементы, лежащие между двумя плоскостями.

Свойства линейности и инвариантности относительно сдвига, определенные в одномерном случае, легко обобщаются на двумерный случай. Выходная функция $f_2(x, y)$ линейной системы связана с ее входной функцией $f_1(x, y)$ интегралом суперпозиции

$$f_2(x, y) = \iint_{-\infty}^{\infty} h(x, y; x', y') f_1(x', y') \mathrm{d}x' \mathrm{d}y', \tag{Б.14}$$

где $h(x, y; x', y')$ — весовая функция, которая описывает влияние входной функции в точке (x', y') на выходную функцию в точке (x, y). Функция $h(x, y; x', y')$ является **функцией отклика системы на импульсное воздействие**, которая известна также как **функция рассеяния точки**, или **аппаратная функция**.

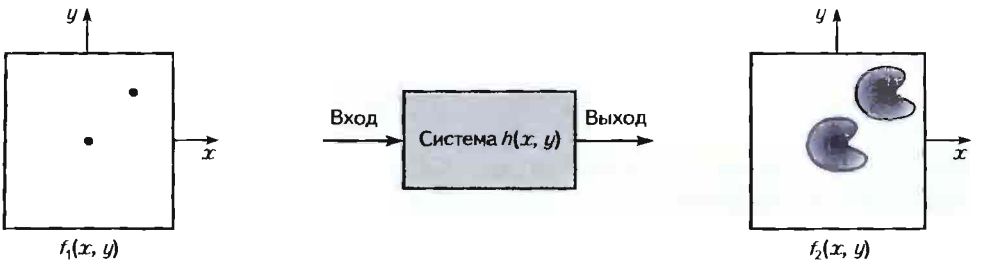


Рис. Б.3. Отклик двумерной линейной системы, инвариантной относительно сдвига на гармоническое воздействие

Система называется **инвариантной относительно сдвига**, или **изопланарной**, если сдвиг входной функции в некотором направлении приводит к такому же сдвигу выходной функции без каких-либо ее других изменений (рис. Б.3). В этом случае функция отклика на импульсное воздействие зависит только от разностей координат, т. е. $h(x, y; x', y') = h(x - x', y - y')$. Формула (Б.14) тогда становится двумерной сверткой функции $h(x, y)$ с $f_1(x, y)$:

$$f_2(x, y) = \int_{-\infty}^{\infty} h(x - x', y - y') f_1(x', y') \mathrm{d}x' \mathrm{d}y'. \qquad \text{(Б.15)}$$

Применяя двумерную теорему свертки, которая обсуждалась в разд. А.3 приложения А, получим

$$F_2(\nu_x, \nu_y) = H(\nu_x, \nu_y) F_1(\nu_x, \nu_y), \qquad \text{(Б.16)}$$

где $F_2(\nu_x, \nu_y)$, $H(\nu_x, \nu_y)$ и $F_1(\nu_x, \nu_y)$ — Фурье-образы $f_2(x, y)$, $h(x, y)$ и $f_1(x, y)$ соответственно.

Следовательно, гармоническая функция на входе с комплексной амплитудой $F_1(\nu_x, \nu_y)$ создает на выходе гармоническую функцию на той же пространственной частоте, но с комплексной амплитудой $F_2(\nu_x, \nu_y) = H(\nu_x, \nu_y)F_1(\nu_x, \nu_y)$, как показано на рис. Б.4. Множитель $H(\nu_x, \nu_y)$ — **передаточная функция** системы. Передаточная функция представляет собой Фурье-образ функции отклика на импульсное воздействие. Каждая из этих функций полностью характеризует систему и дает возможность определить выходную функцию по произвольной входной функции.

Итак, двумерная линейная система, инвариантная относительно сдвига, характеризуется своей функцией отклик на импульсное воздействие $h(x, y)$ или передаточной функцией $H(\nu_x, \nu_y)$. Например, система с $h(x, y) = \mathrm{circ}\,(x/\rho_s, y/\rho_s)$

отображает каждую точку на входе в пятно в виде круга радиусом ρ_s. Она имеет передаточную функцию

$$H(\nu_x, \nu_y) = \rho_s J_1 \frac{2\pi\rho_s \nu_\rho}{\nu_\rho},$$

где

$$\nu_\rho = \left(\nu_x^2 + \nu_y^2\right)^{1/2}$$

Форма этой функции показана на рис. А.2. Система жестко подавляет пространственные частоты выше, чем $0{,}61/\rho_s$ штрихов/мм.

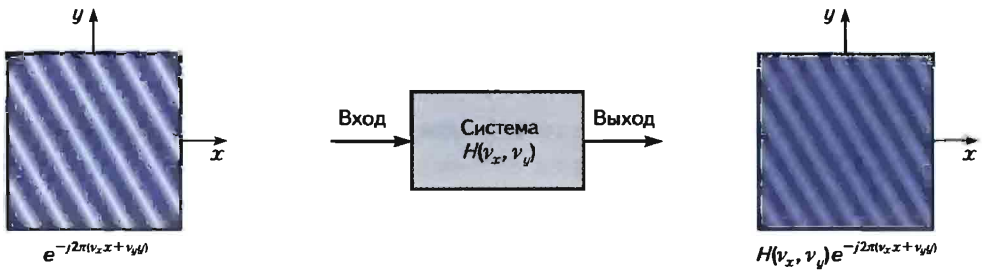


Рис. Б.4. Отклик двумерной линейной системы, инвариантной относительно сдвига, на гармоническое воздействие

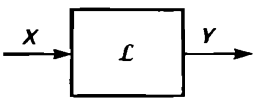
Рекомендуемая литература

См. список литературы к приложению А.

ПРИЛОЖЕНИЕ В МОДЫ ЛИНЕЙНЫХ СИСТЕМ

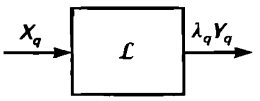
Это приложение — краткий обзор мод линейной системы, описываемой явно соотношением между входом и выходом в матричном или интегральном виде либо с помощью линейного дифференциального уравнения в частных производных.

Во-первых, рассмотрим линейную систему, описываемую явным соотношением между входом и выходом с помощью линейного оператора $\mathcal{L}$, действующего на входной вектор X и генерирующего соответствующий выходной вектор



$$Y = \mathcal{L}X. \tag{B.1}$$

Вектор X может представлять собой массив комплексных чисел, записываемый в виде матрицы-столбца, либо комплексную функцию одной или нескольких переменных. **Моды** такой системы — это особые входные векторы, которые не меняются (с точностью до постоянного множителя) при прохождении через систему, т. е.



$$\mathcal{L}X_q = \lambda_q X_q, \tag{B.2}$$

Задача на собственные значения

где q — индекс, нумерующий моды. Вектор X_q называется **собственным вектором**. Постоянный множитель λ_q, называемый **собственным значением**, обычно является комплексным числом. Условие (B.2) определяет **задачу на собственные значения**.

Во-вторых, рассмотрим линейную динамическую систему, состояние которой описывается N непрерывными переменными, образующими вектор $X(t)$. Эволюция любой из N переменных этого N-мерного вектора, вообще говоря, зависит от всех N переменных. Однако ту же систему можно описывать в других координатах, где N новых переменных эволюционируют независимо, так что система распадается на N независимых одномерных систем. Эти расцепленные переменные — моды системы.

В-третьих, рассмотрим линейную систему, описываемую неявно линейным дифференциальным уравнением в частных производных, которое можно пред-

ставить в виде (В.2), где $\mathcal{L}$ — дифференциальный оператор; X — комплексная функция одной или нескольких переменных. В этом случае моды — это просто решения дифференциального уравнения, а собственные векторы называются **собственными функциями**. Понятие входа и выхода в этом случае не имеет смысла.

В данном приложении мы описываем ряд приложений модового анализа в фотонике. Но сначала мы кратко напомним несколько геометрических понятий из линейной алгебры. С каждой парой векторов X и Y связан комплексный скаляр (X, Y), называемый **скалярным произведением**. Квадратный корень из скалярного произведения вектора самого на себя (X, X) есть **норма** вектора X — мера его «длины». Скалярное произведение двух векторов с единичной нормой можно трактовать как косинус «угла» между ними. Говорят, что два вектора ортогональны, если их скалярное произведение равно нулю. Если векторы представляют собой массивы комплексных чисел $\{X_i\}$ и $\{Y_i\}$, $i = 1, 2, ..., N$, то

$$(X,\ Y) = \sum_{i=1}^{N} X_i^* Y_i.$$

Если векторы представляют собой комплексные функции $X(t)$, $Y(t)$, то

$$(X,\ Y) = \int_{-\infty}^{\infty} X(t)^* Y(t) \mathrm{d}t$$

и так далее.

Остановимся на двух классах операторов $\mathcal{L}$, для которых решение задачи на собственные значения обладает особыми свойствами.

Эрмитовы операторы

Эрмитов оператор определяется свойством $(X, \mathcal{L}Y) = (\mathcal{L}X, Y)$, т. е. скалярное произведение не зависит от того, на какой из векторов действует оператор. Собственные значения эрмитовых операторов действительны, а собственные векторы ортогональны. Далее, собственные векторы эрмитова оператора подчиняются вариационному принципу, который основан на использовании скаляра

$$E_v = \frac{1}{2}\frac{(X,\ \mathcal{L}X)}{(X,\ X)},$$

называемого вариационной энергией. Этот принцип утверждает, что собственный вектор X_1 с наименьшим собственным значением минимизирует E_v; собственный вектор X_2 со следующим в порядке возрастания собственным значением минимизирует E_v при условии его ортогональности X_1 и так далее.

Унитарные операторы

Пассивные физические системы без потерь описываются унитарными операторами, которые характеризуются свойством сохранения нормы $(\mathcal{L}X, \mathcal{L}X) = (X, X)$. Примером является операция «поворота». Собственные значения унитарных операторов унимодулярны ($|\lambda_q| = 1$), т. е. представляют собой чисто фазовые множители.

B.1. МОДЫ ДИСКРЕТНОЙ ЛИНЕЙНОЙ СИСТЕМЫ

Дискретная линейная система описывается матричным соотношением $\boldsymbol{Y} = \mathbf{M}\boldsymbol{X}$, где входной вектор $\boldsymbol{X}$ — набор из N комплексных чисел $(X_1, X_2, ..., X_N)$, расположенных в виде матрицы-столбца; $\mathbf{M}$ — матрица $N \times N$, описывающая линейную систему; $\boldsymbol{Y}$ — выходной вектор, также представляющий собой матрицу-столбец размерностью N. Моды — это те входные векторы, которые остаются параллельными себе после прохождения через систему, так что выполняется матричное равенство

$$\mathbf{M}\boldsymbol{X}_q = \lambda_q \boldsymbol{X}_q. \tag{B.3}$$

Таким образом, моды — это собственные векторы матрицы $\mathbf{M}$, а скаляры λ_q — соответствующие собственные значения, определяемые решением алгебраического уравнения $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$, где $\mathbf{I}$ — единичная матрица. Существует N таких мод, нумеруемых индексом $q = 1, 2, ..., N$.

Частный случай бинарной системы ($N = 2$) особенно важен для оптики. В бинарной системе каждый вектор — пара комплексных чисел (X_1, X_2), расположенных в виде матрицы-столбца $\boldsymbol{X}$. Система характеризуется квадратной матрицей $\mathbf{M}$ размерностью 2×2, элементы которой обозначаются A, B, C и D. Соотношение $\boldsymbol{Y} = \mathbf{M}\boldsymbol{X}$ означает

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

Собственные значения определяются путем решения алгебраического уравнения $(A - \lambda)(D - \lambda) - BC = 0$, имеющего два корня λ_1 и λ_2.

Вот примеры оптических систем, описываемых бинарными линейными системами.

Приложение: поляризационная матричная оптика

В поляризационной матричной оптике (см. подразд. 6.1.2) вектор (X_1, X_2) представляет две ортогональные компоненты входящего электромагнитного поля (вектор Джонса), а (Y_1, Y_2) — то же самое для выходящего поля. Матрица $\mathbf{M}$ — матрица Джонса системы. В этом случае модами являются состояния поляризации, сохраняющиеся при прохождении света через оптическую систему.

Приложение: лучевая матричная оптика

В геометрической параксиальной оптике (см. разд. 1.4) положение и угол наклона луча описываются вектором (X_1, X_2), а действие оптических элементов, таких как линзы и зеркала, описывается матрицей $\mathbf{M}$, которая в этом случае называется матрицей $ABCD$. Для замкнутой оптической системы, такой как резонатор, модами являются такие положения и углы наклона лучей, которые воспроизводят себя при полном обходе, так что лучи остаются внутри резонатора.

Приложение: многослойная матричная оптика

В матричной оптике многослойных сред (см. подразд. 7.1.1) свет отражается и преломляется на каждой границе, так что в каждой плоскости имеются волны, бегущие вперед и назад с амплитудами, описываемыми вектором $X = (X_1, X_2)$. Система, содержащая набор границ между входной и выходной плоскостями, описывается матрицей передачи M. Моды такой системы — векторы, воспроизводящие себя при передаче через систему, так что если система периодически повторяется, как в одномерных фотонных кристаллах (см. разд. 7.2), то моды распространения — это моды системы M.

B.2. МОДЫ НЕПРЕРЫВНОЙ СИСТЕМЫ, ОПИСЫВАЕМОЙ ИНТЕГРАЛЬНЫМ ОПЕРАТОРОМ

Линейные системы, представляемые интегральными операторами, обсуждаются в приложении Б. Рассмотрим функцию времени $f(t)$, описывающую, например, оптический импульс или широкополосное оптическое поле, пропускаемое через линейную систему, не меняющуюся со временем, например оптический фильтр. Система описывается операцией свертки

$$g(t) = \int_{-\infty}^{\infty} h(t-\tau) f(\tau) \mathrm{d}\tau. \tag{B.4}$$

В этой системе векторы X и Y — это функции $f(t)$ и $g(t)$, а $\mathcal{L}$ — интегральный оператор. Моды такой системы — гармонические функции $\exp(j2\pi\nu t)$. Это очевидно, так как входная функция $\exp(j2\pi\nu t)$ порождает на выходе другую гармоническую функцию $H(\nu)\exp(j2\pi\nu t)$, где $H(\nu)$ — Фурье-образ $h(t)$. В этом случае имеется континуум мод с непрерывным множеством собственных значений $H(\nu)$. Роль индекса q здесь играет частота ν, принимающая непрерывный набор значений.

Другой пример — линейная система, инвариантная относительно сдвига, действующая на двумерную функцию $f(x, y)$ координат точки (x, y), как в (Б.14)

$$g(x, y) = \iint_{-\infty}^{\infty} h(x-x', y-y') f(x', y') \mathrm{d}x' \mathrm{d}y' \tag{B.5}$$

Собственными функциями являются двумерные гармонические функции $\exp[j2\pi(\nu_x x + \nu_y y)]$, а собственные значения — это $H(\nu_x, n_y)$, двумерный Фурье-образ функции $h(x, y)$. И в этом случае имеется непрерывное множество собственных функций, нумеруемых пространственными частотами (ν_x, ν_y).

Трансляционная симметрия и гармонические моды

Не удивительно, что гармонические функции являются собственными модами системы, инвариантной относительно сдвига. Поскольку гармоническая функция инвариантна относительно временно́го сдвига, т. е. остается гармонической при переносе во времени, она является собственной функцией стационарной (инвариантной во времени) линейной системы. Аналогично, поскольку двумерные гармонические функции инвариантны относительно сдвигов на плоскости, они являются собственными функциями пространственно-инвариантной (однородной) линейной системы.

Если линейная система не является пространственно-инвариантной, т. е. не обладает свойством трансляционной симметрии, то она представляется (в двумерном случае) более общей линейной операцией:

$$g(x,\ y) = \iint_{-\infty}^{\infty} h(x,\ y;\ x',\ y')\, f(x',\ y')\,\mathrm{d}x'\mathrm{d}y'. \qquad \text{(B.6)}$$

В этом случае собственные функции не обязательно гармонические. Их можно определить путем решения задачи на собственные значения (B.2), которая теперь принимает вид интегрального уравнения

$$\iint_{-\infty}^{\infty} h(x,\ y;\ x',\ y')\, f_q(x',\ y')\,\mathrm{d}x'\mathrm{d}y' = \lambda_q f_q(x,\ y), \quad q = 1,\ 2, \qquad \text{(B.7)}$$

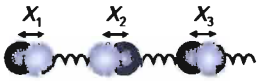
Функции $f_q(x,\ y)$ и постоянные λ_q представляют собой собственные функции и собственные значения системы соответственно, а индекс q нумерует дискретный набор мод.

Приложение: моды оптического резонатора

Примером служит обсуждаемый в подразд. 10.2.5 свет, распространяющийся между двумя параллельными зеркалами лазерного резонатора. Распределения оптического поля по поперечной плоскости в начале и в конце одного полного обхода — это в данном случае вход и выход системы. Моды резонатора — это те распределения поля, которые поддерживают свою форму после полного обхода. Ядро $h(x,\ y;\ x',\ y')$ в (B.7) представляет распространение в свободном пространстве, отражение от первого зеркала, обратное распространение в свободном пространстве и отражение от второго зеркала. Очевидно, присутствие искривленных зеркал или зеркал конечной протяженности нарушает сдвиговую инвариантность системы. Если зеркала сферические и предполагается, что они модулируют приходящий свет путем введения фазового множителя, у которого фаза является квадратичной функцией радиального расстояния, то модами резонатора являются функции Эрмита—Гаусса от x и y. В присутствии апертур уравнение (B.7) можно решить только численно (см. подразд. 10.2.5).

В.3. МОДЫ СИСТЕМЫ, ОПИСЫВАЕМОЙ ОБЫКНОВЕННЫМИ ДИФФЕРЕНЦИАЛЬНЫМИ УРАВНЕНИЯМИ

Динамика некоторых физических систем описывается набором связанных обыкновенных дифференциальных уравнений. Например, динамика N связанных осцилляторов описывается N дифференциальными уравнениями, которые в матричном виде записываются как



$$\ddot{\mathbf{X}} = -\mathbf{M}\mathbf{X}, \tag{B.8}$$

где $\mathbf{X}$ — матрица-столбец с компонентами $(X_1, X_2, ..., X_N)$;

$$\ddot{\mathbf{X}} = \frac{\mathrm{d}^2\mathbf{X}}{\mathrm{d}t^2};$$

$\mathbf{M}$ — матрица размерности $N \times N$ с независящими от времени элементами, так что система обладает временнóй инвариантностью.

Эта инвариантность требует, чтобы моды были гармоническими функциями вида $\exp(j\omega t)$, т. е. вектор $X(t)$ имеет вид $X(t) = X(0)\exp(j\omega t)$. Подставляя это в (В.8), получаем

$$\mathbf{M}\mathbf{X} = \omega^2\mathbf{X}. \tag{B.9}$$

Это уравнение имеет вид задачи на собственные значения для дискретной системы. Ее собственные значения дают резонансные частоты мод $\omega_1, \omega_2, ..., \omega_N$, а собственные векторы называются **нормальными модами**. Все компоненты собственного вектора $\mathbf{X}_q$ моды с номером q осциллируют на одной и той же резонансной частоте ω_q без изменения их относительных амплитуд или фаз. В этом смысле моды являются стационарными решениями, независимыми друг от друга.

В.4. МОДЫ СИСТЕМЫ, ОПИСЫВАЕМОЙ ДИФФЕРЕНЦИАЛЬНЫМ УРАВНЕНИЕМ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Поля и волны описываются дифференциальными уравнениями в частных производных, такими как уравнения Максвелла, которые описывают динамику электрического и магнитного полей в диэлектрической среде, и уравнение Шредингера, которое описывает динамику волновой функции частицы под действием некоторого потенциала. Если эти физические системы стационарны, т. е. диэлектрическая среда и распределение потенциала не зависят от времени, то каждая мода должна быть гармонической функцией времени вида

$\exp(j\omega t)$ с некоторой частотой ω. Следовательно, волновое уравнение преобразуется в обобщенное уравнение Гельмгольца

$$\nabla \times \left[\eta(\boldsymbol{r}) \nabla \times \boldsymbol{H}\right] = \frac{\omega^2}{c_0^2} \boldsymbol{H}, \tag{B.10}$$

где $\eta(\boldsymbol{r}) = e_0/e(\boldsymbol{r})$ — диэлектрическая непроницаемость среды [см. (7.2)]. Аналогично из уравнения Шредингера получается стационарное уравнение Шредингера

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\right]\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{B.11}$$

где $E = \hbar\omega$, а $V(\boldsymbol{r})$ — распределение потенциала [см. (13.3)]. Каждое из этих уравнений теперь имеет вид задачи на собственные значения (B.2), где $\mathcal{L}$ — эрмитов дифференциальный оператор, характеризуемый функциями $\eta(\boldsymbol{r})$ или $V(\boldsymbol{r})$. Действительные собственные значения дают частоты мод ω_q (и следовательно, соответствующие значения энергии E_q в случае уравнения Шредингера). Собственные функции — пространственные распределения электромагнитного поля (или волновой функции) каждой моды. Заметим, что поле (или волновая функция) моды с номером q меняется со временем как $\exp(j\omega_q t)$ во всех точках, так что каждая мода стационарна, что и требовалось.

Моды поля/волны в однородной среде с граничными условиями

Если диэлектрическая среда однородна, т. е. $\eta(\boldsymbol{r})$ постоянна, то система инвариантна относительно сдвига. Для согласования с этой трансляционной симметрией моды электромагнитной системы должны быть гармоническими функциями координат, т. е. плоскими волнами. Аналогично, если потенциал $V(\boldsymbol{r})$ постоянен, то моды — волновые функции в виде плоских волн, так что частица с одинаковой вероятностью может быть обнаружена в любой точке пространства.

В других случаях $\eta(\boldsymbol{r})$ и $V(\boldsymbol{r})$ постоянны внутри конечной области, ограниченной поверхностью, что накладывает некоторые граничные условия. Например, электромагнитные моды объемного резонатора с идеально проводящими стенками могут быть получены из требования того, чтобы параллельные компоненты электрического поля на границе были равны нулю. Для прямоугольного резонатора моды являются гармоническими функциями координат — стоячими волнами, осциллирующими в синхронно (см. подразд. 10.3.3). Аналогично, моды частицы в квантовом ящике (квантовой точке) получаются из требования того, чтобы волновая функция на границах обращалась в нуль (см. подразд. 16.1.7).

В еще одной геометрии однородная диэлектрическая среда может быть ограничена в одном направлении, например, посредством двух параллельных плоских зеркал. Здесь граничные условия соответствуют дискретному набору

стоячих волн в направлении, ортогональном к поверхности зеркал (поперечном направлении), а в продольном (аксиальном) направлении получаются бегущие волны, так что моды распространяются в таком оптическом волноводе как гармонические функции в аксиальном направлении, не меняя своего поперечного распределения (см. разд. 8.1). Если β_q — постоянная распространения моды q, то собственное значение есть фазовый множитель $\exp(-j\beta_q z)$.

Моды полей/волн в периодической среде

Как очевидно из предыдущих примеров, моды системы, описываемой дифференциальным уравнением в частных производных, определяются пространственным распределением среды, т. е. функцией $\eta(\boldsymbol{r})$ или $V(\boldsymbol{r})$. Если эта функция постоянна, моды должны быть инвариантны по отношению к произвольной трансляции. Если она периодическая, то моды должны быть инвариантны относительно трансляции на период. Такой тип трансляционной симметрии требует, чтобы моды были волнами Блоха (см. подразд. 7.2.1). Например, если среда однородна в направлениях x и y, но периодична в направлении z, то мода Блоха имеет вид гармонической функции $\exp(-jKz)$, модулированной периодической стоячей волной $p_K(z)$ с периодом, равным периоду среды; зависимость от x и y, конечно, гармоническая. Для данного значения K частоты мод и форма соответствующих стоячих волн $p_K(z)$ зависят от вида периодической функции $\eta(\boldsymbol{r})$ или $V(\boldsymbol{r})$. Этот тип трансляционной симметрии приводит к спектру собственных значений (и следовательно, частот ω или энергий $E = \hbar\omega$) в виде разрешенных зон, разделенных запрещенными зонами, внутри которых моды не существуют. Так, например, энергия электрона в периодическом потенциале демонстрирует хорошо известную зонную структуру, типичную для твердого тела (см. подразд. 13.3.3). Аналогично, оптическое поле в периодической диэлектрической среде, т. е. фотонном кристалле, демонстрирует такую же зонную структуру с фотонными запрещенными зонами (см. разд. 7.2 и 7.3).

Издательский Дом
ИНТЕЛЛЕКТ

Книги по фундаментальной и прикладной оптике и фотонике

В.А. Астапенко

Взаимодействие излучения с атомами и наночастицами

496 с., формат 60x90/16, тв. переплет

Серия «Физтеховский учебник»

Учебник посвящен систематическому изложению физики основных элементарных процессов, возникающих при взаимодействии электромагнитного поля с атомами и наночастицами. Рассмотрены такие фундаментальные явления как поглощение излучения, люминесценция, индуцированное излучение, фотоионизация, тормозное излучение, включая поляризационный канал, различные виды рассеяния фотона на атомах и мезообъектах, в том числе с участием поверхностных плазмонов, а также фотопереходы в молекулярных системах и наноструктурах. Когерентные нестационарные процессы – самоиндуцированная прозрачность, оптическая нутация и фотонное эхо – описаны в рамках формализма вектора Блоха, позволяющего дать их наглядную геометрическую интерпретацию.

Основному материалу предпослано рассмотрение электромагнитного взаимодействия на примере простейших классических систем – гармонического осциллятора и осциллятора Морзе. Особое внимание уделено фазовому контролю фотовозбуждения классического осциллятора в поле ультракоротких импульсов и в бихроматическом поле, играющему важную роль в современной лазерной физике и химии.

Представлено классическое описание взаимодействия электромагнитного излучения с веществом на основе спектроскопического принципа соответствия, включая вывод выражений для коэффициентов Эйнштейна. Подробно рассмотрена динамическая поляризуемость атома – величина, занимающая центральное место в описании отклика вещества на электромагнитное воздействие.

Рассмотрены неупругие процессы при взаимодействии электронов с атомными частицами.

Для студентов и преподавателей, научных работников в областях атомной физики и фотоники.

Г.Р. Локшин

Основы радиооптики

344 с., формат 60x90/16, тв. переплет

Серия «Физтеховский учебник»

Многие задачи обработки информации не могут быть решены без привлечения оптических методов.

Преобразование света в оптических системах изучается с использованием закономерностей, управляющих работой линейных колебательных систем.

Вопросы дифракции и формирования оптического изображения, а также принципы пространственной фильтрации и оптической обработки информации рассмотрены на основе естественного обобщения принципов линейной фильтрации электрических сигналов. Обсуждаются методы улучшения качества изображений и наблюдения фазовых объектов; согласованная фильтрация и проблема распознавания образов; способы получения голограмм без опорного пучка, основанные на идеях корреляционной фильтрации, и ряд других задач.

Книга предназначена для студентов и преподавателей радиофизических и оптических специальностей, инженеров-разработчиков и исследователей.

Полные оглавления этих и многих других книг на сайте www.id-intellect.ru

В.А. Астапенко

Электромагнитные процессы в среде, наноплазмоника и метаматериалы

584 с., формат 60x90/16, тв. переплет

В книге рассмотрены как хорошо известные явления, так и ряд важных электромагнитных процессов, не описывавшихся в традиционных курсах электродинамики сплошных сред, но приобретших актуальность в контексте современного развития физики. К ним относятся: поляризационное тормозное излучение в плазме, конденсированном веществе и наноструктурах, рассеяние ультракоротких импульсов в плазме, на атомах наночастицах.

Рассмотрены основные характеристики и методы возбуждения поверхностных плазмонов, как на плоских поверхностях, так и в наночастицах. Значительное внимание уделено средам с отрицательным преломлением и способам создания метаматериалов такого рода. В качестве примеров практического использования наноплазмоники представлены работы по наноантеннам, поверхностно-усиленной рамановской спектроскопии, спазерам, фотодетекторам и солнечным батареям.

Наряду с традиционными подходами рассматриваются малоизвестные модели и приближения, хорошо зарекомендовавшие себя в практическом использовании, такие как вращательное приближение крамерсовой электродинамики, метод локальной плазменной частоты в описании излучательных процессов и приближение Борна-Комптона в теории столкновительной ионизации атомов. В книге использованы современные экспериментальные данные.

Учебное пособие адресовано студентам старших курсов, аспирантам, преподавателям физических и инженерно-физических факультетов, исследователям, разработчикам и научным работникам.

М. Бертолотти

История лазера, *пер. с англ.*

336 с., формат 60x90/16, обложка

Многие задачи обработки информации не могут быть решены без привлечения оптических методов.

Преобразование света в оптических системах изучается с использованием закономерностей, управляющих работой линейных колебательных систем.

Вопросы дифракции и формирования оптического изображения, а также принципы пространственной фильтрации и оптической обработки информации рассмотрены на основе естественного обобщения принципов линейной фильтрации электрических сигналов. Обсуждаются методы улучшения качества изображений и наблюдения фазовых объектов; согласованная фильтрация и проблема распознавания образов; способы получения голограмм без опорного пучка, основанные на идеях корреляционной фильтрации, и ряд других задач.

Книга предназначена для студентов и преподавателей радиофизических и оптических специальностей, инженеров-разработчиков и исследователей.

Н.Я. Молотков

Учебные эксперименты по волновой оптике. СВЧ демонстрации

352 с. формат 60x90/16, обложка

Пособие содержит описание лекционных демонстраций по основным и принципиально важным оптическим явлениям в диапазоне СВЧ (λ=3.2см). Демонстрационные опыты включают учебные исследования свойств электромагнитных волн и охватывают все основные разделы волновой оптики: интерференцию, дифракцию, поляризацию и дисперсию. Большое внимание уделяется моделированию кристаллооптических явлений и иллюстрации основных методов и принципов рентгеноструктурного анализа.

Описание каждой демонстрации помимо рассмотрения физического явления включает рекомендации по методике и технике лекционного эксперимента. Даются указания по изготовлению необходимого учебного оборудования.

Для преподавателей физики университетов, институтов и техникумов, работников физических кабинетов, студентов физических специальностей, учителей средних школ.

Полные оглавления этих и многих других книг на сайте www.id-intellect.ru

В. Демтрёдер

Современная лазерная спектроскопия, *пер. с англ.*

1072 с., формат 70x100/16, тв. переплет

Книга известного специалиста в области оптики и лазерной физики, профессора университета Кайзерслаутерн (Германия) Вольфганга Демтрёдера широко известна во всем мире среди студентов, аспирантов, преподавателей и ученых как уникальное издание, вобравшее в себя практически полную информацию о методах лазерной спектроскопии. Адресованная прежде всего физикам-экспериментаторам, она содержит весьма детальное описание разнообразных методов, приборов и схем спектроскопии с помощью лазеров, которому предшествует изложение основных физических принципов. Изложение вопросов отличается емкостью и информационной насыщенностью, и при этом достаточно кратко. Поэтому по своему целевому назначению книга одновременно выполняет несколько ролей. В первую очередь, это великолепный учебник для студентов различных уровней – от ознакомительного на младших курсах до основательного на старших.

С этой целью в конце каждого раздела имеется продуманный набор задач. В этом же качестве книга незаменима для преподавателей оптики, физики лазеров и спектроскопии, которые найдут в книге много полезного, а для курса лазерной спектроскопии она может быть рекомендована в качестве основного учебника. Для научного работника это великолепный справочник. Цитирование этой книги в научных статьях в качестве энциклопедии лазерной спектроскопии стало традицией.

В 2008 году на английском языке вышло 4-е издание книги, существенно переработанное и значительно увеличившееся в объеме. В связи с последним книга в оригинале разделена на два тома, в первом из которых излагаются основные принципы лазерной спектроскопии, а во втором – методы и техника эксперимента. При переработке книги автор сохранил ее положительные стороны, но существенно обогатил новыми достижениями в таких областях, как спектроскопия с помощью коротких и ультракоротких лазерных импульсов, спектроскопия высокого разрешения с помощью перестраиваемых лазеров, спектроскопия одиночных частиц, лазерное охлаждение и лазерные ловушки, новые метрологические возможности, открываемые с появлением фемтосекундной техники, новые типы полупроводниковых лазеров и многое другое. Можно с уверенностью сказать, что 4-е издание книги дает достаточно объективное представление о современном состоянии лазерной спектроскопии.

Книга будет востребована широким кругом читателей – студентов, преподавателей, научных работников, чьи интересы связаны с оптикой, лазерной физикой и спектроскопией.

П.Г. Крюков

Лазеры ультракоротких импульсов и их применения

248 с., формат 60x90/16, обложка

Книга посвящена проблеме получения лазерного излучения в виде ультракоротких импульсов, длительность которых приближается к периоду световой волны, т.е. составляет несколько фемтосекунд. Это одно из важнейших и актуальных направлений современной лазерной физики.

Изложена краткая история исследований, приведших к созданию лазеров фемтосекундных импульсов. Обеуждаются принципы работы лазеров, позволяющих генерировать импульсы фемтосекундной длительности и усиливать их мощность вплоть до петаваттного уровня. Показано как измеряются длительности лазерных импульсов, столь коротких. Описаны конкретные системы лазеров.

Рассматриваются некоторые наиболее яркие применения в области научных исследований, в технике и медицине, основанные как на предельно короткой длительности лазерных импульсов, так и на сверхвысокой интенсивности лазерного излучения. В частности, рассматривается новейшее применение фемтосекундных лазеров – прецизионное измерение оптических частот и возможность создания сверхточных и компактных оптических часов на этой основе.

Книга адресована студентам и аспирантам, изучающим лазерную физику, а также специалистам, работающим в этой области.

Полные оглавления этих и многих других книг на сайте www.id-intellect.ru

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 2

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. 49 л. Зак. М-1030.
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

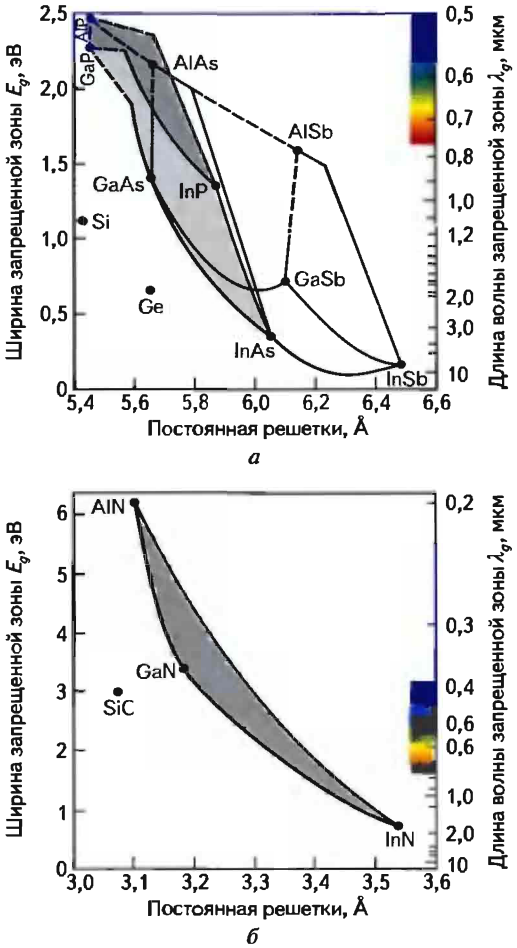


Рис. 16.7. Ширины и длины волн запрещенной зоны и постоянные решетки для Si, Ge, SiC и 12 двухкомпонентных соединений $A^{III}B^{V}$ Сплошные и штриховые линии соответствуют прямозонным и непрямозонным полупроводникам. Один и тот же материал может быть прямозонным или непрямозонным в зависимости от соотношения компонентов. Трехкомпонентные материалы представлены линиями, соединяющими два двухкомпонентных материала. Четырехкомпонентным материалам соответствуют участки поверхности, образованные их двухкомпонентными составляющими:

а – $In_{1-x}Ga_xAs_{1-y}P_y$ представлен участком, окрашенным в светло-серый цвет, вершины его границы представляют InP, InAs, GaAs и GaP. Участок, изображающий $(Al_xGa_{1-x})_yIn_{1-y}P$, окрашен в темно-серый цвет и имеет вершины, представляющие AlP, InP и GaP. Оба являются важными четырехкомпонентными соединениями для фотоники, первый – в ближнем ИК, а второй — в видимом диапазоне. $Al_xGa_{1-x}As$ представлен точками вдоль линии, соединяющей GaAs и AlAs. При изменении x от 0 до 1 изображающая точка движется вдоль линии от GaAs к AlAs. Поскольку эта линия практически вертикальна, решетка $Al_xGa_{1-x}As$ такая же, как у GaAs; *б* — хотя III-нитрид $In_xGa_{1-x}N$ в принципе может за счет изменения состава перекрыть весь видимый спектр, этот материал трудно выращивать, когда содержание In становится заметным. $In_xGa_{1-x}N$ главным образом используется в зеленой, синей и фиолетовой области, тогда как $Al_xGa_{1-x}N$ и $Al_xIn_yGa_{1-x-y}N$ — в ультрафиолетовой области. Все эти соединения являются прямозонными полупроводниками

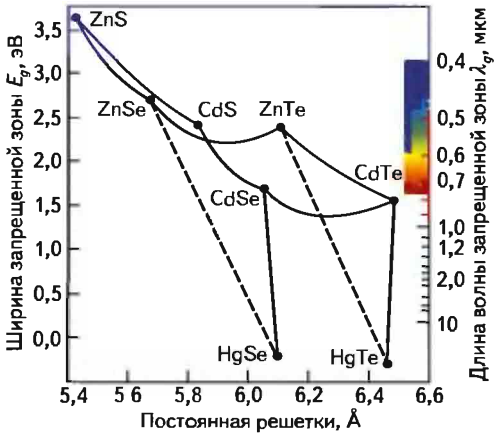


Рис. 16.8. Ширины и длины волн запрещенной зоны и постоянные решетки для различных полупроводников II –VI групп (HgSe и HgTe являются полуметаллами с малой отрицательной шириной запрещенной зоны). HgTe и CdTe имеют почти одинаковые решетки, что видно по вертикальности соединяюшей их линии, поэтому трехкомпонентный полупроводник $Hg_xCd_{1-x}Te$ можно выращивать без напряжений на подложке CdTe. Это важный материал для фотоприемников в средней ИК-области спектра

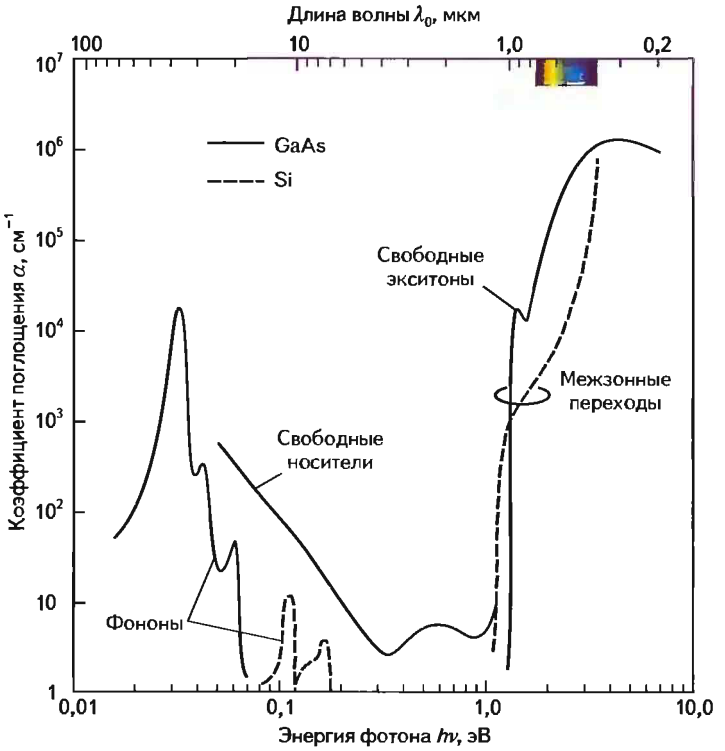


Рис. 16.31. Наблюдаемые оптические коэффициенты поглощения α в зависимости от энергии и длины волны фотона для Si и GaAs в состоянии теплового равновесия при $T = 300$ K. Ширина запрещенной зоны E_g составляет 1,12 эВ для Si и 1,42 эВ для GaAs. Кремний сравнительно прозрачен в области λ_0 от 1,1 до 12 мкм, тогда как собственный GaAs – в области от 0,87 до 12 мкм (см. рис. 5.84)

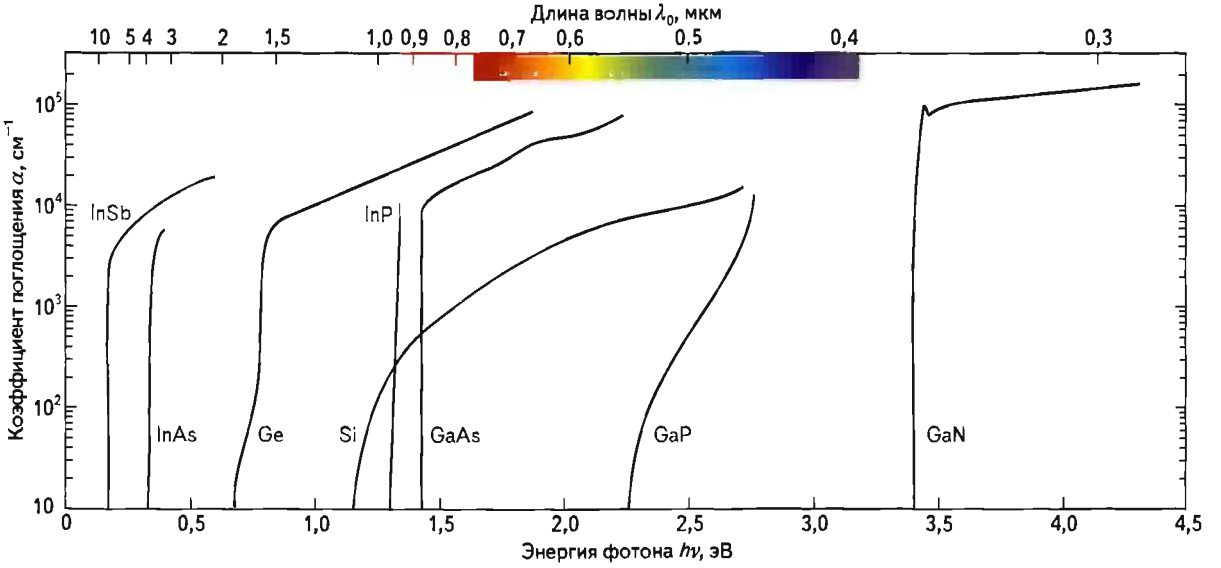


Рис. 16.32. Коэффициент поглощения в зависимости от энергии и длины волны фотона для Ge, Si, GaAs, GaN и некоторых других двухкомпонентных полупроводников III—V групп при $T = 300$ K в увеличенном масштабе

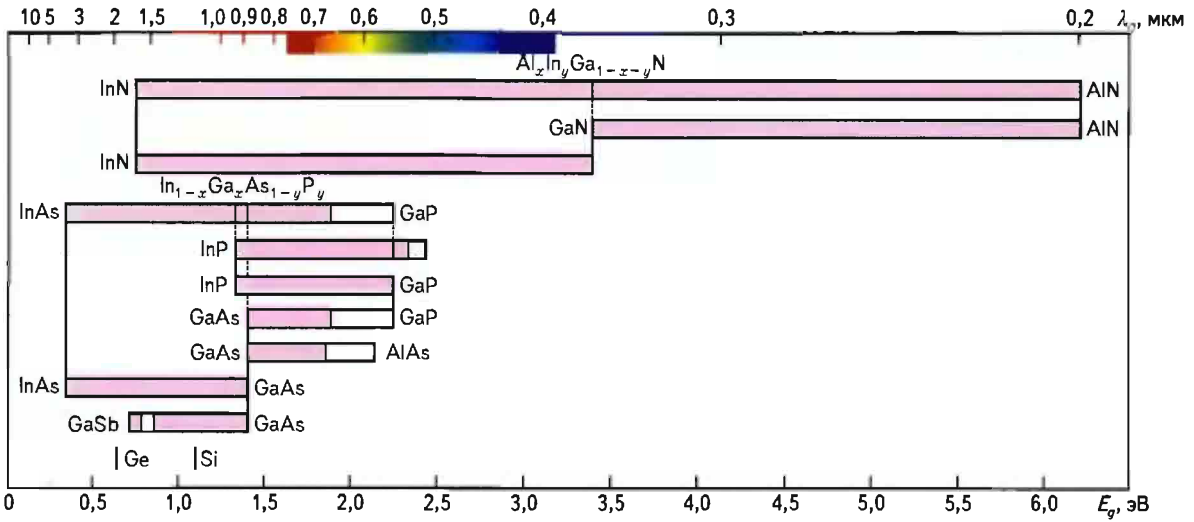


Рис. 16.33. Ширина запрещенной зоны E_g и соответствующая длина волны λ_g для некоторых одно-, двух-, трех- и четырехкомпонентных полупроводников III—V групп. Начиная сверху, полосы представляют AlInGaN, AlGaN, InGaN, InGaAsP, AlInGaP, InGaP, GaAsP, AlGaAs, InGaAs и GaAsSb. Закрашены области, в которых материалы являются прямозонными полупроводниками

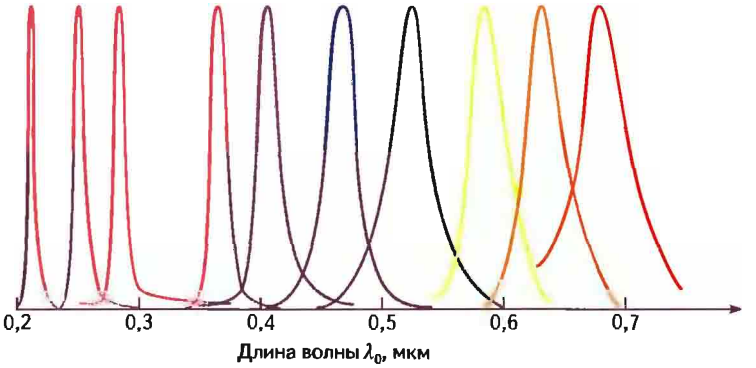


Рис. 17.15. Спектральная интенсивность как функция длины волны для СИД, работающих в ультрафиолетовом и видимом диапазоне спектра. Пиковые интенсивности сделаны равными путем нормировки. Для инфракрасных СИД результаты представлены на рис. 17.6

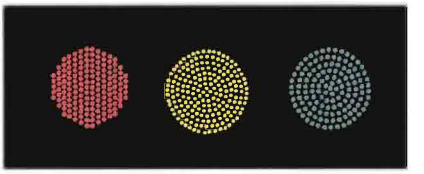
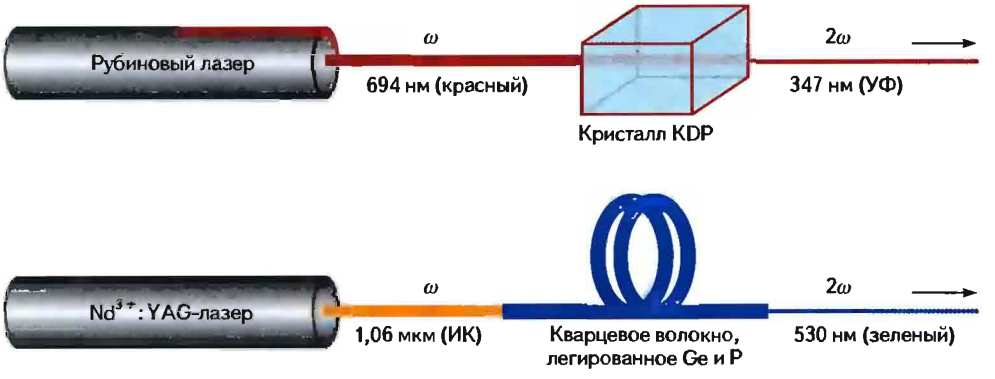
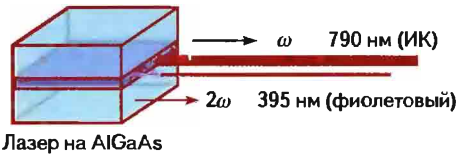


Рис. 17.17. Дорожный светофор на основе соединений III V групп



б



в

Рис. 21.5. Генерация второй гармоники света в объемном кристалле (*а*), в стеклянном волокне (*б*) и в резонаторе лазерного диода (*в*)

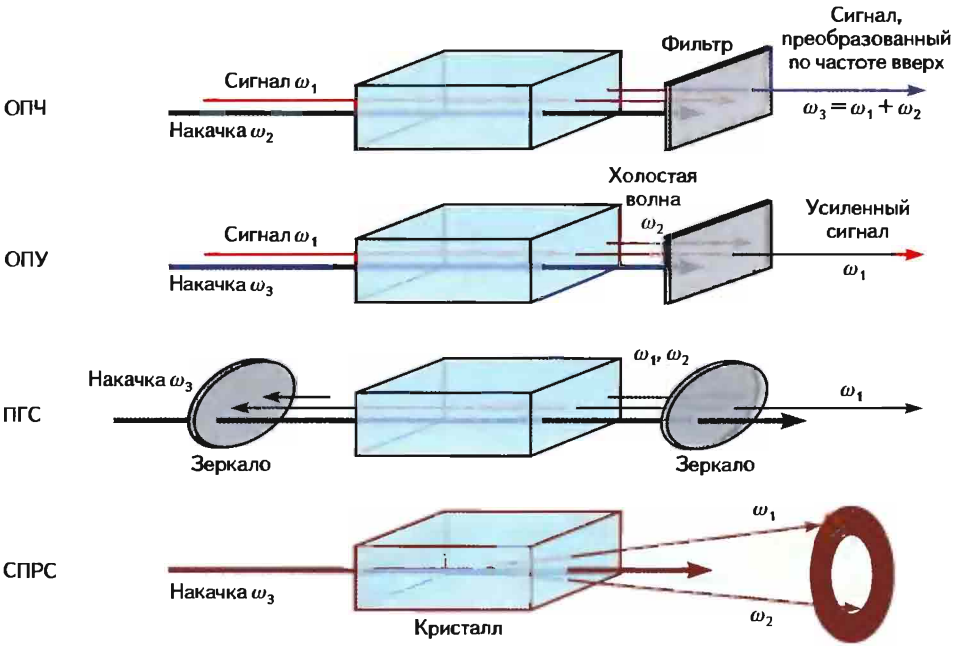


Рис. 21.10. Оптические параметрические приборы: оптический преобразователь частоты (ОПЧ), оптический параметрический усилитель (ОПУ), параметрический генератор света (РГС) и преобразователь частоты вниз на основе спонтанного параметрического рассеяния света (СПРС)

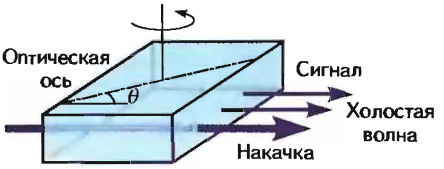
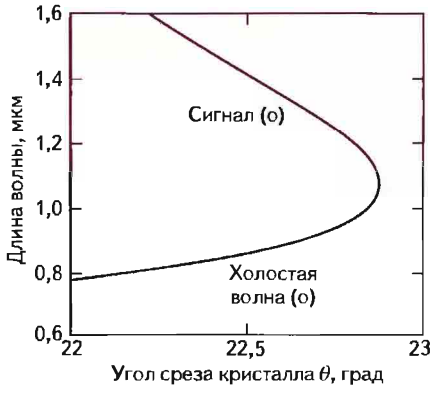
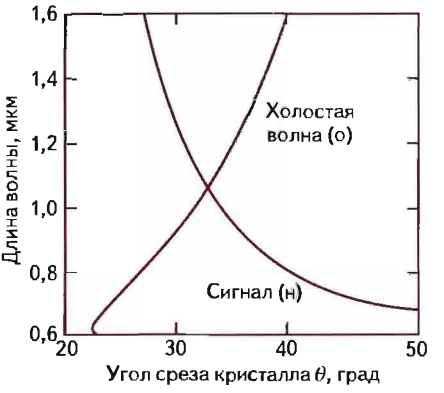


Рис. 21.13. Кривые настройки для коллинеарного ПГС на кристалле BBO с накачкой на 532 нм, которая легко получается удвоением частоты Nd YAG-лазера:

а – типа I; *б* – типа II



а



б

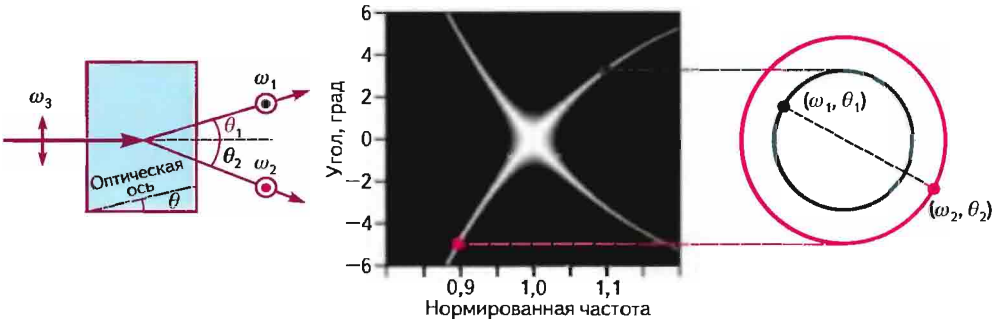


Рис. 21.15. Кривые настройки для неколлинеарного о-о-н типа I спонтанного параметрического рассеяния света с понижением частоты в кристалле BBO под углом 33,53° при накачке на 351,5 нм (аргоновый ионный лазер). Каждая точка светлой области на среднем рисунке соответствует частоте ω_1 и углу θ_1 для возможной волны с пониженной частотой. Каждой такой точке соответствует точка с дополнительной частотой $\omega_2 = \omega_3 \quad \omega_1$ и углом θ_2. Частоты нормированы на величину $\omega_0 = \omega_3/2$. Например, две показанные точки представляют пару волн с частотами $0{,}9\omega_0$ и $1{,}1\omega_0$. Из-за осевой симметрии каждая точка на самом деле представляет кольцо точек: всем им соответствует одинаковая частота, но каждая точка на кольце согласована только с одной диаметрально противоположной точкой на соответствующем кольце, как показано на правом графике

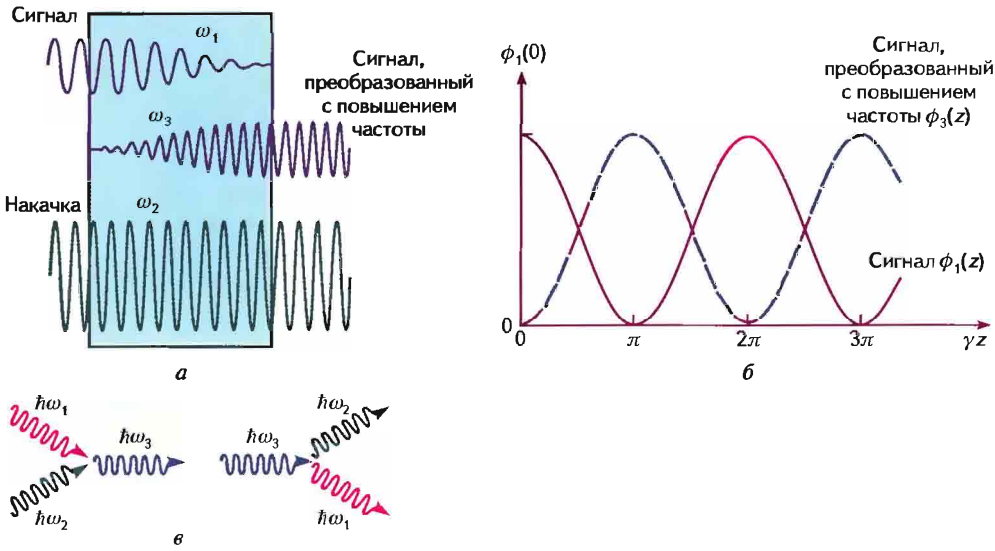
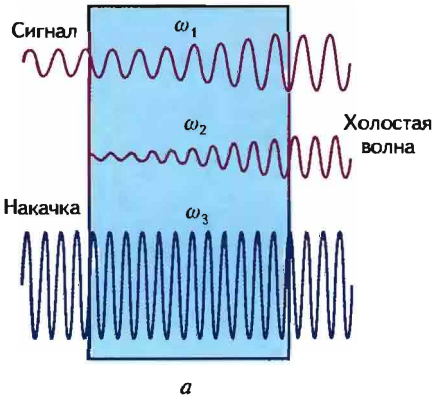
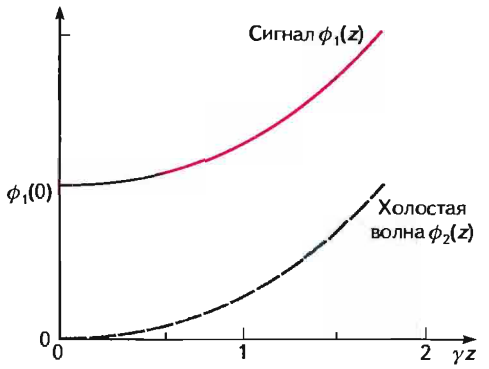


Рис. 21.32. Преобразователь частоты вверх:

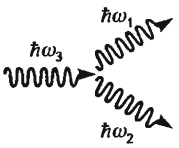
а — смешение волн; *б* — эволюция плотностей потока входной ω_1-волны и преобразованной по частоте вверх ω_3-волны. Волна накачки ω_2 предполагается постоянной; *в* — взаимодействие фотонов



а



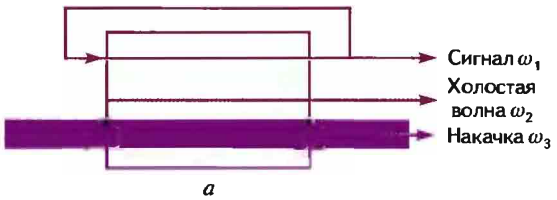
б



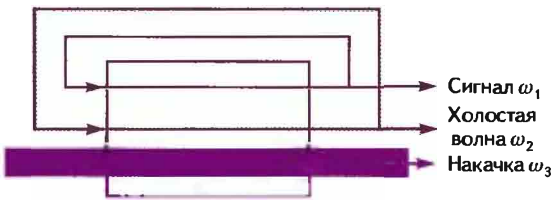
в

Рис. 21.33. Оптический параметрический усилитель:

а — смешение волн; *б* — плотности потока фотонов для сигнальной и холостой волн (плотность потока фотонов накачки считается постоянной); *в* — взаимодействие фотонов



а



б

Рис. 21.34. Параметрический генератор генерирует свет на частотах ω_1 и ω_2. Накачка с частотой $\omega_3 = \omega_1 + \omega_2$ служит источником энергии:

а — ПГС с одним резонансом; *б* — ПГС с двойным резонансом

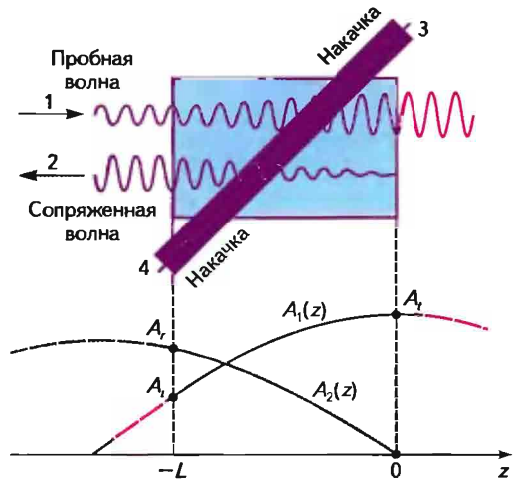


Рис. 21.35. Вырожденное четырехволновое смешение. Волны 3 и 4 — интенсивные волны накачки, распространяющиеся в противоположных направлениях. Пробная волна 1 и сопряженная волна 2 также распространяются в противоположных направлениях и имеют нарастающие амплитуды

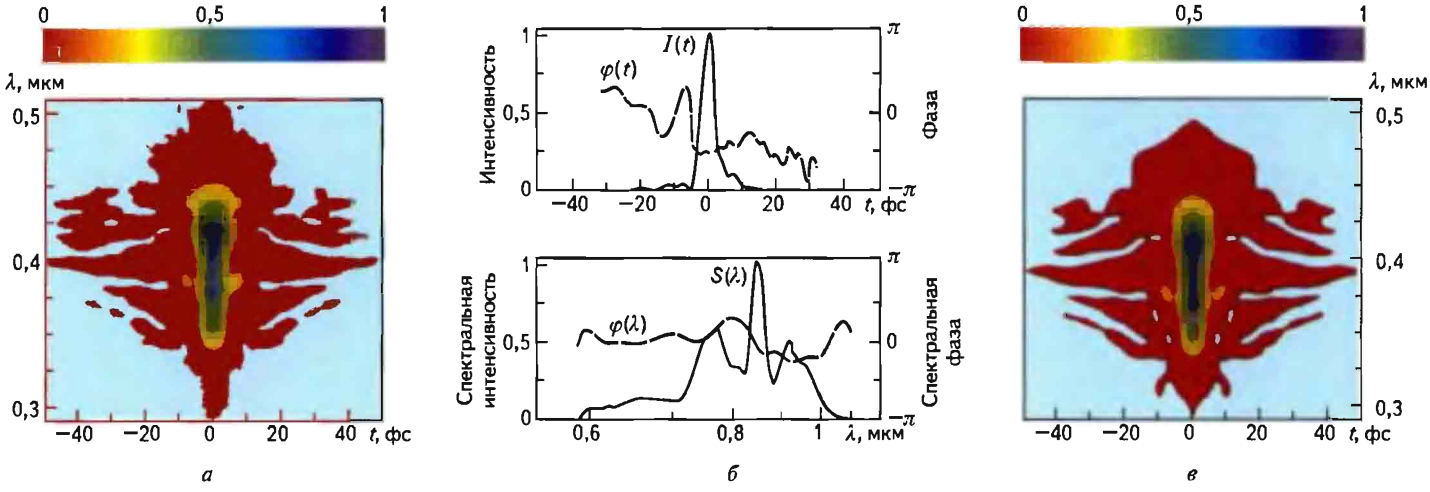


Рис. 22.63. Измеренная методом ГВГ-FROG спектрограмма импульса длительностью 4,5 фс, содержащего 2,5 оптических периода (*а*). Рассчитанные временные и спектральные характеристики импульса (*б*). ГВГ-FROG-спектрограмма, рассчитанная для импульса (*б*). примерно совпадает с измеренной (*а*) (*в*). (Заимствовано из *Baltuška A.*, *Pshenichnikov M.S.*, *Wiersma D.A.* IEEE Journal of Quantum Electronics. Vol. 25. P. 459—478, рис. 17, *a*, *b* и 18. ©1999 IEEE; R. Trebino, ed. Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses. Kluwer, 2000, рисунок на прилагаемом CD-ROM.)



Серьезная разработка оптических межсоединений и фотонных коммутаторов началась в 1980-х гг. под руководством Лабораторий Белла, организованных в 1925 г. компанией АТТ. Лаборатории Белла стали частью компании Люсент Текнолоджиз в 1996 г, а затем влились в компанию Алкател в 2006 г.

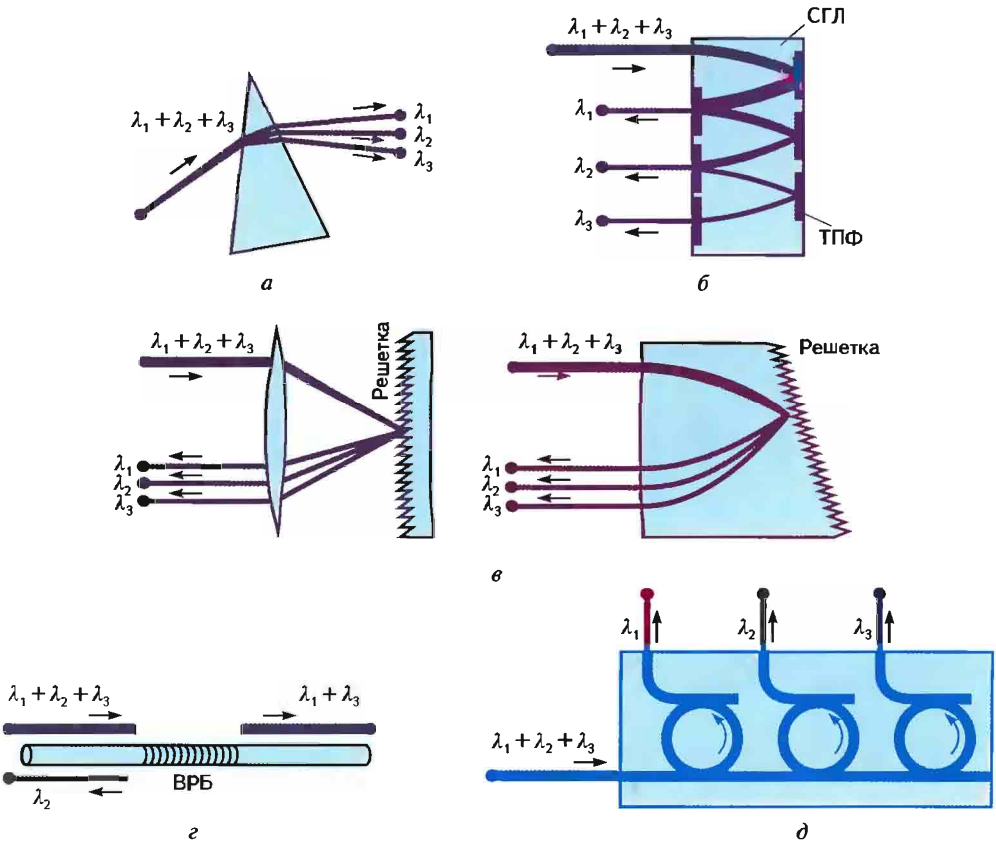


Рис. 23.18. Демультиплексоры с разделением по длине волны:

а – призма; *б* — диэлектрические интерференционные тонкопленочные фильтры (ТПФ); *в* — дифракционная решетка с обычной или стрежневой градиентной линзой (СГЛ); *г* — волоконная решетка Брэгга (ВРБ); *д* — фильтр на кольцевых микрорезонаторах

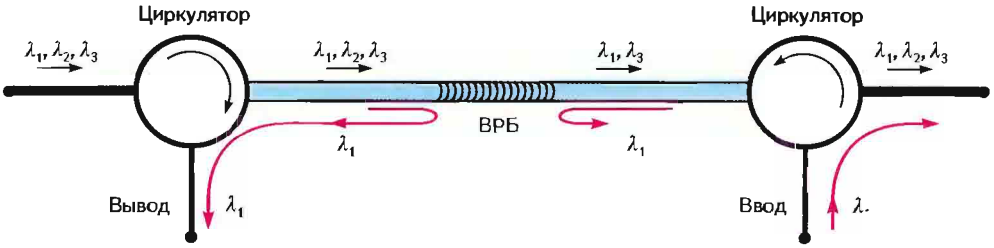


Рис. 23.19. В оптическом мультиплексоре ввода-вывода (ОМВВ) волоконная решетка Брэгга (ВРБ) отражает выводимую компоненту с длиной волны λ_1, а циркулятор направляет ее на приемник. Другие компоненты (λ_2 и λ_3) проходят через решетку. Другой циркулятор служит для ввода света на длине волны λ_1, промодулированного новыми данными. ВРБ отражает распространяющийся обратно свет на длине волны λ_1

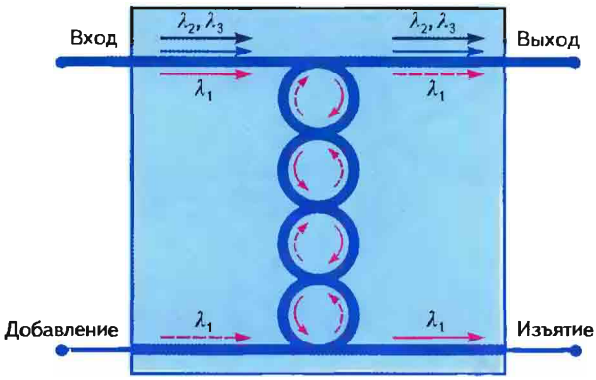


Рис. 23.20. В оптическом мультиплексоре ввода-вывода (ОМВВ) используется множество кольцевых микрорезонаторов для изъятия канала λ_1 из входящего пучка многоканальной связи и направления его на приемник. Другие компоненты (λ_2 и λ_3) пропускаются. Новые данные на λ_1 выбираются фильтром и добавляются в выходной пучок. Множество кольцевых микрорезонаторов имеют большую селективность по длине волны (т. е. более узкую спектральную ширину и больший коэффициент подавления, чем одиночный кольцевой микрорезонатор

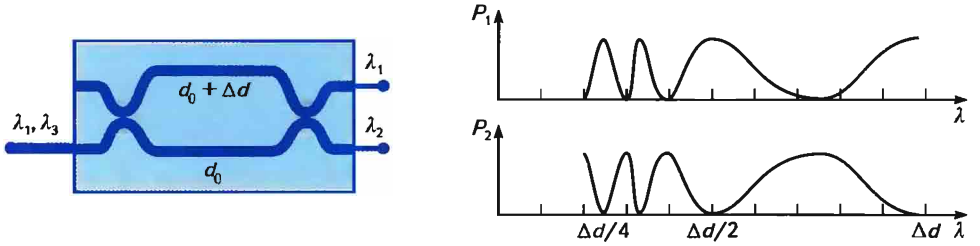


Рис. 23.21. Маршрутизация с разделением по длине волны (демультиплексирование) с использованием интегрально-оптического интерферометра Маха Цендера

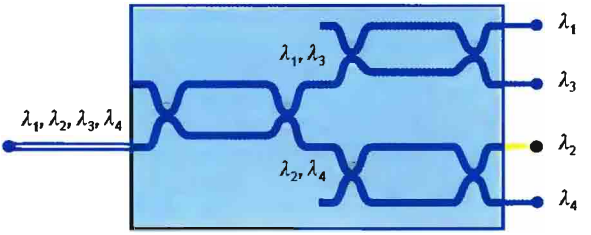
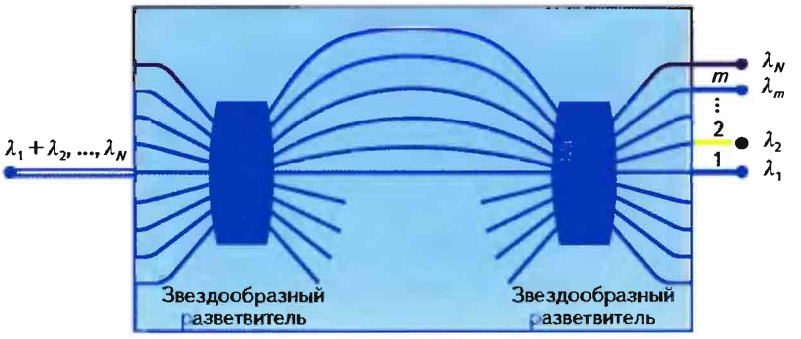


Рис. 23.22. Демультиплексирование с разделением по длине волны с использованием интегрально-оптического каскада интерферометров Маха Цендера



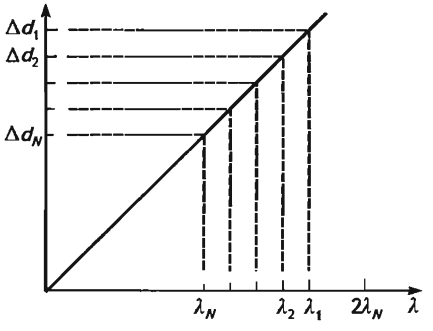


Рис. 23.24. Демультиплексирование с разделением по длине волны с помощью маршрутизатора на решетке волноводов (WGR)

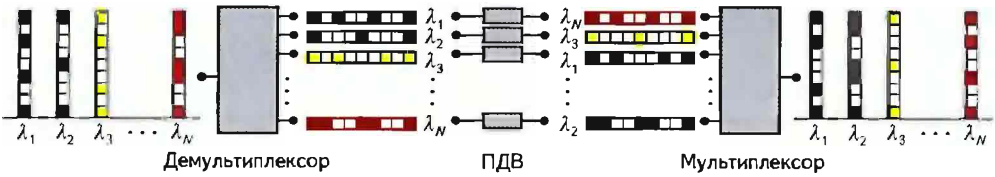


Рис. 23.52. Реализация обмена между каналами с разделением по длине волны (WCI). В данном примере данные в канале с длиной волны 2 (зеленый цвет) входящего пучка направляются в канал 3 с другой длиной волны (желтый цвет) выходящего пучка. Биты данных изображены как цветные и белые квадратики. Такой коммутатор реализуется с помощью демультиплексора с разделением по длинам волн, который разделяет каналы и направляет их на набор преобразователей длины волны. Собирающий $N \times 1$-мультиплексор обеспечивает объединение переключенных каналов в единый пучок

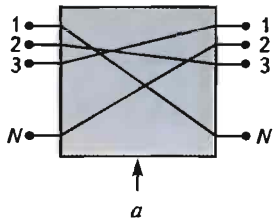
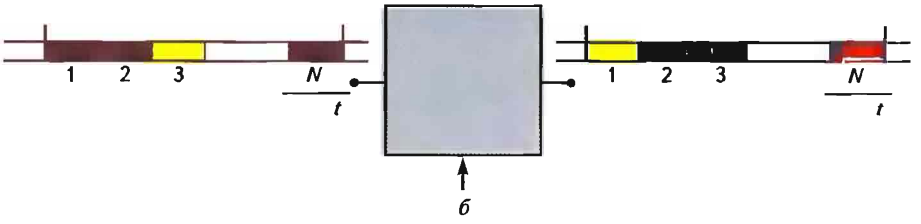


Рис. 23.55. Соответствие между временны́м и пространственным коммутатором:

а — пространственный коммутатор. В показанном примере данные из линии 2 направляются в линию 3; *б* — коммутатор с разделением по времени, реализующий межинтервальный обмен (TSI). В показанном примере данные из временно́го кванта 2 направляются во временно́й квант 3 в каждом кадре



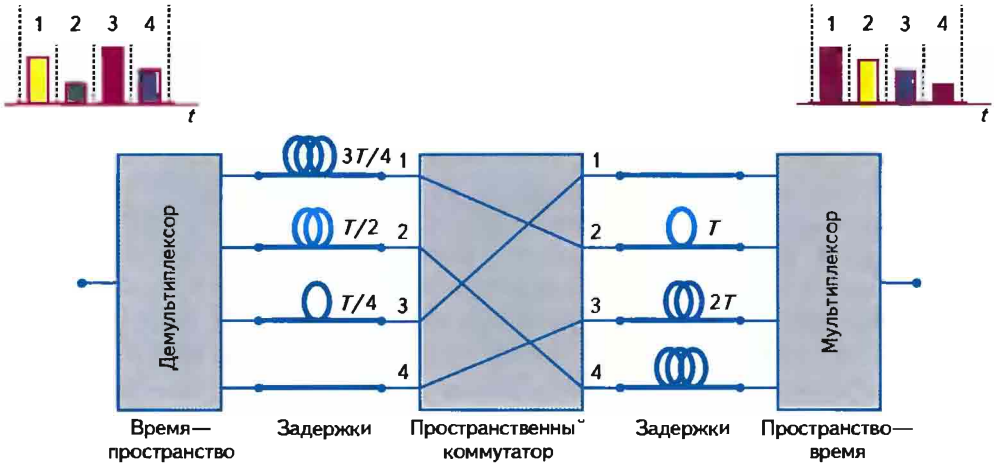
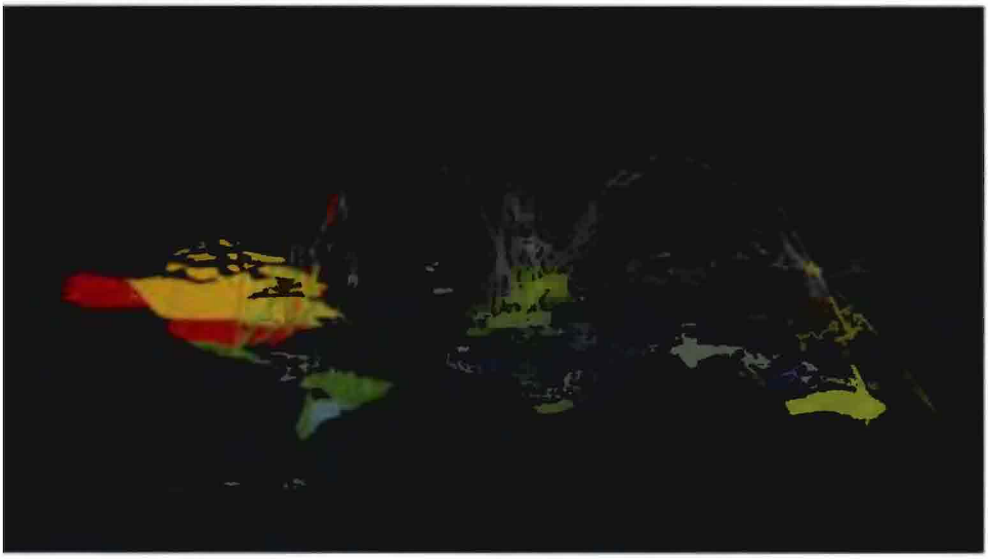


Рис. 23.59. Межинтервальный обмен (TSI)



Межконтинентальная сеть волоконно-оптической связи

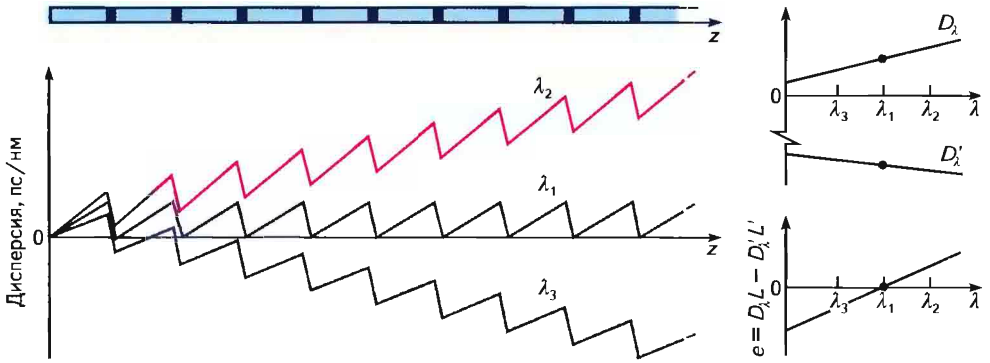


Рис. 24.23. Полная компенсация дисперсии на λ_1 и неполная компенсация с положительной и отрицательной результирующей дисперсией на λ_2 и λ_3 соответственно. Ошибка e_λ исчезает, если наклоны D_λ и D'_λ равны