

Федеральная целевая программа книгоиздания России.

Рецензенты: кафедра «Физика диэлектриков и полупроводников» ГГУ (зав. кафедрой д-р наук, проф. И. А. Карпович); д-р наук, проф. А. А. Блистанов

Пихтин, А. Н.

П 35 **Оптическая и квантовая электроника: Учеб. для вузов / А. Н. Пихтин. — М.: Высш. шк., 2001. — 573 с.: ил.**

ISBN 5-06-002703-1

В книге изложены физические основы оптической электроники, включая процессы взаимодействия электромагнитного излучения с веществом, оптические явления в твердых телах, а также принципы работы, особенности и основные характеристики приборов и методов квантовой электроники и оптоэлектроники.

Для студентов вузов, обучающихся по направлению «Электроника и микроэлектроника».

УДК 621.38
ББК 32.86

ISBN 5-06-002703-1

© ФГУП «Издательство «Высшая школа», 2001

Оригинал-макет данного издания является собственностью издательства «Высшая школа», и его репродуцирование (воспроизведение) любым способом без согласия издательства запрещается.

ПРЕДИСЛОВИЕ

Успехи квантовой электроники определили новые возможности оптики, приведя к возникновению и быстрому развитию нового направления — оптической электроники. Соединив в себе возможности как оптики, так и электроники, оптическая электроника способна решать задачи, не посильные по отдельности ни оптике, ни электронике. Появление лазеров, их непрерывное совершенствование и быстрое освоение в промышленном масштабе привело к созданию принципиально новых приборов и методов, их широкому использованию в самых разнообразных областях науки и техники (физика, химия, медицина, электроника, связь, информатика и вычислительная техника, приборостроение, энергетика, машиностроение, строительство и т. п.). Применение приборов и методов оптической и квантовой электроники во всех этих областях (и даже в искусстве) позволило получить не только количественно, но и качественно новые результаты, способствуя тем самым дальнейшему прогрессу.

Настоящая книга написана в соответствии с учебной программой для вузов и рассчитана на студентов, имеющих подготовку по общей физике, статистической физике и квантовой механике, физике твердого тела. Основная её цель — дать представления о фундаментальных физических процессах, лежащих в основе оптической и квантовой электроники, рассмотреть принцип действия и определить возможности приборов и устройств оптической электроники, подготовить читателя к дальнейшему изучению специальной литературы по отдельным вопросам данной области и к технически грамотному применению приборов оптической электроники. Поскольку книга в первую очередь написана как учебник по направлениям «Электроника» и «Техническая физика» и инженерным специальностям 200.100 и 200.200, то основное внимание в ней уделено рассмотрению физических процессов в активных материалах и приборах, а также свойствам приборов и компонентов оптической электроники. Книга состоит из двух разделов. В первом (гл. 1 — 5) рассмотрены физические основы квантовой и оптической электроники. Этот раздел в зна-

чительной мере базируется на книге автора «Физические основы квантовой электроники и оптоэлектроники» (М., Высшая школа, 1983 г.). Второй раздел (гл. 6—12) посвящен конкретным приборам и методам. Основное внимание уделено приборам оптического диапазона, что определяется отчетливо выраженной тенденцией к возрастанию роли именно оптических методов в электронной технике. Автор основное внимание намеренно уделит рассмотрению физических явлений, лежащих в основе работы приборов оптической электроники, не рассматривая их конструктивные особенности. Практика чтения соответствующих курсов показала, что при хорошем усвоении принципов построения приборов и понимании физических процессов, лежащих в основе их работы, технические подробности усваиваются довольно легко.

Автор выражает благодарность коллективу кафедры физики диэлектриков и полупроводников Нижегородского государственного университета, возглавляемой доктором физ.-мат. наук, профессором Карповичем И. А., и доктору физ.-мат. наук, профессору МИСИС Блистанову А. А. за ценные замечания и советы, высказанные при рецензировании рукописи.

Существенную помощь в работе над книгой оказали своими советами, обсуждением рукописи и поддержкой акад. Ж. И. Алферов и проф. Ю. М. Таиров. Автор выражает им искреннюю признательность.

Надеюсь, что книга будет полезна не только студентам, обучающимся по направлениям «Электроника» и «Техническая физика», но и студентам и аспирантам других специальностей, а также преподавателям, инженерам и научным работникам смежных областей.

Замечания и предложения по содержанию и оформлению книги просьба направлять по адресу: 101430, ГСП-4, Москва, Неглинная ул., 29/14, издательство «Высшая школа».

Автор

ВВЕДЕНИЕ

§ В.1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Прежде чем переходить к рассмотрению сущности предмета, целесообразно договориться о терминологии.

Квантовая электроника — область науки и техники, исследующая и применяющая квантовые явления для усиления, генерации и преобразования когерентных электромагнитных волн.

Оптоэлектроника — область науки и техники, исследующая и применяющая процессы взаимодействия оптического излучения с веществом для передачи, приема, переработки, хранения и отображения информации.

Интегральная оптика — раздел оптоэлектроники, изучающий и применяющий оптические явления в тонкопленочных полупроводниковых и диэлектрических волноводах и структурах, изготовленных на единой подложке методами групповой (интегральной) технологии.

Оптическое излучение — электромагнитное излучение оптического диапазона.

Оптический диапазон спектра составляют электромагнитные колебания, длина волны которых лежит в пределах от 1 м до 1 нм*. Внутри оптического диапазона выделяют *видимое* ($\lambda = 0,38 \dots 0,78$ мкм), *инфракрасное* ($\lambda = 0,78 \dots 1000$ мкм) и *ультрафиолетовое* ($\lambda = 0,001 \dots 0,38$ мкм) излучения (рис. В.1).

Световые волны — электромагнитные волны оптического диапазона.

Монохроматическое излучение — оптическое излучение, характеризующееся какой-либо одной частотой (одной длиной волны) световых колебаний.

*Этот интервал определен Международной комиссией по освещению (МКО) и является условным. Иногда длинноволновую границу оптического диапазона, прилегающую к радиодиапазону, относят к радиоволнам, а коротковолновую — к мягкому рентгеновскому излучению.

Квантовый усилитель — усилитель электромагнитных волн, основанный на использовании вынужденного излучения.

Квантовый генератор — источник когерентного излучения, основанный на использовании вынужденного излучения.

Лазер* (оптический квантовый генератор) — квантовый генератор (усилитель) оптического излучения.

Мазер — квантовый генератор (усилитель) электромагнитного излучения радиодиапазона.

Вынужденное излучение — когерентное электромагнитное излучение, возникающее в результате вынужденного испускания.

Вынужденное испускание — когерентное испускание фотона при квантовом переходе системы в результате взаимодействия с внешним электромагнитным полем.

Когерентность — согласованное протекание во времени и в пространстве колебательных или волновых процессов. Электромагнитная волна называется когерентной, если её амплитуда, частота, фаза, направление распространения и поляризация постоянны или изменяются по определенному закону (упорядоченно).

§ В.2. ОСОБЕННОСТИ ОПТИЧЕСКОЙ ЭЛЕКТРОНИКИ

Современная оптическая и квантовая электроника определяют новые возможности как электроники, так и оптики, но не пересматривают их фундаментальные положения. Необходимость дальнейшего освоения оптического диапазона с использованием хорошо развитых методов радиофизики, радиотехники и электроники определяется рядом принципиальных обстоятельств.

1. Частота электромагнитных колебаний (несущая частота ν_0) в оптическом диапазоне существенно выше, чем в радиодиапазоне (рис. В.1). Например, частота световых колебаний в наиболее освоенной видимой и ближней инфракрасной областях спектра ($\sim 10^{15} \dots 10^{13}$ Гц) в миллионы раз превышает частоту радиоволн в областях радио- и телевещания. Это определяет высокую информационную емкость оптического канала связи. Напомним, что для передачи обычного телевизионного изображения требуется полоса частот $\Delta\nu \approx 5$ МГц. Поэтому в метровом диапазоне (при $\lambda = 1$ м $\nu_0 = 300$ МГц) можно передать лишь около десятка телевизионных программ. В оптическом диапазоне при том же отношении $\Delta\nu/\nu_0$ это число возрастает в миллионы раз.

* Термины «лазер» (от англ. laser) и мазер (от англ. maser) образованы от начальных букв фразы Light (Microwave) Amplification by Stimulated Emission of Radiation — усиление света (СВЧ-волн) с помощью индуцированного излучения. В этих названиях по существу определен принцип работы квантовых усилителей и генераторов.

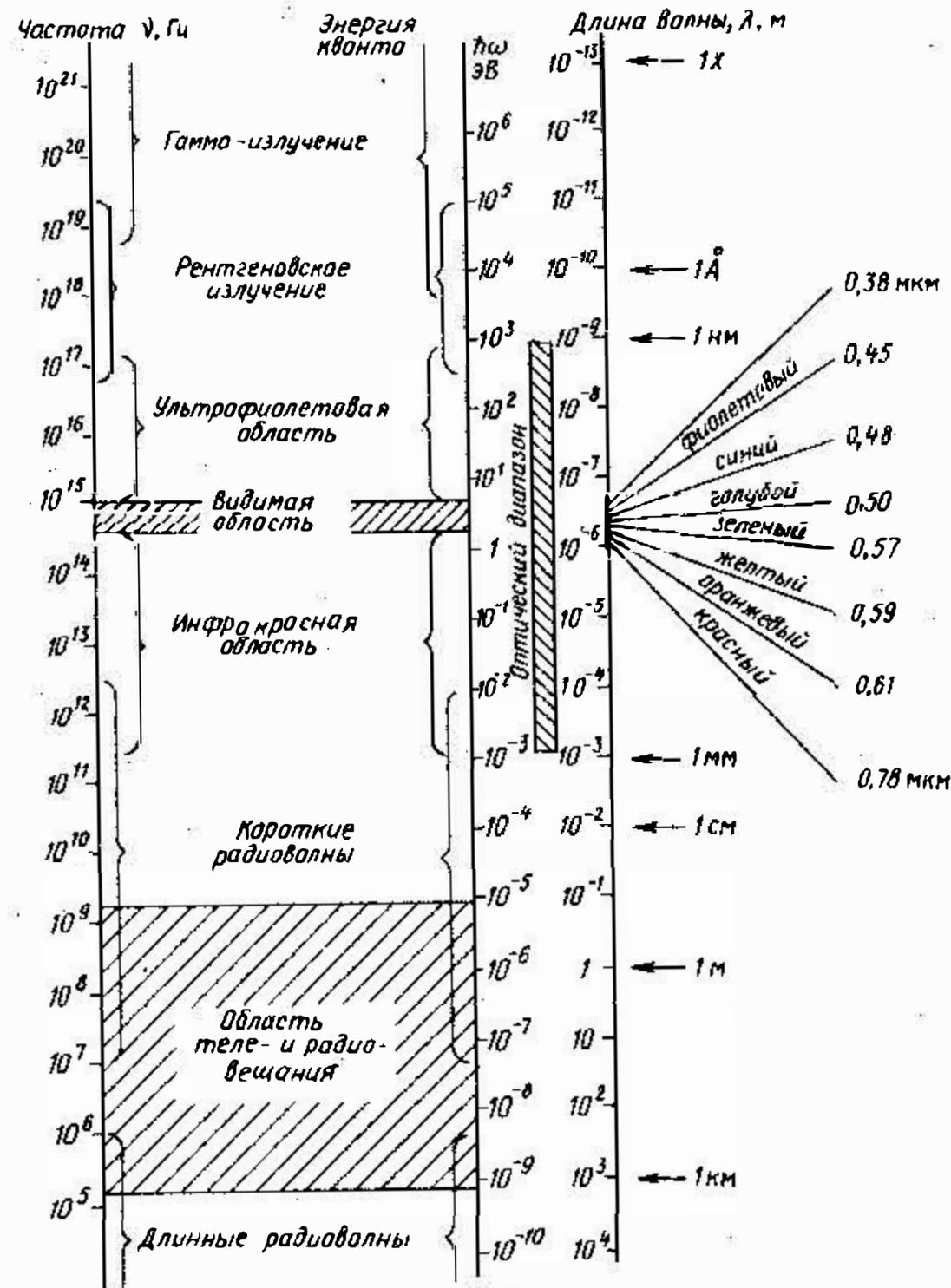


Рис. В.1. Шкала электромагнитных волн

2. Длина световых волн существенно меньше, чем длина радиоволн. Это позволяет получить высокую концентрацию оптического излучения в пространстве, поскольку минимальный объем, в котором можно сфокусировать электромагнитное излучение, имеет размеры порядка длины волны. Поэтому оптические волноводы (световоды) при прочих равных условиях обладают существенно меньшими (на несколько порядков) размерами по сравнению с СВЧ-волноводами, что важно с точки зрения микроминиатюризации электронной аппаратуры. И, наконец, в оптическом диапазоне не трудно сформировать узкую

диаграмму направленности излучения с углом расходимости $0,1^\circ$ и менее. Для формирования подобной диаграммы в радиодиапазоне (при $\lambda = 1$ м) потребовалась бы антенна диаметром порядка сотен метров. В оптическом диапазоне функцию такой антенны способны выполнить, например, сферическое зеркало или линза умеренных размеров, поскольку для получения одинаковой диаграммы направленности размер антенны пропорционален длине волны.

3. Передача информации осуществляется фотонами. В отличие от электронов, которые служат основными носителями информации в обычных электронных приборах, фотоны являются электрически нейтральными частицами, не взаимодействующими между собой и с внешними электрическим и магнитным полями. Это определяет возможность идеальной гальванической развязки входа и выхода, однонаправленность потока информации, высокую помехозащищенность, исключение взаимных наводок и паразитных связей между различными элементами схемы. Поэтому использование оптических методов в современной микроэлектронике заметно расширяет её функциональные возможности, позволяя выполнять многоканальные сложные связи и осуществлять «оптический монтаж», исходя лишь из требуемых функциональных задач. Так как фотон в оптоэлектронных системах является основным носителем информации, то, по аналогии с электронной, оптоэлектронную называют также фотоникой.

4. Применение оптических методов записи, хранения и обработки информации открывает новые возможности для построения ЭВМ. Это обусловлено, с одной стороны, возможностью реализации новых принципов параллельной обработки информации (например, на основе голографических методов), а с другой — возможностью достижения высокой плотности записи ($\sim 10^8$ бит/см²) в оптических запоминающих устройствах.

§ В.3. ИСТОРИЧЕСКАЯ СПРАВКА

Оптические методы передачи информации существовали с незапамятных времен. Первобытный человек использовал для этих целей ночью зажженные факелы или костры, а днем — сигнальные дымки. Такие способы были хорошо известны в Древнем Китае, у ассирийцев, в Египте, начиная с первых цивилизаций. Согласно преданию, в 1084 г. до н.э. Агамемнон, предводитель ахейского войска, с помощью девяти эстафет из огней на холмах передал своей жене весть о взятии Трои и о своем скором возвращении. Оракул этой «оптической тринивязи» составила 800 км и была превзойдена лишь в наше время; спустя почти 3000 лет. Правда, использована эта информация была весьма своеобразно: в день возвращения из-под Трои в Аргос Агамемнон был

убит своей любимой женой Клитемнестрой и её любовником Эгистом. Справедливость всё же восторжествовала: Орест, сын Агамемнона и Клитемнестры, возмужав, отомстил за смерть отца.

Оптика является одной из древнейших наук. В освоении оптического излучения можно выделить два исторических этапа. Первый связан в основном с изучением наблюдаемого непосредственно глазом видимого света и соответствует развитию классической оптики. Эволюция взглядов на природу света иллюстрирует диалектический характер познания. Великие мыслители древности считали, что световые лучи исходят из глаз (Платон). На основе учения о зрительных лучах, исходящих из глаз, Эвклид, Птоломей и другие мыслители древности создали теорию отражения света от плоских и сферических зеркал, им же было положено начало развития геометрической оптики. Теория зрительных лучей в древности противопоставлялась еще более фантастической теории Эпикура и Лукреция о «слепках» с предметов, летящих во всех направлениях и попадающих в глаз. Затем появилась корпускулярная концепция, или теория истечения, поддерживаемая вначале Декартом, а затем Ньютоном. Согласно этой теории свет представляет совокупность мельчайших частиц — корпускул, движущихся по определенной траектории — световому лучу. Эта теория наглядно и естественно объясняла такие явления, как прямолинейность распространения света в однородных средах, отражение света от поверхности зеркал, преломление светового луча на границе двух сред, а также его искривление при распространении в неоднородных средах. Одновременно Х. Гюйгенсом была предложена интерпретация тех же явлений на основе понятий «световая волна» и «волновая поверхность», согласно которым свет представляет волны, распространяющиеся в пространстве. Лучи света являются чисто абстрактным понятием и определяются как кривые, ортогональные волновым поверхностям.

Главным доводом Ньютона против волновой теории было отсутствие вещественной среды («эфира») в мировом пространстве. Явления, связанные с периодичностью (кольца Ньютона), в корпускулярной теории объяснялись тем, что частицы вращаются. Пространство, пробегаемое такой частицей за один оборот, сопоставлялось с некоей «длиной волны». Полярность (в современной терминологии поляризацию) Ньютон считал свойством только твердых частиц. Огибание (дифракцию) он пытался истолковать «отталкивательным» и «притягательным» действием вещества на световые корпускулы. Вслед за Ньютоном в XVIII в. большинство ученых стали склоняться к корпускулярной теории, и у волновой теории, блестяще развитой Гюйгенсом, сохранялось лишь немного последователей.

Только на рубеже XVIII — XIX веков англичанином Томасом Юнгом были начаты серьезные исследования интерференции и дифракции, а французом Огюстом Френелем дано их полное теоретическое объяснение на основе волновой теории Гюйгенса. Кроме того, Френель показал, что представления о волновой природе не противоречат факту прямолинейности распространения света в однородной среде. Точностью своих предсказаний волновая теория в первой половине XIX в. победила теорию истечения. К концу XIX в. Максвелл дал волнам Френеля электромагнитную интерпретацию и показал, что всякая световая волна является электромагнитным возмущением особого рода. Опыты Г. Герца и А. С. Попова экспериментально подтвердили это.

Электромагнитная теория, обобщенная в виде системы дифференциальных уравнений Максвелла, явилась вершиной первого «классического» этапа развития оптики и наших представлений о природе света.

Второй этап развития оптики тесно связан с революционными открытиями в физике в начале XX в. Характерно, что именно изучение оптических спектров поглощения и испускания привело к необходимости введения понятий о квантовых скачках и кванте действия h как минимальном действии, которые ввел в 1900 г. М. Планк для объяснения спектра излучения черного тела. Впоследствии постоянная h , имеющая размерность «действия» (Дж·с), была названа постоянной Планка. В 1905 г. А. Эйнштейн на основе теории Планка возродил в новой форме корпускулярную теорию света, предположив, что планковские кванты энергии $E = h\nu$ существуют в виде реальных частиц, названных им световыми квантами. Таким образом, Эйнштейну удалось объяснить открытый ранее фотоэффект. Применяя эти понятия к атому, Нильс Бор в 1913 г. объяснил просгую связь частоты излучения ν с разницей энергий между уровнями n и m : $\nu_{nm} = \left| \frac{E_n - E_m}{h} \right|$.

Фундаментальную роль для последующего развития квантовой электроники сыграла работа А. Эйнштейна (1917 г.), в которой он, рассмотрев термодинамическое равновесие системы молекул, ввел понятие индуцированного излучения. На возможность использования индуцированного излучения для наблюдения отрицательного поглощения (усиления) впервые указал в 1940 г. В. А. Фабрикант.

В конце 1954 г. независимо и почти одновременно в лаборатории колебаний Физического института им. П. Н. Лебедева в Москве под руководством А. М. Прохорова был разработан конкретный проект, а в лаборатории излучений физического факультета Колумбийского университета в Нью-Йорке под руководством Ч. Таунса был создан действующий мазер на пучке моле-

кул аммиака. Это был первый прибор, работавший на квантовых принципах, в основе которого лежало явление усиления электромагнитных колебаний с помощью индуцированного излучения. За эти работы Н. Г. Басову и А. М. Прохорову была присуждена Ленинская премия, а затем совместно с американским физиком Ч. Таунсом — Нобелевская премия. Таким образом, 1954 год может быть назван годом рождения квантовой электроники как самостоятельной науки.

Методы, развитые первоначально для радиодиапазона (первый аммиачный мазер работал на длине волны 1,25 см), затем были перенесены в оптический диапазон, и в 1960 г. был создан рубиновый лазер, а в 1961 г. — газовый лазер, работающий на смеси гелия и неона. Наука и техника впервые получили когерентный источник световых волн. Это дало толчок развитию таких новых областей науки, как нелинейная оптика, голография. Основные работы по голографии были выполнены Д. Габором и Ю. Н. Денисюком.

Параллельно с развитием квантовой электроники быстрыми темпами развивалась физика полупроводников и полупроводниковая электроника. Успехи в этих областях привели к созданию эффективных полупроводниковых фотоприемников и генераторов света — приборов, составляющих основу полупроводниковой оптоэлектроники. Назовем главные вехи в развитии данной отрасли. В 1873 г. В. Смит обнаружил в слоях селена фотопроводимость — изменение сопротивления под действием освещения, открыв тем самым внутренний фотоэффект. Внешний фотоэффект был открыт в 1888 г. А. Г. Столетовым. В 1923 г. О. В. Лосев наблюдал свечение кристаллов карборунда (карбида кремния) под действием электрического тока и дал правильное объяснение этому явлению, которое легло в основу действия современных электролюминесцентных источников света. В 1960 г. в Физическом институте им. П. Н. Лебедева в Москве был рассмотрен принцип работы полупроводникового инжекционного лазера, а в 1962 г. в Физико-техническом институте им. А. Ф. Иоффе в Ленинграде было обнаружено эффективное рекомбинационное излучение и наблюдалось вынужденное излучение в кристаллах арсенида галлия. Параллельно в Физико-техническом институте Ж. И. Алферовым с сотрудниками успешно проводились работы по получению гетеропереходов и созданию на их базе эффективных приемников и источников света. В 1968 — 1970 гг. ими были созданы низкороговые полупроводниковые лазеры, в том числе работающие в непрерывном режиме при комнатной температуре. За эти работы акад. Ж. И. Алферову в 2000 г. была присуждена Нобелевская премия по физике. Применение квантово-размерных слоев и сверхрешеток позволило создать полупроводниковые лазеры с характеристиками, близкими к теоретическому пределу. Важным моментом в развитии

оптической электроники явилось получение оптических волокон с низкими потерями (< 1 дБ/км), что обеспечило возможность их применения в качестве эффективных оптических волноводов. Разработка эффективных полупроводниковых лазеров, работающих в непрерывном режиме, и создание оптических волокон с малыми потерями привели к рождению и быстрому развитию новой системы передачи информации через волоконно-оптические линии связи (ВОЛС), общая протяженность которых составляет миллионы километров. Разрабатываются оптические волокна с потерями до 0,005 дБ/км. Отчетливо просматриваются тенденции к увеличению скорости передачи информации, которая в современных ВОЛС уже достигает 2,4 Гбит/с (до 10 Гбит/с). В лабораторных условиях продемонстрирована возможность передачи оптической информации на скоростях 1000 Гбит/с. Это требует новых подходов в электронике. Методами интегральной оптики создаются оптические аналоги электронных схем, работающие на частотах до 10^{12} Гц. Интенсивно ведутся работы по созданию устройств оптической памяти. В настоящее время стоит вопрос о разработке оптических вычислительных машин, работающих на новых принципах. Но это лишь начало. По мнению большинства специалистов, оптическая электроника будет в значительной мере определять технику завтрашнего дня.

ГЛАВА 1 ВЗАИМОДЕЙСТВИЕ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ С АТОМНЫМИ СИСТЕМАМИ

Элементарными процессами, определяющими взаимодействие оптического излучения с веществом, являются поглощение, испускание и рассеяние света атомными системами. Эти процессы настолько сложны, что их детальное рассмотрение выходит за рамки квантовой механики и может быть выполнено лишь в рамках квантовой электродинамики, где рассматривается взаимодействие квантованного электромагнитного поля с атомными системами. Но в подавляющем большинстве случаев нет необходимости использовать весьма громоздкий математический аппарат квантовой электродинамики. Достаточно ограничиться соответствующим приближением. Часто используют полуфеноменологический подход, при котором электромагнитное излучение описывается на языке классической электродинамики, а энергетические состояния объектов, с которыми взаимодействует это излучение, — на языке квантовой механики.

§ 1.1. ОПТИЧЕСКОЕ ИЗЛУЧЕНИЕ

Определение оптического излучения как электромагнитного излучения оптического диапазона было дано во введении. В радиодиапазоне в большинстве случаев проявляются волновые свойства излучения и его удобно рассматривать в виде определенного набора электромагнитных волн. В диапазоне жесткого рентгеновского и гамма-излучения (см. рис. В.1), наоборот, проявляются преимущественно корпускулярные свойства излучения, а его волновые свойства часто можно не учитывать. В оптическом диапазоне в одинаковой мере проявляются как волновые, так и корпускулярные свойства света. Квантово-волновой дуализм выступает здесь во всей своей полноте. Поэтому оптическое излучение, строго говоря, необходимо рассматривать как электромагнитное поле, подчиняющееся квантовым закономерностям. Однако при обсуждении многих практически важных

вопросов можно ограничиться приближенными представлениями и рассматривать оптическое излучение в виде: световых лучей (или пучков); световых (электромагнитных) волн; фотонных коллективов. Рассмотрим эти приближения.

Световые лучи. Простейшее представление оптического излучения в виде светового луча (или совокупности световых лучей, образующих световой пучок) соответствует приближению *геометрической оптики*. В этом приближении пренебрегают конечностью длины световой волны и оптические законы формулируют на языке геометрии. Считают, что световая энергия распространяется вдоль определенных кривых, называемых *световыми лучами*. В однородной среде световые лучи являются прямыми линиями. Пучки световых лучей не взаимодействуют друг с другом и после пересечения распространяются независимо.

Основу геометрической оптики составляет *принцип Ферма*, называемый также *принципом наименьшего оптического пути* или *принципом наименьшего времени*. Согласно этому принципу оптическая длина пути, по которому распространяется свет между двумя точками A и B , короче оптической длины любой другой кривой, соединяющей эти точки. Напомним, что *оптическая длина* dS равна произведению показателя преломления n на геометрическую длину dl , так что оптическая длина кривой AB есть интеграл

$$\int_A^B n dl. \quad (1.1)$$

Для однородной среды $n = \text{const}$ и, очевидно, $S = nl$. Так как $dl = \frac{c}{n} dt$, то

$$\int_A^B n dl = c \int_A^B dt. \quad (1.2)$$

Отсюда видно, что принцип Ферма можно сформулировать как принцип *наименьшего времени*: из всех возможных путей, соединяющих две точки A и B , свет выбирает тот путь, который требует наименьшего времени для его прохождения. Из этого принципа вытекают прямолинейность распространения света в однородной среде, а также законы преломления и отражения.

Геометрическая оптика является предельным случаем волновой оптики и соответствует $\lambda \rightarrow 0$.

Электромагнитные волны. Явления дифракции и интерференции не могут быть объяснены в рамках геометрической оптики и свидетельствуют о волновой природе света. В приближении классической *волновой оптики* свет представляет собой электромагнитные волны (колебания), распространяющиеся в вакууме

с постоянной и наибольшей известной скоростью $c = 299792 \pm 0,5$ км/с. Волновая оптика базируется на уравнениях классической *электродинамики*, основу которой составляют уравнения Максвелла. Напомним, что теория Максвелла является феноменологической, в ней последовательно обобщены основные законы, описывающие электрические и магнитные явления и установленные экспериментально: закон Кулона, закон Био — Савара — Лапласа и закон электромагнитной индукции. Электрические и магнитные свойства среды учитываются с помощью трех величин: относительной диэлектрической проницаемости ϵ_r , относительной магнитной проницаемости μ_r и удельной электрической проводимости σ . Важно, что при распространении электромагнитной волны в различных средах ее частота остается неизменной.

Полная система уравнений Максвелла в дифференциальной форме имеет вид

$$\begin{aligned} \text{rot } \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, & \text{div } \mathbf{D} &= \rho, \\ \text{rot } \mathbf{H} &= \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, & \text{div } \mathbf{B} &= 0. \end{aligned} \quad (1.3)$$

Эта система дополняется уравнениями, связывающими векторы \mathbf{E} , \mathbf{D} , \mathbf{B} , \mathbf{H} , \mathbf{j} с величинами, описывающими электрические и магнитные свойства среды (иногда их называют материальными уравнениями):

$$\mathbf{D} = \epsilon_r \epsilon_0 \mathbf{E}, \quad \mathbf{B} = \mu_r \mu_0 \mathbf{H}, \quad \mathbf{j} = \sigma \mathbf{E}, \quad (1.4)$$

где ϵ_r и μ_r — относительные диэлектрическая и магнитная проницаемости; ϵ_0 и μ_0 — электрическая и магнитная постоянные ($\epsilon_0 = 10^7/4\pi c^2$ Ф/м, $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м); σ — удельная электрическая проводимость.

Величины ϵ_r , μ_r и σ отражают реакцию среды на внешнее электромагнитное возмущение.

На границе раздела двух сред выполняются граничные условия для тангенциальных (индекс τ) и нормальных (индекс n) составляющих

$$\begin{aligned} E_{\tau 1} &= E_{\tau 2}, & D_{n1} - D_{n2} &= \sigma_{\text{пов}}, \\ H_{\tau 1} - H_{\tau 2} &= j_{\text{пов}}, & B_{n1} &= B_{n2}. \end{aligned} \quad (1.5)$$

Здесь $\sigma_{\text{пов}}$ — поверхностная плотность свободных зарядов; $j_{\text{пов}}$ — проекция вектора плотности поверхностных токов проводимости на направление $[\mathbf{tn}]$.

Система уравнений, включающая в себя уравнения электромагнитного поля (1.3), уравнения (1.4) и граничные условия (1.5), называется *системой уравнений Максвелла* и играет в классичес-

кой электродинамике ту же роль, что и аксиоматика уравнений Ньютона в классической механике. Эта система является полной, т. е. из нее можно получить все свойства электромагнитного поля.

Из системы (1.3) видно, что переменное магнитное поле порождает переменное электрическое поле, и наоборот. В результате этого в пространстве распространяется переменное электромагнитное поле, называемое *электромагнитными волнами*. Такие волны, в частности, могут распространяться в вакууме. Действительно, для вакуума $j=0$, $\sigma=0$, $\mu_r=1$, $\epsilon_r=1$ и система (1.3) имеет вид:

$$\operatorname{rot} \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \quad \operatorname{div} \mathbf{E} = 0, \quad (1.6)$$

$$\operatorname{rot} \mathbf{H} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad \operatorname{div} \mathbf{H} = 0.$$

Отсюда находим

$$\operatorname{rot} \operatorname{rot} \mathbf{E} = -\mu_0 \frac{\partial}{\partial t} \operatorname{rot} \mathbf{H} = -\epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad (1.7)$$

Из векторного анализа известно соотношение $\operatorname{rot} \operatorname{rot} \mathbf{X} = \operatorname{grad} \operatorname{div} \mathbf{X} - \Delta \mathbf{X}$, справедливое для любого вектора \mathbf{X} . Используя его, имеем

$$\operatorname{rot} \operatorname{rot} \mathbf{E} = -\Delta \mathbf{E},$$

откуда получаем волновое уравнение для вектора напряженности электрического поля:

$$\Delta \mathbf{E} - \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad (1.8)$$

(напомним, что волновое уравнение для продольной плоской волны, распространяющейся в направлении z со скоростью c , имеет вид $\frac{\partial^2 \varphi}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2}$).

Аналогично для вектора напряженности магнитного поля можно написать

$$\Delta \mathbf{H} - \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0. \quad (1.8a)$$

Для решения системы уравнений Максвелла часто вместо векторов \mathbf{E} и \mathbf{H} вводят скалярный φ и векторный \mathbf{A} потенциалы:

$$\mathbf{B} = \operatorname{rot} \mathbf{A}, \quad \mathbf{E} = -\operatorname{grad} \varphi - \frac{\partial \mathbf{A}}{\partial t}. \quad (1.9)$$

Как известно, такое определение не однозначно. Потенциалы \mathbf{A} и φ могут быть выбраны так, чтобы выполнялись условия $\operatorname{div} \mathbf{A} = 0$, $\varphi = 0$. Тогда

$$\mu_0 \mathbf{H} = \operatorname{rot} \mathbf{A}, \quad \mathbf{E} = -\frac{\partial}{\partial t} \mathbf{A}. \quad (1.10)$$

Подставляя (1.10) в (1.6), находим

$$\operatorname{rot} \operatorname{rot} \mathbf{A} = \mu_0 \operatorname{rot} \mathbf{H} = \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} = -\epsilon_0 \mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2},$$

откуда аналогично (1.8) получаем, что векторный потенциал \mathbf{A} удовлетворяет, как и векторы \mathbf{E} и \mathbf{H} , волновому уравнению

$$\Delta \mathbf{A} - \epsilon_0 \mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \quad (1.11)$$

Рассмотрим основные свойства электромагнитных волн.

1. Электромагнитные волны распространяются в вакууме со скоростью $v = 1/\sqrt{\epsilon_0 \mu_0} = c \approx 3 \cdot 10^8$ м/с; постоянная c называется *электродинамической постоянной* или *скоростью света*.

2. Электромагнитные волны являются поперечными. Векторы \mathbf{v} , \mathbf{E} и \mathbf{H} взаимно перпендикулярны и образуют правовинтовую систему: из конца вектора \mathbf{v} вращение от \mathbf{E} и \mathbf{H} происходит по кратчайшему расстоянию против часовой стрелки.

3. Энергия электромагнитного поля в единице объема, называемая *объемной плотностью энергии излучения*,

$$\rho = \epsilon_0 E^2/2 + \mu_0 H^2/2. \quad (1.12)$$

Распространение электромагнитной энергии в пространстве характеризуется *вектором Пойнтинга* \mathbf{S} :

$$\mathbf{S} = [\mathbf{E} \times \mathbf{H}]. \quad (1.13)$$

Его направление совпадает с направлением распространения энергии излучения, а абсолютное значение в единицах Дж/(м²с) = Вт/м² равно количеству энергии, переносимой электромагнитной волной в единицу времени через единицу поверхности, перпендикулярную направлению вектора \mathbf{S} , т. е.

$$|\mathbf{S}| = \rho c. \quad (1.13a)$$

Отношение энергии, переносимой излучением, ко времени переноса, значительно превышающему период световых колебаний, называется *поток излучения* или *световым потоком*. Он измеряется в ваттах [Вт].

В качестве энергетической характеристики часто пользуются понятием интенсивности излучения I , под которой понимают величину $I = |\mathbf{S}|$.

Линия, касательная в каждой точке которой совпадает с направлением переноса энергии (т. е. с направлением распространения волны в этой точке), называется *световым лучом*.

4. Наряду с энергией электромагнитное поле переносит импульс (количество движения), распределенный в пространстве с объемной плотностью P :

$$P = [E \times H]/c^2 = S/c^2. \quad (1.14)$$

В общем случае решения волновых уравнений (1.8) имеют вид:

$$\begin{aligned} E &= E_1(1 - rn/c) + E_2(t + rn/c), \\ H &= H_1(t - rn/c) + H_2(t + rn/c), \end{aligned} \quad (1.15)$$

где r — радиус-вектор, проведенный из начала координат в рассматриваемую точку волны; n — единичный вектор нормали к поверхности волны.

Из этих соотношений видно, что в каждой точке $r = \text{const}$ поле изменяется во времени и в каждый данный момент времени поле различно для различных r . Очевидно, поле имеет одинаковое значение для координат r и моментов времени t , удовлетворяющих соотношению $t \mp rn/c = \text{const}$ (знак «-» для E_1 и H_1 , знак «+» для E_2 и H_2).

Таким образом, векторы E_1 и H_1 определяют волны, распространяющиеся в направлении возрастания r , а векторы E_2 и H_2 — в обратном направлении. Не нарушая общности, можно ограничиться рассмотрением волны, распространяющейся только в сторону положительных значений r , т. е. учесть лишь первые слагаемые в (1.15). В общем случае величины E_1 и H_1 могут выражаться сложными функциями координат и времени.

Частный случай представляют плоские монохроматические (гармонические) волны. Электромагнитная волна называется *плоской*, если векторы E и H зависят только от времени и одной декартовой координаты. Например, для плоской волны, распространяющейся в направлении z ($E_x = H_x = 0$), волновые уравнения (1.8) принимают вид:

$$\begin{aligned} \frac{\partial^2}{\partial z^2} E(z, t) - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} E(z, t) &= 0, \\ \frac{\partial^2}{\partial z^2} H(z, t) - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} H(z, t) &= 0. \end{aligned} \quad (1.16)$$

В плоской волне все лучи параллельны друг другу. Энергия электрического поля световой волны равна энергии магнитного поля: $\varepsilon_0 E^2/2 = \mu_0 H^2/2$, поэтому

$$H = \sqrt{\frac{\varepsilon_0}{\mu_0}} [n \times E], \quad (1.17)$$

где n — единичный вектор в направлении распространения волны.

Подставляя (1.17) в (1.13), с учетом (1.12) находим

$$S = n \rho c. \quad (1.18)$$

Электромагнитная волна называется *монохроматической*, если компоненты векторов E и H электромагнитного поля совершают гармонические колебания. Их зависимости от времени определяются множителем вида $\cos(\omega t + \varphi)$, где ω — *угловая частота* (или просто частота) волны. Вторая производная от поля по времени, например, для вектора E : $\partial^2 E/\partial t^2 = -\omega^2 E$, и распределение поля по пространству в монохроматической волне соответствует уравнению

$$\Delta E + \frac{\omega^2}{c^2} E = 0, \quad (1.19)$$

которое вытекает из (1.8). Аналогичные соотношения получаются из уравнений (1.8a), (1.11) для векторов H и A .

В плоской волне, распространяющейся в направлении z , поле является функцией только $(t - z/c)$. Поэтому плоскую монохроматическую волну можно представить в виде

$$E = E_0 \cos \omega(t - z/c), \quad (1.20)$$

или

$$E = \text{Re} \left\{ E_0 e^{-i\omega(t - z/c)} \right\} = \text{Re} \left\{ E_0 e^{i\omega(t - z/c)} \right\}, \quad (1.21)$$

где Re — вещественная часть комплексного выражения, стоящего в фигурных скобках, а E_0 — некоторый постоянный комплексный вектор, называемый *амплитудным вектором*. Укажем, что выражения (1.20) и (1.21) эквивалентны. Обычно замена выражения (1.20) комплексной формой (1.21) облегчает преобразование*.

Величина $\lambda = 2\pi c/\omega$ называется *длиной волны* монохроматического излучения. Это есть расстояние, на которое смещается поверхность равной фазы волны за один период колебаний.

Интервал времени T , за которое совершается одно колебание, т. е. за которое волна проходит расстояние, равное длине волны, а фаза гармонического колебания изменяется на 2π , называется *периодом колебаний*:

$$T = 2\pi/\omega = 1/\nu.$$

Вектор

$$k = \frac{\omega}{c} n = \frac{2\pi}{\lambda} n \quad (1.22)$$

называется *волновым вектором*. С его помощью выражение (1.21) удобно представить в виде, не зависящем от выбора осей координат:

*Исключения составляют операции вычисления производных синусоидальных функций, т. е. мощностей. В этом случае лучше использовать действительные функции (1.20).

$$E = E_0 e^{i(kr - \omega t)} \quad (1.23)$$

Аналогично для вектора напряженности магнитного поля можно написать

$$H = H_0 e^{i(kr - \omega t)} \quad (1.23a)$$

и для векторного потенциала

$$A = A_0 e^{i(kr - \omega t)} \quad (1.23b)$$

Величину, стоящую с множителем i в показателе экспоненты, называют *фазой волны*. Скорость c в выражениях (1.20), (1.21) есть фазовая скорость света.

Рассмотрим подробнее вопрос о направлении поля монохроматической волны. Известно, что векторы E и H и соответственно амплитудные векторы E_0 и H_0 должны быть расположены в плоскости, перпендикулярной направлению распространения света. Для более точного описания ориентации плоской волны служит понятие поляризации. Свет, у которого существует упорядоченность ориентации векторов электрического и магнитного полей, называют *поляризованным*. *Линейно поляризованным* (или *плоскополяризованным*) называют свет, у которого направления колебаний векторов E и H в любой точке пространства остаются неизменными во времени. *Плоскостью поляризации* при этом является плоскость, проходящая через вектор E , и направление распространения электромагнитной волны. Для линейно поляризованного света, распространяющегося в направлении r :

$$E = E_0 e^{i(kr - \omega t)}, \quad (1.24)$$

где E_0 — амплитуда волны.

«Мгновенная фотография» плоской линейно поляризованной монохроматической волны представлена на рис. 1.1.

Свет, у которого векторы электрического и магнитного полей в любой точке пространства вращаются, а концы этих векторов описывают эллипсы, называется *эллиптически поляризованным*. Его частным случаем является свет, поляризованный по кругу, так называемый *циркулярно поляризованный свет*. В зависимости от направления вращения различают правую и левую поляризацию (соответственно по часовой стрелке и против нее, если

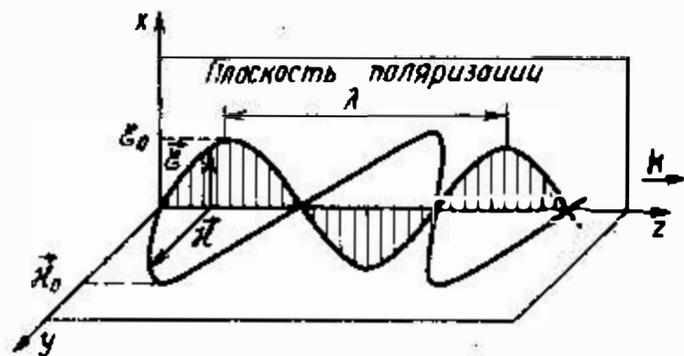


Рис. 1.1. «Мгновенная фотография» плоской линейно поляризованной электромагнитной волны

описывают эллипсы, называется *эллиптически поляризованным*. Его частным случаем является свет, поляризованный по кругу, так называемый *циркулярно поляризованный свет*. В зависимости от направления вращения различают правую и левую поляризацию (соответственно по часовой стрелке и против нее, если

смотреть навстречу волне). Эллиптически поляризованную волну можно рассматривать как наложение двух линейно поляризованных волн.

Свет, у которого векторы электрического и магнитного полей хаотически изменяют свое направление, называют *неполяризованным* или *естественным*.

Частота электромагнитных колебаний в оптическом диапазоне очень велика. Все известные приемники оптического излучения реагируют не на электрическое или магнитное поле волны, а на плотность светового потока, усредненную за время наблюдения. *Интенсивность I электромагнитной волны* (света) есть величина, численно равная энергии, переносимой волной за единицу времени через единицу площади поверхности, перпендикулярной направлению распространения волны, т. е. она определяется модулем вектора Пойнтинга (1.13), (1.18):

$$I = \langle S \rangle = \frac{1}{T} \int_0^T S dt = \rho c, \quad (1.25)$$

где угловые скобки $\langle \rangle$ показывают усреднение по времени.

Для плоской линейно поляризованной монохроматической волны в формуле (1.24) $I \sim E_0^2$.

Понятие плоской монохроматической волны, описываемой соотношениями (1.23), (1.24), является идеализированным (*абстрактным*). В природе таких волн не существует, так как они должны простирались во времени и в пространстве от $-\infty$ до $+\infty$. Даже если имеется идеальная синусоида, которая «начинается» в момент времени t_1 и «оканчивается» в момент времени t_2 , как изображено на рис. 1.2, то она не является монохроматической волной (соотношения (1.23), (1.24) должны быть справедливы при любых значениях t) и ее можно представить состоящей из совокупности плоских монохроматических волн, т. е. в виде разложения Фурье. Совокупность монохроматических волн, образующих реальную волну, называется *группой волн* или *волновым пакетом*. Важность понятия плоской поляризованной монохроматической волны заключается в том, что любую реальную волну можно представить набором (суперпозицией) идеализированных монохроматических волн с разными частотами ω . Действительно, функцию $f(t)$ можно представить в виде *разложения Фурье по монохроматическим (гармоническим) волнам*

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{-i\omega t} d\omega. \quad (1.26)$$

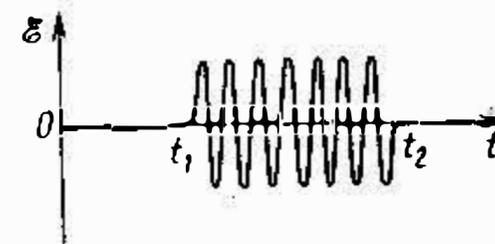


Рис. 1.2. Пример немонохроматической волны — пакет волн

Здесь под $f(t)$ понимают любую из компонент векторов E , H или A . Интеграл Фурье (1.26) представляет функцию $f(t)$ в виде суммы бесконечно большого числа гармонических колебаний с непрерывно изменяющейся частотой. Разложение в интеграл Фурье называется также *спектральным разложением*. Частоте ω соответствует плотность спектра $F(\omega)$ — так называемая *фурье-компонента*. Она выражается через $f(t)$ известным соотношением

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \quad (1.27)$$

и имеет смысл амплитуды монохроматической волны с частотой ω .

Таким образом, согласно классическим представлениям, световая волна в общем случае является суперпозицией монохроматических волн. Каждая плоская монохроматическая волна характеризуется амплитудой, волновым вектором k и поляризацией.

Для того чтобы охарактеризовать распределение энергии по чистоте, т. е. спектральное распределение, вводят функцию $\rho(\omega)$, называемую *спектральной плотностью излучения* и связанную с интегральной плотностью излучения ρ [см. выражение (1.12)] соотношением *

$$\rho = \int_0^{\infty} \rho(\omega) d\omega. \quad (1.28)$$

Отсюда следует, что $\rho(\omega) d\omega$ есть энергия электромагнитного поля в единице объема в интервале частот от ω до $\omega + d\omega$.

Фотоны. Ряд экспериментальных фактов свидетельствует, что кроме волновых свойств свету (электромагнитному излучению) присущи корпускулярные свойства. На это указывают явления внешнего и внутреннего фотоэффектов, эффект Комптона (упругое рассеяние света на свободных электронах, сопровождающееся увеличением длины световой волны), законы излучения абсолютно черного тела (в частности, спектр излучения) и др. Элементарная частица (квант) электромагнитного излучения называется *фотоном*** . В настоящее время существуют фотоприемники, способные регистрировать (считать) отдельные фотоны.

Перечислим основные свойства фотона.

*Под спектральной плотностью какой-либо величины понимают отношение её среднего значения в рассматриваемом малом спектральном интервале к ширине этого интервала. Поэтому в более полном смысле $\rho(\omega)$ есть «спектральная плотность объёмной плотности энергии излучения», но для краткости её называют просто спектральной плотностью излучения.

**От греч. *phos* — свет; термин «фотон» был введен в физику Г. Н. Льюисом в 1929 г.

1. Энергия E и импульс P фотона связаны с частотой и волновым вектором k эквивалентной плоской монохроматической волны соотношениями:

$$E_{\text{фот}} = \hbar\omega, \quad (1.29)$$

$$P_{\text{фот}} = \hbar k, \quad (1.30)$$

где $\hbar = h/2\pi = 1,05 \cdot 10^{-34}$ Дж·с — постоянная Планка.

2. Масса покоя фотона равна нулю (экспериментально доказано, что у фотона, как элементарной частицы, масса покоя, по крайней мере, меньше $10^{-21} m_0$, где m_0 — масса свободного электрона). Скорость фотона равна скорости света. Не существует системы отсчета, в которой фотон покоится.

3. Фотон является электрически нейтральной частицей и не имеет электрического заряда.

4. Спин фотона равен единице (в единицах \hbar).

5. Каждый фотон может находиться в некотором состоянии поляризации. Например, линейно поляризованное электромагнитное излучение можно рассматривать состоящим из фотонов, каждый из которых линейно поляризован в том же направлении.

Подчеркнем, что представления оптического излучения (света) в виде волн или корпускул не исключает одно другое. Это приближенное описание одних и тех же процессов. Квантово-волновой дуализм присущ природе в целом. Но в одних случаях в большей мере проявляются корпускулярные свойства вещества (микрочастиц) и удобнее пользоваться понятием частицы (или квазичастицы). В других случаях, наоборот, сильнее проявляются волновые свойства материи и целесообразнее пользоваться волновыми представлениями. Иногда волновые и корпускулярные свойства проявляются в одинаковой мере (например, оптическое излучение). В области длинных радиоволн (см. рис. В.1), где энергии квантов очень малы, корпускулярные свойства электромагнитного излучения практически не проявляются, так что в радиотехнике используют исключительно волновые представления. Наоборот, в области гамма-излучения, где длина волны менее 10^{-10} м, а энергия кванта более 10^4 эВ, в основном проявляются корпускулярные свойства электромагнитного излучения.

Общий для квантовой механики принцип суперпозиции состояний, согласно которому любое состояние можно рассматривать как результат наложения двух или многих состояний, применим и к фотонам. Напомним, что представление состояния системы (или частицы) в виде результата суперпозиции некоторого числа других состояний является чисто математической процедурой и формально она возможна всегда. Будет ли такое разложение полезно, зависит от конкретной физической задачи. В частности, для фотона эта процедура может быть аналогична разложению

волны на компоненты Фурье по плоским монохроматическим волнам.

Принцип суперпозиции позволяет объяснить такие явления, как интерференция, дифракция, поляризация отдельных фотонов, и тем самым примирить противоречия между корпускулярными и волновыми свойствами света. Каждое из состояний фотона связывается с некоторой волновой функцией, подобно тому как каждое из состояний электрона описывается электронной волновой функцией (волной де Бройля). В случае фотонов параметры волны де Бройля (частота, волновой вектор и поляризация) совпадают с соответствующими параметрами классической электромагнитной волны.

Таким образом, с точки зрения квантовых представлений оптическое излучение можно рассматривать как фотонный коллектив, состоящий из $N_{k\alpha}$ фотонов в $k\alpha$ -состояниях, т. е. как газ, состоящий из частиц с энергией $\hbar\omega$ и импульсом $\hbar k$. Принципиально важным моментом квантовой теории электромагнитного излучения (света) является то, что обмен энергией и импульсом между фотоном и атомной системой (электрон, атом, молекула и т. д.) происходит путем рождения одних и исчезновения других квантов света. Это отличает фотонный газ от газа, состоящего из «обычных» частиц — электронов, атомов, молекул, ионов и др. Указанное свойство наиболее полно отражается в законах сохранения энергии и импульса при рассмотрении взаимодействия света с какой-либо квантовой системой.

Пусть E и P — энергия и импульс системы до взаимодействия с квантом света, а E' и P' — после взаимодействия. Аналогично, $\hbar\omega$ и $\hbar k$ — энергия и импульс фотона до взаимодействия, а $\hbar\omega'$ и $\hbar k'$ — после взаимодействия. В результате взаимодействия, которое иногда называют столкновением, квант света ($\hbar\omega$, $\hbar k$) изменил (или не изменил) свою энергию и импульс до ($\hbar\omega'$, $\hbar k'$). Законы сохранения энергии и импульса при этом запишем в следующем виде:

$$\hbar\omega + E = \hbar\omega' + E', \quad (1.31)$$

$$\hbar k + p = \hbar k' + p'. \quad (1.32)$$

В этих соотношениях, как и в (1.29), (1.30), заложена основная связь между волновыми и корпускулярными свойствами света.

Если $N_{k\alpha}$ соответствует числу фотонов, находящихся в единице объема, то плотность энергии ρ светового поля можно представить в виде суммы энергий фотонов:

$$\rho = \sum_{\alpha=1}^2 \sum_k \hbar\omega N_{k\alpha}. \quad (1.33)$$

Аналогично полный импульс

$$P = \sum_{\alpha=1}^2 \sum_k \hbar k N_{k\alpha}. \quad (1.34)$$

Сравнение этих соотношений с аналогичными классическими выражениями (1.12) и (1.14) позволяет перейти от числа фотонов в $k\alpha$ -состоянии к амплитуде электромагнитной волны с поляризацией α и волновым вектором k .

При рассмотрении свойств большого числа фотонов принципиально важную роль начинает играть их поведение в коллективе себе подобных, т. е. их статистические свойства. Спин фотона равен единице, поэтому фотон относится к бозе-частицам (бозонам), т. е. описывается статистикой Бозе — Эйнштейна. В противоположность частицам с полуполым спином — ферми-частицам (фермионам), для которых справедлив принцип запрета Паули, бозе-частицы способны находиться в одинаковых квантовых состояниях в неограниченном числе. Для бозонов вероятность заполнения состояния оказывается тем выше, чем сильнее оно заселено: бозоны как бы обнаруживают тенденцию к взаимному объединению, не взаимодействуя при этом между собой. Именно это обстоятельство в конечном итоге дает возможность поставить в соответствие коллективу фотонов классическую электромагнитную волну.

Квантовомеханическое описание электромагнитного излучения является более полным, чем классическое. Но во многих практически важных случаях проще и удобнее пользоваться понятием электромагнитной волны. Такое классическое описание является частным случаем квантовомеханического. При классическом описании энергия светового поля ρ и его импульс P изменяются непрерывно [см. (1.12) и (1.14)], а при квантовомеханическом — скачками [см. (1.33) и (1.34)]. Квантованием поля, т. е. его дискретностью, можно пренебречь, если полная энергия ρ существенно больше энергии кванта $\hbar\omega$, т. е. если

$$N_{k\alpha} \gg 1. \quad (1.35)$$

Это условие называется *условием классичности*. Оно определяет условие перехода от квантовомеханического описания электромагнитного поля к классическому. Отметим, что для фермионов, в том числе для электронов, вследствие принципа запрета Паули всегда $N_{k\alpha} \leq 1$, поэтому с электронным коллективом никаких классических волн сопоставить нельзя.

§ 1.2. ЭНЕРГЕТИЧЕСКИЕ СОСТОЯНИЯ АТОМОВ И МОЛЕКУЛ

Важнейшим свойством атомов и молекул как квантовых систем, состоящих из связанных между собой микрочастиц, является то, что их внутренняя энергия (энергия, не определяемая

движением системы как целого) при определенных условиях может принимать лишь разрешенные дискретные значения E_j . Каждому из этих разрешенных значений энергии соответствует одно или несколько устойчивых состояний движения частиц в системе: энергия квантована. Квантование энергии является следствием волновых свойств микрочастиц, образующих атом или молекулу.

Атом водорода. Атом водорода в квантовой механике является простейшим «модельным» объектом, который служит основой для описания свойств более сложных атомов, молекул и твердых тел. Задача об атоме водорода сводится к задаче о движении двух взаимодействующих друг с другом частиц: ядра (протона) с массой M_p и зарядом $+e$ и электрона с массой m и зарядом $-e$. Взаимодействие осуществляется по закону Кулона

$$V(r) = -e^2/4\pi\epsilon_0 r. \quad (1.36)$$

Приведением к центру масс оператор Гамильтона (гамильтониан) разбивается на две независимые части. Первая описывает движение центра инерции с массой $M = M_p + m$, а вторая — внутреннее движение как движение частицы с массой m , в центросимметричном поле (1.36).

Поскольку $M_p \gg m$, то в атоме водорода полная масса $M \approx M_p$, приведенная масса $m_r \approx m$, а координаты центра масс R практически совпадают с координатами ядра.

Стационарное уравнение Шредингера

$$\hat{H}\Psi = E\Psi \quad (1.37)$$

для внутреннего движения атома водорода имеет вид

$$-\frac{\hbar^2}{2m} \nabla^2 \Psi(r) - \frac{e^2}{4\pi\epsilon_0 r} \Psi(r) = E \Psi(r). \quad (1.38)$$

Здесь ∇^2 — оператор Лапласа: $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, а коэффициент E является параметром, называемым энергией. Его значение не зависит от r и определяет состояние системы.

Напомним, что для решения уравнения (1.38) переходят от декартовых координат x, y, z к сферическим r, θ, φ и используют метод разделения переменных. Получаемые при этом решения $\Psi_{nlm_l}(r, \theta, \varphi)$ и значения E_{nlm_l} зависят от трех целочисленных индексов n, l, m_l , появляющихся в процессе разделения переменных. Это *квантовые числа*, с помощью которых удобно обозначать собственные состояния атома.

Для атома водорода понятия «собственное состояние атома» и «состояние электрона в атоме» эквивалентны, поскольку в нем имеется только один электрон. Поэтому квантовые числа в атоме

водорода однозначно характеризуют как состояние внутреннего движения атома, так и состояние электрона.

Три квантовых числа n, l и m_l , называемые *главным, орбитальным и магнитным квантовыми числами*, определяются тремя степенями свободы электрона. Напомним физический смысл этих квантовых чисел.

1. *Главное квантовое число n* определяет собственное значение энергии для связанного состояния электрона в атоме водорода (и только в атоме водорода!) и может принимать лишь целые положительные значения: $n=1, 2, 3, \dots$ (в противном случае уравнение (1.38) не имеет решений). Собственные значения внутренней энергии атома водорода для дискретной его части определяются соотношением

$$E_{nlm_l} = E_n = -\frac{me^4}{8(\epsilon_0 \hbar)^2} \frac{1}{n^2} = -\frac{R_\infty}{n^2}, \quad (1.39)$$

где

$$R_\infty = \frac{me^4}{8(\epsilon_0 \hbar)^2} = 2,18 \cdot 10^{-18} \text{ Дж} \approx 13,6 \text{ эВ} \quad (1.40)$$

— постоянная Ридберга.

В атоме водорода энергия не зависит от квантовых чисел l и m_l : энергетические состояния вырождены по этим квантовым числам, что является следствием центральной симметрии поля.

2. *Орбитальное квантовое число l* определяет в единицах \hbar величину орбитального момента количества движения электрона M_l , движущегося вокруг ядра. Квантовомеханическое рассмотрение показывает, что значение вектора M_l в единицах \hbar должно равняться не l , а $\sqrt{l(l+1)}$, т. е.

$$|M_l| = \sqrt{l(l+1)} \hbar \quad (1.41)$$

При заданном n квантовое число l может принимать целочисленные значения от $l=0$ до $(n-1)$ через единицу. По исторически сложившейся традиции состояния с различными значениями момента M_l принято обозначать буквами латинского алфавита со следующим соответствием:

$$\begin{array}{l} l=0, 1, 2, 3, 4, 5 \dots \\ s, p, d, f, g, h \dots \end{array} \quad (1.42)$$

3. *Магнитное квантовое число m_l* определяет в единицах \hbar проекцию орбитального момента количества движения на направление z , задаваемое внешним воздействием:

$$(M_l)_z = m_l \hbar \quad (1.43)$$

Квантовое число m_l может принимать значения от $+l$ до $-l$ через единицу, т. е. всего $(2l+1)$ значений: $m_l = 0, \pm 1, \pm 2, \dots, \pm l$.

В уравнении (1.38) не учитывается собственный момент количества движения электрона — спин. Его наличие приводит к появлению дополнительных степеней свободы. Поэтому при описании движения электрона необходимо помимо его координат рассматривать еще две величины, характеризующие собственный внутренний момент и его проекцию на некоторое избранное направление в пространстве. Эти величины могут принимать ограниченное число дискретных значений, которым соответствуют определенные квантовые числа. По аналогии с l и m_l эти числа обозначают s и m_s и называют *спиновыми квантовыми числами*.

4. Спиновое квантовое число s определяет в единицах \hbar спиновый момент количества движения электрона:

$$|M_s| = \sqrt{s(s+1)} \hbar \quad (1.44)$$

Для отдельного электрона s может принимать только одно значение: $s = 1/2$.

5. Спиновое квантовое число m_s определяет в единицах \hbar z -компоненту спинового момента количества движения электрона

$$(M_s)_z = m_s \hbar \quad (1.45)$$

Как и m_l , m_s может принимать дискретные значения от $+s$ до $-s$ через единицу, т. е. для отдельного электрона $m_s = \pm 1/2$.

Рассмотренные физические величины, определяемые квантовыми числами n , l , m_l , s и m_s , образуют полную систему одновременно измеряемых параметров. Задание этих пяти квантовых чисел полностью определяет состояние атома водорода. В такую систему невозможно включить какие-либо дополнительные независимые квантовые числа. При необходимости ее можно характеризовать и некоторыми другими физическими величинами с соответствующими квантовыми числами, но лишь ценой отказа от тех или иных квантовых чисел, перечисленных ранее. В частности, вместо рассматриваемых по отдельности независимых моментов количества движения — орбитального и спинового — можно ввести третий момент количества движения, равный векторной сумме первых двух моментов, а также проекцию этого суммарного момента на ось z . Этому моменту, называемому полным моментом количества движения, соответствуют два новых квантовых числа: j и m_j .

6. Квантовое число j . Это число определяет в единицах \hbar полный момент количества движения как векторную сумму орбитального и спинового моментов

$$\mathbf{M}_j = \mathbf{M}_l + \mathbf{M}_s \quad (1.46)$$

Как и ранее,

$$|M_j| = \sqrt{j(j+1)} \hbar \quad (1.47)$$

Исходя из классических представлений векторы \mathbf{M}_l и \mathbf{M}_s могут иметь любую взаимную ориентацию, а вектор \mathbf{M}_j — любое направление. На самом деле, согласно правилам квантовой механики сложения, значение полного момента $|M_j|$, как и других моментов, квантовано так, что квантовое число j в формуле (1.47) может принимать только целые или полуцелые значения и изменяться через единицу от состояния, соответствующего «параллельной» ориентации векторов \mathbf{M}_l и \mathbf{M}_s , при которой $j = l + s$, до состояния с антипараллельной ориентацией ($j = |l - s|$). Таким образом, при заданных l и s допустимые значения j образуют следующую систему:

$$j = l + s; l + s - 1; l + s - 2; \dots |l - s|. \quad (1.48)$$

В атоме водорода для единственного электрона $s = 1/2$, так что квантовое число j может принимать только два значения $j = l + 1/2$ и $j = l - 1/2$ при всех $l > 0$ и одно значение $j = 1/2$ при $l = 0$.

7. Квантовое число m_j . Оно определяет в единицах \hbar z -компоненту полного момента количества движения электрона

$$(M_j)_z = m_j \hbar \quad (1.49)$$

и по общим правилам может принимать $2j+1$ разрешенных значений от $+j$ до $-j$ через единицу.

Система квантовых чисел (n, l, m_l, s, m_s) удобна при рассмотрении отдельного электрона в атоме, в то время как для сложных атомов более удобна новая система (n, l, s, j, m_j) .

Многоэлектронные атомы. В многоэлектронных атомах состояние каждого из электронов зависит не только от его «собственных» параметров, но и от состояний всех других электронов. Отдельный электрон не является независимой системой в атомах с порядковыми номерами $z \geq 2$, поэтому рассматривать электроны по отдельности, приписывая им квантовые числа, строго говоря, не совсем правильно. Необходимо учитывать квантовые состояния атома в целом. Тем не менее в качестве первого приближения вводят понятие о состояниях каждого электрона в отдельности как о стационарных состояниях движения электрона в некотором эффективном поле, созданном ядром вместе со всеми остальными электронами. Такое поле называют *самосогласованным*, а приближение — *одноэлектронным**.

Самосогласованное поле обладает центральной симметрией, поэтому каждое состояние электрона можно характеризовать опре-

*Это приближение по своей сущности аналогично «одноэлектронному» приближению, являющемуся фундаментальным в физике твердого тела.

деленным орбитальным и спиновым моментами количества движения, а следовательно, квантовыми числами l , m_l , s и m_s , которые в сферически симметричном поле полностью сохраняют свой физический смысл. Как и в атоме водорода, энергетические состояния вырождены по квантовым числам m_l и m_s , но вырождение по l в многоэлектронных атомах снимается. По аналогии с атомом водорода состояния электрона с заданным l в многоэлектронных атомах нумеруются с помощью квантового числа n , являющегося аналогом главного квантового числа и принимающего значения $n = l + 1, l + 2, \dots$ по мере возрастания энергии. Но характер возрастания энергии в многоэлектронных атомах отличен от энергии атома водорода. Например, если в атоме водорода энергия не зависит от l , то в сложных атомах она возрастает по мере увеличения l .

Так как электроны обладают полуцелым спином, т. е. являются фермионами, то к ним применим принцип запрета Паули, согласно которому в замкнутой квантовой системе (в данном случае в атоме) не может быть двух электронов в одинаковых состояниях, т. е. двух электронов, характеризующихся одинаковым набором четырех квантовых чисел n , l , m_l и m_s (напомним, что пятое квантовое число для отдельного электрона фиксировано: $s = 1/2$).

Состояния электронов с заданными n и l образуют электронную оболочку, например $1s$ -оболочку, $2p$ -оболочку. Цифра указывает значение главного квантового числа, а следующая за ней буква — значение l в соответствии с символикой (1.42). Число электронов в оболочке обозначают для краткости показателем у символа. Например, символ $2p^4$ означает четыре электрона в состоянии $2p$. О распределении электронов в атоме по состояниям с различными l и n говорят как об электронной конфигурации.

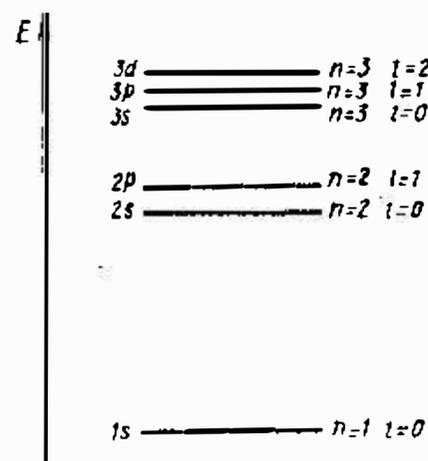


Рис. 1.3. Уровни энергий многоэлектронного атома (одноэлектронное приближение)

В одноэлектронном приближении уровни энергии остаются вырожденными по m_l и m_s , а вырождение по l снимается, как показано на рис. 1.3. В атоме может быть не более $2(2l + 1)$ электронов с заданными n и l . Максимальное количество электронов в s -оболочках — 2, в p -оболочках — 6, d -оболочках — 10 и в f -оболочках — 14.

Рассмотренные электронные конфигурации весьма приближенно отражают энергетическое состояние атома в целом. В отличие от электрона свободный атом представляет собой полную и независимую систему. Поэтому говорить о квантовых состояниях отдельного атома впо-

лно корректно. Энергетическое состояние атома также удобно описывать с помощью квантовых чисел, набор которых определяется характером взаимодействия, присущим рассматриваемому атому. Во многих случаях в атомах преобладающим является нерелятивистское межэлектронное взаимодействие. Такой совокупности электронов можно приписать в отдельности суммарный орбитальный и суммарный спиновый моменты. Соответствующие им квантовые числа обозначают аналогично квантовым числам для отдельного электрона, но не строчными, а прописными буквами. Суммарное орбитальное квантовое число L для двух электронов в состояниях l_1 и l_2

$$L = [l_1 + l_2], \quad (1.50)$$

а для i электронов

$$L = \sum_i [l_i].$$

Аналогично для суммарного спинового квантового числа можно записать

$$S = [s_1 + s_2], \quad (1.51)$$

$$S = \sum_i [s_i].$$

Квадратные скобки указывают, что в соотношениях (1.50)—(1.51) проводится квантовое суммирование. Это означает, что, например, L в (1.50) может принимать значения от $l_1 + l_2$ до $|l_1 - l_2|$ через единицу. Это же справедливо для S : в (1.51) $S = s_1 + s_2; s_1 + s_2 - 1; s_1 + s_2 - 2; \dots |s_1 - s_2|$.

Такая классификация атомных уровней основана на представлении, что орбитальные моменты электронов складываются в полный орбитальный момент L атома, а их спиновые моменты — в полный спин S^* . Соответствующий тип взаимодействия, при котором данная классификация оказывается справедливой, называют нормальной или LS -связью. Энергия атома определяется в основном его суммарными орбитальным и спиновым моментами.

Энергетическое состояние атома с заданными квантовыми числами L и S называют спектральным термом. Каждый терм вырожден соответственно различным возможным направлениям векторов L и S в пространстве. Кратность вырождения равна $(2L + 1)(2S + 1)$.

*Моменты количества движения атома обозначают теми же символами, что и определяющие их квантовые числа, но в векторной записи. Например, S — спиновый момент количества движения атома, а L — его орбитальный момент.

В действительности внутренняя энергия атома зависит не только от абсолютных значений векторов орбитального и спинного моментов L и S , но и от их взаимного расположения. Эта зависимость определяется налучнем релятивистского электромагнитного взаимодействия, зависящего от спинов электронов. В результате этого взаимодействия важный смысл приобретает полный момент $J = L + S$, который является сохраняющейся величиной, и ему может быть приписано квантовое число J . Аналогично (1.48) квантовое число J , определяющее в единицах \hbar полный момент количества движения атома J , может принимать значения от $(L + S)$ до $(L - S)$ через единицу:

$$J = [L + S]. \quad (1.52)$$

В большинстве случаев релятивистские эффекты относительно малы, и их можно учитывать как малое возмущение. Под действием этого возмущения происходит частичное снятие вырождения. Терм с заданными L и S расщепляется на ряд энергетических подуровней с различными значениями J . Об этом расщеплении говорят как о тонкой структуре, или мультиплетном расщеплении уровня. На основании (1.52) терм с данными L и S расщепляется на $(2S + 1)$ при $L > S$ или $(2L + 1)$ при $L < S$ различных подуровней. Каждый из этих подуровней остается вырожденным по направлениям вектора J , т. е. по квантовому числу M_J . Кратность вырождения равна $2J + 1$.

Для обозначения атомных уровней энергии в приближении LS -связи принята следующая символика:

$$\boxed{\begin{array}{c} 2S+1 \\ L_J \end{array}} \quad (1.53)$$

Состояния с различными значениями полного орбитального момента L обозначают прописными буквами латинского алфавита посередине аналогично (1.42):

$$\begin{array}{cccccccc} L=0 & 1 & 2 & 3 & 4 & 5 & 6 & 7, \dots \\ S & P & D & F & G & H & I & K, \dots \end{array} \quad (1.54)$$

Слева сверху у этого символа указывают число $2S + 1$, называемое мультиплетностью термина, справа внизу — квантовое число J . Например, символы 1S_0 , $^2P_{3/2}$ обозначают состояния с $L = 0$, $S = 0$, $J = 0$ и $L = 1$, $S = 1/2$, $J = 3/2$ соответственно. Если необходимо задать проекцию полного момента на выделенное направление z , то справа сверху в символе (1.53) пишут индекс M_J .

Например, уровень с $L = 2$, $S = 1$, $J = 2$ и $M_J = -1$ запишется как $^3D_{2,-1}$.

Таким образом, задание электронной конфигурации и квантовых чисел L , S , J , M_J полностью определяет состояние атома. Главное квантовое число N как аналог n в атоме теряет физический смысл.

Вычисление квантовых чисел L и S атома облегчается тем, что для полностью заполненных оболочек $L = S = 0$ и на основании (1.52) $J = 0$. Поэтому при нахождении L и S всего электронного облака необходимо учитывать только не полностью заполненные оболочки. Для полностью заполненных оболочек с конфигурациями s^2 , p^6 , d^{10} , ... всегда имеется только один терм 1S_0 .

Обратим внимание на следующее обстоятельство: термы, отвечающие электронным конфигурациям, из которых одна имеет столько электронов, сколько не хватает другой для заполнения оболочки, совпадают. Например, термы для p^1 -конфигурации будут аналогичны термам для p^5 -конфигурации. Поэтому отсутствие электрона в оболочке атома можно рассматривать как дырку, состояние которой определяется теми же квантовыми числами, что и состояние отсутствующего электрона.

В качестве примера рассмотрим возможные энергетические состояния атомов двух инертных газов — гелия и неона. Основное состояние атома гелия ($z = 2$) имеет электронную конфигурацию $1s^2$, которой соответствует единственный терм 1S_0 . Наиниже возбужденные состояния соответствуют случаю, когда один электрон из $1s$ -оболочки перейдет в $2s$ -оболочку (см. рис. 1.3), т. е. электронной конфигурации $1s^1 2s^1$. Для определения термов возбужденных состояний удобно рассматривать отдельно состояние одного возбужденного электрона и оставшихся электронов, которые образуют так называемый атомный остаток. Квантовые числа, характеризующие энергетическое состояние атомного остатка, снабжаются штрихом. В нашем случае остаток $1s$ характеризуется квантовыми числами $S' = 1/2$, $L' = 0$, а возбужденному электрону в $2s$ -состоянии можно приписать только значения $s = 1/2$ и $l = 0$. Отсюда для конфигурации $1s^1 2s^1$ согласно (1.51) возможны два значения спинового числа S атома гелия в возбужденном состоянии: $S = 0$ и $S = 1$ и единственное значение $L = 0$ [см. выражение (1.50)]. Каждому из этих состояний соответствуют единственные значения $J = 0$ и $J = 1$. Им отвечают два возможных термина: 1S_0 и 3S_1 . Чтобы указать, что эти термы

*Иногда в символе (1.53) справа сверху указывают четность состояния: нечетные состояния отмечают индексом «0», а четные состояния не отмечают. Четность многоэлектронного состояния совпадает с четностью суммарной волновой функции и в атоме определяется суммой орбитальных квантовых чисел для всех электронов, т. е. $\sum_{i=1}^Z l_i$.

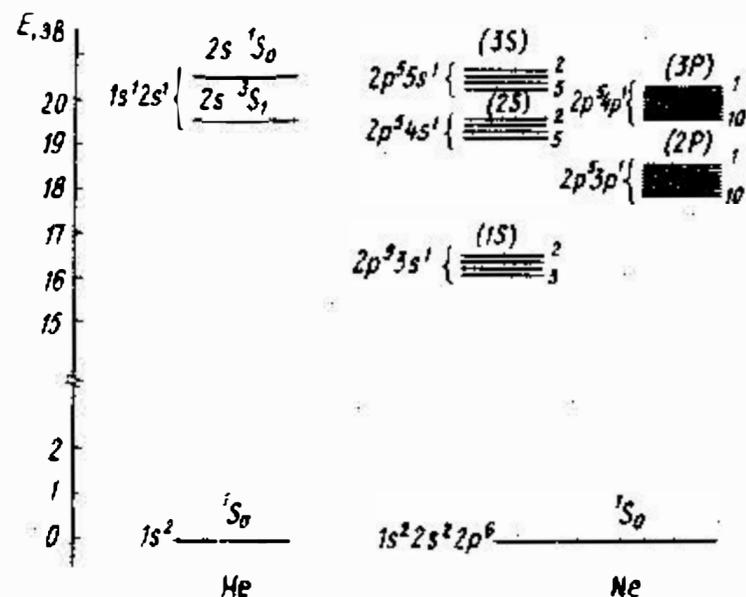


Рис. 1.4. Энергетические состояния атомов гелия и неона. За начало отсчета энергии принята энергия основного состояния

принадлежат возбужденному состоянию с электронной конфигурацией $1s^1 2s^1$, перед символом термина (1.51) указывают главное и орбитальное числа возбужденного электрона. Таким образом, переходу одного электрона с $1s$ -оболочки атома гелия в $2s$ -оболочку отвечают два разрешенных энергетических состояния (терма): $2s^1 S_0$ и $2s^1 S_1$. Состояние $2s^1 S_1$ по правилам, аналогичным правилам Хунда, должно быть расположено ниже состояния $2s^1 S_0$, как показано на рис. 1.4. Нетрудно убедиться, что аналогичные термы будут для электронных конфигураций $1s^1 3s^1$, $1s^1 4s^1$ и т. д.

Рассмотрим теперь энергетические состояния атома неона ($z=10$). Основному состоянию $1s^2 2s^2 2p^6$ здесь, как и в атоме гелия, соответствует единственный терм $1S_0$. Наимизим возбужденным состоянием отвечает электронная конфигурация $1s^2 2s^2 2p^5 3s^1$. Квантовые числа атомного остатка — $S'=1/2$, $L'=1$. Квантовые числа возбужденного электрона $s=1/2$, $l=0$. На основании (1.50) и (1.51) возможны следующие комбинации квантовых чисел для атома в целом:

$$\begin{aligned} S=1, L=1, J=2, \text{ терм } 3s^3 P_2, \\ J=1, \text{ терм } 3s^3 P_1, \\ J=0, \text{ терм } 3s^3 P_0, \\ S=0, L=1, J=1, \text{ терм } 3s^1 P_1. \end{aligned}$$

Им соответствуют четыре уровня энергии, как показано на рис. 1.4. Очевидно, каждой из электронных конфигураций $2p^5 4s^1$, $2p^5 5s^1$, $2p^5 6s^1$ и т. д. также будут соответствовать по четыре разрешенных состояния.

Нетрудно убедиться, что переходу электрона в возбужденные состояния атома неона, т. е. электронным конфигурациям $2p^5 np^1$ ($n=3, 4, 5, \dots$), будут соответствовать по 10 разрешенных уровней энергии.

Молекулы. Полную энергию молекулы, аналогично тому, как это было сделано для атома водорода, можно представить в виде кинетической энергии, связанной с движением центра масс, и внутренней энергии молекулы. В отличие от атома в молекуле появляются две дополнительные степени свободы, связанные с колебательным и вращательным движением. Так как массы ядер атомов, составляющих молекулу, велики по сравнению с массой электронов, а скорости движения ядер малы по сравнению со скоростями электронов, то внутреннюю энергию молекул можно считать состоящей из трех частей: 1) электронной энергии $E_{эл}$, связанной с движением электронов вокруг ядер; 2) колебательной энергии $E_{кол}$, связанной с колебаниями ядер относительно центра масс; 3) вращательной энергии $E_{вр}$, обусловленной вращением молекулы относительно центра масс

$$E = E_{эл} + E_{кол} + E_{вр}. \quad (1.55)$$

Все эти виды движения, а следовательно, и энергетические уровни квантованы. Порядок электронной энергии $E_{эл}$ в молекуле такой же, как в атоме, т. е. единицы и десятки электрон-вольт. Отношение $E_{кол}/E_{эл}$ имеет порядок $(m/M)^{1/2} \approx 10^{-2}$. В свою очередь $E_{вр}/E_{кол} \approx (m/M)^{1/2}$, т. е. расстояние между колебательными уровнями энергии имеет порядок десятых, а между вращательными — тысячных долей электрон-вольта.

Рассмотрим в качестве примера молекулу, состоящую из двух одинаковых атомов. Внутренняя энергия такой системы, очевидно, будет зависеть от расстояния R между ядрами атомов в молекуле. Для определения собственных значений энергии необходимо решить стационарное уравнение Шредингера, которое даже в простейшем случае двухатомной молекулы является очень сложным. Однако качественные результаты можно получить из общих соображений. Зависимость энергии от R имеет вид, показанный на рис. 1.5 для ос-

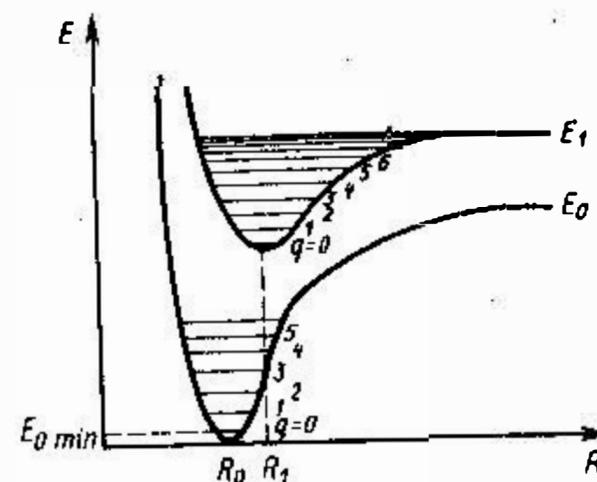


Рис. 1.5. Потенциальные кривые для основного (E_0) и возбужденного (E_1) состояний и уровни энергии двухатомной молекулы (R — расстояние между ядрами атомов; q — колебательное квантовое число)

новного E_0 и первого возбужденного E_1 состояний. Действительно, при очень больших R ($R \rightarrow \infty$) уровни E_0 и E_1 соответствуют изолированным атомам. При их сближении потенциальная энергия сначала плавно уменьшается, а при $R < R_0$ быстро возрастает. Это означает, что при больших R преобладают силы притяжения между атомами, а при малых R — силы отталкивания. В противном случае просто не происходило бы образования молекулы (напомним, что производная от энергии по R с отрицательным знаком есть сила, с которой атомы действуют друг на друга). Расстояние R_0 соответствует равновесному состоянию молекулы с энергией E_0 . В возбужденном состоянии E_1 точка равновесия R_1 сдвинута относительно R_0 . Как правило, возбужденное электронное состояние соответствует более разрыхленной молекуле. Поэтому среднее межатомное расстояние для молекулы в возбужденном состоянии больше, чем в основном. Энергия $D = E_{R \rightarrow \infty} - E_{R=R_0} = E_{\infty} - E_{\text{мин}}$ называется энергией диссоциации и равна энергии разрыва молекулы на атомы.

Вблизи положения равновесия, например вблизи R_0 , возвращающая «упругая» сила F пропорциональна отклонению $F = -\chi(R - R_0)$, а потенциальная энергия имеет вид параболы

$$U(R) = 0,5\chi(R - R_0)^2 + \text{const}, \quad (1.56)$$

как для гармонического осциллятора. Здесь χ — упругая постоянная, а константа определяет начало отсчета. Собственные значения энергии гармонического осциллятора, как известно, квантованы и могут принимать значения

$$E_{\text{хол}} = \hbar\Omega_0(q + 1/2), \quad (1.57)$$

где $q = 0, 1, 2, \dots$ — колебательное квантовое число, а

$$\Omega_0 = \sqrt{\chi/M}$$

— частота собственных колебаний осциллятора.

Величина

$$E_0 = 0,5\hbar\Omega_0 \quad (1.58)$$

называется энергией нулевых колебаний.

Приближение гармонического осциллятора дает серию эквидистантно расположенных уровней энергии (1.57) с расстоянием между ними $\hbar\Omega_0$. При больших отклонениях от положения равновесия, как видно из рис. 1.5, зависимость $U(R)$ отклоняется от закона (1.56), и колебательные уровни начинают сближаться, сходясь к границе диссоциации. В этом случае энергетический колебательный спектр двухатомной молекулы можно представить в виде

$$E_{\text{хол}} = \hbar\Omega_0(q + 1/2) - \hbar\Omega_0 x(q + 1/2)^2, \quad (1.59)$$

где

$$x = \frac{\hbar\Omega_0}{4D} \ll 1$$

— постоянная ангармоничности.

Выше границы диссоциации, как обычно, образуется непрерывный энергетический спектр, отвечающий свободному движению.

Вращение двухатомной молекулы происходит вокруг центра масс, находящегося на оси симметрии молекулы. Энергетический спектр определяют путем решения уравнения Шредингера для жесткого ротатора, предполагая, что расстояния между ядрами атомов остаются неизменными. Квантование вращательной энергии для двухатомной молекулы описывается выражением

$$E_{\text{вр}} = \frac{\hbar^2}{2I} K(K+1) = BK(K+1), \quad (1.60)$$

где $I = MR_0^2$ — момент инерции молекулы; B — вращательная постоянная молекулы, а $K = 0, 1, 2, 3, \dots$ — вращательное квантовое число, определяющее в единицах \hbar возможные значения вращательного момента количества движения молекулы

$$|M_{\text{вр}}| = \sqrt{K(K+1)}\hbar.$$

Расстояние между соседними вращательными уровнями энергии

$$E_{\text{вр}} = E_{K+1} - E_K = 2B(K+1) \quad (1.61)$$

увеличивается с увеличением K , т. е. с увеличением $E_{\text{вр}}$.

Помимо квантового числа K вращательное движение двухатомной молекулы характеризуется вторым квантовым числом M_K , которое аналогично m_l в (1.43) определяет в единицах \hbar возможные значения проекции вращательного момента на выделенное направление z

$$(M_{\text{вр}})_z = M_K \hbar,$$

где $M_K = K, K-1, \dots, -K$.

Электронные термы молекул по своей природе похожи на электронные термы изолированных атомов, но число их в молекуле значительно больше, чем в атоме. Каждый из атомов в молекуле приближенно можно рассматривать как находящийся в электрическом поле остальных атомов. Это внутримолекулярное электрическое поле вызывает расщепление «атомных» уровней в результате эффекта Штарка (на рис. 1.5 для простоты взяты невырожденные уровни E_0 и E_1) и значительно усложняет электронные спектры молекул.

Классификация электронных термов двухатомной молекулы основана на тех же принципах, что и для многоэлектронного атома. Однако в молекуле в отличие от атома электрическое поле нескольких ядер не обладает центральной симметрией, поэтому закон сохранения полного орбитального момента электронов L в молекулах не имеет места, и классификация термов по квантовому числу L лишена физического смысла. В то же время в двухатомных молекулах электрическое поле обладает аксиальной симметрией, и в этом случае должна сохраняться проекция орбитального момента на ось, проходящую через оба ядра. Соответственно электронные термы такой молекулы можно классифицировать по значению проекции орбитального момента на эту ось. Абсолютное значение проекции орбитального момента обозначают $|\Lambda|$, а соответствующее квантовое число Λ называют *орбитальным квантовым числом молекулы*. Оно может принимать значения 0, 1, 2, ... Термы с различными значениями Λ обозначают прописными буквами греческого алфавита со следующим соответствием:

$$\Lambda=0 \quad 1 \quad 2 \quad 3 \quad 4, \dots$$

$$\Sigma \quad \Pi \quad \Delta \quad \Phi \quad \Gamma, \dots \quad (1.62)$$

Каждое электронное состояние молекулы также характеризуется полным спином S всех электронов в молекуле. Подобно атому при отличном от нуля квантовом числе S наблюдается вырождение уровня кратности $2S+1$. Число $2S+1$, как и в атомах, называют мультиплетностью терма.

По аналогии с (1.47) и (1.52) вводят *внутреннее квантовое число молекулы*:

$$\Omega = [\Lambda + S].$$

Систематика электронных термов молекулы по числам Λ , S и Ω является обобщением систематики электронных термов атома по числам L , S и J . Аналогично (1.53) молекулярный терм обозначается символом

$$\boxed{2S+1 \Lambda_{\Omega}} \quad (1.63)$$

Для двухатомных симметричных молекул вместо Ω в правом нижнем углу указывают индекс g для четных и u для нечетных состояний.

Возможность перехода молекулы из одного колебательно-вращательного состояния в другое определяется симметрией этих состояний. Например, если при каком-либо определенном колебании молекулы ее электрический дипольный момент не изменяется, то в дипольном приближении соответствующий пе-

реход будет запрещен. Отсюда вытекают правила отбора для колебательных и вращательных переходов. Переходы между колебательными уровнями разрешены в дипольном приближении, если выполняется правило отбора

$$\Delta q = q' - q'' = \pm 1. \quad (1.64)$$

Для переходов между вращательными подуровнями правило отбора следующее:

$$\Delta K = K' - K'' = 0; \pm 1. \quad (1.65)$$

Переходы между состояниями с $\Delta K = -1; 0$ и $+1$ в спектрах поглощения и излучения молекул дают так называемые P -, Q - и R -ветви. Для линейных молекул существует дополнительный запрет на переход с $\Delta K = 0$, и Q -ветвь отсутствует.

Энергетический спектр многоатомной молекулы также может быть представлен в виде (1.55). В общем случае объемная молекула, состоящая из N атомов, обладает $3N-6$ колебательными степенями свободы. У линейной молекулы, ядра всех атомов в которой расположены на одной прямой, как у молекулы CO_2 , число колебательных степеней свободы равно $3N-5$. Каждой степени свободы отвечают свои колебательные уровни энергии вида (1.57) или (1.59) со своими частотами нормальных колебаний ν_{oi} .

Схема вращательных уровней энергии для линейных многоатомных молекул аналогична рассмотренной ранее для двухатомной молекулы [см. соотношение (1.60)]. Для нелинейных многоатомных молекул по мере снижения их симметрии схема вращательных уровней энергии усложняется.

Эффект Зеемана и эффект Штарка. Если поместить атом или молекулу во внешнее поле, то их уровни энергии изменятся. Расщепление и смещение энергетических уровней под действием магнитного поля называется *эффектом Зеемана*, а под действием электрического поля — *эффектом Штарка**.

Сущность этих эффектов с квантовомеханической точки зрения состоит в следующем. Если атом поместить в однородное магнитное или электрическое поле, то появится «выделенное» направление z : электроны будут двигаться не в сферически симметричном поле, а в поле, обладающем аксиальной симметрией. Это приведет к тому, что полный момент количества движения атома J уже не будет сохраняться, а будет сохраняться лишь проекция M_z полного момента на выделенное внешним полем

*Впервые эти эффекты наблюдались как расщепление спектральных линий под действием магнитного (П. Зееман, 1896) и электрического (М. Штарк, 1913) полей. В настоящее время под эффектами Зеемана и Штарка понимают не только расщепление спектральных линий в магнитном или электрическом поле, но также смещение и расщепление на несколько подуровней самых уровней энергии.

направление. Поэтому состояния с разными значениями M , будут обладать различными энергиями, т. е. под действием внешнего магнитного или электрического поля произойдет снятие вырождения уровня — его расщепление (напомним, что энергетические состояния в атоме вырождены по квантовому числу M_j).

§ 1.3. КВАНТОВЫЕ ПЕРЕХОДЫ

Квантовым переходом называют переход системы из одного квантового состояния в другое. Согласно общим принципам квантовой механики невозможно провести наблюдения за системой в процессе такого перехода, следовательно, его необходимо рассматривать как происходящий скачкообразно. Основной задачей теории квантовых переходов является вычисление вероятности перехода под воздействием внешних полей или из-за внутренних причин. В большинстве случаев возмущение, под воздействием которого совершается квантовый переход, можно считать малым по сравнению с внутренней энергией системы. Тогда решать задачу можно с помощью хорошо разработанных методов теории возмущений.

Требуется рассмотреть изменение состояния системы во времени. Соответствующую волновую функцию, зависящую от времени, обозначим $\varphi(r, t)$, а не зависящую от времени — $\Psi(r)$, как и ранее. Волновая функция $\varphi(r, t)$ описывает состояние квантовой системы и зависит от совокупности координат частиц r и от времени t . Эта волновая функция должна удовлетворять уравнению Шредингера

$$i\hbar \frac{\partial \varphi^{(0)}}{\partial t} = \hat{H}^{(0)} \varphi^{(0)}, \quad (1.66)$$

где $\hat{H}^{(0)}$ — оператор энергии (гамильтониан) системы; верхние индексы (0) означают, что на систему пока что не действует возмущение и она находится в стационарном состоянии. В этом случае можно найти уровни энергии E_n , а решением уравнения (1.66) являются функции вида

$$\varphi^{(0)}(r, t) = e^{-\frac{i}{\hbar} E_n t} \Psi_n^{(0)}(r), \quad (1.67)$$

где $\Psi_n^{(0)}(r)$ — собственные функции гамильтониана, т. е. они удовлетворяют стационарному уравнению Шредингера

$$\hat{H}^{(0)} \Psi_n^{(0)} = E_n \Psi_n^{(0)}. \quad (1.68)$$

Индекс n здесь указывает квантовое число или набор квантовых чисел.

Состояние системы, описываемое волновой функцией $\varphi^{(0)}(r, t)$ (1.67), называется стационарным, хотя, как видно из формулы, волновая функция φ зависит от времени. Но непосредственный физический смысл имеет не сама функция φ или Ψ , а квадрат ее модуля, который для данного состояния не зависит от времени

$$|\varphi^{(0)}(r, t)|^2 = \varphi^{(0)} \varphi^{(0)*} = |\Psi_n^{(0)}(r)|^2. \quad (1.69)$$

Пусть теперь на систему действует внешнее возмущение V . Тогда к невозмущенному гамильтониану $\hat{H}^{(0)}$ добавится энергия взаимодействия с внешним полем, и уравнение Шредингера при наличии возмущения запишется в виде

$$i\hbar \frac{\partial \varphi(r, t)}{\partial t} = (\hat{H}^{(0)} + V) \varphi(r, t). \quad (1.70)$$

Функции $\varphi^{(0)}$ (1.67) уже не будут удовлетворять этому уравнению, решение которого удобно искать в виде

$$\varphi(r, t) = \sum_n C_n(t) \varphi_n^{(0)}(r, t) = \sum_n C_n(t) e^{-\frac{i}{\hbar} E_n t} \Psi_n^{(0)}(r), \quad (1.71)$$

т. е. новое состояние можно представить как суперпозицию невозмущенных состояний. Коэффициенты $C_n(t)$ представляют собой амплитуды состояний $\Psi_n^{(0)}$. Квадрат модуля амплитуды состояния $|C_n(t)|^2$ дает вероятность того, что система находится в состоянии $\Psi_n^{(0)}$.

На коэффициенты C_n накладывается условие нормировки

$$\sum_n |C_n|^2 = 1, \quad (1.72)$$

означающее, что система с вероятностью, равной единице, находится в каком-либо из разрешенных состояний n .

Будем считать, что в момент времени $t=0$ только один из коэффициентов C_n при $n=a$ не равен нулю, т. е. $C_a \neq 0$. Остальные коэффициенты C_n в формуле (1.71), для которых $n \neq a$, равны нулю. Это означает, что при $t=0$ система находится в «чистом» состоянии Ψ_a . Под воздействием возмущения она может перейти из состояния Ψ_a в другие состояния, т. е. амплитуда C_a будет уменьшаться, а амплитуды $C_{n \neq a}$ возрастать, становясь отличными от нуля. Тогда вероятность того, что к моменту времени t система окажется в состоянии b , будет равна квадрату модуля амплитуды состояния b

$$W_{ba}(t) = |C_b(t)|^2 \quad (1.73)$$

при условиях

$$\begin{aligned} |C_a(0)|^2 &= 1, \\ |C_{n \neq a}(0)| &= 0. \end{aligned} \quad (1.74)$$

Таким образом, определение вероятности перехода W_{ba} из состояния a в состояние b сводится к вычислению амплитуды $C_b(t)$ с начальными условиями (1.74).

Для нахождения C_b обычно пользуются теорией возмущений. Подставляя (1.71) в (1.70), с учетом (1.67) получаем

$$\sum_n \varphi_n^{(0)} \frac{\partial C_n}{\partial t} = -\frac{i}{\hbar} \sum_n C_n V \varphi_n^{(0)}. \quad (1.75)$$

Умножая обе стороны этого равенства на $\varphi_m^{(0)*}$ и интегрируя, т. е. проводя обычные в теории возмущений преобразования, получаем уравнение для коэффициентов $C_n(t)$:

$$\frac{\partial C_m}{\partial t} = -\frac{i}{\hbar} \sum_n V_{mn} e^{i\omega_{mn}t} C_n, \quad (1.76)$$

где

$$V_{mn} = \int \Psi_m^{(0)*} V \Psi_n^{(0)} d\tau \quad (1.77)$$

— матричный элемент возмущения между состояниями m и n . Частота $\omega_{mn} = (E_m - E_n)/\hbar$ в (1.76) есть боровская частота перехода $m \rightarrow n$. Физический смысл матричных элементов поясним в § 1.5, сопоставляя классический и квантовый осцилляторы.

Систему уравнений (1.76) удобно решать методом последовательных приближений. Подставляя в правую часть исходные значения коэффициентов C_n (1.74), находим уравнение для первого приближения

$$\frac{\partial C_m^{(1)}}{\partial t} = -\frac{i}{\hbar} \sum_n V_{mn} e^{i\omega_{mn}t} C_n^{(0)}, \quad (1.78)$$

откуда

$$C_m^{(1)}(t) = -\frac{i}{\hbar} \int_0^t V_{mn}(\tau) e^{i\omega_{mn}\tau} d\tau. \quad (1.79)$$

Подставляя это первое приближение $C_m^{(1)}(t)$ в правую часть уравнения (1.76), находим второе приближение и т. д. В результате при достаточно больших t (по сравнению с периодом колебаний) получаем следующее выражение для коэффициента $C_b(t)$:

$$C_b(t) = \frac{M_{ba}}{\hbar\omega_{ba}} (1 - e^{i\omega_{ba}t}), \quad (1.80)$$

где

$$M_{ba} = V_{ba} + \sum_{n_1} \frac{V_{bn_1} V_{n_1a}}{(E_a - E_{n_1})} + \sum_{n_1, n_2} \frac{V_{bn_1} V_{n_1n_2} V_{n_2a}}{(E_a - E_{n_1})(E_a - E_{n_2})} + \dots \quad (1.81)$$

— матричный элемент перехода. Первому слагаемому в (1.81) соответствует первое приближение теории возмущений, второму — второе и т. д. Структуру матричного элемента M_{ba} рассмотрим несколько позже. Сейчас лишь отметим, что в большинстве случаев достаточно ограничиться первым или вторым приближением.

Таким образом, согласно (1.73) и (1.80) вероятность перехода

$$W_{ba} = \frac{|M_{ba}|^2 \sin^2(\omega_{ba}t/2)}{\hbar^2 (\omega_{ba}/2)^2}. \quad (1.82)$$

Рассмотрим функцию $f(x) = \sin^2 xt/x^2$, где $x = \omega_{ba}/2$. График этой функции изображен на рис. 1.6. Амплитуда в максимуме при $x=0$ пропорциональна t^2 , а ширина пика — $1/t$. При больших t функцию $f(x)$ можно заменить дельта-функцией $\delta(x)$:

$$f(x) = A\delta(E_b - E_a). \quad (1.83)$$

Постоянный множитель A определяется из условия нормировки

$$\int_0^\infty f\left(\frac{\omega_{ba}}{2}\right) dE_b = \int_0^\infty A\delta(E_b - E_a) dE_b. \quad (1.84)$$

Правый интеграл равен A из определения δ -функции, а левый — $2\pi\hbar t$. Следовательно,

$$A = 2\pi\hbar t \quad (1.85)$$

и вероятность перехода

$$W_{ba} = \frac{2\pi}{\hbar} |M_{ba}|^2 \delta(E_b - E_a) t. \quad (1.86)$$

Отсюда видно, что полная вероятность перехода пропорциональна времени. Введем вероятность перехода в единицу времени:

$$w_{ba} = \frac{2\pi}{\hbar} |M_{ba}|^2 \delta(E_b - E_a). \quad (1.87)$$

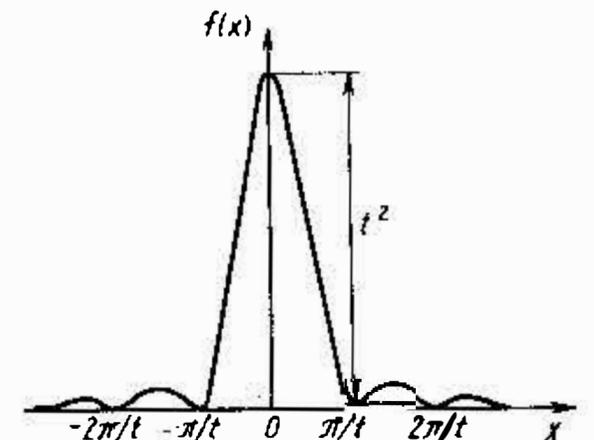


Рис. 1.6. График функции $f(x) = (\sin^2 xt)/x^2$

Это важное соотношение связывает вероятность квантового перехода из состояния a в состояние b под воздействием возмущения V с соответствующим матричным элементом (1.81). Такая вероятность отлична от нуля лишь в случае $E_b = E_a$, т. е. когда полная энергия системы в начальном состоянии E_a равна полной энергии системы в конечном состоянии E_b .

Рассмотрим теперь матричный элемент (1.81). Он имеет простую структуру и его удобно представить графически в виде набора следующих диаграмм, имеющих реальный физический смысл и отражающих развитие процесса во времени. Точкой на графике будем условно изображать некоторое взаимодействие, а непрерывной линией — состояние системы. Первому слагаемому в (1.81) сопоставим график

$$\begin{array}{c} a \quad b \\ \text{---} \cdot \text{---} \end{array} = V_{ba} \quad (1.88)$$

Он означает, что в результате взаимодействия с некоторым возмущением V система переходит непосредственно из одного разрешенного состояния a в другое разрешенное состояние b . Второе слагаемое изобразим в виде

$$\begin{array}{c} a \quad n_1 \quad b \\ \text{---} \cdot \text{---} \cdot \text{---} \end{array} = \sum_{n_1} \frac{V_{bn_1} V_{n_1a}}{(E_a - E_{n_1})} \quad (1.89)$$

Этот график описывает следующий процесс: под воздействием возмущения V система переходит из начального состояния a в конечное состояние b через промежуточное состояние n_1 . В отличие от реальных состояний a и b , в которых система может в принципе находиться сколь угодно долго, состояние n_1 является неустойчивым и называется *виртуальным состоянием*.

Третьему слагаемому сопоставим график

$$\begin{array}{c} a \quad n_1 \quad n_2 \quad b \\ \text{---} \cdot \text{---} \cdot \text{---} \end{array} = \sum_{n_1, n_2} \frac{V_{bn_2} V_{n_2n_1} V_{n_1a}}{(E_a - E_{n_1})(E_a - E_{n_2})} \quad (1.90)$$

Он указывает, что система под воздействием возмущения переходит из состояния a в состояние b через два последовательных промежуточных (виртуальных) состояния n_1 и n_2 . С помощью таких диаграмм, пользуясь определенными правилами, можно не только передать число и характер взаимодействий, но и написать соответствующие данному переходу аналитические выражения. Как нетрудно понять из графиков (1.88) — (1.90), этими правилами являются следующие:

1. Каждой точке соответствует матричный элемент вида

(1.77), например, переходу $\begin{array}{c} n_1 \quad n_2 \\ \text{---} \cdot \text{---} \end{array}$ — матричный элемент возмущения между состояниями n_1 и n_2 , т. е.

$$V_{n_2 n_1} = \int \Psi_{n_2}^* V \Psi_{n_1} dr.$$

2. Каждому участку между точками, например $\begin{array}{c} n_1 \\ \text{---} \cdot \end{array}$, соответствует энергетический множитель $1/(E_a - E_{n_1})$, где E_{n_1} — энергия состояния, соответствующая данному участку, а E_a — начальная энергия системы.

3. По всем промежуточным состояниям проводится суммирование.

В качестве примера на рис. 1.7 изображен график второго порядка, внизу под каждым элементом которого записано соответствующее ему аналитическое выражение.

Таким образом, для вычисления вероятности квантового перехода необходимо знать энергетический спектр невозмущенной системы и матричные элементы возмущения. Непосредственный переход между состояниями a и b , т. е. переход $\begin{array}{c} a \quad b \\ \text{---} \cdot \text{---} \end{array}$ возможен, если матричный элемент возмущения V_{ba} между этими состояниями не равен нулю и не равна нулю δ -функция в (1.87), т. е. если выполняется закон сохранения энергии. В этом случае, т. е. при $V_{ba} \neq 0$, остальные слагаемые в (1.81) для матричного элемента перехода можно не учитывать, так как они обычно на несколько порядков меньше V_{ba} .

Если по каким-либо причинам (например, вследствие особенностей симметрии волновых функций) матричный элемент $V_{ba} = 0$, то это не означает, что такой переход вообще невозможен. Он будет запрещен в первом порядке теории возмущений, но может происходить во втором приближении. При этом следует учитывать второе слагаемое в (1.81), т. е. переходы через промежуточные, виртуальные состояния. Как видно из (1.87), для осуществления перехода из состояния a в состояние b через виртуальное состояние n_1 (для перехода $\begin{array}{c} a \quad n_1 \quad b \\ \text{---} \cdot \text{---} \cdot \text{---} \end{array}$) закон сохранения энергии должен соблюдаться только для всего процесса в целом, т. е. в данном случае должно выполняться условие $E_a = E_b$. Для перехода в виртуальное состояние (виртуального перехода) требуется только отличие от нуля соответствующего матричного элемента. В виртуальных состояниях система может существовать лишь очень короткое время δt , определяемое соотношением неопределенности

$$\delta t \delta E \approx \hbar, \quad (1.91)$$

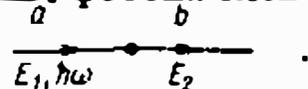
где δE — неопределенность энергии.

$$\begin{array}{c} a \quad n_1 \quad b \\ \text{---} \cdot \text{---} \cdot \text{---} \\ \sum_{n_1} V_{n_1a} \frac{1}{E_{n_1} - E_a} V_{bn_1} \end{array}$$

Рис. 1.7. Диаграмма взаимодействия, изображающая квантовый переход из состояния «а» в состояние «б» через виртуальное состояние «n₁» и соответствующее ей аналитическое выражение

Если неопределенность времени δt мала, то неопределенность энергии может быть большой. Поэтому при переходах в виртуальные состояния закон сохранения энергии как бы не соблюдается. Вообще для виртуальных переходов не выполняются обычные соотношения между энергией, импульсом и массой. Возможность существования таких переходов является чисто квантово-механическим эффектом и обусловлена соотношением неопределенностей Гейзенберга. Виртуальный переход сам по себе не имеет смысла, но используется как условная ступенька к реальному переходу.

Пусть возмущающим фактором является электромагнитное излучение в виде коллектива фотонов. Тогда полная энергия системы может быть представлена состоящей из электронной энергии (внутренней энергии атомов) и энергии электромагнитного поля (энергии фотонов). Процессу поглощения фотона электроном (атомом) можно сопоставить график



В начальном состоянии E_a система состояла из атома с внутренней энергией E_1 и фотона с энергией $\hbar\omega$, а в конечном состоянии E_b остался атом с внутренней энергией E_2 . Но такое изображение не наглядно. При графическом изображении процессов взаимодействия с фотонами гораздо удобнее и нагляднее представлять отдельно электронную и фотонную ветви. Последнюю обычно изображают волнистой линией. Тогда диаграмма взаимодействия, отражающая процесс поглощения фотона, будет иметь вид



$$(1.92)$$

Она означает, что в результате взаимодействия с фотоном (в результате его поглощения) атом перешел из состояния E_1 в состояние E_2 . Начальная энергия системы $E_a = E_1 + \hbar\omega$, а конечная энергия $E_b = E_2$. Согласно общим правилам, изложенным ранее, такой переход будет возможен, если $E_a = E_b$, т. е. если выполняется правило частот Бора $\hbar\omega = E_2 - E_1$ и матричный элемент перехода, определяемый электрон-фотонным взаимодействием ($V_{E_2, E_1}^{\text{фот}}$), отличен от нуля.

Аналогичные результаты получаются, если электромагнитное поле представить в виде электромагнитной волны. Как указывалось, плоскую электромагнитную волну можно представить в виде совокупности линейно поляризованных монохроматических волн, чему соответствует разложение в интеграл Фурье (1.26). Тогда окажется, что вероятность перехода атома из состояния E_1 в состояние E_2 будет отлична от нуля только в том случае, когда в спектре возмущения содержится частота $\omega_{12} = (E_2 - E_1)/\hbar$. Это означает, что переход носит резонансный характер, а квантовая система ведет себя аналогично набору осцилляторов с собствен-

ными частотами, равными боровским частотам ω_{mn} . При воздействии внешнего переменного возмущения возбуждаются только те осцилляторы, частоты которых совпадают с частотами, присутствующими во внешнем воздействии.

§ 1.4. СПОНТАННОЕ И ВЫНУЖДЕННОЕ ИЗЛУЧЕНИЕ. КОЭФФИЦИЕНТЫ ЭЙНШТЕЙНА

Квантовая механика позволяет описывать движение электронов, протонов, позитронов и других частиц, но не их рождение или исчезновение, т. е. она применима для описания систем с неизменным числом частиц. Нас же интересуют как раз процессы испускания (рождения) и поглощения (исчезновения) фотонов. Эти процессы детально обсуждаются в квантовой теории поля, которая является обобщением квантовой механики с учетом взаимного превращения частиц. Свободное электромагнитное поле здесь учитывается путем введения операторов рождения и исчезновения фотонов. Отметим, что появляться и исчезать могут не только фотоны. Взаимная превращаемость частиц является общим свойством микромира. Например, пара электрон — позитрон может аннигилировать, породив фотон. И наоборот, фотон с достаточно большой энергией может породить пару электрон — позитрон. Аналог этого процесса существует в твердых телах, когда в результате поглощения фотона «рождается» пара свободный электрон — дырка.

В подавляющем большинстве случаев процессы испускания и поглощения фотонов можно рассматривать менее строго, опираясь на полуперенормированную теорию излучения, развитую Эйнштейном еще в 1917 г., и на теорию квантовых переходов. Такое представление основывается на законах сохранения энергии (1.31) и импульса (1.32) при взаимодействии квантовых систем с электромагнитным излучением.

Из § 1.3 следует, что непосредственный (разрешенный в первом порядке теории возмущений) квантовый переход из одного состояния в другое возможен, если в электромагнитном излучении присутствуют компоненты с частотами, удовлетворяющими правилу частот Бора. Возьмем для определенности два энергетических состояния E_m и E_n , в которых может находиться квантовая система (например, атом). Это могут быть любые два уровня из неограниченного набора уровней, свойственных данной системе. Пусть $E_m > E_n$. Тогда из закона сохранения энергии (1.31) следуют три возможных процесса взаимодействия.

$$1. E = E_m; E' = E_n < E_m; \hbar\omega' = \hbar\omega_{mn} = E_m - E_n.$$

Это означает, что в результате взаимодействия система перешла из более высокого энергетического состояния в более низкое

с испусканием кванта света. Такой процесс называется *спонтанным испусканием фотона*. Фотоны с энергией $\hbar\omega \neq E_m - E_n$ во взаимодействии не участвуют, поэтому можно положить $\hbar\omega = 0$.

$$2. E = E_n; E' = E_m > E_n; \hbar\omega = \hbar\omega_{mn} = E_m - E_n; \hbar\omega' = 0.$$

В результате взаимодействия система перешла из более *низкого* энергетического состояния в более *высокое* с поглощением кванта света. Данный процесс называется *резонансным поглощением* (или просто *поглощением*) фотона.

$$3. E = E_m; E' = E_n < E_m; \hbar\omega = \hbar\omega' = \hbar\omega_{mn} = E_m - E_n.$$

Из формулы (1.31) следует, что в этом случае фотон $\hbar\omega$ не исчезает, а появляется дополнительный фотон $\hbar\omega' = \hbar\omega$, т. е. в правой части соотношения (1.31) будет стоять $2\hbar\omega$. Это означает, что в результате взаимодействия система перешла из более *высокого* энергетического состояния в более *низкое* с испусканием кванта света. В качестве вынуждающего фактора, определяющего подобный переход, здесь выступает фотон той же частоты, что и испускаемый. Поэтому такой процесс называется *вынужденным (индуцированным) испусканием фотона*. Важно, что после процесса взаимодействия оба эти фотона становятся неразличимы, т. е. тождественны.

На перечисленных трех элементарных процессах взаимодействия электромагнитного излучения с квантовыми системами (схематически они изображены на рис. 1.8, а — в) основана работа большинства приборов квантовой и оптической электроники.

Вынуждающим фактором, приводящим к процессам поглощения и вынужденного испускания, является фотон с частотой $\omega_{mn} = (E_m - E_n)/\hbar$, поэтому оба эти квантовых перехода называются *вынужденными* или *индуцированными*. Вынуждающий фактор, приводящий к спонтанному испусканию фотона, в явном

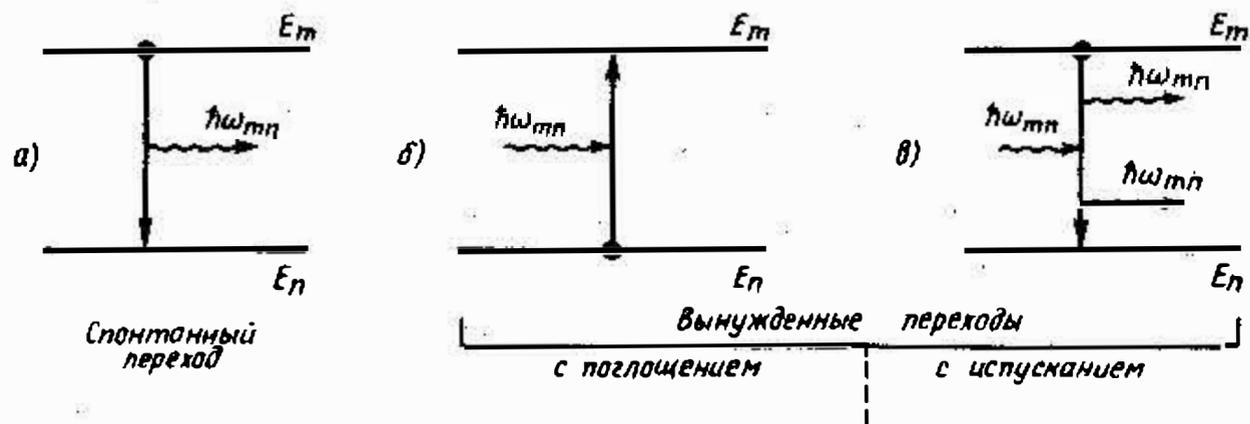


Рис. 1.8. Квантовые переходы при взаимодействии с фотоном: а — спонтанный переход с испусканием фотона; б — вынужденный переход с поглощением фотона; в — вынужденный (индуцированный) переход с испусканием фотона

виде не присутствует, и такой квантовый переход называется *спонтанным*.

Спонтанное испускание возникает вследствие спонтанного, т. е. самопроизвольного квантового перехода атома (или другой квантовой системы) из возбужденного состояния в основное или другое возбужденное состояние с меньшей энергией. На первый взгляд такой процесс кажется очевидным, поскольку любая система стремится к минимуму потенциальной энергии. На самом деле все разрешенные состояния, в которых может находиться атом и которые получаются при решении уравнения Шредингера, являются стационарными, т. е. устойчивыми, стабильными. Абсолютно самопроизвольного перехода, т. е. происходящего без каких-либо взаимодействий, в природе быть не может. В квантовой электродинамике доказывается, что спонтанные переходы возникают вследствие взаимодействия электронов в атоме с так называемым фотонным вакуумом или нулевыми колебаниями электромагнитного поля [ср. с E_0 в соотношении (1.57)]. Управлять указанным процессом невозможно, и в этом смысле спонтанный переход происходит самопроизвольно. Предсказать точно момент спонтанного перехода принципиально невозможно. Можно говорить лишь о вероятности перехода. Очевидно, вероятность спонтанного перехода с уровня E_m на уровень E_n за время dt с испусканием кванта света можно выразить соотношением

$$dW_{mn}^{sp} = A_{mn} dt, \quad (1.93)$$

где A_{mn} — коэффициент Эйнштейна для спонтанных переходов. Его значение не зависит от внешних воздействий и определяется только свойствами данной квантовой системы.

Значение коэффициента Эйнштейна A_{mn} для переходов в оптическом диапазоне может изменяться от $\sim 10^8 \text{ с}^{-1}$ для разрешенных (наиболее вероятных) переходов до $\sim 1 \text{ с}^{-1}$ и более для запрещенных переходов.

— Электромагнитное излучение, обусловленное спонтанными переходами в коллективе атомов, называется *спонтанным излучением*. Случайность спонтанных переходов приводит к тому, что различные атомы излучают независимо и несинхронно. Поэтому спонтанное излучение ненаправленно, некогерентно, неполяризовано и немонохроматично. Такое естественное излучение в оптическом диапазоне испускают все известные «классические» источники света (лампы накаливания, газоразрядные лампы, люминесцентные лампы и т. д.).

Вынужденные (индуцированные) квантовые переходы происходят под воздействием внешнего возмущения, которым является электромагнитное излучение. Вероятность таких переходов пропорциональна интенсивности возмущения. Поэтому вероятность поглощения фотона с частотой $\omega_{mn} = (E_m - E_n)/\hbar$

(см. рис. 1.8, б), обусловленная переходом $n \rightarrow m$, за интервал времени dt

$$dW_{mn}^{\text{погл}} = B_{mn} \rho(\omega) dt, \quad (1.94)$$

где B_{mn} — коэффициент Эйнштейна для вынужденных (индуцированных) переходов с поглощением; $\rho(\omega)$ — спектральная плотность излучения [см. (1.28)].

Если атом находится в возбужденном состоянии E_m и на него действует электромагнитное излучение с частотой $\omega_{mn} = (E_m - E_n)/\hbar$, то это излучение способствует переходу атома в низшее состояние (см. рис. 1.8, в). В результате такого вынужденного перехода атом отдает энергию электромагнитной волне, амплитуда которой (или число фотонов) увеличивается. Этот процесс представляет собой вынужденное (индуцированное) испускание. Вероятность индуцированного испускания фотона за интервал времени dt , как и вероятность поглощения, пропорциональна интенсивности возмущающего фактора, т. е. $\rho(\omega)$

$$dW_{mn}^{\text{исп}} = B_{mn} \rho(\omega) dt. \quad (1.95)$$

Здесь B_{mn} — коэффициент Эйнштейна для вынужденных (индуцированных) переходов с испусканием.

Принципиально важным является то, что фотон, появившийся в результате вынужденного испускания, неразличим с фотоном, индуцировавшим этот процесс, т. е. оба фотона имеют одну и ту же частоту, фазу, поляризацию и распространяются в одном направлении. Поэтому вынужденное излучение, т. е. электромагнитное излучение, возникающее в результате процессов вынужденного испускания в коллективе атомов, является когерентным. Такое излучение в оптическом диапазоне испускают только оптические квантовые генераторы — лазеры.

Определим связь между коэффициентами A_{mn} , B_{mn} и B_{nm} . Для этого, следуя Эйнштейну, рассмотрим совокупность атомов (молекул), находящихся в термодинамическом равновесии со стенками окружающего объема при температуре T . Пусть на уровне E_m находится N_m частиц, а на уровне E_n — N_n частиц. Тогда число поглощенных квантов света за интервал времени dt

$$N_n dW_{nm}^{\text{погл}} = N_n B_{nm} \rho(\omega) dt.$$

Число квантов света, испущенных в результате спонтанных переходов

$$N_m dW_{mn}^{\text{сп}} = N_m A_{mn} dt,$$

а число квантов света, испущенных в результате индуцированного испускания

$$N_m dW_{mn}^{\text{исп}} = N_m B_{mn} \rho(\omega) dt.$$

Условие термодинамического равновесия означает, что суммарное число квантов света, испущенных системой, равно числу поглощенных квантов света

$$N_m (dW_{mn}^{\text{сп}} + dW_{mn}^{\text{исп}}) = N_n dW_{nm}^{\text{погл}},$$

или

$$N_m (A_{mn} + B_{mn} \rho(\omega)) = N_n B_{nm} \rho(\omega). \quad (1.96)$$

В условиях равновесия распределение атомов (молекул) по энергетическим уровням подчиняется распределению Больцмана:

$$N_m = \frac{N}{\Sigma} q_m \exp(-E_m/kT), \quad (1.97)$$

$$N_n = \frac{N}{\Sigma} q_n \exp(-E_n/kT),$$

где N — полное число частиц в системе; q_m и q_n — статический вес уровней E_m и E_n .

Для невырожденных уровней $q_m = q_n = 1$. Знаком Σ обозначена статистическая сумма

$$\Sigma = \sum q_i \exp(-E_i/kT),$$

где суммирование проводится по всем энергетическим состояниям i . Из (1.97) следует, что

$$\frac{N_m/q_m}{N_n/q_n} = e^{-\frac{E_m - E_n}{kT}}. \quad (1.98)$$

Величина, равная отношению числа частиц в единице объема на данном энергетическом уровне к его статистическому весу, называется населенностью энергетического уровня или просто населенностью. Для невырожденных уровней населенность совпадает с числом частиц, находящихся на этом уровне в единице объема. Подставляя (1.97) в (1.96), получаем

$$A_{mn} + B_{mn} \rho(\omega) = \frac{q_n}{q_m} B_{nm} \rho(\omega) e^{\frac{E_m - E_n}{kT}} \quad (1.99)$$

или

$$\frac{A_{mn}}{\rho(\omega)} + B_{mn} = \frac{q_n}{q_m} B_{nm} e^{\frac{E_m - E_n}{kT}}. \quad (1.100)$$

Эти соотношения справедливы при любых температурах, в том числе при $T \rightarrow \infty$. При условии $\hbar\omega \ll kT$, т. е. при малых частотах и больших температурах, спектральная плотность излучения определяется классической формулой Рэля — Джинса

$$\rho(\omega) = \frac{\omega^2}{\pi^2 c^3} kT. \quad (1.101)$$

Если $T \rightarrow \infty$, то $\rho(\omega) \rightarrow \infty$ и $\exp\left(\frac{E_m - E_n}{kT}\right) \rightarrow 1$, и из (1.100) получаем первое соотношение между коэффициентами Эйнштейна для вынужденных переходов

$$q_m B_{mn} = q_n B_{nm}. \quad (1.102)$$

В случае, если кратности вырождения уровней E_m и E_n равны,

$$B_{mn} = B_{nm},$$

т. е. коэффициенты Эйнштейна для вынужденных переходов с поглощением и испусканием фотона равны.

Найдем теперь соотношение между коэффициентами A_{mn} и B_{mn} . Для этого из (1.100) получим

$$\rho(\omega) = \frac{A_{mn}}{B_{mn}} \frac{1}{\frac{E_m - E_n}{kT} - 1}. \quad (1.103)$$

При больших температурах, когда $E_m - E_n \ll kT$, можно разложить экспоненту в ряд и ограничиться первым после единицы членом разложения, откуда имеем

$$\rho(\omega) = \frac{A_{mn}}{B_{mn}} \frac{kT}{(E_m - E_n)}. \quad (1.104)$$

В условиях термодинамического равновесия при $kT \gg \hbar\omega$ спектральное распределение $\rho(\omega)$ должно определяться формулой (1.101). Сравнивая формулы (1.101) и (1.104), видим, что эти выражения будут тождественно равны при выполнении условий

$$E_m - E_n = \hbar\omega, \quad (1.105)$$

$$\frac{A_{mn}}{B_{mn}} = \frac{\hbar\omega^3}{\pi^2 c^3}. \quad (1.106)$$

Таким образом, условие частот Бора (1.105) вытекает из проведенного рассмотрения. Второе соотношение между коэф-

фициентами Эйнштейна (1.106) позволяет связать коэффициент спонтанного испускания с показателем поглощения, который может быть найден непосредственно из измерений. Соотношения (1.102) и (1.106) являются общими и не зависят от внешних условий (в частности, от температуры) и типа вещества.

Подставляя (1.105), (1.106) в (1.103), получаем известную формулу Планка для спектрального распределения излучения, находящегося в термодинамическом равновесии со стенками сосуда, т. е. спектр излучения абсолютно черного тела*

$$\rho(\omega) = \frac{\omega^2}{\pi^2 c^3} \frac{\hbar\omega}{\frac{\hbar\omega}{kT} - 1}. \quad (1.107)$$

Обратим внимание, что коэффициент $\omega^2/\pi^2 c^3$ есть число типов колебаний в единичном объеме и в единичном интервале

частот для свободного пространства. Величина $1/\left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1\right]$

представляет собой среднее число фотонов в заданном типе колебаний (напомним, что фотоны подчиняются статистике Бо-

зе — Эйнштейна), а $\hbar\omega/\left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1\right]$ — средней энергией,

приходящую на один тип колебаний. При $\hbar\omega \ll kT$ (например, в радиодиапазоне при комнатной температуре) эта средняя энергия сводится к kT .

Если излучение распространяется не в вакууме, а в среде с показателем преломления n , то в соотношениях (1.101), (1.106) и (1.107) вместо c должно стоять c/n .

В проведенном рассмотрении нас интересовали не направление и поляризация фотона, частотная зависимость коэффициентов A_{mn} и B_{mn} , а сам факт испускания или поглощения фотона. Поэтому введенные коэффициенты A_{mn} , B_{mn} и B_{nm} являются интегральными коэффициентами Эйнштейна. Для учета перечисленных факторов вводят дифференциальные коэффициенты Эйнштейна. Например, вероятность спонтанного испускания фотона в частотном интервале от ω до $\omega + d\omega$ в телесном угле $d\Omega$ с поляризацией α описывается соотношением

$$d\omega_{mn}^{\alpha} = a_{mn\alpha}(\omega, \Omega) d\omega d\Omega. \quad (1.108)$$

Любую поляризацию для заданного направления распространения можно представить как сложение двух независимых взаимно перпендикулярных поляризаций, т. е. $\alpha = 1; 2$. Очевидно,

* Иногда функцию Планка записывают не для $\rho(\omega)$, а для $\rho(\nu)$, при этом $\rho(\omega) = \rho(\nu)/2\pi$, так как $\rho(\omega)d\omega = \rho(\nu)d\nu$.

$$A_{mn} = \sum_{\alpha=1}^2 \int_{\omega} \int_{\Omega} a_{mn\alpha}(\omega, \Omega) d\omega d\Omega. \quad (1.109)$$

Аналогично вводят дифференциальные коэффициенты Эйнштейна для поглощения $b_{mn}(\omega, \Omega)$ и вынужденного испускания

$$d\omega_{mn}^{\text{погл}} = b_{mn\alpha}(\omega, \Omega) \rho_{\alpha}(\omega, \Omega) d\omega d\Omega, \quad (1.110)$$

$$d\omega_{mn}^{\text{исп}} = b_{mn\alpha}(\omega, \Omega) \rho_{\alpha}(\omega, \Omega) d\omega d\Omega. \quad (1.111)$$

Между дифференциальными коэффициентами имеется такая же связь, как и между интегральными коэффициентами Эйнштейна

$$\frac{a_{mn\alpha}(\omega, \Omega)}{b_{mn\alpha}(\omega, \Omega)} = \frac{A_{mn}}{B_{mn}} = \frac{h\omega^3}{\pi^2 c^3}, \quad (1.112)$$

$$q_m b_{mn\alpha}(\omega, \Omega) = q_n B_{mn\alpha}(\omega, \Omega). \quad (1.113)$$

Для характеристики вероятности перехода часто используют понятие времени жизни атома в возбужденном состоянии. Пусть в момент времени $t=0$ имеется N_m^0 атомов в возбужденном состоянии E_m и опустошение этого состояния возможно только за счет спонтанных переходов $m \rightarrow n$. Тогда уменьшение населенности верхнего уровня за время dt

$$-dN_m(t) = A_{mn} N_m(t) dt, \quad (1.114)$$

поскольку A_{mn} есть вероятность того, что один атом покинет это состояние за единицу времени. Так как $A_{mn} = \text{const}$, то решение уравнения (1.114) имеет вид:

$$N_m(t) = N_m^0 e^{-A_{mn} t} = N_m^0 e^{-t/\tau_{mn}}, \quad (1.115)$$

где $\tau_{mn} = 1/A_{mn}$.

Определим среднее время пребывания атома в возбужденном состоянии. По определению среднего,

$$\langle t \rangle = \frac{1}{N_m^0} \int_0^{\infty} t dN(t) = \frac{1}{N_m^0} \int_0^{\infty} t N_m^0 A_{mn} e^{-A_{mn} t} dt = 1/A_{mn} = \tau_{mn}, \quad (1.116)$$

поскольку $\int_0^{\infty} x e^{-x} dx = 1$.

Таким образом, величина $\tau_{mn} = 1/A_{mn}$ выражает среднее время пребывания (время жизни) атома в возбужденном состоянии, ограниченное спонтанными переходами $m \rightarrow n$.

Из (1.114), (1.115) получаем закон затухания мощности спонтанного излучения

$$P(t) = P_0 e^{-t/\tau_{mn}}, \quad (1.117)$$

где $P_0 = A_{mn} N_m^0 h\omega_{mn}$.

Кроме оптических излучательных переходов, обусловленных взаимодействием с электромагнитным излучением, возможны неоптические квантовые переходы из одного состояния в другое, которые называются *безызлучательными*. Такие переходы, в частности, могут происходить при столкновениях атомов и молекул газа как друг с другом, так и с электронами или со стенками сосуда. Особенно эффективны безызлучательные процессы в твердом теле при взаимодействии с колебаниями кристаллической решетки. В общем случае, если опустошение возбужденного состояния m происходит за счет различных независимых процессов, характеризуемых постоянными времени $\tau_1, \tau_2, \dots, \tau_i, \dots$ (или вероятностями $w_1 = 1/\tau_1, w_2 = 1/\tau_2, \dots, w_i = 1/\tau_i, \dots$), то согласно теореме о сложении вероятностей среднее время жизни атома в состоянии m

$$1/\tau_m = 1/\tau_1 + 1/\tau_2 + \dots + 1/\tau_i + \dots \quad (1.118)$$

Вероятности переходов связаны с соответствующими им матричными элементами. Для оптических переходов эта связь наиболее отчетливо проявляется при сравнении квантового и классического осцилляторов, при этом используют так называемое дипольное приближение.

§ 1.5. ДИПОЛЬНОЕ ПРИБЛИЖЕНИЕ

С точки зрения излучения и поглощения электромагнитной энергии атом можно рассматривать как мультипольную систему. По определению, *мультиполь* представляет собой систему парных, разноименных зарядов, обладающую определенной симметрией. Система двух зарядов — это диполь, четырех — квадруполь, восьми — октополь и т. д. Каждый мультиполь характеризуется своим моментом, порядок n которого связан с полным числом зарядов N соотношением $N = 2^n$. Произвольное распределение зарядов в общем случае можно представить в виде ряда, члены которого составляют моменты различных порядков. Такое представление оказывается удобным для описания излучающих свойств атомов и молекул.

В оптическом диапазоне длина электромагнитной волны много больше размеров атома и ряд будет быстро сходиться с увеличением порядка мультиполя. Как показывают расчеты, для видимой области ($\lambda \approx 0,5 \cdot 10^{-6}$ м) квадрупольное излучение слабее дипольного примерно в 10^6 раз.

Нашей целью является определение вероятностей переходов атомов или молекул из одного состояния в другое при взаимодействии с электромагнитным полем, т. е. нахождение коэффициентов Эйнштейна исходя из свойств атомов (молекул). Зная одни из коэффициентов, например A_{mn} или B_{nm} , и воспользовавшись полученными в § 1.4 соотношениями между ними, можно без особых затруднений определить интенсивности спонтанного и вынужденного излучения или поглощения. Как следует из теории квантовых переходов (см. § 1.3), для этого необходимо прежде всего определить матричные элементы возмущения (1.77). Чтобы пояснить смысл этого важного понятия, полезно рассмотреть свойства осциллирующего диполя с классической и квантовой точек зрения, а затем сравнить полученные результаты.

С точки зрения классических представлений простейшим источником электромагнитного излучения является точечный заряд, движущийся с ускорением. Из уравнения Максвелла (1.4) следует, что энергия, излучаемая зарядом e в единицу времени, пропорциональна квадрату ускорения a :

$$\frac{dE}{dt} = \frac{e^2}{6\pi\epsilon_0 c^3} a^2. \quad (1.119)$$

Если заряд совершает гармонические колебания с частотой ω и амплитудой r_m , т. е.

$$r(t) = r_m \cos \omega t, \quad (1.120)$$

то из (1.119) получаем мгновенную мощность излучения

$$P = -\frac{dE}{dt} = \frac{e^2 r_m^2 \omega^4}{6\pi\epsilon_0 c^3} \cos^2 \omega t. \quad (1.121)$$

Усредненная по времени за период колебаний мощность излучения

$$\langle P \rangle = \frac{e^2 r_m^2 \omega^4}{12\pi\epsilon_0 c^3} \quad (1.122)$$

поскольку $\langle \cos^2 \omega t \rangle = 1/2$.

Эта формула справедлива и для излучения, создаваемого системой из многих зарядов. Простейшей системой является *электрический диполь* — совокупность двух одинаковых по значению и противоположных по знаку зарядов, находящихся на расстоянии L друг от друга. Момент электрического диполя D , или дипольный момент, численно равен произведению зарядов e на расстояние L между ними: $D = eL$ и направлен от отрицательного полюса к положительному (рис. 1.9). Если дипольный момент D гармонически

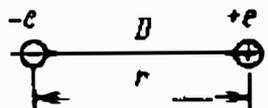


Рис. 1.9. Электрический диполь

изменяется с частотой ω , то такой диполь называют *осциллирующим диполем* или *осциллятором*. Средняя мощность излучения такого осциллятора получается из выражения (1.122)

$$\langle P \rangle = \frac{\omega_0^4}{12\pi\epsilon_0 c^3} D_0^2, \quad (1.123)$$

где D_0 — амплитуда изменения дипольного момента.

Многие оптические свойства излучающих систем можно получить, моделируя такие системы совокупностью гармонических осцилляторов, собственные частоты которых совпадают с частотами рассматриваемых переходов. С точки зрения классической электродинамики осциллирующий диполь излучает энергию непрерывно. Амплитуда колебаний для классического осциллятора может принимать любые значения. Пусть в начальный момент времени $t=0$ осциллятор обладает запасенной энергией E_0 , чему соответствует амплитуда колебаний r_0 , причем

$$E_0 = 0,5 m \omega_0^2 r_0^2. \quad (1.124)$$

С течением времени такой классический осциллятор передает энергию поля по закону (1.121). Этому соответствует уменьшение амплитуды колебаний r_m . При $t > 0$

$$\langle E(t) \rangle = 0,5 m \omega_0^2 r_m^2(t). \quad (1.125)$$

Сопоставляя это выражение с (1.122), получаем

$$\langle P \rangle = -\frac{d\langle E \rangle}{dt} = \frac{e^2 \omega_0^4}{6\pi\epsilon_0 c^3 m} \langle E \rangle. \quad (1.126)$$

Вводя обозначения

$$\gamma = \frac{e^2 \omega_0^4}{6\pi\epsilon_0 c^3 m}, \quad (1.127)$$

получаем уравнение для энергии излучателя

$$\frac{d\langle E \rangle}{dt} + \gamma \langle E \rangle = 0.$$

Отсюда

$$\langle E(t) \rangle = E_0 e^{-\gamma t}, \quad (1.128)$$

где $E_0 = E_{t=0}$ определяется соотношением (1.124).

Сопоставляя выражения (1.128) и (1.125), находим изменение амплитуды колебаний классического осциллятора во времени

$$r_m = r_0 e^{-\gamma t/2}, \quad (1.129)$$

а с учетом (1.120) получаем

$$r(t) = r_0 e^{-\gamma t/2} \cos \omega_0 t. \quad (1.130)$$

График этой функции представлен на рис. 1.10.

Из (1.128) находим, что средняя за период мощность излучения электрического диполя изменяется во времени по экспоненциальному закону:

$$\langle P \rangle = -\frac{d\langle E \rangle}{dt} = \frac{e^2 \omega_0^4}{12\pi \epsilon_0 c^3} r_0^2 e^{-\gamma t} = \frac{\omega_0^4}{12\pi \epsilon_0 c^3} D_0^2 e^{-\gamma t} = P_0 e^{-\gamma t}, \quad (1.131)$$

где $P_0 = \frac{\omega_0^4}{12\pi \epsilon_0 c^3} D_0^2$.

Это выражение полностью совпадает с выражением для мощности спонтанного излучения (1.117), если положить $\gamma = A_{\text{сп}}$. Указанное совпадение позволяет продолжить аналогию между спонтанными переходами и излучением классического осциллятора, в частности в отношении спектрального состава излучения. Из сказанного ранее видно, что с точностью до постоянного множителя мгновенная мощность излучения диполя пропорциональна $r^2(t)$. Очевидно, такая система излучает немонохроматическую волну (см. рис. 1.10). Для определения спектра излучения $r(t)$ следует представить в виде интеграла Фурье, т. е. функцию (1.130), изображенную на рис. 1.10, представить в виде суммы (интеграла) гармонических составляющих. Аналогично (1.26) и (1.27) имеем

$$r(t) = \int_{-\infty}^{\infty} r(\omega) e^{i\omega t} d\omega, \quad (1.132)$$

где амплитуды гармонических составляющих

$$r(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(t) e^{-i\omega t} dt. \quad (1.133)$$

Зависимость (1.130) можно записать в экспоненциальной форме:

$$r(t) = 0,5 r_0 e^{-\gamma t/2} (e^{i\omega_0 t} + e^{-i\omega_0 t}).$$

Подставляя это выражение в интеграл Фурье (1.133), производя интегрирование и отбросив малые члены, получаем

$$r(\omega) = \frac{r_0}{2\pi} \frac{1}{i(\omega_0 - \omega) + \gamma/2}$$

Интересующая нас величина $|r(\omega)|^2 = r^*(\omega)r(\omega)$

$$|r(\omega)|^2 = \frac{r_0^2}{4\pi} \frac{1}{(\omega_0 - \omega)^2 + (\gamma/2)^2}. \quad (1.134)$$

Напомним, что ω_0 — собственная (резонансная) частота осциллятора, а r_0 — начальная амплитуда колебаний при $t=0$. Зависимость

$$g_L(\omega) = \frac{A}{(\omega_0 - \omega)^2 + (\gamma/2)^2} \quad (1.135)$$

называется *кривой Лоренца*, где A — нормировочный множитель. Эта зависимость определяет распределение энергии по частоте.

Из условия нормировки $\int_{-\infty}^{\infty} g_L(\omega) d\omega = 1$ определяем нормировочный множитель $A = \gamma/2\pi$.

Функция Лоренца представлена на рис. 1.11. Она имеет максимум при $\omega = \omega_0$, который из условия нормировки равен $g(\omega_0) = 2/\pi\gamma$. На расстояниях $\omega_0 - \omega = \pm \gamma/2$ спектральная плотность убывает в два раза. Величина $2(\omega_0 - \omega) = \Delta\omega = \gamma$, называемая *полушириной линии*, представляет собой естественную ширину спектральной линии. Как было показано, $\gamma = 1/\tau$, где τ имеет смысл времени релаксации. Отсюда вытекает фундаментальное соотношение спектрального анализа

$$\Delta\omega\tau \approx 1. \quad (1.136)$$

Сравнив его с (1.91), нетрудно убедиться, что оно эквивалентно соотношению неопределенностей в квантовой физике.

Рассмотрим теперь квантовый осциллятор. Его энергия в отличие от классического осциллятора может принимать только определенные дискретные значения. Здесь не может быть плавного изменения амплитуды колебаний, а наблюдается скачкообразный переход из одного разрешенного состояния в другое.

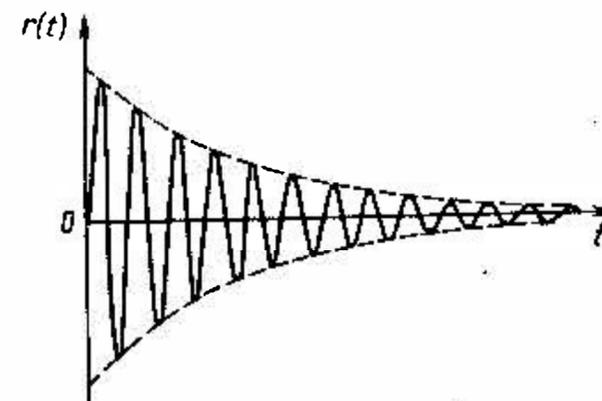


Рис. 1.10. График функции (1.130)

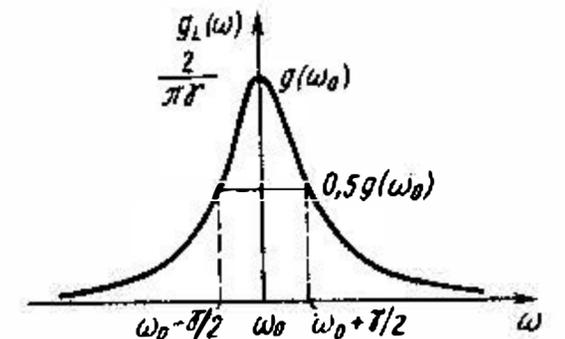


Рис. 1.11. Функция Лоренца

Частота излучения квантового осциллятора определяется через энергию перехода: $\omega_{mn} = (E_m - E_n)/\hbar$. Такому переходу соответствует возникновение осциллирующего электрического момента атома. Электрический дипольный момент атома с точки зрения квантовой теории имеет следующий физический смысл.

Для заданного стационарного состояния n распределение заряда электрона в атоме (распределение «электронного облака») задается в виде $e\varphi_n^* \varphi_n dV$. Эта величина представляет собой эффективную плотность заряда в объеме dV . Тогда среднее значение электрического дипольного момента

$$D_{nn} = e \int \varphi_n^* r \varphi_n dV, \quad (1.137)$$

а его проекции:

$$D_{nn}^x = e \int \varphi_n^* \varphi_n x dV,$$

$$D_{nn}^y = e \int \varphi_n^* \varphi_n y dV,$$

$$D_{nn}^z = e \int \varphi_n^* \varphi_n z dV.$$

Здесь r — радиус-вектор, проведенный из начала координат, где помещено ядро атома.

Поскольку зависимость волновой функции состояния от времени имеет вид (1.67), произведение $\varphi_n^* \varphi_n$ не зависит от времени и, следовательно, распределение заряда в атоме, находящемся в стационарном состоянии, не изменяется. Такая система согласно классической электродинамике не должна излучать энергию.

При переходе атома из состояния m в состояние n распределение заряда определяется волновыми функциями обоих состояний:

$$\rho_{mn} = e\varphi_n^* \varphi_m.$$

С учетом (1.67) произведение $\varphi_n^* \varphi_m$ будет зависеть от времени по закону

$$\varphi_n^* \varphi_m = \Psi_n^* \Psi_m e^{i \frac{(E_m - E_n)t}{\hbar}}.$$

Это означает, что объемная плотность заряда при переходе из одного квантового состояния в другое осциллирует с характерной частотой $\omega_{mn} = (E_m - E_n)/\hbar$. Такое распределение заряда можно характеризовать интегральным дипольным моментом

$$D(t) = \int r \rho_{mn} dV.$$

Для стационарного состояния дипольный момент постоянен во времени. При переходе из одного состояния в другое возникают осциллирующие дипольного момента

$$D_{mn}(t) = D_{mn}^0 \exp\left(i \frac{E_m - E_n}{\hbar} t\right). \quad (1.138)$$

Таким образом, квантовому переходу из состояния m в состояние n можно сопоставить возникновение осциллирующего диполя с собственной частотой колебаний ω_{mn} .

Строгое рассмотрение вопроса в рамках квантовой электродинамики показывает, что оптические свойства квантового перехода в основном совпадают с оптическими свойствами эквивалентного осциллятора. Количественные соотношения при этом совпадают с формулами, которые дает классическая электродинамика для осциллирующего электрического диполя, если под амплитудой дипольного момента понимать величину

$$D_{mn}^0 = e \int \Psi_m^* r \Psi_n dV. \quad (1.139)$$

Эта величина количественно характеризует вероятность перехода и называется *дипольным матричным элементом перехода* $m \rightarrow n$. Совокупность квантовых переходов в квантовой системе характеризуется двухмерной совокупностью чисел D_{mn} . Эту совокупность принято записывать в виде бесконечной матрицы

$$\begin{vmatrix} D_{11}, D_{12}, \dots, D_{1n}, \dots \\ D_{21}, D_{22}, \dots, D_{2n}, \dots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ D_{n1}, D_{n2}, \dots, D_{nn}, \dots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \end{vmatrix}. \quad (1.140)$$

Числа, составляющие матрицу, называются *матричными элементами*. Нondiagonальные элементы матрицы D_{ik} при $i \neq k$ являются функциями времени и соответствуют излучению или поглощению света на частотах, определяемых правилом частот Бора. Совокупность этих частот обусловлена спектром разрешенных значений энергии данной квантовой системы и также образует матрицу

$$\begin{vmatrix} 0, \omega_{12}, \dots, \omega_{1n}, \dots \\ \omega_{21}, 0, \dots, \omega_{2n}, \dots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \omega_{n1}, \omega_{n2}, \dots, \omega_{nn}, \dots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \end{vmatrix}. \quad (1.141)$$

Частоты ω_{ii} , расположенные по диагонали, тождественно равны нулю. Им соответствуют диагональные элементы D_{ii} матрицы

(1.140), которые либо также равны нулю (для систем с центральной симметрией), либо являются постоянными величинами. Например, мощность, излучаемую при спонтанном переходе $m \rightarrow n$, можно вычислить с помощью полученного для классического осциллятора выражения (1.123) или (1.131), если вместо амплитуды классического дипольного момента D_0 подставить удвоенный матричный элемент D_{mn}^0 (множитель 2 возникает при переходе от показательной формы записи к тригонометрической). Считая, что в единице объема имеется N_m диполей, получаем

$$\langle P \rangle = \frac{\omega_{mn}^4 N_m}{3\pi \epsilon_0 c^3} (D_{mn}^0)^2. \quad (1.142)$$

Величина N_m имеет смысл населенности верхнего энергетического уровня. Учитывая связь между излучательной мощностью и вероятностью спонтанного перехода

$$P_{mn} = A_{mn} N_m \hbar \omega_{mn},$$

имеем

$$A_{mn} = \frac{\omega_{mn}^3}{3\pi \epsilon_0 \hbar c^3} (D_{mn}^0)^2. \quad (1.143)$$

Это выражение соответствует излучению осциллятора с дипольным моментом

$$D_{mn} = 2e r_{mn} \cos \alpha_{mn} t,$$

где $r_{mn} = \int_V \Psi_m^* r \Psi_n dV$.

Воспользовавшись соотношениями (1.102) и (1.106) между коэффициентами Эйнштейна, получим

$$B_{mn} = \frac{\pi}{12\hbar^2 \epsilon_0} (D_{mn}^0)^2. \quad (1.144)$$

Соотношения (1.143) и (1.144) позволяют рассчитать коэффициенты Эйнштейна, если известны характеристики (волновые функции и энергии) состояний, между которыми происходит оптический переход.

Аналогичные результаты дают соотношения для вероятности квантового перехода под воздействием внешнего возмущения (см. § 1.3). При этом необходимо учитывать, что вынужденный переход (например, определяющий поглощение) вызывается взаимодействием между электрическим полем электромагнитной волны и электрическим дипольным моментом атома. Тогда возмущение

$$V = e E(r, t) r.$$

Поскольку длина электромагнитной волны существенно больше размеров атома, можно считать, что $E(r, t) \approx E(0, t)$ и

$$V_{mn} = e \int \Psi_m^* E r \Psi_n dV \approx e E(0, t) \int \Psi_m^* r \Psi_n dV = E(0, t) D_{mn}. \quad (1.145)$$

Подставляя это выражение в (1.79), с учетом (1.87), (1.93) получаем формулу (1.143).

Для характеристики вероятности перехода наряду с матричным элементом, коэффициентом Эйнштейна и временем жизни иногда используют понятие «сила осциллятора», подчеркивая тем самым связь свойств классического осциллятора и атома как излучающих систем. В классической теории сила осциллятора f имеет смысл как доля элементарных осцилляторов, принимающих участие в макроскопической поляризуемости среды на частоте ω . В квантовой теории сила осциллятора характеризует как бы степень «дипольности» системы оптический электрон — ионный остов на рассматриваемой частоте ω , определяющей возможность совершить оптический переход из одного состояния в другое. Сила осциллятора f_{mn} связана с коэффициентом Эйнштейна A_{mn} для спонтанных переходов соотношением

$$f_{mn} = \frac{3mc^2}{2\hbar^2 \omega^2} A_{mn}. \quad (1.146)$$

Проведем оценку, например, коэффициента Эйнштейна A_{mn} . Дипольный момент D_{mn} имеет порядок величины ea_B , где a_B — боровский радиус атома. Положив $a_B \approx 1 \text{ \AA} = 10^{-10} \text{ м}$, по формуле (1.143) находим для видимой области спектра ($\lambda \approx 0,5 \text{ мкм}$) $A_{mn} \approx 10^8 \text{ с}^{-1}$.

Таким образом, коэффициенты Эйнштейна, вероятности спонтанных и вынужденных переходов определяются через соответствующие дипольные матричные элементы, которые являются недиагональными элементами матрицы (1.140). Некоторые из этих матричных элементов могут оказаться равными нулю. Это означает, что такой переход не может происходить в дипольном приближении, поэтому его называют *запрещенным*. Переходы, для которых $D_{mn} \neq 0$, являются *разрешенными*. Принадлежность перехода к запрещенным или разрешенным определяется *правилами отбора*.

Правила отбора указывают допустимые изменения квантовых чисел, при которых матричный элемент D_{mn} отличен от нуля. Для данного перехода они определяются в основном симметрией волновых функций в начальном и конечном состояниях. В частности, для квантовых систем, обладающих центральной симметрией, например для изолированных атомов, все энергетические уровни можно разделить на четные, для которых $\Psi(r) = \Psi(-r)$,

я нечетные, для которых $\Psi(r) = -\Psi(-r)$. Например, для атома водорода уровни с $l=0, 2, \dots$ (*s*-, *d*-уровни и т. д.) являются четными, а уровни с $l=1, 3, \dots$ (*p*-, *f*-уровни) являются нечетными. Дипольный матричный элемент (1.139) отличен от нуля только для уровней с различной четностью (для уровней с одинаковой четностью вклад, получаемый при вычислении подынтегральной части выражения (1.139) в точках r и $-r$, одинаков по значению, но противоположен по знаку). Поэтому дипольные переходы возможны лишь между уровнями энергии с разной четностью.

Для атомных уровней, характеризуемых квантовыми числами n , l и m_l , правила отбора для дипольных переходов следующие.

1. Изменение главного квантового числа может быть любым: $\Delta l = 0; 1; 2; \dots$

2. Орбитальное квантовое число l может изменяться только на ± 1 : $\Delta l = \pm 1$.

Это правило по существу отражает закон сохранения момента количества движения для системы электрон + фотон, так как момент количества движения фотона равен \hbar .

3. Магнитное квантовое число m_l может изменяться только на 0 или ± 1 : $\Delta m_l = 0; \pm 1$.

Аналогичные правила существуют и для квантовых чисел L , J и S . В частном случае одновременного выполнения условий $\Delta J = 0$ и $\Delta M_J = 0$ переход также запрещен.

Правила отбора для переходов между колебательными и вращательными уровнями энергии молекул указаны в § 1.2 [соотношения (1.64) и (1.65)].

Переход будет разрешен, если выполняются все правила отбора. Например, все переходы $s \rightarrow s$ и $p \rightarrow p$ запрещены, а переходы $p \rightarrow d$ разрешены только для $\Delta m_l = 0; \pm 1$.

Если переходы разрешены в дипольном приближении, то для них A_{mn} имеет порядок величины, оценка которой была проведена выше, т. е. $A_{mn} \approx 10^8 \text{ с}^{-1}$. Соответственно время жизни системы в таком состоянии $\sim 10^{-8} \text{ с}$, если релаксация из возбужденного состояния определяется только спонтанными излучательными переходами, или $< 10^{-8} \text{ с}$, если имеются какие-то другие (например, безызлучательные) процессы опустошения уровня. Такие уровни с малыми временами жизни называются *лабильными*.

Если переходы запрещены в дипольном приближении $D_{mn} = 0$, то это не значит, что они вообще не могут произойти. Кроме электрического дипольного момента и связанного с ним дипольного излучения атому можно приписать электрический квадрупольный (октупольный) или магнитный дипольный (квадрупольный) момент. Но это будет следующим приближением. Матричные элементы и соответственно вероятности электрического квадрупольного и магнитного дипольного переходов приблизительно в 10^6 раз меньше, чем для электрического дипольного прибли-

жения (если и те и другие разрешены правилами отбора). Вероятность октупольных переходов, т. е. переходов с изменением момента третьего порядка, еще меньше.

Возбужденное энергетическое состояние системы, для которого все переходы в более низкие состояния запрещены при электрических дипольных взаимодействиях, называются *метастабильным уровнем*. Время жизни атомов в этом состоянии порядка 10^{-3} с и больше.

§ 1.6. УШИРЕНИЕ СПЕКТРАЛЬНЫХ ЛИНИЙ

При рассмотрении энергетических состояний квантовых систем в § 1.2 нас не интересовал вопрос о ширине дискретных энергетических уровней атомов и молекул, и эти уровни принимались бесконечно узкими. Излучательным переходам между такими идеализированными уровнями энергии соответствовала бы бесконечно узкая спектральная линия излучения (или поглощения) на строго фиксированной частоте, т. е. идеальная монохроматическая волна. В действительности подобная ситуация невозможна хотя бы по той причине, что время жизни τ в возбужденном состоянии конечно. В силу соотношения неопределенностей $\delta E \tau \approx \hbar$ это приводит к неопределенности в энергии состояния, т. е. к «размытию» энергетического уровня на δE . Для изолированных атомов и молекул это уширение очень мало ($\delta E \ll E$), но энергетические уровни имеют конечную ширину.

Степень и характер уширения энергетических уровней квантовых систем наиболее отчетливо проявляются при изучении формы *спектральных линий*, т. е. характерных узких линий испускания или поглощения, отвечающих определенному излучательному квантовому переходу.

Распределение интенсивности излучения (или поглощения) по частоте в пределах данной линии характеризуется функцией $g(\omega)$, которая называется *форм. фактором спектральной линии* или просто *формой линии*. Эта функция нормирована

$$\int_{-\infty}^{\infty} g(\omega) d\omega = 1. \quad (1.147)$$

Для характеристики относительной ширины спектральной линии используют понятие *добротности спектральной линии*, которая численно равна отношению резонансной частоты ω_0 в максимуме линии к ее ширине $\Delta\omega$ на уровне половинной интенсивности.

В реальных квантовых системах существует ряд факторов, приводящих к уширению их энергетических уровней и соответственно к уширению спектральных линий. Рассмотрим наиболее характерные из них.

Естественное уширение. Это уширение связано с конечностью времени пребывания атомов и молекул в возбужденном состоянии, что, как отмечалось, приводит к «размытию» энергетического уровня на величину порядка $\delta E \approx \hbar/\tau$. Отметим, что «естественная ширина линии» определяет тот предел, уже которого спектральная линия в естественных условиях быть не может. Форма спектральной линии, определяемая конечным временем жизни в возбужденном состоянии, была получена в § 1.4, где было показано, что спектр излучения затухающего осциллятора описывается функцией Лоренца (1.135), или, другими словами, спектральная линия имеет лоренцеву форму. Выражение (1.135) перепишем в виде

$$g_L(\omega) = \frac{\Delta\omega}{2\pi} \frac{1}{(\omega_0 - \omega)^2 + (\Delta\omega/2)^2}, \quad (1.148)$$

где $\Delta\omega = \gamma$ — ширина линии на уровне 0,5 от максимума, а ω_0 — резонансная частота. Если переход осуществляется в основное состояние 0, то $\Delta\omega = \gamma = A_{m0}$. Если нижний уровень n не является основным, то необходимо учитывать возможность спонтанного распада не только верхнего m , но и нижнего n уровней.

В этом случае $\Delta\omega = (A_{m0} + A_{n0}) = \left(\frac{1}{\tau_m} + \frac{1}{\tau_n}\right)$. Квантовая электродинамика позволяет вычислить естественную форму спектральной линии, которая в точности описывается выражением (1.135), полученным для классического осциллятора.

Конечность времени жизни частицы на уровне E_m формально можно учесть, введя затухание в волновые функции невозмущенного состояния. В этом случае выражение (1.67) можно записать в виде

$$\varphi_m(\mathbf{r}, t) = \Psi_m(\mathbf{r}) e^{i E_m t / \hbar} e^{-t/2\tau}.$$

Множитель $e^{-t/2\tau}$ учитывает конечность времени жизни частицы на уровне E_m , поскольку вероятность ее нахождения в этом состоянии пропорциональна $\varphi_m^* \varphi_m$, т. е. $\exp(-t/\tau)$. Нетрудно заметить, что реальная часть функции $\varphi_m(\mathbf{r}, t)$ изменяется во времени по закону, соответствующему рис. 1.10. Поэтому разложение по монохроматическим волнам (разложение Фурье) даст спектр вида (1.135), т. е. функцию Лоренца (1.148). Это еще раз указывает на правомерность представления квантовых излучающих систем в виде набора осцилляторов с собственными частотами, удовлетворяющими правилу частот Бора.

Заметим, что чем больше вероятность спонтанного испускания A_{m0} , тем больше естественная ширина спектральной линии.

Типичное значение естественной ширины линии для разрешенных в дипольном приближении переходов в видимой области спектра $\Delta\omega = A_{m0} \approx 10^8 \text{ с}^{-1}$, т. е. $\Delta\nu$ порядка 20 МГц при $\nu_0 \approx 5 \cdot 10^{14} \text{ Гц}$. Для переходов с метастабильных уровней естественная ширина линии существенно меньше и при тех же условиях имеет порядок сотен герц.

Так как коэффициент Эйнштейна A_{m0} пропорционален кубу частоты [см. выражение (1.143)], то в радиодиапазоне естественная ширина линии существенно меньше, чем в оптическом диапазоне. Например, для газообразного аммиака NH_3 спектральная линия перехода с частотой $\nu_0 = 24870 \text{ МГц}$ ($\lambda_0 = 1,25 \text{ см}$) имеет естественную ширину $\sim 10^{-3} \text{ Гц}$.

Форму спектральной линии можно выразить не только как функцию частоты $g(\omega)$, но и как функцию длины волны $g(\lambda)$, учитывая простую связь $\omega = 2\pi c/\lambda$. Ширина спектральной линии, выраженная в длинах волн

$$\Delta\lambda = \frac{2\pi c}{\omega^2} \Delta\omega.$$

Примечательно, что для классического осциллятора естественная ширина линии, выраженная в длинах волн и называемая также радиационной шириной, является универсальной постоянной. Действительно, подставляя в последнее соотношение величину $\Delta\omega = \gamma$ из (1.127), получим

$$\Delta\lambda = \frac{e^2}{3\epsilon_0 c^2 m} \approx 1,2 \cdot 10^{-14} \text{ м},$$

что приблизительно равно двум классическим диаметрам электрона.

Доплеровское уширение. Эффект Доплера есть изменение частоты (длины волны), наблюдаемое при движении источника воли относительно приемника. Этот эффект вытекает из теории относительности и характерен для любых волн, в том числе для звуковых, электромагнитных и т. д. Если источник излучает электромагнитные волны с собственной частотой ω_0 , то для наблюдателя, относительно которого он движется со скоростью v (рис. 1.12), воспринимаемая частота излучения

$$\omega = \omega_0 \frac{\sqrt{1 - v^2/c^2}}{1 - v/c \cos\theta}.$$

При $v \ll c$ (c — скорость света)

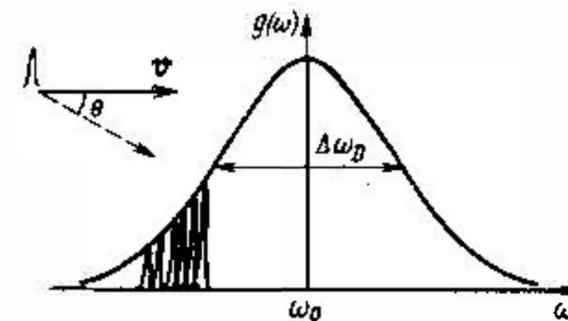


Рис. 1.12. Доплеровское уширение спектральной линии

$$\omega = \omega_0 \left(1 + \frac{v}{c} \cos \theta\right). \quad (1.149)$$

Частота излучения увеличивается, если источник приближается к наблюдателю, и уменьшается, если он удаляется от него. Хаотичность теплового движения атомов и молекул в газе приводит к тому, что вместо одной резонансной линии с частотой ω_0 приемник воспринимает совокупность плотно расположенных линий, огибающая которых дает наблюдаемый контур спектральной линии, как показано на рис. 1.12. Ширина этого спектра пропорциональна наиболее вероятной скорости частиц в газе (т. е. зависит от температуры) и собственной частоте перехода ω_0 . Зная распределение частиц по скоростям (распределение Максвелла)

$$f(v)dv = \frac{1}{\sqrt{\pi}v_0} \exp[-(v/v_0)^2] dv, \quad (1.150)$$

где $v_0 = \sqrt{2kT/M}$ — наиболее вероятная скорость движения частиц в газе, и учитывая выражение (1.149), можно получить форму спектральной линии, обусловленную доплеровским уширением

$$g_G(\omega) = \frac{2\sqrt{\ln 2}}{\sqrt{\pi}\Delta\omega_D} \exp\left\{-\left[\frac{2\sqrt{\ln 2}(\omega - \omega_0)}{\Delta\omega_D}\right]^2\right\}. \quad (1.151)$$

Здесь

$$\Delta\omega_D = 2\sqrt{\ln 2} \frac{\omega_0}{c} \sqrt{\frac{kT}{M}} \quad (1.152)$$

— ширина доплеровской линии на уровне 0,5 от максимального значения при $\omega = \omega_0$. Подставляя численные значения констант в (1.152), получаем формулу, удобную для численных расчетов

$$\frac{\Delta\omega_D}{\omega_0} = \frac{\Delta\nu_D}{\nu_0} = \frac{-\Delta\lambda_D}{\lambda_0} \approx 7 \cdot 10^{-7} \sqrt{\frac{T}{M}},$$

где M — относительная молекулярная масса, а температура T выражена в К.

Функция (1.151) есть функция Гаусса, поэтому говорят, что линия, уширенная за счет эффекта Доплера, имеет гауссову форму. Множитель перед экспонентой определяется условием нормировки.

На рис. 1.13 для сравнения приведены гауссова и лоренцева формы спектральных линий при одинаковых параметрах уширения $\Delta\omega$.

В условиях газового разряда для легких атомов величина $\sqrt{2kT/M} \approx 10^3$ м/с, что для видимой области спектра при $\nu_0 \approx 5 \cdot 10^{14}$ Гц дает $\Delta\nu_D = \Delta\omega_D/2\pi \approx 1$ ГГц. В радиодиапазоне, например, для линии аммиака $\nu_0 = 24870$ МГц, при комнатной температуре доплеровское уширение $\Delta\nu_D \approx 70$ кГц, т. е. в миллионы раз превышает естественную ширину линии.

Обратим внимание на следующее обстоятельство: если излучатель движется внутри области меньше длины волны, то его движение не влияет на наблюдаемую частоту излучения, т. е. эффект Доплера не проявляется.

Уширение вследствие столкновений. Столкновения атомов с другими атомами, ионами, свободными электронами или стенками сосуда в газе, а также взаимодействие атомов с решеткой в твердых телах (которое можно рассматривать как столкновение с фононом) приводят к увеличению скорости обмена энергией между частицами и соответственно к уменьшению времени жизни атома в возбужденном состоянии. Форма спектральной линии, уширенной за счет столкновений, будет описываться функцией Лоренца, как и при естественном уширении. Но вместо $\Delta\omega = 1/\tau_{\text{ест}} = A_{\text{ест}}$ в (1.148) должно стоять $\Delta\omega = 1/\tau_p$, где τ_p — время релаксации, определяемое процессами столкновений, так называемыми газокинетическими соударениями.

Оценим величину τ_p , например, для атомов неона при давлении $P \approx 60$ Па $\approx 0,5$ мм рт. ст. (характерное давление в He-Ne лазере). Время релаксации будет приблизительно равно интервалу времени $\tau_{\text{ст}}$ между двумя столкновениями атомов Ne. Величину $\tau_{\text{ст}}$ в свою очередь можно оценить как отношение средней длины свободного пробега атома $\bar{\lambda}$ к средней тепловой скорости v_T . Из кинетической теории газов имеем

$$\tau_p \approx \tau_{\text{ст}} \approx \frac{\sqrt{MkT}}{16\sqrt{\pi}a_p^2 P}, \quad (1.153)$$

где a_p — эффективный борковский радиус атома (молекулы); P — давление газа. При комнатной температуре $\tau_{\text{ст}} \approx 5 \cdot 10^{-7}$ с, чему соответствует ширина линии $\Delta\nu = \frac{\Delta\omega}{2\pi} = \frac{1}{2\pi\tau_{\text{ст}}} \approx 0,3$ МГц.

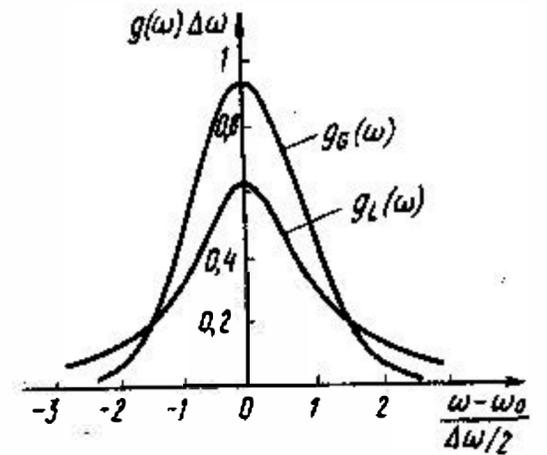


Рис. 1.13. Лоренцева $g_L(\omega)$ и гауссова $g_G(\omega)$ формы спектральных линий при одинаковой их ширине на уровне 0,5

Отметим, что $\tau_{\text{ст}}$ обратно пропорциональна, а $\Delta\omega$ прямо пропорциональна давлению газа. Поэтому если уширение линии за счет столкновений в рассмотренном случае меньше доплеровского уширения, то для CO_2 -лазеров, в которых рабочий газ находится при атмосферном давлении, столкновительное уширение может преобладать над доплеровским.

Особенно эффективно процессы релаксации из возбужденных состояний могут осуществляться в твердых телах при взаимодействии с колебаниями решетки. При этом τ_r может составлять $10^{-9} \dots 10^{-12}$ с, приводя к существенным уширениям соответствующих спектральных линий.

Уширение за счет влияния внутренних (внутрикристаллических) и внешних электрического и магнитного полей. Этот механизм уширения определяется рассмотренными в § 1.2 эффектами Зеемана и Штарка. Если зеемановское или штарковское расщепление меньше ширины каждого подуровня, то рядом расположенные подуровни частично перекрываются, вызывая уширение соответствующих спектральных линий.

В качестве активных материалов твердотельных лазеров широко используют различные кристаллы, в решетку которых введены специальные ионы — активаторы. К ним, например, относятся рубин — кристалл окиси алюминия Al_2O_3 , легированный хромом. Ионы хрома Cr^{3+} , являющиеся рабочими ионами, замещают ионы алюминия Al^{3+} и находятся под воздействием сильного внутрикристаллического электрического поля $E_{\text{кр}}$ со стороны ближайших к ним ионов алюминия и кислорода, которое сдвигает и расщепляет энергетические уровни Cr^{3+} . Тепловые колебания решетки приводят к тому, что внутрикристаллическое поле флуктуирует вокруг некоторого среднего значения. Это вызывает хаотическое смещение энергетических уровней Cr^{3+} и соответственно «флуктуации» спектральной линии, что эквивалентно ее уширению. Понижение температуры уменьшает флуктуации и ширину спектральной линии. Кроме того, вследствие возможного неоднородного окружения активного иона, обусловленного неоднородностями кристалла или посторонними дефектами, возникает дополнительное уширение суммарной линии. Последнее наиболее сильно проявляется в активированных стеклах, где окружение каждого из активных ионов различно. Могут быть и другие механизмы уширения спектральных линий, например вызванные неоднородностями среды.

Различают однородное и неоднородное уширение. Уширение называется *однородным*, если линии каждого отдельного атома и системы в целом уширяются одинаково. К однородному уширению относятся естественное уширение, уширение за счет процессов релаксации, в том числе столкновений, и т. д. Уширение

называется *неоднородным*, если резонансные частоты отдельных атомов не совпадают и распределяются в некоторой полосе частот, приводя к уширению линии системы в целом при существенно меньшем уширении линии отдельных атомов. К неоднородному уширению относятся доплеровское уширение, уширение за счет неоднородностей среды и т. д.

Форма однородно уширенных спектральных линий обычно хорошо описывается функцией Лоренца $g_L(\omega)$ (1.148), а форма неоднородно уширенных линий — функцией Гаусса $g_G(\omega)$ (1.151). Подчеркнем, что в чистом виде спектральная линия будет уширена однородно или неоднородно в случае, когда преобладает какой-то один механизм уширения. Нередко общее уширение спектральной линии определяется одновременно несколькими механизмами. В этом общем случае форма линии будет определяться сверткой функций $g_1(\omega)$ и $g_2(\omega)$, описывающих форму линии для различных процессов:

$$g(\omega) = g_1(\omega) * g_2(\omega) = \int_{-\infty}^{\infty} g_1(\omega') g_2(\omega - \omega') d\omega' \quad (1.154)$$

(операция свертки двух функций $g_1(\omega)$ и $g_2(\omega)$ обозначается звездочкой *).

Свертка двух лоренцевых линий шириной $\Delta\omega_1$ и $\Delta\omega_2$ приводит снова к лоренцевой линии шириной $\Delta\omega = \Delta\omega_1 + \Delta\omega_2$, т. е.

$$g_{L1}(\omega, \Delta\omega_1) * g_{L2}(\omega, \Delta\omega_2) = g_L(\omega, \Delta\omega_1 + \Delta\omega_2).$$

Это по существу эквивалентно тому, что при однородном уширении времена жизни определяются соотношением (1.118).

Свертка двух гауссовых линий шириной $\Delta\omega_1$ и $\Delta\omega_2$ приводит также к гауссовой линии шириной $\Delta\omega = \sqrt{\Delta\omega_1^2 + \Delta\omega_2^2}$, т. е.

$$g_{G1}(\omega, \Delta\omega_1) * g_{G2}(\omega, \Delta\omega_2) = g_G(\omega, \sqrt{\Delta\omega_1^2 + \Delta\omega_2^2}).$$

Если однородное и неоднородное уширения одного порядка, то задача определения формы линии сводится к нахождению свертки лоренцевой и гауссовой функций (соответствующие интегралы табулированы и в математике известны как интегралы Войта).

Таким образом, форма спектральной линии $g(\omega)$ известна. Это позволяет определить в явном виде спектральную зависимость коэффициентов Эйнштейна, которая с точностью до постоянного множителя должна определяться функцией $g(\omega)$. Например, для вероятности спонтанного перехода

$$a_{mn}(\omega) = \text{const } g(\omega).$$

Так как функция $g(\omega)$ нормирована, то

$$\text{const} = \int_0^{\infty} a_{mn}(\omega) d\omega = A_{mn},$$

т. е.

$$a_{mn}(\omega) = A_{mn} g(\omega). \quad (1.155)$$

Аналогично для остальных коэффициентов Эйнштейна имеем:

$$b_{mn}(\omega) = B_{mn} g(\omega), \quad (1.155a)$$

$$b_{nm}(\omega) = B_{nm} g(\omega), \quad (1.155b)$$

где A_{mn} , B_{mn} и B_{nm} — интегральные коэффициенты Эйнштейна, определяемые через дипольные матричные элементы соотношениями (1.143) и (1.144).

Для однородного уширения функция $g(\omega)$ имеет лоренцеву форму (1.148), а для неоднородного — гауссову (1.151). В общем случае $g(\omega)$ дается сверткой (1.154).

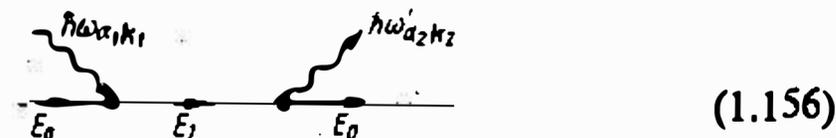
§ 1.7. РАССЕЯНИЕ СВЕТА

Ранее всюду принималось, что взаимодействие света с квантовой системой возможно только в случае соблюдения правила частот Бора. Рассмотрим теперь взаимодействие атомов и молекул с фотонами, когда правило частот Бора не выполняется даже приближенно. Как следует из теории квантовых переходов (см. § 1.3), в первом приближении теории возмущений такое взаимодействие отсутствует. Соответствующие процессы могут происходить лишь во втором и более высоких порядках теории возмущений, т. е. они имеют существенно меньшую вероятность. Одним из таких процессов является рассеяние света.

Рассеянием света называется явление, при котором распространяющийся в среде направленный световой пучок отклоняется по всевозможным направлениям. Рассеяние света как макроскопическое явление может быть обусловлено различными макроскопическими неоднородностями среды. Подобное рассеяние происходит, например, в так называемых мутных средах — эмульсиях, коллоидных растворах, аэрозолях (туман, дым) и т. д. Эти процессы описываются обычными методами классической оптики.

Будем интересоваться процессами рассеяния света, происходящими на молекулярном уровне в макроскопически однородной среде. Когда частота внешнего излучения далека от частоты линии поглощения, т. е. от резонансной частоты эквивалентного осциллятора, квант энергии все же может быть передан от поля

излучения атому, который перейдет в виртуальное состояние (см. § 1.3). Затем атом из этого промежуточного виртуального состояния, время жизни в котором очень мало и определяется соотношением неопределенности, возвращается в исходное состояние, вновь испуская излучение той же частоты, что и падающее. При этом направление распространения и поляризация испущенной волны (фотона) могут быть другими, чем у падающей волны (фотона). В простейшем случае падающая волна — плоская, а испущенная — сферическая. Этот фундаментальный процесс лежит в основе молекулярного рассеяния. Ему соответствует следующая диаграмма взаимодействия:

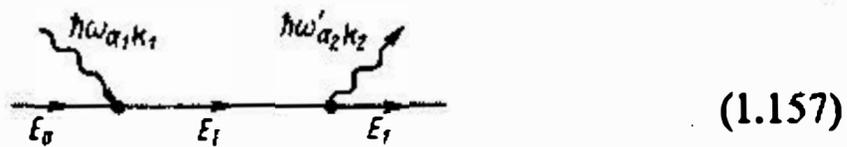


Фотон с энергией $\hbar\omega$, поляризацией α_1 и волновым вектором k_1 взаимодействует с системой (атомом, молекулой), находящейся в состоянии E_0 , переводя ее в виртуальное состояние E_1 . Переходя из этого неустойчивого состояния в начальное, система испускает фотон с энергией $\hbar\omega'$, поляризацией α_2 и волновым вектором k_2 . Как и для любых других процессов, при рассеянии света должны выполняться законы сохранения энергии (1.31) и импульса (1.32). Но они должны соблюдаться для процесса в целом. При переходе в виртуальные состояния, как отмечалось в § 1.3, закон сохранения энергии не соблюдается. Для осуществления такого перехода необходимо лишь чтобы матричный элемент электрон-фотонного взаимодействия $V_{E_1, E_0}^{e, \text{фот}}$ был отличен от нуля.

В оптическом диапазоне импульс фотона $\hbar k$ существенно меньше импульса электрона p , а тем более атома или молекулы. Поэтому из законов сохранения (1.31) и (1.32) для процесса (1.156) получаем: $\hbar\omega = \hbar\omega'$, $\alpha_1 \neq \alpha_2$, $k_1 \neq k_2$ (но $|k_1| \approx |k_2|$). Рассеяние, при котором изменяется только поляризация и направление распространения волны, а энергия фотона (длина волны) остается неизменной, называется *рэлеевским рассеянием*. В рентгеновской области спектра и в области гамма-излучения импульсом фотона пренебречь нельзя. В этом случае для процесса (1.156) $\hbar\omega \approx \hbar\omega'$, а величина $\Delta\hbar\omega = (\hbar\omega' - \hbar\omega)$ мала и зависит от угла рассеяния, т. е. от угла между k_1 и k_2 . Изменение энергии фотона при этом определяется «эффектом отдачи» аналогично тому, как это происходит при упругих столкновениях двух частиц. Рассеяние, при котором наблюдается малое изменение энергии фотона (длины волны), зависящее от угла рассеяния, называется *комptonовским рассеянием* или *эффектом Комптона*.

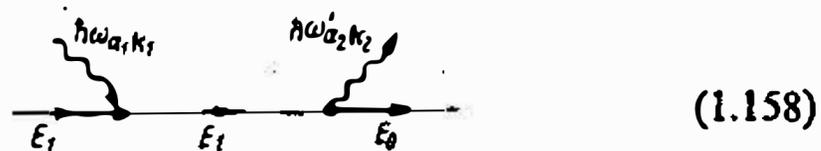
При некоторых условиях атом может перейти из виртуального состояния в конечное, которое отличается от исходного

состояния. Такому процессу соответствует диаграмма взаимодействия



Очевидно, в этом случае испущенный фотон будет отличаться не только по поляризации и направлению распространения, но и по частоте. При этом из закона сохранения энергии имеем $\hbar\omega' = \hbar\omega - (E_1 - E_0)$. Рассеяние света со сравнительно большим изменением энергии фотона (длины волны), не зависящим для изотропных сред от угла рассеяния, называется *комбинационным рассеянием* *. При комбинационном рассеянии частоты рассеянного света представляют собой комбинации (суммы и разности) частот колебаний падающей волны с частотами собственных колебаний рассеивающей системы. Энергетические диаграммы, поясняющие рассмотренные процессы рассеяния, изображены на рис. 1.14.

При комбинационном рассеянии могут быть два случая: 1) энергия начального состояния (обычно основного состояния E_0) меньше энергии конечного (возбужденного) состояния (рис. 1.14, б). При этом частота рассеянного фотона смещена в область меньших частот на $\Delta\omega = (E_1 - E_0)/\hbar$ от частоты падающего фотона. Такое смещение в область меньших частот называется *стоксовым смещением*; 2) энергия начального состояния больше энергии конечного состояния (рис. 1.14, в). Этому процессу соответствует диаграмма взаимодействия



т. е. начальным состоянием является возбужденное состояние E_1 , а конечным — основное состояние E_0 . При этом смещение частоты будет происходить в область больших частот на ту же величину $\Delta\omega = (E_1 - E_0)/\hbar$. Такое смещение называется *антистоксовым смещением* **. Очевидно, интенсивность антистоксовых компонент рассеяния будет в значительной мере определяться населенностью возбужденного состояния, т. е. будет зависеть от

*Эффект комбинационного рассеяния света наблюдался одновременно русскими физиками Г. С. Ландсбергом и Л. И. Мандельштамом в твердых телах и индийским физиком Ч. Раманом в жидкостях. В зарубежной литературе комбинационное рассеяние иногда называют рамановским.

**Названия «стоксово» и «антистоксово» смещения связаны с тем, что первое удовлетворяет правилу, сформулированному Дж. Стоксом еще в XIX в., согласно которому длина волны люминесценции больше длины волны возбуждающего света, а второе — не удовлетворяет этому правилу.

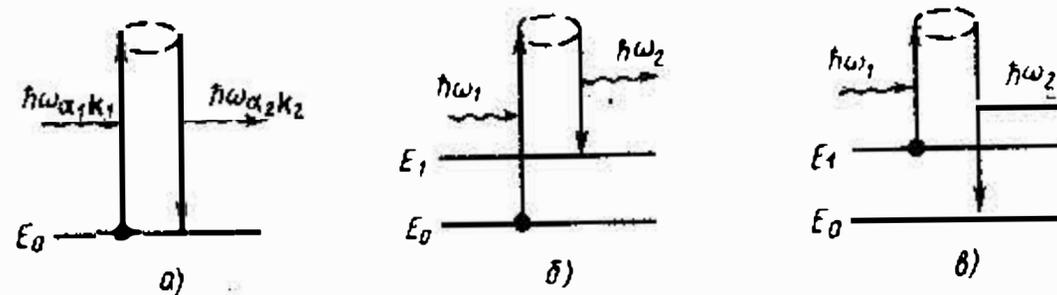


Рис. 1.14. Энергетические диаграммы, поясняющие эффекты рассеяния: а — рэлеевское рассеяние; б, в — комбинационное рассеяние (б — стоксова компонента; в — антистоксова компонента)

температуры. Стоксовы и антистоксовы компоненты в спектре рассеяния будут располагаться симметрично относительно возбуждающей линии. Так как в термодинамически равновесных условиях населенность возбужденного состояния меньше населенности основного состояния, то интенсивность антистоксова рассеяния всегда меньше интенсивности стоксова рассеяния.

Напомним, что вероятности рассмотренных процессов рассеяния довольно малы, поскольку они пропорциональны произведению вероятностей переходов из начального состояния в промежуточное и из промежуточного состояния в конечное, как следует из диаграмм взаимодействия (1.156) — (1.158). Для процессов рассеяния, в том числе и для комбинационного рассеяния, существуют свои правила отбора, причем иные, чем для эквивалентного излучательного перехода. Например, вероятность комбинационного рассеяния света для процесса, изображенного на рис. 1.14, б и на диаграмме взаимодействия (1.157), не будет зависеть от матричного элемента перехода $V_{E_1, E_0}^{e-фот}$ между состояниями E_1 и E_0 . Согласно диаграмме (1.157) она будет определяться лишь произведением матричных элементов $V_{E_1, E_0}^{e-фот} V_{E_1, E_1}^{e-фот}$. Поэтому комбинационное рассеяние света является удобным методом изучения состояний, непосредственные оптические переходы в которые запрещены правилами отбора.

Рассмотренные процессы рассеяния следует отличать от двухступенчатых процессов, при которых атом, поглощая фотон, сначала переходит в разрешенное (т. е. реальное, а не виртуальное) состояние, а затем, испуская или снова поглощая фотон, — во второе разрешенное состояние. Эти процессы ответственны за люминесценцию и двухступенчатое поглощение соответственно.

§ 1.8. ОПТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВЕЩЕСТВА

При рассмотрении распространения электромагнитного излучения в веществе нас обычно не интересует каждый из элементарных актов взаимодействия в отдельности. Интерес представ-

ляет реакция среды в целом. В этом случае естественно рассматривать среду как континуум и воспользоваться макроскопической электромагнитной теорией Максвелла. Материальные уравнения (1.4), как говорилось в § 1.1, как раз и учитывают реакцию среды на электромагнитное возмущение.

Рассмотрим однородную изотропную среду. В оптическом диапазоне частот для большинства применяемых материалов относительную магнитную проницаемость μ можно положить равной единице, а объемную плотность зарядов для однородного материала $\rho = 0$. В этом случае соотношения (1.4) переписываются в виде

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}; \quad \mathbf{B} = \mu_0 \mathbf{H}; \quad \mathbf{j} = \sigma \mathbf{E}, \quad (1.159)$$

где относительная диэлектрическая проницаемость ϵ , и удельная электрическая проводимость σ являются скалярными величинами. Отметим, что здесь σ есть проводимость на оптической частоте ω , которая отнюдь не равна проводимости на низкой или нулевой частоте.

Подставляя (1.159) в уравнения Максвелла (1.3) и выполняя преобразования, аналогичные проведенным в § 1.1, получаем волновое уравнение, например, для вектора \mathbf{E}

$$\Delta \mathbf{E} - \sigma \mu_0 \frac{\partial \mathbf{E}}{\partial t} - \epsilon_0 \epsilon_r \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0. \quad (1.160)$$

Оно отличается от волнового уравнения (1.8) для электромагнитной волны в вакууме дополнительным вторым слагаемым и множителем ϵ , в третьем слагаемом. Дополнительное слагаемое

$-\sigma \mu_0 \frac{\partial \mathbf{E}}{\partial t}$ пропорционально первой производной по времени и учитывает затухание в среде. Оно является аналогом «трения» для колебательных процессов.

Решения уравнения (1.160) хорошо известны и для одной из компонент вектора \mathbf{E} могут быть записаны в виде

$$E_x = E_0 e^{i\omega(t - Az/c)} \quad (1.161)$$

Это выражение описывает волну, распространяющуюся в направлении z с постоянной распространения \hat{n} , которая в общем случае является комплексной величиной и называется комплексным показателем преломления.

Решение (1.161) удовлетворяет уравнению (1.160) при условии

$$(\hat{n})^2 = c^2 (\epsilon_0 \epsilon_r \mu_0 - i \sigma \mu_0 / \omega). \quad (1.162)$$

Разобьем \hat{n} на вещественную и мнимую части:

$$\hat{n} = n - i\chi. \quad (1.163)$$

Подставляя (1.163) в (1.161), получаем

$$E_x = E_0 e^{-\frac{\omega \chi z}{c}} e^{i\left(\omega t - \frac{\omega n z}{c}\right)} \quad (1.164)$$

или то же самое в другой форме записи

$$E_x = E_0 e^{-\gamma z/2} \cos(\omega t - \beta z),$$

где $\gamma/2 = \omega \chi / c$; $\beta = \omega n / c$.

Взяв решение (1.164) и подставив его во второе уравнение Максвелла (1.4) с учетом (1.159), получим решение для \mathbf{H} ,

$$H_y = H_0 e^{-\frac{\omega \chi z}{c}} e^{i\left(\omega t - \frac{\omega n z}{c}\right)}. \quad (1.165)$$

Величины H_0 и E_0 связаны между собой

$$H_0 = \frac{\hat{n}}{c \mu_0} E_0. \quad (1.166)$$

Анализ выражений (1.164) — (1.165) дает ясное представление о свойствах распространяющихся волн. Они описывают волну с угловой частотой ω , распространяющуюся с фазовой скоростью c/n и испытывающую ослабление. Первый множитель в (1.164) — (1.165) определяет амплитуду волны в точке $z=0$, второй — затухание волны с расстоянием z , а третий — фазовый множитель.

Наглядная картина распространения плоской волны может быть получена, если построить зависимости (1.164) — (1.165) как функции координаты z при различных фиксированных

моментах времени t . На рис. 1.15 такая зависимость изображена для момента времени $t=0$.

На практике прибор измеряет всегда не амплитуду и фазу волны, а ее интенсивность, энергию или мощность, которые пропорциональны $|E|^2$. Средний поток мощности через единицу площа-

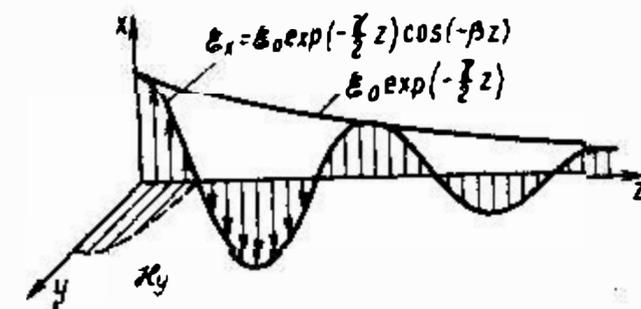


Рис. 1.15. «Мгновенная фотография» плоской линейно поляризованной волны, распространяющейся в поглощающей среде

ди, определенный ранее в § 1.1 как интенсивность излучения I , есть действительная часть вектора Пойнтинга (1.25). Используя выражения (1.164) для E и (1.165) для H , выражение для интенсивности волны можно записать в виде

$$I(z) = I_0 e^{-\frac{2\omega\chi}{c}z} = I_0 e^{-k_\omega z}, \quad (1.167)$$

где в I_0 входят все сомножители, стоящие перед $\exp(-k_\omega z)$. Очевидно, I_0 есть интенсивность волны в точке $z=0$.

Из (1.164), (1.165), (1.167) видно, что коэффициенты χ , γ и k_ω характеризуют свойства среды ослаблять (поглощать) электромагнитное излучение. Безразмерный коэффициент χ , введенный в (1.163) как мнимая часть комплексного показателя преломления, называется *главным показателем поглощения* или *коэффициентом экстинкции*. Он связан с введенным в (1.167) коэффициентом k_ω соотношениями

$$k_\omega = \frac{2\omega\chi}{c} = \frac{4\pi\nu}{c} \chi = \frac{4\pi}{\lambda_0} \chi, \quad (1.168)$$

вытекающими из сравнения (1.167) и (1.164). Этот коэффициент k_ω называют *натуральным показателем поглощения* на частоте ω или просто *показателем поглощения**. Он наиболее часто используется для характеристики поглощающих способностей среды.

Экспоненциальный закон уменьшения интенсивности излучения в поглощающей среде (1.167) есть не что иное, как хорошо известный закон Бугера — Ламберта. Он может быть получен весьма просто следующим образом.

Пусть в поглощающей среде в направлении z распространяется плоская волна с интенсивностью $I(z)$, как показано на рис. 1.16. Выделим слой толщиной dz . Очевидно, изменение интенсивности волны $dI(z)$ при прохождении этого слоя пропорционально его толщине dz , интенсивности света $I(z)$ и коэффициенту пропорциональности k_ω

$$-dI(z) = k_\omega I(z) dz. \quad (1.169)$$

Знак «-» указывает на ослабление волны. Элементарное интегрирование этого уравнения с граничным условием $I(0) = I_0$ дает закон изменения интенсивности

*Часто коэффициент k_ω называют коэффициентом поглощения, а показателем поглощения называют безразмерный коэффициент χ (главный показатель поглощения). В настоящей книге мы будем придерживаться терминологии, установленной ГОСТ 7601—78 «Оптическая оптика. Термины и определения» и коэффициент k_ω называть показателем поглощения, а χ — главным показателем поглощения.

$$I(z) = I_0 e^{-k_\omega z}, \quad (1.170)$$

в точности совпадающей с (1.167).

Из (1.169) видно, что показатель поглощения k_ω представляет собой относительное уменьшение интенсивности излучения dI/I на бесконечно малой длине dz , отнесенное к этой длине

$$k_\omega = -(dI/I)/dz. \quad (1.170a)$$

Он численно равен обратному расстоянию, на котором интенсивность уменьшается в e раз. Обычно k_ω измеряется в см^{-1} .

Величина, обратная показателю поглощения, равна средней длине пробега фотона в веществе

$$l_\phi = (k_\omega)^{-1}.$$

Естественно, что показатель поглощения k_ω (как и главный показатель поглощения χ) зависит от частоты (длины волны) электромагнитного излучения. Об этом напоминает нижний индекс у k_ω . Эта зависимость определяет спектр оптического поглощения вещества.

Довольно часто оптическое поглощение на данной частоте (длине волны) вызвано присутствием в веществе частиц определенного сорта. Например, в газе или смеси газов это могут быть те или иные атомы. При отсутствии сильного взаимодействия между ними показатель поглощения k_ω пропорционален их концентрации N . Поэтому для характеристики способности одного центра поглотить фотон с энергией $\hbar\omega$ вводят понятие *оптического сечения поглощения* σ_ω

$$\sigma_\omega = k_\omega / N. \quad (1.171)$$

Эта величина имеет размерность площади и иногда называется *поперечником поглощения*. Ее физический смысл вытекает из размерности: атом как бы заменяется непрозрачной мишенью площадью σ . Чем больше размер этой «мишени», тем больше вероятность того, что фотон частоты ω поглотится.

Наряду с сечением поглощения σ_ω характеризующим способность каждой из частиц поглощать электромагнитную энергию на частоте ω , вводят понятие *интегрального сечения поглощения* χ , характеризующее способность каждой из частиц поглощать электромагнитную

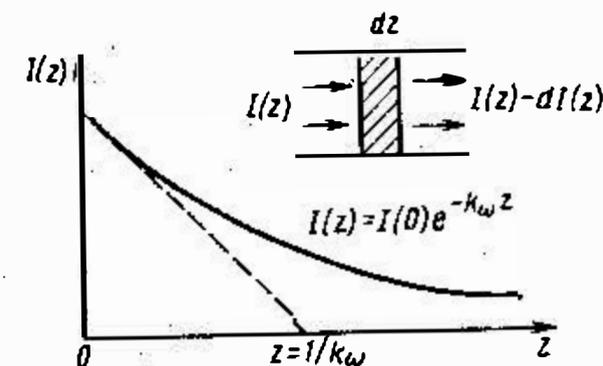


Рис. 1.16. Изменение интенсивности электромагнитной волны в поглощающей среде

энергию в пределах характерного для неё контура спектральной линии

$$\chi = \int \sigma_{\omega} d\omega = \frac{1}{N} \int k_{\omega} d\omega. \quad (1.172)$$

Интегрирование проводится по контуру $g(\omega)$. Поскольку χ определяется интегральной вероятностью поглотить фотон атомом или молекулой в заданной спектральной области, то она, очевидно, связана с интегральным коэффициентом Эйнштейна B_{nm} , в то время как сечение поглощения σ_{ω} должно быть связано с дифференциальным коэффициентом Эйнштейна $b_{nm}(\omega)$.

Наряду с комплексным показателем преломления \hat{n} оптические свойства вещества можно характеризовать комплексной относительной диэлектрической проницаемостью $\hat{\epsilon}_r$. По определению

$$\hat{\epsilon}_r = (\hat{n})^2 = (n - i\chi)^2 = n^2 - \chi^2 - 2in\chi. \quad (1.173)$$

С учетом соотношения $\epsilon_0 \mu_0 = 1/c^2$ из (1.162) получаем

$$\hat{\epsilon}_r = \epsilon - i \frac{\sigma}{\epsilon_0 \omega}. \quad (1.174)$$

Разобьем $\hat{\epsilon}_r$ по аналогии с (1.163) на вещественную и мнимую части:

$$\hat{\epsilon}_r = \epsilon_1 - i\epsilon_2. \quad (1.175)$$

Сравнение соотношений (1.173), (1.174) и (1.175) дает простую связь между n и χ , с одной стороны, вещественной ϵ_1 и мнимой ϵ_2 частями комплексной относительной диэлектрической проницаемости — с другой

$$\epsilon_1 = \epsilon_r = n^2 - \chi^2, \quad (1.176)$$

$$\epsilon_2 = \frac{\sigma}{\epsilon_0 \omega} = 2n\chi. \quad (1.177)$$

Введение \hat{n} или $\hat{\epsilon}_r$ равноценно по своей общности, т. е. оптические характеристики, называемые также оптическими константами n и χ , а также ϵ_1 и ϵ_2 являются равноценными макроскопическими параметрами, определяющими взаимодействие электромагнитной волны с веществом. Но n и χ могут быть измерены независимо и непосредственно оптическими методами. Поэтому в оптическом диапазоне их используют чаще, чем ϵ и σ .

Таким образом, мнимая часть χ комплексного показателя преломления характеризует способность среды поглощать (или усиливать при определенных условиях) электромагнитное излуче-

ние. На микроскопическом уровне процессы поглощения определяются квантовыми переходами из более низкого в более высокое энергетическое состояние.

Действительная часть n комплексного показателя преломления определяет скорость распространения электромагнитной волны в среде: $v = c/n$. Подчеркнем, что в общем случае скорость v равна c/n , а не $c/\sqrt{\epsilon_r}$, поскольку в проводящей (поглощающей) среде $n^2 \neq \epsilon_r$, как видно из соотношения (1.176). В непроводящей (непоглощающей) среде проводимость σ равна нулю, величина \hat{n} вещественная и отождествляется с «обычным» показателем преломления среды n .

Может возникнуть вопрос, каким образом происходит уменьшение скорости света в веществе и какими элементарными процессами на микроскопическом уровне обусловлено это явление. Ответить на такой вопрос довольно трудно, но в общих чертах физическая природа возникновения показателя преломления сводится к следующему. Электрическое поле световой волны поляризует молекулы вещества, создавая осциллирующие дипольные моменты. В свою очередь осциллирующие заряды, движущиеся с ускорением, излучают новое электромагнитное поле, которое, интерферируя со старым, изменяет его. Изменение поля эквивалентно тому, что происходит фазовый сдвиг (запаздывание) первоначальной волны. Вследствие того, что суммарный фазовый сдвиг пропорционален толщине материала, эффект в целом оказывается эквивалентным изменению фазовой скорости света в веществе.

Коэффициент пропорциональности, связывающий индуцированный дипольный момент с возмущающим электрическим полем световой волны, зависит от частоты. В макроскопическом масштабе это приводит к частотной зависимости показателя преломления, которая определяет дисперсию показателя преломления, т. е. зависимость $n(\omega)$ или $n(\lambda)$. Очевидно, наиболее сильной эта зависимость будет вблизи резонансных частот, когда частота изменения электромагнитного поля совпадает с собственными частотами колебаний атомов или молекул (или близка к ним). Но на резонансной частоте, как известно, максимально также поглощение. Это означает, что наиболее сильная дисперсия показателя преломления будет вблизи областей интенсивного поглощения.

Из сказанного ясно, что оптические константы n и χ не являются полностью независимыми. Связь между ними имеет довольно глубокую основу и может быть выведена из принципа причинности. Эта связь устанавливается соотношением Крамерса — Кронига

$$n(\omega) - 1 = \frac{2}{\pi} \int_0^{\infty} \frac{\omega'}{\omega'^2 - \omega^2} \chi(\omega') d\omega', \quad (1.178)$$

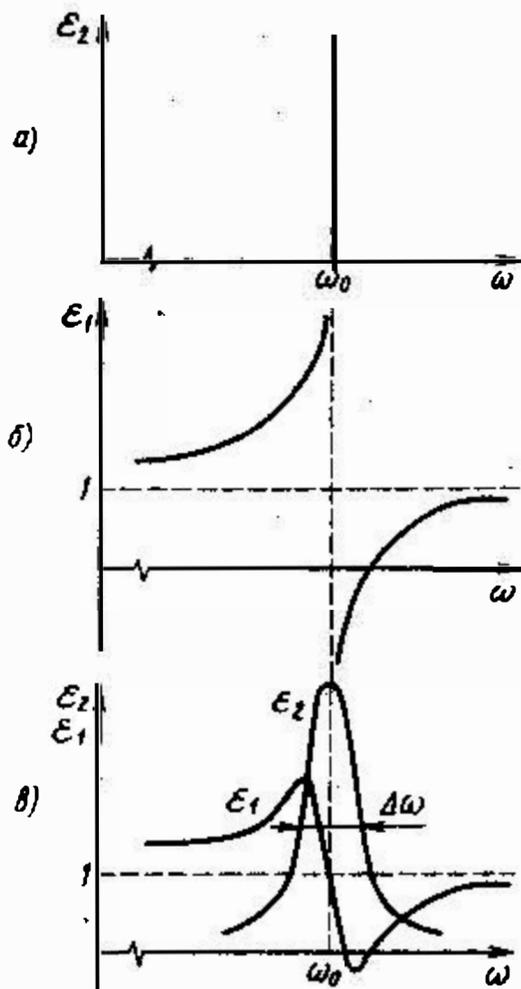


Рис. 1.17. Действительная ϵ_1 и мнимая ϵ_2 части комплексной диэлектрической проницаемости как функции частоты для классического осциллятора без затухания (а, б) и при наличии затухания (в)

из которого видно, что $n(\omega)$ полностью определяется зависимостью $\chi(\omega)$ (и наоборот). Следовательно, если одна из величин, например $\chi(\omega)$, известна во всем диапазоне частот от 0 до ∞ , то другую можно вычислить для любой частоты.

Аналогичным образом связаны между собой вещественная и мнимая части комплексной диэлектрической проницаемости

$$\epsilon_1(\omega) - 1 = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' \epsilon_2(\omega')}{(\omega')^2 - \omega^2} d\omega' \quad (1.179)$$

Отметим, что в соотношениях (1.178) и (1.179) имеется в виду главное значение интеграла.

В качестве примера рассмотрим оптические свойства гармонического осциллятора с собственной частотой ω_0 . Как указывалось в § 1.5, многие оптические явления в атомах и молекулах могут быть смоделированы с помощью набора таких осцилляторов. Поэтому рассматриваемая задача может служить своего рода моделью для определения свойств более сложных объектов.

Для начала рассмотрим идеальный осциллятор с бесконечно малым затуханием $\gamma \approx 0$ [см. (1.127), (1.135)]. Это соответствует бесконечно узкой спектральной линии, т. е. переходу между бесконечно узкими уровнями энергии E_1 и E_2 . В этом случае спектральная зависимость ϵ_2 может быть представлена δ -функцией на резонансной частоте ω_0 , как показано на рис. 1.17, а:

$$\epsilon_2 = G \delta(\omega/\omega_0 - 1). \quad (1.180)$$

Подставляя (1.180) в (1.179) и интегрируя, получаем

$$\epsilon_1 - 1 = n^2 - \chi^2 - 1 = \frac{2}{\pi} G \frac{\omega_0^2}{\omega_0^2 - \omega^2}. \quad (1.181)$$

Эта функция приведена на рис. 1.17, б. Сильная дисперсия диэлектрической проницаемости ϵ_1 и показателя преломления n прояв-

ляется вблизи резонансной частоты ω_0 , а при $\omega = \omega_0$ наблюдается разрыв. При $\omega \rightarrow \infty$ $\epsilon_1 \rightarrow 1$, а при $\omega \rightarrow 0$ $\epsilon_1(0) = \left(\frac{2}{\pi} G + 1\right) > 1$. Это означает, что при $\omega \gg \omega_0$ осциллятор с резонансной частотой ω_0 не вносит вклада в поляризуемость вещества.

Поскольку при всех $\omega \neq \omega_0$ $\epsilon_2 = 0$ и $\chi = 0$, то (1.181) можно переписать в виде

$$n^2 = 1 + \frac{2}{\pi} G \frac{\omega_0^2}{\omega_0^2 - \omega^2} \quad (1.181a)$$

или

$$n^2 = 1 + A \frac{\lambda_0^2}{\lambda^2 - \lambda_0^2}, \quad (1.181б)$$

где $A = \frac{2}{\pi} G$.

Последнее соотношение известно как формула Зельмейера. Ею часто пользуются для аппроксимации спектральной зависимости показателя преломления. В газах, где $n \approx 1$, раскладывая n^2 в ряд, из (1.181б) получаем

$$n(\lambda) = 1 + A \frac{\lambda_0^2}{\lambda^2 - \lambda_0^2}. \quad (1.181в)$$

Разрыв в диэлектрической проницаемости ϵ_1 при $\omega = \omega_0$ обусловлен тем, что в (1.181) не учитывалось затухание и $\epsilon_2(\omega)$ было представлено δ -функцией в виде (1.180). Учет затухания для гармонического осциллятора (см. § 2.5) приводит к тому, что δ -функцию надо заменить функцией Лоренца (1.135), (1.148). Подставляя $\epsilon_2(\omega)$ в виде (1.148) в (1.179), получим результаты, представленные на рис. 1.17, в. Отметим, что внутри спектральной линии, т. е. в области шириной $\Delta\omega$ вблизи резонансной частоты ω_0 , дисперсия $\frac{1}{\epsilon_1} \frac{d\epsilon_1}{d\omega} < 0$ и $\frac{1}{n} \frac{dn}{d\omega} < 0$. Спектральная область, где

$\frac{1}{n} \frac{dn}{d\omega} < 0$, называется областью аномальной дисперсии, а где $\frac{1}{n} \frac{dn}{d\omega} > 0$ — областью нормальной дисперсии. Из рис. 1.17 видно, что аномальная дисперсия наблюдается в области интенсивного поглощения.

ми (1.95) и (1.155a). Сечение слоя в плоскости, перпендикулярной оси z , для простоты примем равным единице.

Поскольку спонтанное излучение ненаправленно, его вкладом в изменение интенсивности I_ω пренебрежем. Тогда получим изменение интенсивности монохроматической волны в слое толщиной dz , обусловленное как процессами вынужденного излучения, так и процессами поглощения

$$dI_\omega = \hbar\omega [N_2 b_{21}(\omega) - N_1 b_{12}(\omega)] \rho(\omega) dz = \frac{\hbar\omega n}{c} g(\omega) [N_2 B_{21} - N_1 B_{12}] I_\omega dz.$$

Здесь учтена связь (1.155) между коэффициентами $b_{mn}(\omega)$ и B_{mn} через форму спектральной линии $g(\omega)$. Величина $v = c/n$ есть скорость распространения света в среде с показателем преломления n . Принимая во внимание связь между коэффициентами Эйнштейна для вынужденных переходов с поглощением и испусканием (1.102), получаем окончательно

$$dI_\omega = \frac{\hbar\omega n B_{12} g(\omega)}{c} \left(\frac{g_1}{g_2} N_2 - N_1 \right) I_\omega dz. \quad (2.1)$$

Отсюда видно, что если населенность нижнего уровня N_1/g_1 больше населенности верхнего уровня N_2/g_2 , то изменение интенсивности электромагнитной волны при прохождении через слой вещества толщиной dz будет отрицательным, т. е. волна будет ослабляться. Наоборот, если населенность верхнего уровня N_2/g_2 больше населенности нижнего уровня N_1/g_1 , то $dI_\omega > 0$ и волна будет усиливаться. Состояние вещества, при котором населенность верхнего энергетического уровня превышает населенность нижнего уровня, называется *инверсией населенностей*.

Таким образом, среда, в которой создана инверсия населенностей между уровнями E_2 и E_1 , способна усиливать электромагнитное излучение с частотой $\omega_{21} = (E_2 - E_1)/\hbar$. Если статистические веса уровней 1 и 2 равны, т. е. $g_1 = g_2$, то условием инверсии будет $N_2 > N_1$. Инверсия населенностей является необходимым, но недостаточным условием для получения усиления в среде. Достаточным условием будет превышение усиления, достигнутого за счет процессов вынужденного испускания, над всеми возможными потерями. Если частота перехода ω_{21} лежит в оптическом диапазоне, то соответствующий усилитель называется *лазерным*, если в СВЧ-диапазоне — *мазерным*.

Обозначим коэффициент пропорциональности в (2.1), стоящий перед $I_\omega dz$, через $-k_\omega$

$$k_\omega = \frac{\hbar\omega n B_{12} g(\omega)}{c} \left(N_1 - \frac{g_1}{g_2} N_2 \right). \quad (2.2)$$

ГЛАВА 2

УСИЛЕНИЕ И ГЕНЕРАЦИЯ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

Из § 1.4 непосредственно следует, что при вынужденном испускании появляется фотон, неразличимый с фотоном, индуцировавшим соответствующий квантовый переход. Этот процесс может служить основой для усиления и генерации электромагнитного излучения, т. е. для создания мазеров и лазеров.

§ 2.1. ПРИНЦИП РАБОТЫ КВАНТОВЫХ УСИЛИТЕЛЕЙ И ГЕНЕРАТОРОВ

Рассмотрим прохождение плоской монохроматической волны с частотой ω через среду, в единице объема которой N_1 атомов находятся в энергетическом состоянии E_1 , а N_2 атомов — в состоянии E_2 . Пусть для определенности $E_2 > E_1$ и в пределах спектральной линии $\hbar\omega = E_2 - E_1$. Будем считать, что волна имеет интенсивность $I_\omega = \rho(\omega)v$ и распространяется в направлении z со скоростью v , как показано на рис. 2.1.

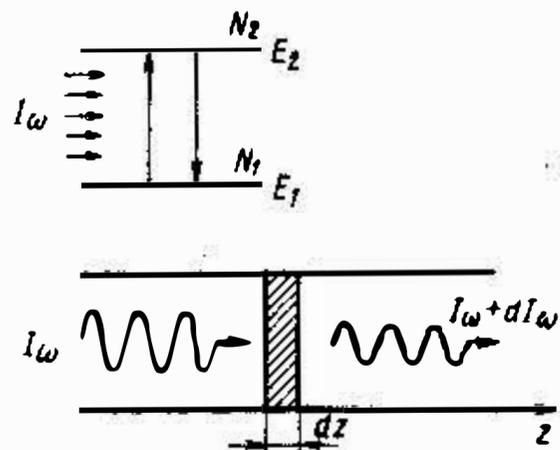


Рис. 2.1. Пролетание плоской электромагнитной волны через слой вещества толщиной dz

Определим изменение интенсивности волны dI_ω при прохождении через слой вещества толщиной dz . Число квантов, поглощенных в этом слое за счет переходов $E_1 \rightarrow E_2$, будет равно $N_1 b_{12}(\omega) \rho(\omega) dz$, где $b_{12}(\omega)$ — дифференциальный коэффициент Эйнштейна (1.155,б). Число квантов, испущенных за счет индуцированных переходов $E_2 \rightarrow E_1$, будет равно $N_2 b_{21}(\omega) \rho(\omega) dz$, где $b_{21}(\omega)$ определяется соотношения-

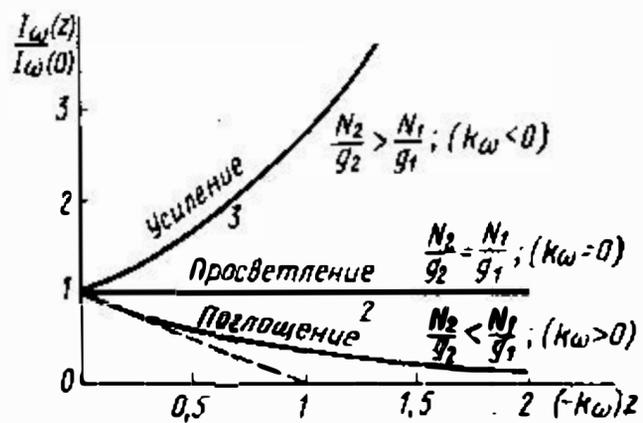


Рис. 2.2. Изменение интенсивности света при прохождении через поглощающую (усиливающую) среду

При $N_1/g_1 > N_2/g_2$ показатель поглощения положителен и интенсивность волны уменьшается с расстоянием по экспоненте, как показано на рис. 2.2 (кривая 1), в полном соответствии с рис. 1.16.

При инверсии населенностей, т. е. при $N_1/g_1 < N_2/g_2$, показатель поглощения отрицателен и интенсивность волны возрастает по экспоненциальному закону (см. рис. 2.2, кривая 3) с показателем экспоненты, пропорциональным инверсии

$$\Delta N = N_2/g_2 - N_1/g_1, \quad (2.4)$$

которую иногда называют *степенью инверсии* или *инверсной перенаселенностью*.

Поскольку отрицательное поглощение есть усиление, то отрицательный показатель поглощения называют показателем усиления (точнее, натуральным показателем усиления) и обозначают

$$\alpha_\omega = -k_\omega.$$

Если населенности уровней равны, то $k_\omega = 0$. Этот случай соответствует *просветлению* среды.

Поскольку функция $g(\omega)$ нормирована ($\int g(\omega) d\omega = 1$), то

$$K = \int k_\omega d\omega = \frac{\hbar\omega n B_{12}}{c} \left(N_1 - \frac{g_1}{g_2} N_2 \right). \quad (2.5)$$

Здесь учтено, что ширина спектральной линии мала и в пределах спектральной линии ω изменяется слабо по сравнению с $g(\omega)$, поэтому при интегрировании (2.2) множитель $\hbar\omega$ можно вынести из-под интеграла. Величина $K = \int k_\omega d\omega$ называется *интегральным показателем поглощения* для данного типа переходов.

Тогда (2.1) можно переписать в знакомом для нас по § 1.8 виде

$$-dI_\omega/I_\omega = k_\omega dz,$$

где коэффициент k_ω был назван показателем поглощения. Его связь с коэффициентом Эйнштейна и населенностью уровней устанавливается соотношением (2.2). В случае, если k_ω не зависит от интенсивности I_ω и координаты z , элементарное интегрирование дает экспоненциальный закон изменения интенсивности (1.167) (закон Бугера — Ламберта)

$$I_\omega(z) = I_\omega(0) e^{-k_\omega z}. \quad (2.3)$$

Очевидно, максимальное усиление будет наблюдаться на частоте $\omega = \omega_0$ вблизи максимума спектральной линии. Тогда для лоренцевой формы спектральной линии (1.148) имеем показатель усиления в максимуме спектральной линии при $\omega = \omega_{21} = \omega_0$:

$$\alpha_{\max}(\omega_0) = -k_{\max}(\omega_0) = -\frac{2}{\pi\Delta\omega} K = \frac{2\hbar\omega_0 n B_{12} g_1}{\pi c \Delta\omega} \left(\frac{N_2}{g_2} - \frac{N_1}{g_1} \right), \quad (2.6)$$

где $\Delta\omega$ — ширина линии на уровне 0,5 от максимума, определяемая процессами уширения (см. § 1.6).

Пусть система находится в условиях термодинамического равновесия. Тогда населенности энергетических уровней определяются распределением Больцмана (1.97), так что в обычных условиях всегда $N_2/g_2 < N_1/g_1$, т. е. среда поглощает. Рассмотрим случай, когда уровень 1 является основным и $kT \ll E_2 - E_1$, так что в условиях равновесия $N_2 \ll N_1 \approx N$, где $N = N_1 + N_2$ — полное число частиц в единице объема. Тогда показатель поглощения

$$k_\omega = \frac{\hbar\omega n B_{12}}{c} g(\omega) N \quad (2.7)$$

пропорционален числу поглощающих центров в единице объема. Поэтому величина $\sigma(\omega) = k_\omega/N$ характеризует способность каждой из частиц поглощать электромагнитную энергию на частоте ω . Она была введена в § 1.8 и названа *поперечным сечением поглощения*. Для данного типа переходов $n \rightarrow m$ σ связана с коэффициентами Эйнштейна

$$\sigma_{nm}(\omega) = \frac{\hbar\omega n}{c} b_{nm}(\omega) = \frac{\hbar\omega n}{c} B_{nm} g(\omega). \quad (2.8)$$

Аналогично для *интегрального сечения поглощения* (1.172) имеем

$$\chi = \int \sigma_{nm}(\omega) d\omega = \frac{\hbar\omega n}{c} B_{nm}. \quad (2.9)$$

Таким образом, система, находящаяся в термодинамическом равновесии, не может усиливать электромагнитное излучение. Для получения усиления необходимо создать в среде инверсию населенностей, для чего нужно вывести систему из состояния равновесия, т. е. возбудить. Естественно, что не любое возбуждение и не в каждом веществе приведет к инверсии населенностей. Среду, в которой при определенных условиях может быть создана инверсия населенностей, называют *лазерной* (или *мазерной*) *активной средой*, а соответствующий рабочий элемент усилителя — *активным элементом*. Уровни энергии, между которыми может быть создана инверсия населенностей, называют *рабочими лазерными уровнями энергии*. Процесс возбуждения активной сре-

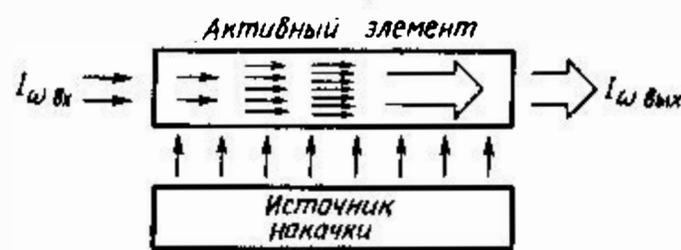


Рис. 2.3. Общая схема квантового усилителя

накачки создает в активном элементе инверсию населенностей между рабочими уровнями E_m и E_n . Входной сигнал $I_{\omega_{вх}}$ на частоте $\omega = (E_m - E_n)/\hbar$, проходя через активный элемент, усиливается за счет процессов индуцированного испускания, так что на выходе получается усиленный сигнал $I_{\omega_{вых}}$. Коэффициент усиления усилителя, равный отношению $I_{\omega_{вых}}/I_{\omega_{вх}}$ (его не надо путать с показателем усиления среды $\alpha_{\omega} = -k_{\omega}$), экспоненциально возрастает согласно (2.3) в соответствии с длиной активного элемента и степенью инверсии $\Delta N = (N_2/g_2 - N_1/g_1)$, которая определяется числом активных частиц в единице объема вещества и интенсивностью накачки.

Состояние вещества с инверсией населенностей иногда называют состоянием с отрицательной температурой (рис. 2.4). Распределение атомов по энергетическим уровням в равновесном состоянии (рис. 2.4, а) определяется статистикой Больцмана. Чисто формально состояние с инверсией населенностей (рис. 2.4, б) можно описать, используя соотношение (1.98), положив $T < 0$. Подчеркнем, что понятие «отрицательная абсолютная температура» указывает лишь на инверсию населенностей и имеет скорее формально математический, нежели физический смысл, поскольку само понятие «температура» применимо только к термодинамически равновесным системам.

Если в активном элементе усилителя (см. рис. 2.3) создана большая степень инверсии ΔN и показатель усиления α_{ω} велик, то

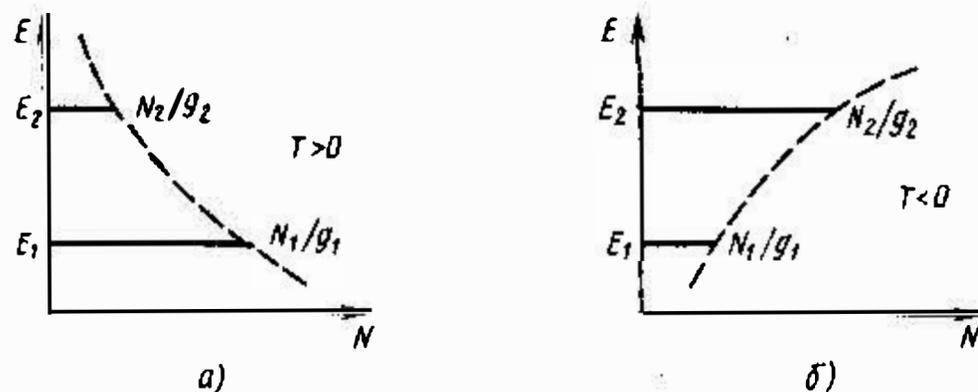


Рис. 2.4. Населенность энергетических уровней: а) в состоянии термодинамического равновесия; б) в состоянии с инверсией населенностей (с отрицательной температурой)

ды (активного вещества) с целью получения инверсии населенностей называют накачкой, а источник этого возбуждения — источником накачки.

Общая схема квантового усилителя представлена на рис. 2.3. Источник

по достижении некоторого критического значения ΔN усилитель может превратиться в генератор: вдоль оси активного элемента будет генерироваться так называемое усиленное спонтанное излучение. Но легче всего усилитель превратить в генератор, введя положительную обратную связь. Напомним, что в электронных усилителях это достигается тем, что часть сигнала с выхода (например, с анода лампы или с коллектора транзистора) подают на вход (например, на управляющую сетку лампы или эмиттер транзистора). Аналогично для превращения квантового усилителя в квантовый генератор следует часть усиленной мощности с выхода подать на вход, осуществляя положительную обратную связь. В лазерах эту функцию выполняют оптические резонаторы: активный элемент размещают между двумя строго параллельными друг другу зеркалами, как показано на рис. 2.5. Плоская электромагнитная волна, распространяющаяся в направлении, перпендикулярном зеркалам резонатора, будет поочередно отражаться от них, усиливаясь при каждом последующем прохождении через активную среду. Одно из зеркал делают полупрозрачным для вывода лазерного излучения. Изменяя коэффициент отражения этого зеркала, можно изменять величину обратной связи.

В квантовых генераторах СВЧ-диапазона (мазерах) обратную связь осуществляют путем помещения активной среды в объемный резонатор.

Как и в обычных генераторах, процесс генерации в лазерах и мазерах развивается из некоторого «затравочного» излучения, которым, как правило, является шум. Шумом, с которого развивается генерация, в нашем случае является спонтанное излучение: всегда найдется «затравочный» фотон, испущенный спонтанно в направлении, близком к направлению, параллельному оси резонатора.

Как и в обычных генераторах, в лазерах и мазерах генерация возможна лишь при выполнении некоторого порогового условия. Необходимо, чтобы усиление за один проход в активном элементе превышало все потери, в том числе потери за счет пропускания одного из зеркал. Так как коэффициент усиления согласно (2.2) зависит от степени перенаселенности рабочих уровней, которая в свою очередь определяется интенсивностью накачки, то существует некоторая критическая (пороговая) мощность накачки, при которой начинается процесс генерации. Пороговой мощности на-

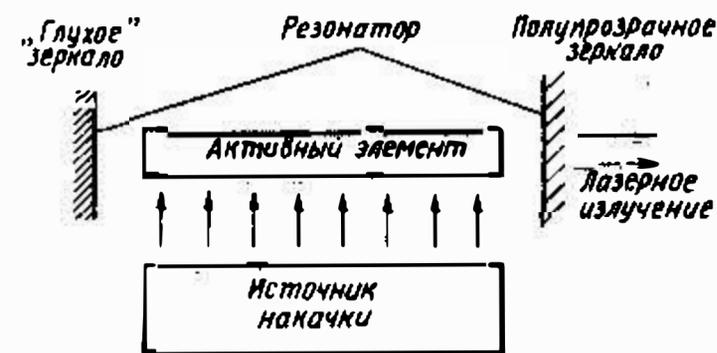


Рис. 2.5. Общая схема лазера

качки соответствует критическая инверсия $N_2/g_2 - N_1/g_1$. Достигнув критической инверсии, генерация разовьется из спонтанного излучения. В отличие от излучения «классических» источников в оптическом диапазоне лазерное излучение: 1) когерентно, поскольку в основе усиления лежит процесс вынужденного испускания; 2) направленно, так как положительная обратная связь максимальна в определенном направлении; 3) монохроматично, ибо излучение определяется переходами в пределах узкой спектральной линии и усиление максимально вблизи максимума спектральной линии, а генерация осуществляется на частотах, близких к собственным частотам резонатора.

Ранее мы предполагали, что показатель поглощения k_ω (или показатель усиления α_ω) не зависит от интенсивности излучения I_ω . Только в этом случае справедлив закон Бугера — Ламберта (1.167), (2.3). Как видно из (2.2), это предположение справедливо до того момента, пока излучение I_ω не вызывает заметных отклонений в распределении частиц N_i по энергетическим уровням, т. е. в случае малых интенсивностей I_ω .

На самом деле излучение I_ω обязательно нарушает равновесие в системе. Если вероятность переходов под влиянием поля излучения (которая пропорциональна I_ω) меньше вероятности релаксационных переходов, то этим нарушением можно пренебречь. В случае, когда эти вероятности становятся сравнимы, происходит изменение населенностей уровней. Под действием мощного излучения на частоте ω_{12} населенности уровней E_1 и E_2 стремятся выравняться и показатель поглощения k_ω (или усиления α_ω) уменьшается. Этот нелинейный процесс, играющий важную роль в развитии и поддержании генерации лазера, приводит к так называемому эффекту насыщения, который будет рассмотрен в § 2.3, 2.4.

§ 2.2. ВОЗБУЖДЕНИЕ АКТИВНОГО ВЕЩЕСТВА (НАКАЧКА)

Усиление электромагнитных колебаний с помощью вынужденного излучения возможно лишь в среде с инверсией населенностей энергетических уровней (см. § 2.1). Такое состояние активного вещества достигается в термодинамически неравновесных системах с помощью накачки. Накачку можно осуществлять различными методами. Перечислим наиболее распространенные из них.

1. Накачка вспомогательным излучением (оптическая накачка). Этот метод является наиболее универсальным и широко используется для накачки твердотельных лазеров на диэлектриках, жидкостных лазеров, может применяться в полупроводниковых и газовых лазерах. Сущность метода состоит в том, что активное вещество облучают мощным электромагнит-

ным излучением, называемым вспомогательным излучением или излучением накачки. Это излучение выбирают так, чтобы оно поглощалось активным веществом, переводя активные центры из основного в возбужденное состояние. В качестве источников накачки в оптическом диапазоне могут служить разнообразные источники света, в том числе обычные лампы накаливания, специальные мощные ксеноновые лампы-вспышки, ртутные лампы, полупроводниковые светодиоды, солнечное излучение и др. В СВЧ-диапазоне источниками накачки могут быть генераторы СВЧ-колебаний, например клистронные или магнетронные генераторы.

2. Накачка с помощью газового разряда. Этот метод применяется в газоразрядных лазерах, где возбуждение активных атомов и молекул осуществляется за счет неупругих столкновений, приводящих к обмену энергией частиц (свободных электронов, атомов, молекул, ионов) в облаке газового разряда.

3. Сортировка частиц. Данный метод используется в приборах СВЧ-диапазона — так называемых пучковых лазерах. Молекулы рабочего вещества, находящиеся в термодинамическом равновесии в основном и возбужденном состояниях ($\Delta E = (E_2 - E_1) \ll kT$), пространственно разделяют так, чтобы в рабочий объем попадали только возбужденные молекулы. Не возбужденные молекулы выводятся из рабочего пучка. Такое разделение возможно вследствие того, что некоторые молекулы, находящиеся в основном и возбужденном состояниях, по-разному реагируют с внешними электрическим и магнитным полями.

4. Инжекция неосновных носителей заряда через p - n -переход. Этот метод применяется в полупроводниковых инжекционных лазерах и позволяет непосредственно, без промежуточных процессов, преобразовывать электрическую энергию источника в когерентное электромагнитное излучение лазера.

5. Возбуждение частицами высоких энергий, например ускоренными электронами. Такой метод используется в полупроводниковых лазерах с электронной накачкой, а также может применяться в других типах лазеров. Пучок предварительно ускоренных электронов (β -частиц) направляется на рабочую мишень из активного вещества, вызывая возбуждение и ионизацию активных центров.

6. Химическая накачка. Данный метод применяется в газовых лазерах. Используется ряд химических реакций, протекающих между газообразными веществами, в результате которых конечный продукт реакции оказывается в возбужденном состоянии.

7. Газодинамическая накачка. Этот метод применяется в газовых лазерах и сводится к тому, что рабочий газ, нагретый до высокой температуры, резко охлаждается. Переходя в равно-

весное состояние, частицы (молекулы) газа задерживаются в наиболее долгоживущих (метастабильных) состояниях, в результате чего может быть достигнута инверсия населенностей.

Для того чтобы определить, при каких условиях можно получить инверсию населенностей между конкретными энергетическими уровнями в активной среде, следует знать кинетику заполнения энергетических состояний при наличии накачки. Для этого необходимо составить и решить соответствующие кинетические уравнения, называемые также уравнениями баланса.

Пусть известен спектр разрешенных энергетических состояний системы и вероятности переходов w_{mn} между любыми уровнями энергии m и n . Тогда изменение числа частиц на уровне E_m выразится соотношением

$$\frac{dN_m}{dt} = \sum_{n \neq m} N_n w_{nm} - N_m \sum_{n \neq m} w_{mn}. \quad (2.10)$$

Здесь первое слагаемое учитывает переход частицы на уровень m со всех остальных уровней n , второе слагаемое — уменьшение (знак минус) населенности m -го уровня за счет переходов из этого состояния во все остальные состояния n . В общем случае вероятность перехода $m \rightarrow n$ может определяться различными излучательными и безызлучательными элементарными процессами. Обозначив скорость безызлучательных переходов S_{mn} , получим

$$w_{mn} = A_{mn} + B_{mn} \rho_{mn} + S_{mn}, \quad (2.11)$$

где A_{mn} и B_{mn} — коэффициенты Эйнштейна для спонтанных и вынужденных излучательных процессов.

Поясним смысл величины $B_{mn} \rho_{mn}$, определяющей вероятность переходов, индуцированных внешней радиацией с плотностью энергии ρ_{mn} . Коэффициент Эйнштейна зависит от частоты: $b_{mn}(\omega)$. Спектр внешней радиации тоже зависит от частоты: $\rho_{mn}(\omega)$. Поэтому в (2.11), как и во многих последующих формулах, вероятность $B_{mn} \rho_{mn}$ представляет собой сокращенную запись интеграла

$$B_{mn} \rho_{mn} = \int b_{mn}(\omega) \rho_{mn}(\omega) d\omega, \quad (2.12)$$

где $\rho_{mn}(\omega)$ — спектральная плотность излучения (интегрирование проводится в пределах спектральной линии).

Если вещество с относительно узкими спектральными линиями облучается электромагнитным излучением широкого спектрального состава, то $\rho_{mn}(\omega)$ можно вынести из-под знака интеграла. Тогда получится обычное выражение для вероятности индуцированных переходов

$$\int b_{mn}(\omega) \rho_{mn}(\omega) d\omega = \rho_{mn}(\omega_{mn}) \int b_{mn}(\omega) d\omega = \rho_{mn}(\omega_{mn}) B_{mn}, \quad (2.13)$$

часто применяемое для расчета вероятностей перехода в канале накачки.

В канале генерации, наоборот, ширина генерируемой линии обычно значительно меньше ширины соответствующей линии поглощения или люминесценции. Поэтому для расчета вероятности перехода в канале генерации можно воспользоваться соотношением

$$\int b_{mn}(\omega) \rho_{mn}(\omega) d\omega = b_{mn}(\omega_r) \int \rho_{mn}(\omega) d\omega \approx b_{mn}(\omega_{ген}) \rho_{mn}(\omega_r) \Delta\omega_{ген}, \quad (2.14)$$

где ω_r — частота, на которой происходит генерация; $\int \rho_{mn}(\omega) d\omega$ — интегральная плотность генерируемого излучения; приближенно она равна произведению плотности энергии в максимуме (т. е. на частоте генерации) на полуширину линии генерации $\Delta\omega_{ген}$.

Так как общее число частиц N в единице объема в стационарных условиях остается неизменным, то

$$\sum_{m=1}^K N_m = N \quad (2.15)$$

(суммирование проводится по всем K энергетическим уровням).

Для каждого уровня может быть записано свое уравнение баланса, так что получится система из K уравнений (2.10), из которых $K-1$ уравнений линейно независимы. Опуская одно из уравнений (2.10) и учитывая (2.15), получим систему из K уравнений, содержащую K неизвестных N_m .

Достигнув стационарного состояния, когда число частиц на каждом из уровней останется неизменным во времени, $dN_m/dt = 0$, и получится система линейных однородных уравнений:

$$N_m \sum_{n \neq m} w_{mn} - \sum_{n \neq m} N_n w_{nm} = 0. \quad (2.16)$$

Эти уравнения совместно с (2.15) можно решить относительно N_m в общем виде с помощью хорошо разработанной в математике теории определителей.

Элементарные процессы, приводящие к образованию инверсии на рабочих уровнях, определяются переходами между рядом энергетических состояний. В большинстве случаев нет необходимости рассматривать возможные переходы между всеми уровнями (число их, как известно, может быть бесконечным). Разумно учесть только те переходы, которые вносят наиболее существенный вклад в изменение населенности рабочих уровней под воздействием внешнего возбуждения (накачки). Более того, при анализе условий возникновения инверсии группу уровней (например,

группу из четырех близко расположенных уровней $2S$ в атоме неона на рис. 1.4) можно рассматривать как один эквивалентный уровень (или полосу) с эффективным временем жизни τ_m . В результате таких упрощений можно говорить о двух-, трех- и четырехуровневой схеме возбуждения активного вещества в зависимости от количества принципиально необходимых энергетических уровней (или групп уровней), участвующих в создании инверсии населенностей. Ей соответствует двух-, трех- и четырехуровневая схема работы квантовых усилителей и генераторов. Очевидно, нижний уровень должен быть основным, а остальные уровни — возбужденными. Рассмотрим особенности указанных схем.

Двухуровневая схема. Рассмотрим систему из двух энергетических уровней E_1 и E_2 , из которых E_1 является основным, т. е. заполненным в условиях термодинамического равновесия. Для простоты будем считать уровни невырожденными, т. е. $g_1 = g_2 = 1$. В такой системе возможны спонтанные и индуцированные оптические переходы, как показано на рис. 2.6, а. Будем осуществлять оптическую накачку за счет процессов поглощения фотонов на частоте перехода ω_{12} . Плотность излучения накачки $\rho_n = \rho_{12}$. Тогда уравнения баланса в стационарном режиме будут иметь вид

$$\left. \begin{aligned} \frac{dN_1}{dt} &= (A_{21} + \rho_{12} B_{21}) N_2 - \rho_{12} B_{12} N_1 = 0, \\ \frac{dN_2}{dt} &= \rho_{12} B_{12} N_1 - (A_{21} + \rho_{12} B_{21}) N_2 = 0, \\ N_1 + N_2 &= N. \end{aligned} \right\} \quad (2.17)$$

Учитывая, что $B_{12} = B_{21}$, находим населенности уровней:

$$N_2 = \frac{\rho_{12} B_{21}}{A_{21} + 2\rho_{12} B_{21}} N, \quad (2.18)$$

$$N_1 = \frac{A_{21} + \rho_{12} B_{21}}{A_{21} + 2\rho_{12} B_{21}} N. \quad (2.18a)$$

Изменение населенности уровней в зависимости от плотности излучения накачки $\rho_n = \rho_{12}$ в двухуровневой схеме представлено на рис. 2.6, б. В отсутствие накачки все частицы находятся в основном состоянии $E_0 = E_1$ (считаем, что $kT \ll E_2 - E_1$). С увеличением интенсивности накачки происходит перераспределение частиц по состояниям E_1 и E_2 . В предельном случае $\rho_n \rightarrow \infty$ населенности

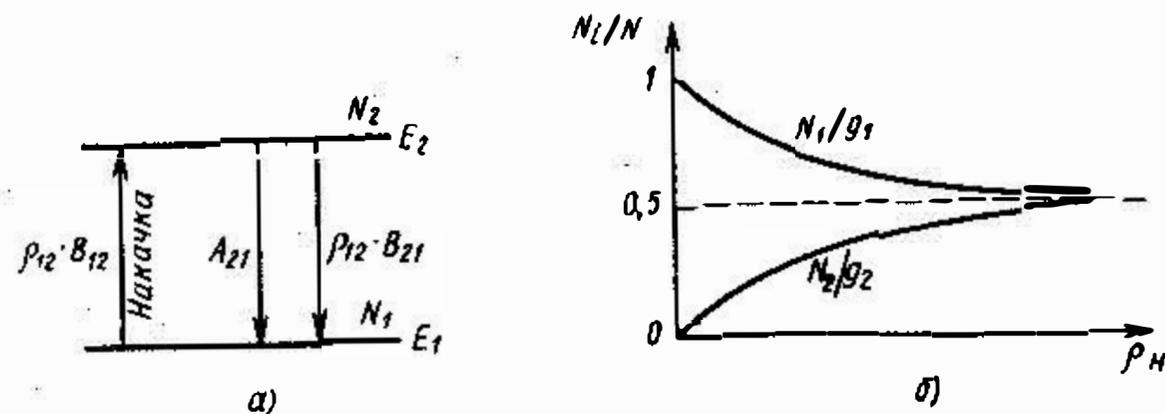


Рис. 2.6. Двухуровневая схема (а) в зависимости относительной населенности уровней от интенсивности накачки (б)

уровней выравниваются. Это явление, называемое *насыщением*, приводит к «просветлению» среды, т. е. при $\rho_n \rightarrow \infty$ показатель поглощения $k_n \rightarrow 0$ (см. кривую 2 на рис. 2.2). Но ни при каких условиях, используя оптическую накачку, в двухуровневой схеме нельзя в стационарных условиях достичь инверсии, а следовательно, и усиления.

Если взять два уровня с разными статистическими весами g_1 и g_2 , то аналогично получим, что в пределе $\rho_n \rightarrow \infty$ число частиц на верхнем уровне $N_2 \rightarrow \frac{g_2}{g_1 + g_2} N$, а на нижнем уровне $N_1 \rightarrow \frac{g_1}{g_1 + g_2} N$.

Хотя при $g_2 > g_1$ число частиц на уровне E_2 может быть больше, чем на уровне E_1 , населенность верхнего уровня N_2/g_2 всегда меньше населенности нижнего уровня N_1/g_1 , т. е. инверсии населенностей и усиления достигнуто не будет.

Полученные результаты однозначно показывают, что по двухуровневой схеме квантовые усилители и генераторы с оптической накачкой работать не могут. Аналогичные результаты получатся, если в системе (2.17) учесть безызлучательные процессы, а также применить накачку с помощью газового разряда, возбуждением частицами высоких энергий или газодинамическую накачку.

Инверсию населенностей в двухуровневой системе можно получить, используя метод сортировки частиц. Именно этот метод привел к созданию первого квантового генератора — мазера на пучке молекул аммиака (см. § 6.1).

Трехуровневые схемы. Предложение использовать для создания инверсии населенностей более сложные трех- и четырехуровневые схемы накачки вызвало дальнейшее развитие квантовой электроники.

Механизм создания инверсии населенностей в трехуровневых схемах поясняет рис. 2.7. В зависимости от того, между какими уровнями достигается инверсия, различают трехуровневые схемы первого и второго типов. В схемах первого типа рабочий переход

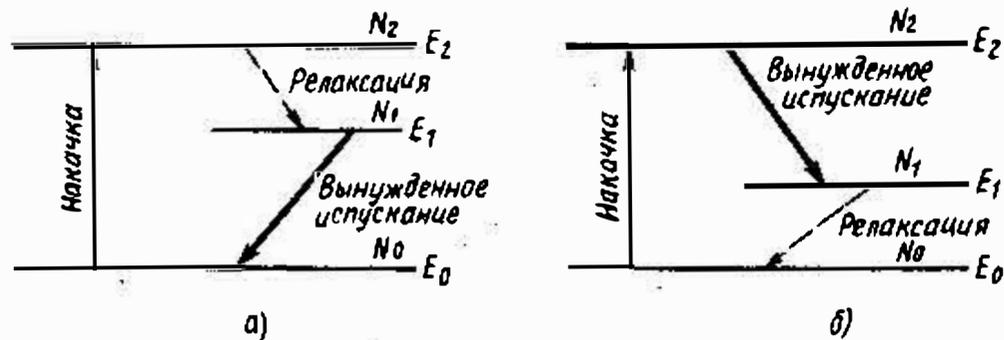


Рис. 2.7. Трехуровневые схемы работы квантовых усилителей и генераторов первого (а) и второго (б) типов

заканчивается в основном состоянии (см. рис. 2.7, а), а в схемах второго типа — в возбужденном (см. рис. 2.7, б). Накачка осуществляется по возможности селективно на уровень E_2 . По трехуровневой схеме первого типа работает рубиновый лазер (см. § 8.2), а по схеме второго типа — гелий-неоновый газовый лазер (§ 7.2).

В трехуровневых схемах канал накачки и канал усиления частично разделены. Это позволяет использовать для достижения инверсии наиболее универсальный метод оптической накачки, а также накачку с помощью газового разряда. Возможность получения инверсии населенностей с помощью оптической накачки в трехуровневой схеме очевидна. Например, если в схеме второго типа осуществить селективный переход $E_0 \rightarrow E_2$, то уровень E_2 окажется инверсно заселенным относительно уровня E_1 (при $kT \ll E_1 - E_0$). Из рис. 2.7 можно заключить, что накопление частиц на верхнем лазерном уровне (E_1 в схеме первого типа и E_2 — в схеме второго типа) будет в том случае, если релаксационные процессы $E_2 \rightarrow E_1$ в схеме на рис. 2.7, а и $E_1 \rightarrow E_0$ в схеме на рис. 2.7, б идут достаточно быстро, а верхний рабочий уровень является метастабильным.

Определим зависимость населенностей уровней от плотности (интенсивности) возбуждающего излучения накачки $\rho_{\pi} = \rho_{02}$. В качестве примера рассмотрим трехуровневую схему первого типа. Предположим, что возбуждение системы (накачка) осуществляется чисто оптическим путем в канале $0 \rightarrow 2$, а внешнее возбуждение в каналах $1 \rightarrow 2$ и $0 \rightarrow 1$ отсутствует (или пренебрежимо мало). Скорость релаксации $2 \rightarrow 1$ обозначим w_{21} . Она может осуществляться за счет излучательных и безызлучательных переходов, так что $w_{21} = A_{21} + S_{21}$. Рассмотрим сначала режим усиления, когда активное вещество не находится в резонаторе. Соответствующие переходы изображены на рис. 2.8, а. Кинетические уравнения (2.10) в этом случае для стационарного режима будут иметь вид

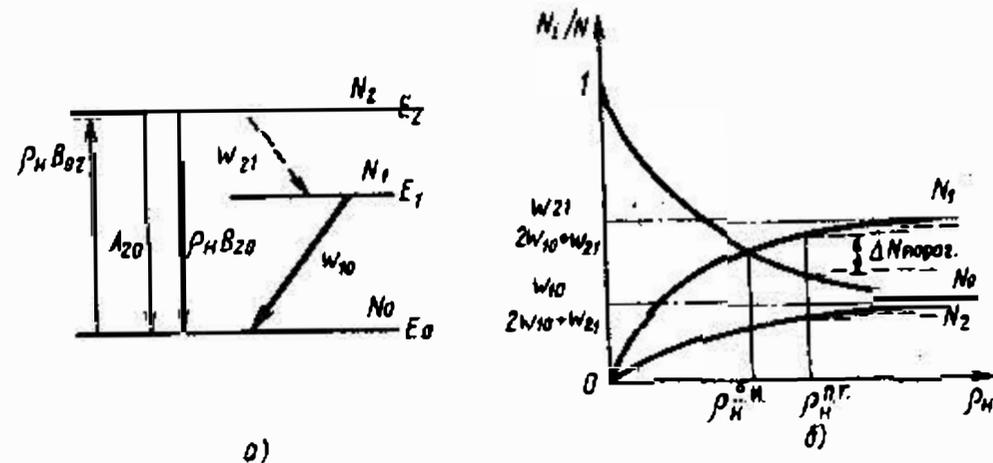


Рис. 2.8. Трехуровневая схема (а) в зависимости относительной населенности уровней от интенсивности накачки (б) при отсутствии (сплошные линии) и наличии (пунктирные) резонатора

$$\left. \begin{aligned} \frac{dN_2}{dt} &= \rho_{\pi} B_{02} N_0 - [\rho_{\pi} B_{20} + (w_{21} + A_{20})] N_2 = 0, \\ \frac{dN_1}{dt} &= w_{21} N_2 - w_{10} N_1 = 0, \\ N_0 + N_1 + N_2 &= N. \end{aligned} \right\} (2.19)$$

Приняв кратности вырождения уровней $q_0 = q_1 = q_2 = 1$ и решая систему уравнений (2.19), находим населенности уровней:

$$N_0 = \left[\frac{w_{10}(A_{20} + w_{21} + \rho_{\pi} B_{20})}{w_{10}(w_{21} + A_{20}) + (2w_{10} + w_{21})\rho_{\pi} B_{20}} \right] N, \quad (2.20)$$

$$N_1 = \left[\frac{w_{21} \rho_{\pi} B_{20}}{w_{10}(w_{21} + A_{20}) + (2w_{10} + w_{21})\rho_{\pi} B_{20}} \right] N, \quad (2.20а)$$

$$N_2 = \left[\frac{w_{10} \rho_{\pi} B_{20}}{w_{10}(w_{21} + A_{20}) + (2w_{10} + w_{21})\rho_{\pi} B_{20}} \right] N. \quad (2.20б)$$

Зависимость относительной населенности уровней N_i/N ($i=0, 1, 2$) от плотности накачки согласно выражениям (2.20) представлена на рис. 2.8, б. При больших плотностях накачки населенности основного и верхнего состояний в пределе $\rho_{\pi} \rightarrow \infty$ стремятся к

$$\lim_{\rho_{\pi} \rightarrow \infty} N_0 = \lim_{\rho_{\pi} \rightarrow \infty} N_2 = \frac{w_{10}}{2w_{10} + w_{21}} N,$$

а населенность уровня E_1 при $\rho_n \rightarrow \infty$ стремится к

$$\lim_{\rho_n \rightarrow \infty} N_1 = \frac{w_{21}}{2w_{10} + w_{21}} N.$$

При $w_{21} > w_{10}$, как видно из рис. 2.8, б и соотношений (2.20), начиная с некоторого значения плотности накачки ρ_n между уровнями E_1 и E_0 будет наблюдаться инверсия населенностей ($N_1 > N_0$). Величина $\rho_n^{\text{пр}}$ называется *пороговой плотностью накачки по инверсии*. С увеличением $\rho_n > \rho_n^{\text{пр}}$ инверсия увеличивается.

Отметим, что пороговая накачка для генерации будет превышать пороговую накачку по инверсии, поскольку для возникновения генерации необходимо выполнить еще условия самовозбуждения (см. § 2.4). Приравняв выражения (2.20) и (2.20а), найдем $\rho_n^{\text{пр}}$:

$$\rho_n^{\text{пр}} = \frac{w_{10}(w_{21} + A_{20})}{B_{20}(w_{21} - w_{10})} \quad (2.21)$$

Из проведенного рассмотрения вытекает, что для накопления частиц на уровне E_1 и создания максимальной инверсии населенностей наиболее выгодны системы с большим значением w_{21} (переход $2 \rightarrow 1$ должен быть быстрым), малым значением w_{10} (уровень E_1 должен быть метастабильным) и большим коэффициентом Эйнштейна B_{02} (оптический переход $0 \rightarrow 2$ должен быть разрешен).

В соотношениях (2.19) и (2.20) вероятность перехода w_{10} считалась постоянной, не зависящей от скорости накачки. Это справедливо в отсутствие генерации, когда опустошением уровня E_1 за счет вынужденных переходов можно пренебречь. После превышения инверсии над некоторым пороговым значением начинает развиваться процесс генерации. Наличие интенсивного излучения на частоте w_{10} и связанных с ним вынужденных переходов $1 \rightarrow 0$ и $0 \rightarrow 1$ вызывает изменение населенности уровней. Возрастание интенсивности накачки ρ_n приводит к увеличению инверсии $\Delta N = N_2 - N_1$ и согласно (2.1) — к увеличению усиления. Это увеличивает ρ_{10} , что в свою очередь обуславливает выравнивание населенностей N_2 и N_0 , т. е. происходит процесс насыщения инверсии населенностей и коэффициента усиления. Увеличение поступления частиц на уровень E_1 , вызываемое ростом накачки, компенсируется возрастанием числа активных переходов $1 \rightarrow 0$. Поэтому в режиме генерации инверсия $\Delta N = N_1 - N_0$ остается приблизительно постоянной, как изображено пунктирными линиями на рис. 2.8, б. Ее значение примерно равно пороговой перенаселенности, при которой усиление превышает потери в генераторе и которая достигается при пороговой накачке, равной $\rho_n^{\text{пр}}$.

Четырехуровневая схема. В этой схеме каналы генерации и накачки полностью разделены, что позволяет получать инверсию населенностей при минимальных уровнях накачки. Очевидно, трехуровневые схемы обоих типов можно рассматривать как частный случай четырехуровневой схемы.

Механизм создания инверсии между рабочими лазерными уровнями E_2 и E_1 в четырехуровневой схеме и переходы, определяющие сущность процесса, изображены на рис. 2.9, а. Так же как и в трехуровневой системе, накачка переводит атомы из основного состояния E_0 в верхнее состояние E_3 . Инверсия достигается между уровнями E_2 и E_1 . Для этого необходимо, чтобы уровень E_2 быстро заполнялся, а уровень E_1 опустошался за счет процессов $E_3 \rightarrow E_2$ и $E_1 \rightarrow E_0$ соответственно. Наличие реально существующих переходов $3 \rightarrow 0$, $3 \rightarrow 1$, $2 \rightarrow 0$, $0 \rightarrow 1$ и $2 \rightarrow 3$ приводит к уменьшению инверсии, т. е. снижает качество усилителя и генератора.

Анализ кинетики заселения уровней в четырехуровневой схеме может быть проведен с помощью кинетических уравнений, аналогично тому, как это было сделано для трехуровневой схемы. В данном случае эти уравнения будут лишь более громоздкими. Зависимость населенности уровней от интенсивности накачки в четырехуровневой схеме представлена на рис. 2.9, б. Инверсия населенностей между уровнями E_2 и E_1 будет достигаться при соблюдении неравенства

$$w_{10} w_{32} > [w_{31}(w_{20} + w_{21}) + w_{21} w_{32}] g_2 / g_1, \quad (2.22)$$

где w_{mn} — вероятности переходов $m \rightarrow n$.

Из рис. 2.9, б видно, что в четырехуровневой схеме (при $kT \ll E_1 - E_0$) ничтожное накопление частиц на третьем уровне приводит к инверсии населенностей, т. е. пороговая плотность излучения накачки по инверсии $\rho_n^{\text{пр}}$ в четырехуровневой схеме

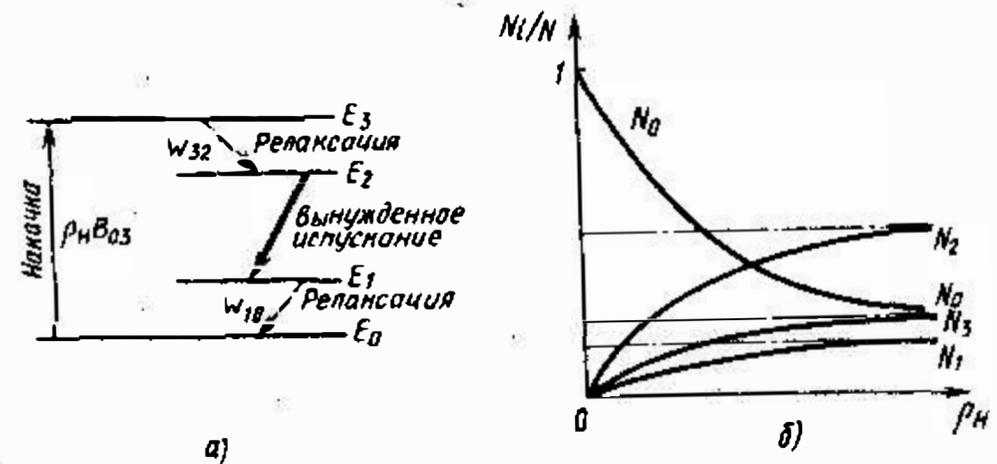


Рис. 2.9. Четырехуровневая схема работы квантовых усилителей и генераторов (а) и распределение населенностей уровней в зависимости от интенсивности накачки (б)

близка к нулю. Пороговая накачка, соответствующая началу генерации — $\rho_n^{кр}$, в активных материалах, работающих по четырехуровневой схеме, больше нуля, но может быть очень мала.

Так же как и в трехуровневой схеме, зависимость населенности уровней в режиме генерации будет отличаться от изображенной на рис. 2.9, б. За счет интенсивных вынужденных переходов с испусканием в канале генерации $2 \rightarrow 1$ значение N_2 будет уменьшаться, а N_1 — возрастать, приводя к насыщению населенностей уровней и показателя усиления.

В заключение отметим следующие моменты, относящиеся как к трехуровневым, так и к четырехуровневым схемам.

а. Для исключения термического заселения необходимо, чтобы энергетические расщепления между уровнями $E_3 - E_2$ и $E_1 - E_0$ были больше kT . Однако они не должны быть слишком большими, поскольку в противном случае большая часть энергии накачки будет расходоваться бесполезно. Это приведет к уменьшению КПД (в пределе $\eta < (E_2 - E_1)/(E_3 - E_0)$) и разогреву активного вещества, поскольку избыточная энергия при релаксационных процессах в конечном итоге выделится в виде тепла.

б. При оптической накачке, когда источник накачки излучает в широкой области спектра, необходимо, чтобы верхний уровень E_3 (или система уровней) был достаточно широким. Это нужно для более полного использования излучения накачки [интеграл (2.12) должен быть максимальным].

в. Для исключения самопоглощения, приводящего к переходам $E_0 \rightarrow E_1$ и $E_2 \rightarrow E_3$, желательно, чтобы релаксационные процессы осуществлялись за счет неоптических (безызлучательных) переходов.

г. Время жизни на верхнем лазерном уровне E_2 должно определяться излучательными процессами, а вероятность безызлучательных переходов с этого уровня должна быть минимальной.

Анализ кинетических уравнений и схем работы квантовых усилителей и генераторов в настоящем параграфе мы для простоты провели для стационарных режимов, полагив в (2.10) $dN_n/dt = 0$. Это справедливо при непрерывном или квазинепрерывном режиме работы лазера, т. е. тогда, когда процессы накачки и генерации продолжают в течение времени, заметно превышающего характерные времена релаксации. Большое значение в лазерной технике и оптоэлектронике имеет также импульсный нестационарный режим. Режим работы считается нестационарным, если длительность воздействия излучения на активную среду мала по сравнению с характерным временем релаксации квантовой системы. Использование нестационарных режимов накачки в ряде случаев позволяет получить инверсию населенностей между такими уровнями энергии, где в стационарных условиях она не может быть достигнута. Примером тому служат

лазеры на самоограниченных переходах, рассмотренные в § 7.1.

Не всякий импульсный режим работы лазера является нестационарным. Если за время импульса успевает установиться равновесие, то такой режим работы будет импульсным, но квазистационарным. Длительность импульса при этом может быть очень мала, достигая 1 нс и менее в полупроводниковых и жидкостных лазерах. В других случаях, напротив, для достижения стационарного режима требуются времена более 1 мс.

§ 2.3. ОПТИЧЕСКИЕ РЕЗОНАТОРЫ

Основные понятия. Важнейшим элементом квантовых генераторов радио- и оптического диапазонов является резонатор (см. общую схему рис. 2.5). За счет многократного отражения в резонаторе достигается наиболее полное взаимодействие электромагнитного излучения с активной средой. Обеспечивая положительную обратную связь, резонатор существенно влияет на свойства генерируемого активной средой излучения, определяя, в частности, его диаграмму направленности и спектральный состав.

В широком смысле *резонатором* называют колебательную систему, в которой возможно накопление энергии акустических, механических или электромагнитных колебаний. Простейшим электромагнитным резонатором, применяемым в радиодиапазоне до частот несколько десятков мегагерц, является колебательный контур. С увеличением частоты размеры контура становятся соизмеримыми с длиной волны, что приводит к резкому возрастанию излучения из контура, и он теряет свои резонансные свойства. Поэтому в СВЧ-радиодиапазоне при частотах порядка $10^9 \dots 10^{11}$ Гц ($\lambda \approx 0,3 \dots 30$ см) применяют объемный резонатор — колебательную систему, представляющую собой полость с проводящими стенками, внутри которой могут возбуждаться электромагнитные колебания. В объемном резонаторе могут возбуждаться колебания только определенных длин волн и определенной структуры, образующие стоячую волну. Частоты этих колебаний называются *резонансными* или *собственными частотами резонатора*, а колебания — *собственными колебаниями резонатора* или *модами*. Таким образом мода — это собственный тип колебаний резонатора. По более точному определению, мода резонатора представляет собой стационарную конфигурацию электромагнитного поля, которая удовлетворяет уравнениям Максвелла и граничным условиям. Электрическое поле такой конфигурации можно записать в виде

$$E(r, t) = E_0 U(r) e^{i\omega t}, \quad (2.23)$$

где ω — собственная частота резонатора.

Например, для объемного резонатора, представляющего собой прямоугольный параллелепипед с размерами ребер L_1 , L_2 и L_3 , длины волн собственных типов колебаний λ_{mq} находят из соотношения

$$\frac{1}{\lambda_{mq}} = \sqrt{\left(\frac{m}{2L_1}\right)^2 + \left(\frac{n}{2L_2}\right)^2 + \left(\frac{q}{2L_3}\right)^2}. \quad (2.24)$$

Здесь m , n и q — целые числа, определяющие число полувольт, укладываемых вдоль сторон L_1 , L_2 и L_3 .

Соотношение (2.24) выражает условие образования стоячей волны при её отражениях от стенок резонатора. Например, для $m=n=0$ имеем $1/\lambda_{00q} = q/2L_3$, или $L_3 = 0,5\lambda q$. На длине L_3 укладывается целое число полувольт, т. е. образуется стоячая волна.

Очевидно, собственные частоты объемного резонатора определяются только его геометрией (в случае незаполненных резонаторов). Отличительной особенностью объемных резонаторов является то, что их размеры соизмеримы с длиной волны. Поэтому число собственных типов колебаний в таком резонаторе мало, а собственные частоты разнесены далеко друг от друга по спектру. Изменяя размеры резонатора, можно изменять (перестраивать) его резонансные частоты.

Электромагнитные волны, распространяющиеся в свободном пространстве, являются поперечными, и векторы E , H , k взаимно перпендикулярны. Электромагнитные колебания в объемном резонаторе не являются поперечными: один из векторов E или H имеет проекцию на направление распространения волны — продольную составляющую. Если продольную составляющую имеет вектор E , то электромагнитное колебание называют электрическим и обозначают E или ТМ. Если продольную составляющую имеет вектор H , то колебание называют магнитным и обозначают H или ТЕ. Каждое колебание, например, в прямоугольном резонаторе характеризуют тремя индексами m , n и q (2.24), соответствующими числу полувольт в направлениях x , y и z . Для цилиндрического объемного резонатора, часто применяемого в СВЧ-диапазоне, индексы m , n и q указывают число полувольт по диаметру, окружности и длине резонатора.

Колебания в объемном резонаторе (который является замкнутым со всех сторон) возбуждают с помощью петлеобразного проводника — *петли связи* или через специальное отверстие (щель). Аналогично обеспечивается вывод электромагнитной энергии наружу.

Кроме собственных частот и собственных типов колебаний объемный резонатор, как и любой другой резонатор, характеризуется потерями энергии (например, тепловыми потерями и потерями на излучение в окружающее пространство). Это свойство

резонатора характеризует добротность. *Добротность резонатора* Q есть отношение энергии, запасенной в резонаторе, к средней энергии, теряемой за $1/2\pi$ часть периода колебаний. Добротность объемных резонаторов в СВЧ-диапазоне имеет порядок 10^3 .

При переходе от радиочастот СВЧ-диапазона к оптическим частотам объемные резонаторы начинают терять свои резонансные свойства: при уменьшении длины волны и неизменных размерах резонатора спектр собственных частот сгущается, так что различные типы колебаний накладываются друг на друга и селективные свойства резонатора исчезают.

Определим число собственных типов колебаний резонатора в единице объема при условии $\lambda \ll L$, т. е. в оптическом диапазоне. Для этого рассмотрим куб с ребром L . Условия образования стоячей волны в кубе запишутся аналогично (2.24):

$$k_x = \pi m/L, \quad k_y = \pi n/L, \quad k_z = \pi q/L,$$

где $k = 2\pi/\lambda$ — волновое число, а m , n и q — целые числа.

Полное число типов колебаний N_k в кубе с ребром L при изменении волнового вектора от 0 до k равно объему сферы радиусом k , деленному на объем, приходящийся на один тип колебаний $(2\pi/L)^3$:

$$N_k = 2 \frac{(4\pi/3)k^3}{(2\pi/L)^3} = \frac{k^3}{3\pi^2} L^3 = \frac{\omega^3}{c^3 3\pi^2} L^3.$$

Множитель 2 учитывает два возможных направления поляризации волны. Отсюда получаем число собственных типов колебаний в единице объема в единичном интервале частот:

$$\frac{1}{V} \frac{dN_k}{d\omega} = \frac{\omega^2}{\pi^2 c^3}. \quad (2.25)$$

Заметим, что именно этот множитель стоит в функции Планка (1.107) и в формуле Рэлея — Джинса (1.101). Оценим число типов колебаний ΔN_k , приходящихся на спектральный диапазон $\Delta\nu = \Delta\omega/2\pi = 1$ ГГц. Величина $\Delta\nu$, как следует из § 1.6, соответствует доплеровскому уширению спектральной линии при комнатной температуре. Приняв $\lambda = 1$ мкм, получим в объеме 1 см^3 $\Delta N_k \approx 10^8$. Из (2.25) определяем, что частотный интервал между ближайшими типами колебаний при тех же условиях имеет значение порядка 10 Гц. Ширина каждой резонансной линии определяется ее добротностью: $\Delta\omega = \omega/Q$. Даже если бы удалось создать резонатор с добротностью 10^6 , то ширина его линии при $\nu \approx 10^{14}$ Гц имела бы порядок 2 МГц. Ясно, что при таких условиях все собственные колебания перекрываются, образуя непрерывный спектр, т. е. в оптическом диапазоне объемный резонатор теряет

свои резонансные свойства. Поэтому оптический резонатор должен быть сконструирован таким образом, чтобы сохранялось лишь небольшое число собственных типов колебаний, а остальные были бы погашены.

Для того чтобы разрядить спектр собственных частот, в оптическом диапазоне вместо объемных резонаторов применяют открытые резонаторы. *Открытым резонатором* называют объемный резонатор, отражающие стенки которого не замкнуты. Резонансными свойствами обладают открытые резонаторы только вполне определенной геометрической конфигурации. Открытый резонатор для оптического диапазона называют *оптическим резонатором*.

Простейший оптический резонатор образуют два плоских зеркала, расположенные строго параллельно друг другу (рис. 2.10). Такой *плоский резонатор**, очевидно, можно рассматривать как открытый резонатор, образованный из прямоугольного объемного резонатора, четыре боковые стенки которого удалены. Прямая OO' , проходящая через центры зеркал перпендикулярно их поверхностям, называется *оптической осью резонатора*. Электромагнитные волны, распространяющиеся вдоль оптической оси, будут отражаться от зеркал и интерферировать между собой. Те из них, для которых выполняется условие резонанса $q\lambda/2 = L$, будут интерферировать на максимум, образуя в резонаторе стоячую волну, как показано на рис. 2.10. За счет многократного прохождения таких волн резко увеличивается эффективность их взаимодействия с рабочим веществом, помещаемым в резонатор. Волны, распространяющиеся под углом (не слишком малым) к оптической оси, после нескольких отражений выходят из резонатора. Этим и достигается разрежение спектра собственных частот открытого резонатора по сравнению с объемным.

Колебания, распространяющиеся строго вдоль оптической оси резонатора, обладают наивысшей добротностью и называются *аксиальными колебаниями*. Определим расстояние $\Delta\nu$ между соседними аксиальными колебаниями при $\lambda = 1$ мкм для резонатора длиной 1 м. На этой длине будет укладываться $2 \cdot 10^6$ полувольт ($q = 2 \cdot 10^6$). Расстояние между ближайшими типами колебаний

$$\Delta\nu_q = \nu_q - \nu_{q-1} = c/\lambda_q - c/\lambda_{q-1} = qc/(2L) - (q-1)c/(2L) = c/(2L). \quad (2.26)$$

Для $L = 1$ м получаем $\Delta\nu_q = 150$ МГц, а для $L = 1$ см $\Delta\nu_q = 1,5 \cdot 10^{10}$ Гц (вместо 10 Гц для объемного резонатора с $L = 1$ см).

* Резонансные свойства системы, состоящей из двух параллельных зеркал, были известны давно в оптике, где подобные системы называются *интерферометрами*. Плоский резонатор иногда называют резонатором Фабри — Перо по имени французских физиков, впервые изучивших плоский интерферометр.

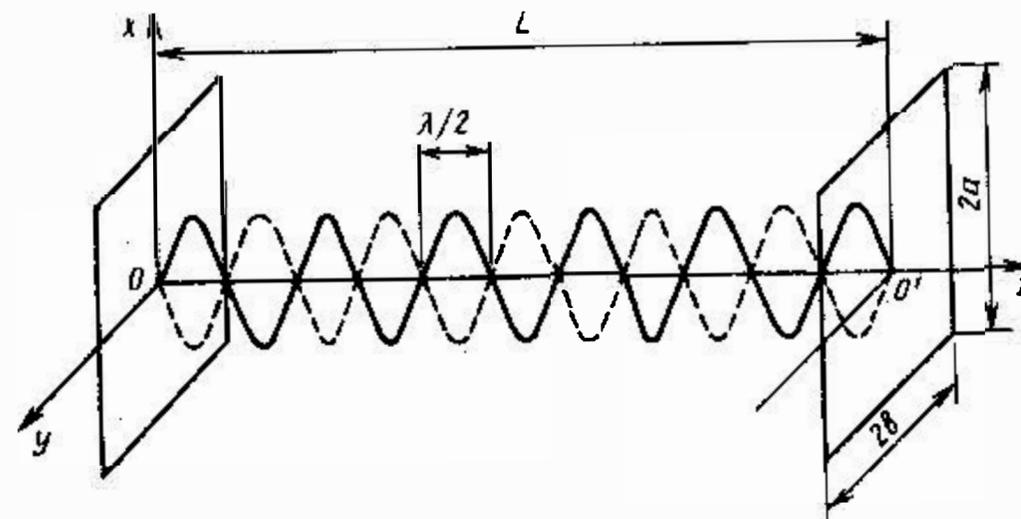


Рис. 2.10. Схема оптического резонатора с плоскими прямоугольными зеркалами (интерферометр Фабри—Перо)

Несмотря на столь существенное разрежение спектра собственных частот, частотный интервал между соседними типами колебаний в оптических резонаторах с размерами $L \approx 0,1 \dots 2$ м значительно меньше ширины спектральной линии даже в газах, где ее ширина минимальна по сравнению с другими веществами. Это означает, что в пределах спектральной линии активной лазерной среды обычно укладывается много (от десятков до десятков тысяч) собственных типов колебаний резонатора (рис. 2.11). Поэтому такой резонатор называется *многомодовым* или *многочастотным*. При этом добротность оптического резонатора существенно выше добротности спектральной линии. Спектр собственных частот в системе (активное вещество + резонатор) в оптическом диапазоне (в лазерах), как видно из рис. 2.11, будет определяться собственными частотами резонатора, лежащими вблизи максимума спектральной линии.

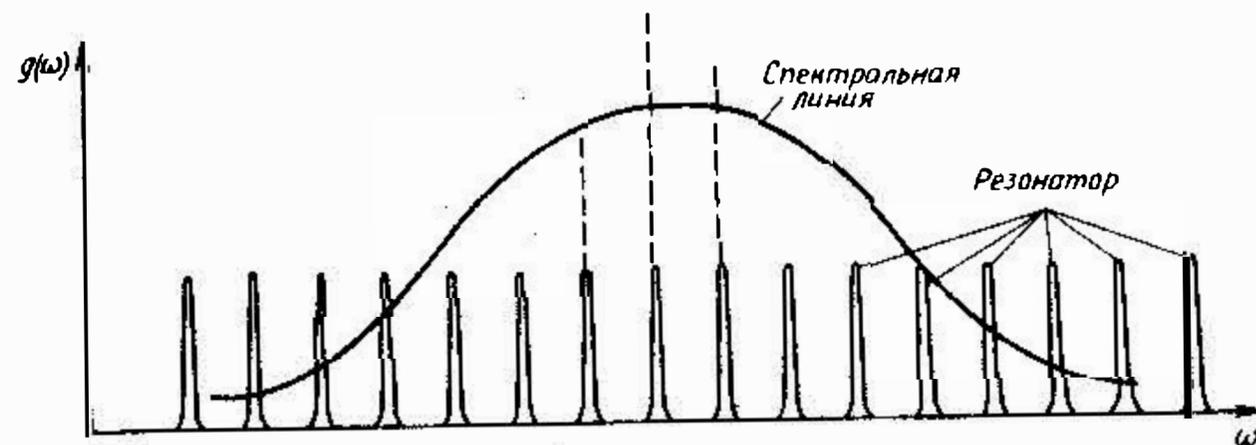


Рис. 2.11. Сравнение формы спектральной линии со спектром собственных частот оптического резонатора. Пунктирными линиями обозначены частоты, наиболее благоприятные для развития генерации

В радиодиапазоне наблюдается обратная картина. Ширина спектральной линии мала (см. § 1.6), расстояние между собственными типами колебаний резонатора велико и добротность спектральной линии существенно выше добротности резонатора. Это соответствует одномодовому резонатору. Спектр собственных частот в такой системе (в лазерах) будет определяться в основном рабочим веществом.

Для определения свойств оптического резонатора в общем случае надо записать уравнения Максвелла и решить их с учетом соответствующих граничных условий. Если проследить за первоначальной волной, претерпевающей многократные отражения, то можно убедиться, что в конце концов в резонаторе устанавливается некоторое стационарное состояние электромагнитного поля. Оно характеризуется тем, что при дальнейших отражениях (после нескольких сотен отражений) установившееся распределение электромагнитного поля не изменяется. Такое стационарное распределение поля и определяет собственные колебания резонатора. Эти колебания, имеющие пренебрежимо малые проекции E и H на ось z , называются *трансверсальными электромагнитными колебаниями* и обозначаются TEM_{mq} . Смысл индексов m , n и q тот же, что в объемных резонаторах.

Величина q , указывающая число полувольт, укладывающихся на длине L резонатора, в оптическом диапазоне очень велика, и при обозначении типов колебаний ее обычно опускают; m и n , наоборот, малы. Для аксиальных колебаний $m=n=0$ и их обозначают TEM_{00} . Кроме TEM_{00} -колебаний возможны TEM_{10} , TEM_{20} , TEM_{01} , TEM_{11} -колебания и т. д., которые называются *неаксиальными колебаниями*. Казалось бы, неаксиальные моды должны быстро выходить из резонатора, поскольку они распространяются не строго параллельно его оптической оси. На самом деле из-за явления дифракции на краях зеркал неаксиальные колебания, дойдя до края зеркала, будут отражаться внутри резонатора. Это явление аналогично отражению электромагнитной волны от границы волновода, как показано на рис. 2.12. Дойдя до открытого конца волновода, часть 2 волны 1 выйдет наружу (излучится), а часть 3 вследствие дифракции на открытом конце волновода отразится обратно. Если угол θ мал, то коэффициент отражения близок единице. Понятно, что рис. 2.12 и проведенные рассуждения отражают процесс условно, поскольку такой «геометрический» подход не объясняет сущность дифракционных

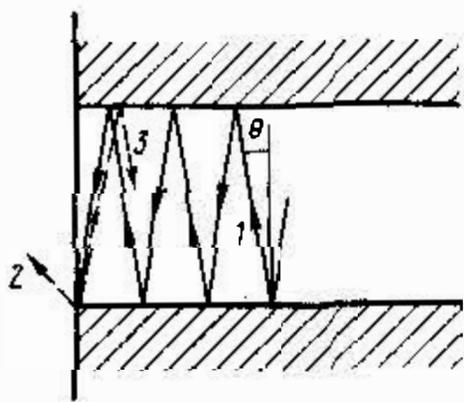


Рис. 2.12. Условное изображение явления дифракции на открытом конце волновода

явлений. Важно, что волны, распространяющиеся под малыми углами θ к оси резонатора (неаксиальные моды с малыми индексами m и n), как и аксиальные TEM_{00} -моды, будут обладать высокими добротностями. С увеличением m и n дифракционные потери будут увеличиваться, а добротность — падать.

Под добротностью оптического резонатора, как и других колебательных систем, понимают величину

$$Q = 2\pi \frac{E_{\text{полн}}}{E_{\text{потерь}}(T)} = \omega \frac{E_{\text{полн}}}{E_{\text{потерь}}(1c)}, \quad (2.27)$$

где $E_{\text{полн}}$ — полная энергия, запасенная в резонаторе; $E_{\text{потерь}}(T)$ — энергия, теряемая за период; $E_{\text{потерь}}(1c)$ — энергия, теряемая за 1 с.

Добротность определяется потерями в резонаторе. Из (2.27) получаем, что изменение энергии E_{ω} , запасенной в моде частотой ω за время dt ,

$$dE_{\omega} = -E_{\omega} \frac{\omega}{Q} dt.$$

Отсюда следует, что запасенная в резонаторе энергия вследствие потерь будет уменьшаться по экспоненциальному закону

$$E_{\omega}(t) = E_{\omega}(0) e^{-(\omega/Q)t}, \quad (2.28)$$

т. е. аналогично изменению во времени мощности излучения классического осциллятора (1.131). Поэтому каждая спектральная линия резонатора будет описываться функцией (1.135) с $\gamma = \omega/Q$, т. е. иметь лоренцеву форму с полушириной $\Delta\omega = \gamma = \omega/Q$. Из (2.28) можно определить постоянную времени затухания моды резонатора

$$t_{\phi} = Q/\omega,$$

которая называется также *временем жизни фотона*. Наряду с добротностью эту величину используют для характеристики потерь в резонаторе.

Для вывода излучения наружу одно из зеркал резонатора должно быть частично пропускающим (полупрозрачным). Это определяет необходимые полезные потери резонатора, называемые *потерями на излучение*. Пусть коэффициент отражения зеркала, через которое выводится излучение, равен R . Тогда коэффициент пропускания этого зеркала $T = (1 - R)$ (потери в современных многослойных диэлектрических зеркалах малы и ими можно пренебречь). Если длина резонатора $L \gg \lambda$, то теряемая за

единицу времени энергия равна $E_0(1-R)c/(2L)$. Из (2.27) получаем добротность резонатора, определяемую потерями на излучение

$$Q = \frac{2\omega L}{(1-R)c} = \frac{2kL}{(1-R)} \quad (2.29)$$

где $k=2\pi/\lambda$. Взяв типичные значения $\lambda=1$ мкм, $L=1$ м и $R=0,9$, получаем $Q \approx 1 \cdot 10^8$, что существенно выше добротности резонаторов радиодиапазона.

Кроме необходимых потерь на излучение в реальных резонаторах существуют дополнительные потери. Перечислим основные из них.

1. Дифракционные потери. Эти потери связаны с дифракцией электромагнитной волны на зеркалах резонатора, имеющих конечные размеры. Дифракционные потери определяются отношением $a^2/(\lambda L)$. Чем меньше поперечные размеры a зеркала, тем больше дифракционные потери. Дифракционные потери зависят также от типа колебаний. Они минимальны для аксиальных TEM_{00} -мод и возрастают с увеличением индексов m и n моды. Это используют для селекции неаксиальных мод в резонаторах.

2. Потери на несовершенствах зеркал. Эти потери обусловлены возможным поглощением в зеркалах (что недопустимо), рассеянием на шероховатостях, отклонением геометрии зеркала от заданной и т. д. Для их исключения к зеркалам резонатора предъявляются исключительно высокие требования. В частности, обработка поверхности зеркала должна проводиться с погрешностью $\sim 0,1\lambda$.

3. Потери на разъюстировку резонатора. Для того чтобы система из двух зеркал обладала резонансными свойствами, необходима очень точная их ориентация относительно друг друга. В частности, в плоском резонаторе отражающие поверхности двух плоских зеркал должны быть строго параллельны друг другу. Для выполнения этого требования зеркала резонатора очень точно настраивают (юстируют). Угол между плоскостями зеркал, называемый углом разъюстировки, в плоском резонаторе не должен превышать нескольких угловых секунд.

4. Потери в активном веществе. Они обусловлены поглощением и рассеянием энергии на различных дефектах в активной среде, заполняющей резонатор.

Свойства плоского резонатора. Рассмотрим резонатор, образованный двумя плоскими квадратными зеркалами с размерами $2a \times 2a$, находящимися на расстоянии L друг от друга, как показано на рис. 2.13. Собственные частоты в первом приближении определяются выражением (2.24). Их можно найти, полагая $L_1=L_2=2a$ и $L_3=L$. Поскольку $(m, n) \ll q$, разлагая (2.24) в степенной ряд, имеем

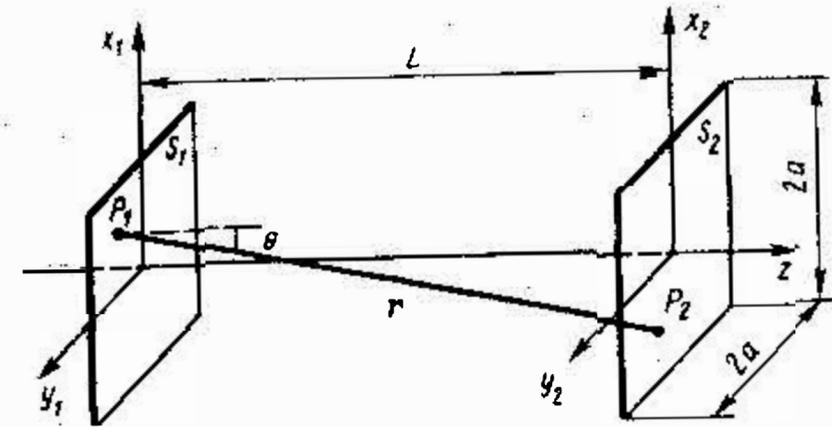


Рис. 2.13. К определению свойств плоского резонатора

$$v_{mq} = \frac{c}{\lambda_{mq}} = \frac{c}{2} \left[\frac{q}{L} + \frac{1}{2} \frac{(m^2 + n^2)}{q} \frac{L}{4a^2} \right] \quad (2.30)$$

Отсюда получаем разность частот между соседними аксиальными колебаниями в виде (2.26). Разность частот между двумя неаксиальными колебаниями, отличающимися друг от друга на единицу по индексу m , получается из (2.30):

$$\Delta v_m = v_{m+1} - v_m = \frac{c}{4} \frac{L[(m+1)^2 - m^2]}{4a^2 q} = \frac{cL}{8a^2 q} \left(m + \frac{1}{2} \right) \quad (2.31)$$

Эту формулу с учетом (2.26) можно переписать в виде

$$\Delta v_m = \Delta v_q \frac{L^2}{4qa^2} \left(m + \frac{1}{2} \right) \quad (2.32)$$

Аналогично находят расстояние по частоте для двух типов колебаний (мод), отличающихся друг от друга на единицу по индексу n :

$$\Delta v_n = \Delta v_q \frac{L^2}{4qa^2} \left(n + \frac{1}{2} \right) \quad (2.32a)$$

Неаксиальные моды, отличающиеся значениями n или m , будут различаться распределением поля в плоскости, ортогональной оси z , т. е. в поперечном направлении. Поэтому неаксиальные моды называют также поперечными модами. Для поперечных мод с различными индексами m структура поля будет различной в направлении x , а с n — в направлении y .

Формулы (2.32) и (2.32a) можно переписать в виде:

$$\Delta v_m = \Delta v_q \frac{m + 1/2}{8N},$$

$$\Delta v_n = \Delta v_q \frac{n + 1/2}{8N},$$

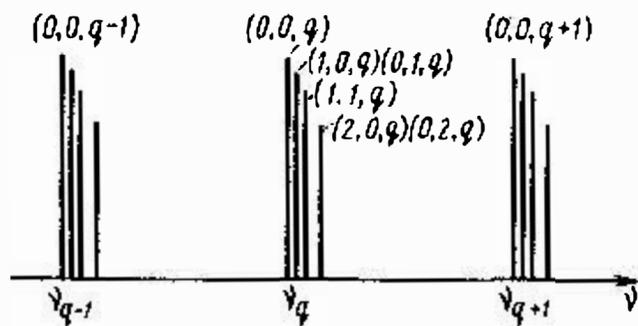


Рис. 2.14. Спектр собственных частот плоского резонатора

где

$$N = a^2 / (\lambda L) \quad (2.33)$$

— безразмерный параметр, называемый числом Френеля.

Оценим соответствующие величины для типичного случая $\lambda = 1$ мкм, $L = 1$ м и $2a = 1$ см. Число Френеля $N = 25 \gg 1$, и расстояние между

неаксиальными колебаниями ($\Delta\nu_m, \Delta\nu_n$) порядка мегагерц, т. е. существенно меньше расстояния между аксиальными колебаниями ($\Delta\nu_q \approx 150$ МГц) [см. (2.26)]. Дифракционные потери для неаксиальных мод выше, чем для аксиальных, поэтому добротность первых меньше.

Спектр собственных частот плоского резонатора как для аксиальных, так и для некоторых неаксиальных колебаний показан на рис. 2.14. Расстояние по частоте между аксиальными модами $\Delta\nu_q$ увеличивается при уменьшении длины резонатора L , а между неаксиальными модами $\Delta\nu_m$ и $\Delta\nu_n$ — при уменьшении числа Френеля N . Из (2.30) видно, что моды с одинаковыми q , но разными m и n , удовлетворяющими условию $m^2 + n^2 = \text{const}$, имеют одну и ту же частоту. Такие моды называют *частотно вырожденными*.

Более строгое рассмотрение свойств оптических резонаторов должно базироваться на решениях уравнений Максвелла. Определение собственных колебаний сводится к отысканию стационарных решений соответствующей краевой задачи. Такой метод является наиболее точным, однако он приводит к чрезвычайно громоздким соотношениям. Более наглядный метод расчета характеристик оптических резонаторов основывается на принципе Гюйгенса — Френеля. Этот принцип был предложен еще в XVII в. Гюйгенсом, а затем дополнен и развит Френелем и Кирхгофом. Согласно принципу Гюйгенса все точки, через которые проходит фронт волны в момент времени t , можно рассматривать как источники вторичных волн, а положение фронта волны в последующий момент времени $t + \Delta t$ дастся поверхностью, огибающей фронты всех вторичных волн. Он позволяет по возмущению, заданному на одной волновой поверхности, определить амплитуду и фазу возмущения в последующие моменты времени и в точках, лежащих в направлении распространения волны.

Зададим произвольное распределение электромагнитного поля u_1 на зеркале S_1 (см. рис. 2.13), индуцирующее какое-то распределение поля на зеркале S_2 . В произвольной точке P_2 этого

зеркала поле можно рассчитать путем суммирования вкладов бесконечно малых гюйгенсовских источников во всех точках P_1 . В результате суммирования поле $u_2(P_2)$ будет определяться дифракционным интегралом Кирхгофа — Френеля:

$$u_2(P_2) = \frac{ik}{4\pi S_1} \int u_1(P_1) \frac{\exp(-ikr)}{r} (1 + \cos\theta) dS_1. \quad (2.34)$$

Здесь $k = 2\pi/\lambda$, а все обозначения приведены на рис. 2.13.

Если известно распределение поля на зеркале S_2 , то оно в свою очередь индуцирует распределение поля на зеркале S_1 , которое может быть найдено аналогично с помощью выражения (2.34). Это распределение поля $u_1(P_1)$ вызовет распределение поля $u_2(P_2)$ на зеркале S_2 и т. д. Если после многократных отражений распределение поля по поверхностям каждого из зеркал приближается к стационарному, т. е. с точностью до постоянного множителя перестает изменяться от отражения к отражению, то такое распределение поля и будет определять собственные типы колебаний резонатора. Естественно, что эти стационарные решения зависят от формы зеркал и геометрических размеров резонатора. При наличии стационарных состояний после p проходов поле около зеркал становится одинаковым с точностью до комплексной постоянной, т. е. можно записать

$$u_p = \left(\frac{1}{\gamma}\right)^p v, \quad (2.35)$$

где v — функция, не изменяющаяся от отражения к отражению; γ — комплексная постоянная, не зависящая от пространственных координат.

Подставляя (2.35) в (2.34), получаем интегральное уравнение

$$\left(\frac{1}{\gamma}\right)^{p+1} v = \frac{ik}{4\pi S} \int \left(\frac{1}{\gamma}\right)^p \frac{v}{r} e^{ikr} (1 + \cos\theta) dS,$$

или

$$v = \gamma \frac{ik}{4\pi S} \int \frac{v}{r} e^{ikr} (1 + \cos\theta) dr. \quad (2.36)$$

Функция распределения v , удовлетворяющая интегральному уравнению (2.36), определяет распределение поля на зеркалах резонатора и соответствует его модам. Логарифм γ определяет затухание и фазовый сдвиг волны в течение каждого прохода. Его можно рассматривать как постоянную распространения, связанную с нормальным типом колебаний.

Интегральное уравнение (2.36) обычно решают на ЭВМ методом последовательных приближений. Отметим, что этот метод довольно наглядно отображает описанный процесс возбуждения начального распределения поля u_1 в резонаторе и его последующее многократное прохождение между зеркалами.

При рассмотрении конкретных резонаторов обычно используют условие малости поперечных размеров зеркал по сравнению с расстоянием между ними. Это позволяет положить $\cos\theta \approx 1$, а отличие r от L учитывать только в фазовом множителе. При подобных допущениях интегральное уравнение (2.36) упрощается:

$$v(x, y) = v \frac{ie^{-ikL}}{\lambda L} \int_S v(x, y) e^{-ik(R-L)} dS. \quad (2.37)$$

Собственные функции $v_{mn}(x, y)$, являющиеся решением этого уравнения при определенных значениях γ_{mn} (собственные значения), характеризуют структуру поля на поверхности зеркал для различных типов колебаний резонатора.

Логарифм γ_{mn} имеет вещественную и мнимую составляющие

$$\ln \gamma_{mn} = \beta_{mn} + i(\alpha_{mn} + kL). \quad (2.38)$$

Для каждого собственного колебания резонатора величина β_{mn} определяет затухание волны за один проход, а α_{mn} — фазовый сдвиг за один проход, добавленный к геометрическому фазовому сдвигу.

Рассмотрим результаты решения, полученные для плоского резонатора, изображенного на рис. 2.13. Если в качестве начального («затравочного») распределения поля u_1 выбрать равномерное распределение по всей поверхности зеркала, т. е. $u_1 = \text{const}$, то примерно после 300 отражений в резонаторе установится стационарное распределение поля, как показано на рис. 2.15, а. Если бы в качестве начального распределения было взято любое другое распределение u_1 , но с условием, чтобы оно не изменяло знака, то получилась бы та же самая картина. Это распределение характеризует наименьший четно-симметричный тип аксиальных TEM_{00} -колебаний.

Если в качестве начального распределения задана нечетно-симметричная функция u_1 с одинарной сменой знака (например, $u_1(x) = -u_1(-x)$, причем $|u_1(x)| = \text{const}$, как показано на рис. 2.15, б), то после нескольких отражений получится стационарное распределение, изображенное на рис. 2.15, б. Это распределение

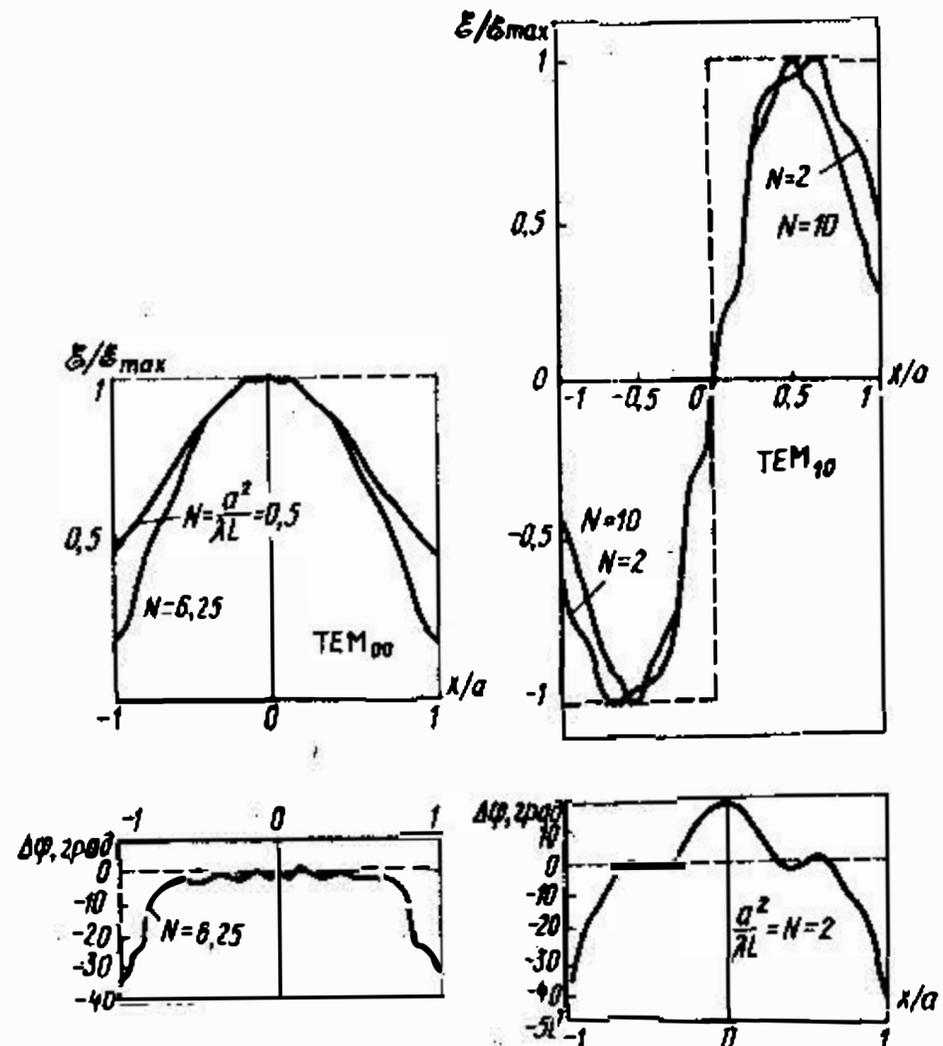


Рис. 2.15. Распределение относительной амплитуды $\varepsilon/\varepsilon_{\max}$ в фазы $\Delta\varphi$ в направлении x для аксиальных TEM_{00} (а)-и неаксиальных TEM_{10} (б)-колебаний в плоском резонаторе. Штриховой линией указано начальное распределение поля

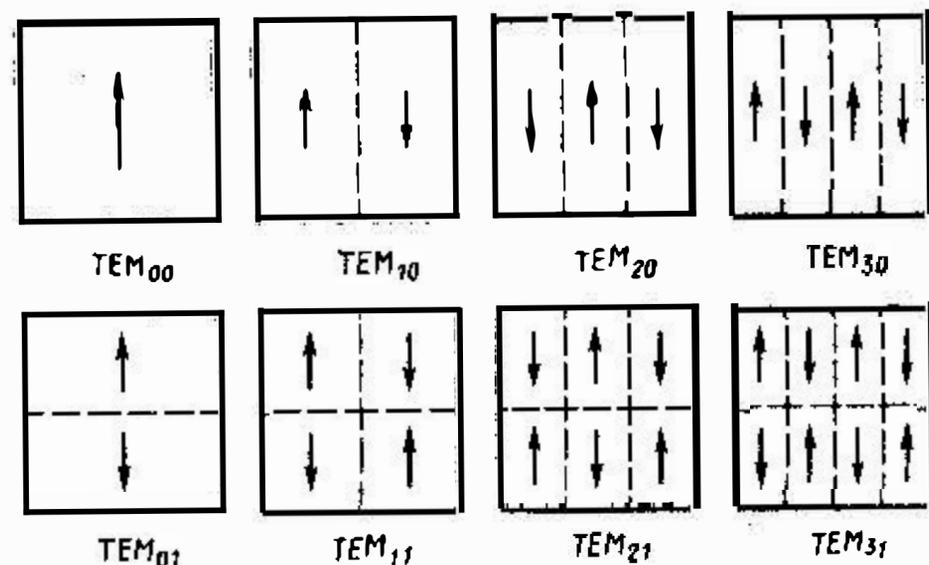


Рис. 2.16. Структура поля для некоторых типов колебаний в плоском резонаторе с квадратными зеркалами

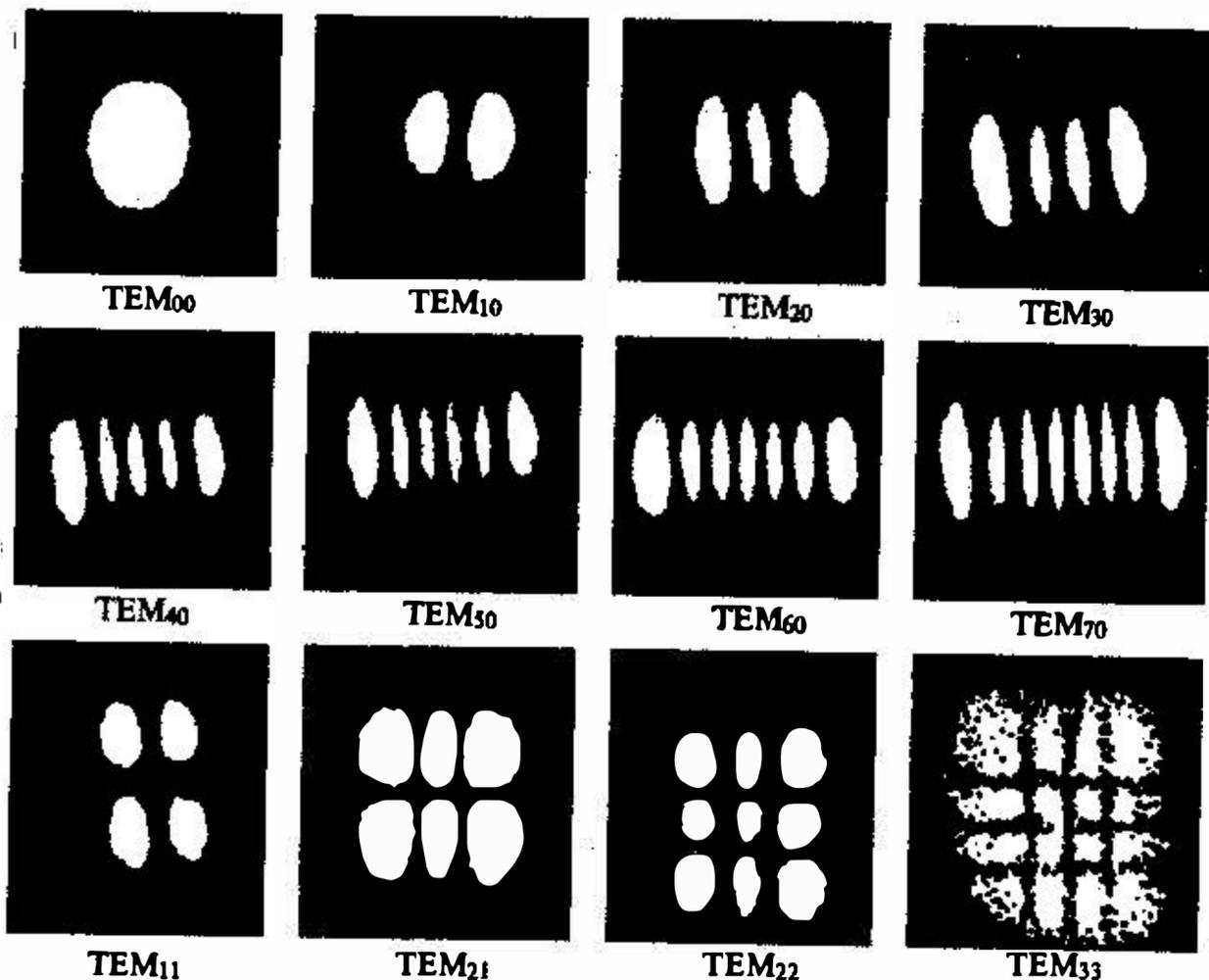


Рис. 2.17. Фотография структуры поля на зеркалах резонатора гелий-неонового газового лазера

характеризует TEM_{10} -моду. Аналогично получают распределения для неаксиальных колебаний более высокого порядка.

Как нетрудно заметить, для резонатора с зеркалами прямоугольной формы индексы m и n характеризуют число изменений направления поля вдоль осей x и y соответственно. На рис. 2.16 показана структура электрического поля для простейших типов колебаний, а на рис. 2.17 — фотографии структуры светового поля на зеркалах резонатора гелий-неонового лазера. С увеличением индексов m и n амплитуда поля на краях зеркала возрастает. Она минимальная для аксиальных TEM_{00} -колебаний.

Для зеркал круглой формы в силу осевой симметрии число m будет характеризовать число изменений направления поля по радиусу, а число n — по азимуту.

Существенным является то, что при выполнении неравенства $a^2/(\lambda L) \ll (L/a)^2$ истинные размеры зеркал и расстояния между ними не имеют значения и единственным важным параметром является число Френеля $N = a^2/(\lambda L)$. С увеличением N амплитуда поля на краях зеркала (а следовательно, и дифракционные потери) уменьшается. Зависимость дифракционных потерь за один проход от N в двойном логарифмическом

масштабе представлена на рис. 2.18.

Распределение поля внутри резонатора с плоскими зеркалами в разных сечениях, перпендикулярных оптической оси, изменяется незначительно. Это позволяет достаточно эффективно использовать активный объем рабочего вещества.

Недостатком плоского резонатора, с точки зрения его практического применения, является трудность юстировки: параллельность зеркал друг относительно друга должна быть выдержана с точностью несколько угловых секунд. Этому недостатка лишены резонаторы со сферическими зеркалами.

Резонаторы со сферическими зеркалами. Широкое распространение в лазерной технике получили оптические резонаторы, образованные двумя сферическими или одним сферическим и одним плоским зеркалами. Различные типы резонаторов со сферическими зеркалами представлены на рис. 2.19, а — г. Особое место среди них занимает конфокальный резонатор, на свойствах которого остановимся подробнее.

1. Конфокальный резонатор (рис. 2.19, а). Конфокальным называют открытый резонатор, образованный одинаковыми сферическими зеркалами, оси и фокусы которых совпадают. Так как фокус сферического зеркала радиусом R расположен на

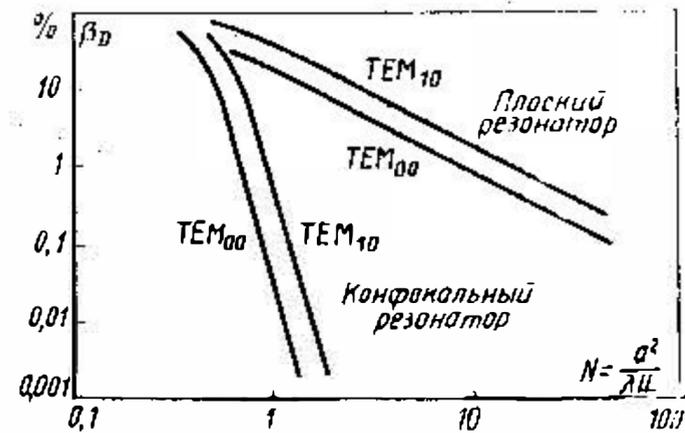


Рис. 2.18. Дифракционные потери β_D за один проход для простейших типов колебаний в зависимости от числа Френеля для круглых зеркал плоскопараллельного и конфокального резонаторов

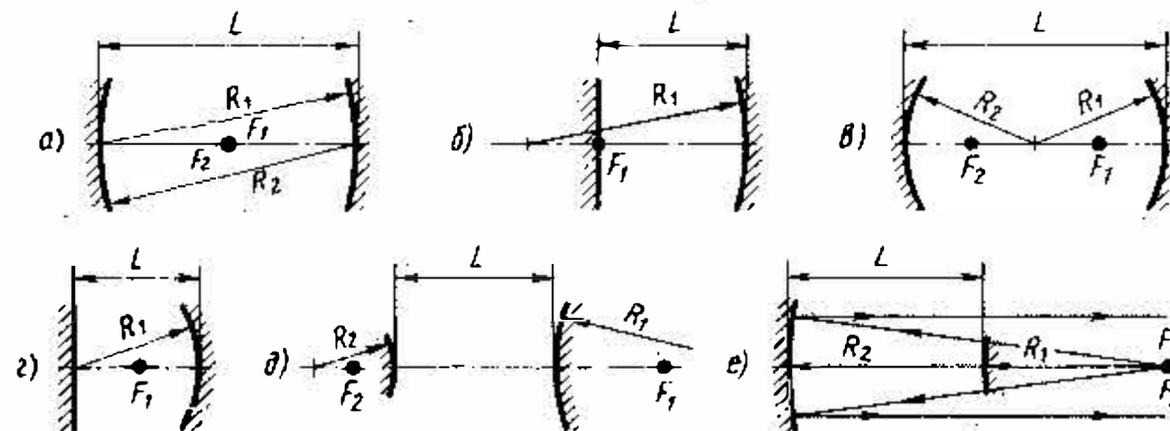


Рис. 2.19. Типы резонаторов со сферическими зеркалами: а) — конфокальный ($R_1 = R_2 = L$); б) — полуконфокальный ($R_1 = 2L$; $R_2 = \infty$); в) — коинцентрический ($R_1 = R_2 = L/2$); г) — полуконцентрический ($R_1 = L$; $R_2 = \infty$); д) — неустойчивый резонатор, общий случай; е) — телескопический неустойчивый резонатор

расстоянии $R/2$, то это означает, что радиусы кривизны зеркал равны длине резонатора.

Расчет поля и определение собственных колебаний в резонаторе произвольной конфигурации могут быть выполнены аналогично тому, как это было сделано для резонатора с плоскими зеркалами, поскольку соотношения (2.34) — (2.36) справедливы для общего случая.

В отличие от плоского резонатора для конфокального резонатора интегральное уравнение (2.36) имеет аналитическое решение. Обычно все размеры резонатора велики по сравнению с длиной волны. Тогда для зеркал круглой или прямоугольной формы уравнение (2.36) допускает разделение переменных относительно поперечных координат и сводится к двум одномерным интегральным уравнениям. Для сферических зеркал квадратного сечения со стороной $2a$ при условии $a \ll L$ и $N > 1$ собственные функции $v(x)$ (или $v(y)$) аппроксимируются произведениями полиномов Эрмита $H_n(x)$ на гауссову функцию $\exp(-x^2/r_{1/2}^2)$. В системе координат, начало которой совпадает с фокальной точкой F резонатора, а ось z — с осью резонатора, поперечное распределение поля дается выражением

$$v_{mn}(x, y) = v_0 H_m(x/r_{1/2}) H_n(y/r_{1/2}) \exp\left(-\frac{x^2 + y^2}{2r_{1/2}^2}\right), \quad (2.39)$$

где H_m и H_n — полиномы Эрмита. Для первых четырех степеней они имеют вид

$$\begin{aligned} H_0(x) &= 1; \quad H_1(x) = 2x; \quad H_2(x) = 4x^2 - 2; \\ H_3(x) &= 8x^3 - 12x; \quad H_4(x) = 16x^4 - 48x^2 + 12. \end{aligned}$$

Выражение (2.39) описывает поперечные распределения поля для TEM_{mn} -мод в конфокальном резонаторе. На рис. 2.20 показаны эти распределения для первых трех мод. Радиус $r_{1/2}$ в (2.39) характеризует поперечный размер пучка.

Для основной TEM_{00} -моды наблюдается гауссово распределение интенсивности:

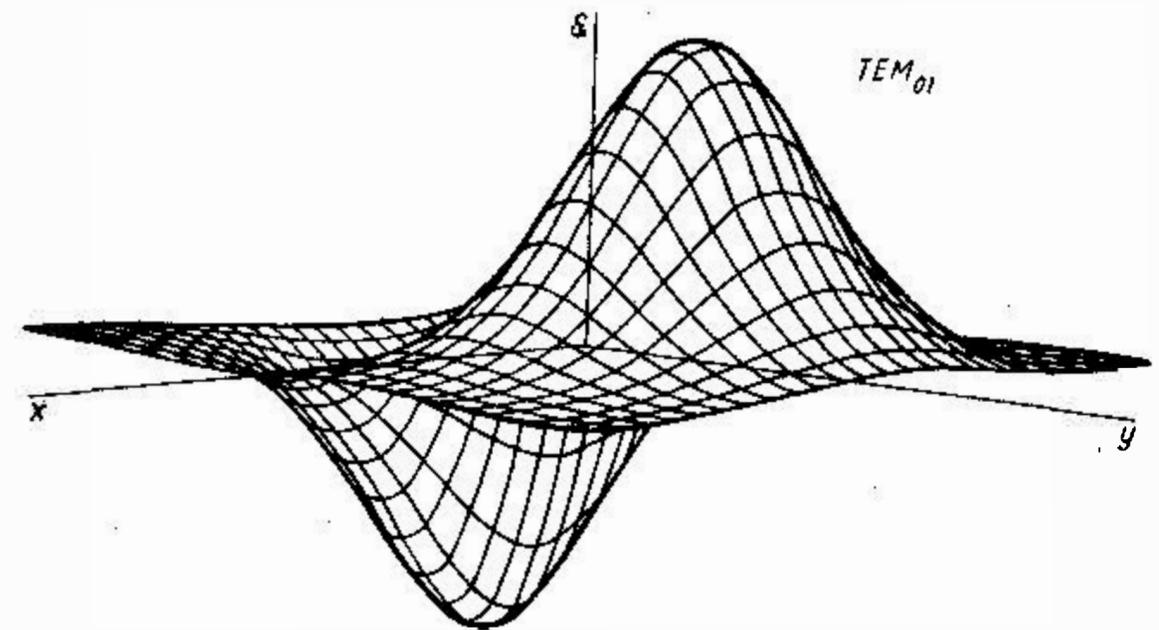
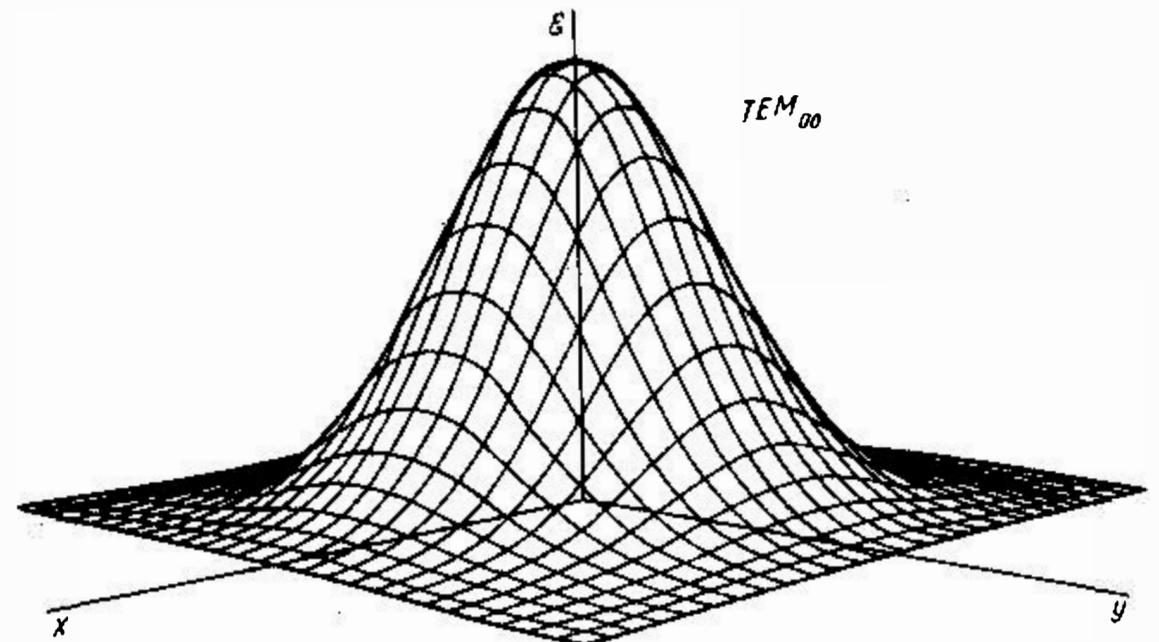
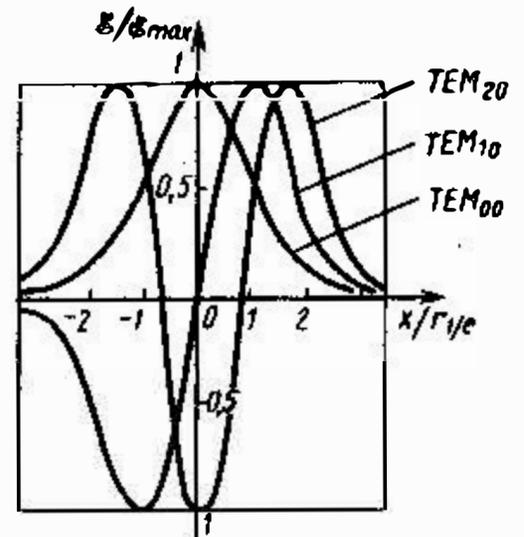
$$I(x, y) \sim v_{00}^2(x, y) = v_0^2 \exp\left(-\frac{x^2 + y^2}{r_{1/2}^2}\right). \quad (2.39a)$$

Значение $r_{1/2}$ определяет размер поперечного сечения, где интенсивность поля падает в e раз. В пятне площадью $\pi r_{1/2}^2$ сосредоточена основная энергия волны, проходящей в направлении z через плоскость xy . Ширина пятна меняется вдоль оси z по закону

$$r_{1/2}^2 = \frac{\lambda L}{4\pi} \left[1 + \left(\frac{2\pi}{L}\right)^2 \right]. \quad (2.40)$$

Рис. 2.20. Распределение поля в конфокальном резонаторе для первых трех мод:

а) аналогично рис. 2.15, но по оси абсцисс отложена величина $x/r_{1/2}$, где радиус пятна $r_{1/2}$ определяется соотношением (2.40); б) и в) распределение поля по плоскости зеркала



Эта зависимость представлена на рис. 2.21 сплошной линией. В фокальной плоскости при $z=0$ радиус пучка минимален. Он называется *радиусом перетяжки* или *радиусом шейки каустики** и равен

$$r_0 = r_{1/e}(0) = \sqrt{\frac{\lambda L}{4\pi}} = \sqrt{\frac{L}{2k}}. \quad (2.41)$$

На поверхности зеркала при $z=L/2$ радиус пучка в $\sqrt{2}$ раз больше, чем в центре.

Пучок, распределение поля в поперечном сечении которого характеризуется функцией Гаусса, называется *гауссовым пучком*. Поверхности равных фаз для гауссова пучка представляют собой сферические поверхности. Радиус кривизны синфазной поверхности, пересекающей оптическую ось резонатора в точке z (напомним, что начало отсчета совпадает с общим фокусом зеркал), определяется соотношением

$$R(z) = z \left(1 + \frac{L^2}{4z^2} \right). \quad (2.42)$$

След этих поверхностей изображен на рис. 2.21 пунктирными линиями. С приближением к центру резонатора радиусы кривизны синфазных поверхностей увеличиваются. При $z=0$ $R \rightarrow \infty$, т. е. синфазная поверхность является плоской. При $z = \pm L/2$ радиус $R = \pm L$, т. е. поверхности зеркал в конфокальном резонаторе являются поверхностями равных фаз. При больших $z \gg L/2$, что соответствует дальней зоне, волновой фронт (2.42) приближается к волновому фронту сферической волны ($R \approx z$), распространяющейся из точки, расположенной на оси пучка в месте перетяжки.

Таким образом, TEM_{00} -мода конфокального резонатора представляет собой сферическую волну, распространяющуюся из его центра и обладающую гауссовым распределением интенсивности в плоскости, перпендикулярной направлению распространения. Основная часть энергии пучка, как следует из (2.39а), сосредоточена в телесном угле

$$\theta = \pi \left(\frac{r_{1/e}}{z} \right)^2 = \frac{\lambda}{L}. \quad (2.43)$$

Обратим внимание, что расходимость пучка в основной моде конфокального резонатора определяется не поперечным, а про-

* *Каустиком* или *каустической поверхностью* (от греч. *kaustikos* — жгучий, паливший) называется поверхность, огибающая световые лучи, испущенные точечным источником после их прохождения (фокусировки) через разнородную оптическую систему, обладающую aberrацией и не собирающую лучи в одну точку.

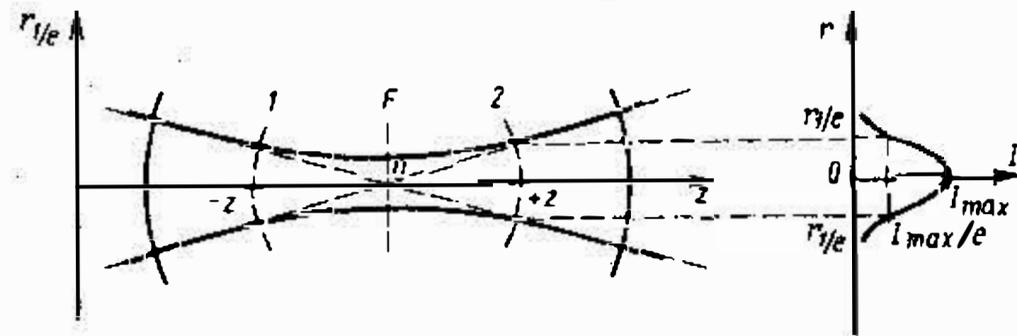


Рис. 2.21. Размер освещенного пятна $r_{1/e}$, синфазные поверхности и распределение интенсивности для аксиальной TEM_{00} -моды в конфокальном резонаторе

дольным размером L резонатора. Аналогично, поперечный размер пучка $r_{1/e}$, определяемый (2.40), не зависит от поперечного размера зеркала $2a$. Это является следствием сделанных выше предположений о малости значения отношения (a/L) и о большом значении числа Френеля $N = a^2/\lambda L$.

Собственным функциям уравнения (2.36), дающим поперечные распределения (2.39), соответствуют собственные частоты, определяемые условием

$$\nu_{mnq} = \frac{\omega_{mnq}}{2\pi} = \frac{c}{4L} (2q + 1 + m + n). \quad (2.44)$$

Отсюда видно, что спектр собственных частот конфокального резонатора сильно вырожден: изменение m и n такое, что $\Delta(m+n) = -2\Delta q$ дает одно и то же значение частоты.

Перечисленные особенности распределения поля в конфокальном резонаторе имеют как преимущества, так и недостатки. Например, по сравнению с плоским резонатором поле в конфокальном резонаторе более плотно сконцентрировано у его оси и спадает на краях значительно быстрее (см. рис. 2.15). Это приводит к тому, что дифракционные потери в конфокальном резонаторе очень малы, в том числе для неаксиальных колебаний высокого порядка (см. рис. 2.18). Поэтому конфокальный резонатор не очень чувствителен к разъюстировке зеркал. Однако эта же причина затрудняет получение одномодового режима генерации и мешает полностью использовать объем активного вещества в резонаторе.

Ниже мы покажем, что конфокальный резонатор лежит на границе, разделяющей устойчивую и неустойчивую области. Поэтому в чистом виде конфокальный резонатор применяют редко. Если желательно сохранить симметрию резонатора, то применяется квазиконфокальный резонатор, расстояние между зеркалами которого отличается от $L = R$ на небольшую величину α :

$$L/R = 1 \pm \alpha, \quad \alpha \ll 1.$$

Небольшие значения α , не изменяя по существу характер распределения поля в резонаторе, делают его устойчивым.

2. Резонаторы с произвольными сферическими зеркалами. Эти резонаторы состоят из двух соосных сферических зеркал радиусами R_1 и R_2 , расположенных на расстоянии L друг от друга. Свойства таких резонаторов легко определить, если найдена соответствующая конфокальная система, в которой две синфазные поверхности совпадают с поверхностями зеркал резонатора. Если известны радиусы R_1 и R_2 зеркал и расстояние между ними, то из выражения (2.42) можно найти длину, а также радиусы кривизны и координаты зеркал эквивалентного конфокального резонатора. Если известно $L_{\text{эк}}$, то распределение поля внутри и вне резонатора будет таким же, как в эквивалентном конфокальном резонаторе. В частности, радиус светового пятна будет определяться соотношением (2.40), где вместо L будут стоять $L_{\text{эк}}$. Если резонатор образован двумя одинаковыми сферическими зеркалами радиусами $R_1 = R_2 = R$, расположенными на расстоянии L друг от друга, то из (2.42) получаем, что длина эквивалентного конфокального резонатора

$$L_{\text{эк}} = \sqrt{L(2R-1)}. \quad (2.45)$$

Эквивалентный конфокальный резонатор определяет только собственные функции произвольного сферического резонатора. Нахождение собственных значений не может быть сведено к задаче о конфокальном резонаторе. В частности, собственные частоты ν_{mn} в общем случае будут определяться не соотношением (2.44), а более сложными выражениями. В обобщенном сферическом резонаторе частотное вырождение мод обычно исчезает.

Сферический резонатор может быть либо устойчивым, либо неустойчивым (в последнем случае резонатор теряет свои резонансные свойства). Например, если резонатор образован двумя одинаковыми сферическими зеркалами радиусами R , причем $L > 2R$, то, как следует из (2.45), ему невозможно подобрать эквивалентный конфокальный резонатор. Это означает, что в таких резонаторах невозможно образование устойчивого стационарного распределения электромагнитного поля с малыми потерями.

Пусть R_1 и R_2 — радиусы зеркал, а L — расстояние между ними. Можно показать, что резонатор будет устойчивым и ему всегда можно подобрать эквивалентный конфокальный резонатор, если выполняется неравенство

$$0 \leq (L/R_1 - 1)(L/R_2 - 1) \leq 1. \quad (2.46)$$

На рис. 2.22 приведена диаграмма, иллюстрирующая условие устойчивости оптических резонаторов. Устойчивые области соот-

ветствуют заштрихованным участкам. Отметим характерные точки на этой диаграмме. Точке A с координатами $(-1; -1)$ соответствует резонатор с плоскими зеркалами, расположенный на границе устойчивости. Точке B с координатами $(0; 0)$ соответствует конфокальный резонатор. Точке C с координатами $(1; 1)$ соответствует резонатор, образованный двумя одинаковыми сферическими зеркалами, оси и центры кривизны которых совпадают, т. е. $R_1 = R_2 = L/2$. Такой *концентрический резонатор* (см. рис. 2.19, в) также лежит на границе, разделяющей устойчивую и неустойчивую области. Дифракционные потери в нем очень быстро возрастают с увеличением типа колебаний, т. е. с увеличением индексов m и n . Поэтому такой резонатор наряду с неустойчивыми резонаторами применяют для селекции (отбора) неаксиальных колебаний.

Точкам D и D' с координатами $(-1; -0,5)$ и $(0,5; -1)$ соответствует *полукофокальный резонатор*, т. е. резонатор, образованный одним плоским и одним сферическим зеркалом, радиус кривизны которого равен удвоенной длине резонатора (см. рис. 2.19, б). Он находится в середине устойчивой области, и его свойства аналогичны конфокальному резонатору с удвоенной длиной. Такой резонатор часто применяют на практике. Точкам E и E' с координатами $(-1; 0)$ и $(0; -1)$ соответствует *полукоцентрический резонатор* (см. рис. 2.19, г). Он обладает особенностями, характерными для концентрического резонатора.

Неустойчивые резонаторы и селекция поперечных типов колебаний. В ряде случаев требуется обеспечить такой режим работы лазера, при котором генерация осуществлялась бы на одной, как правило, продольной моде. Такой режим называется *одномодовым*. Для его осуществления необходимо выделить (селектировать) основную моду при подавлении поперечных мод. Это можно сделать за счет увеличения потерь при увеличении поперечных индексов мод m и n . Но в устойчивых резонаторах дифракционные потери существенно меньше потерь на частичное пропускание зеркал, а потому добротность Q резонатора для аксиальных (продольных) и неаксиальных (поперечных) колебаний низших порядков слабо зависит от m и n . В этом случае разница в дифракционных потерях не может служить основой для селекции

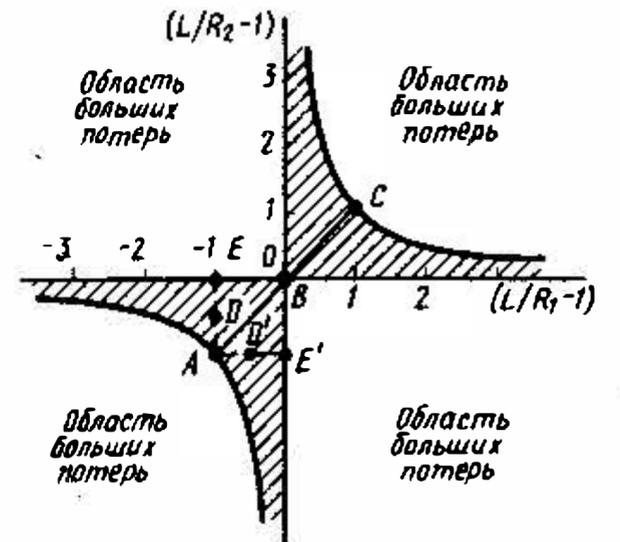


Рис. 2.22. Диаграмма устойчивости оптических резонаторов со сферическими зеркалами

колебаний, но она может быть основана на различии в пространственном распределении поля мод с различными поперечными индексами m и n . Поскольку в резонаторах со сферическими зеркалами основная TEM_{00} -мода имеет гауссово распределение (2.39а) с минимальной шириной пучка $r_{1/2}$, то простейшим и часто применяемым на практике способом селекции является диафрагмирование пучка внутри резонатора. Для этого внутри резонатора помещают диафрагму, размер отверстия которой примерно равен поперечному размеру моды, следующей за основной, т. е. TEM_{01} -моды (см. рис. 2.20). Такая диафрагма будет создавать большие потери для неосновных колебаний, большая часть энергии которых сосредоточена по периферии. Очевидно, что при ~~различии~~ диафрагмы внутри резонатора число Френеля N будет определяться не поперечным размером зеркала в (2.33), а размером отверстия этой диафрагмы.

Недостатками обсуждавшегося выше способа выделения основной TEM_{00} -моды в устойчивом резонаторе являются: а) малость поперечных размеров моды n , как следствие, малость используемого объема активного вещества; б) внесение дополнительных потерь в основную моду.

В мощных лазерах, где необходимо применение большого объема активного вещества, эффективным средством селекции поперечных мод является переход к неустойчивым резонаторам. В неустойчивых резонаторах, параметры которых попадают в незаштрихованные области диаграммы (см. рис. 2.22), дифракционные потери даже основной моды велики и превосходят все остальные виды потерь. Для поперечных мод потери очень быстро нарастают при увеличении индексов m и n . Именно это обстоятельство приводит к эффективному выделению основной моды. В качестве примера на рис. 2.19, *д* и 2.19, *е* показаны два типа неустойчивых резонаторов.

Неустойчивые резонаторы подразделяются на два класса: резонаторы положительной области, для которых

$$(1 - L/R_1)(1 - L/R_2) > 1$$

(на диаграмме устойчивость рис. 2.22 они расположены в первом и третьем квадрантах незаштрихованной области), и резонаторы отрицательной области, для которых

$$(1 - L/R_1)(1 - L/R_2) < 0$$

(они расположены во втором и четвертом квадрантах диаграммы рис. 2.22).

Неустойчивые резонаторы могут быть применены лишь в лазерах с большим показателем усиления. Это обусловлено необходимостью компенсации больших потерь излучения за один проход, принципально присутствующих в таких резонаторах.

Важно, что это излучение, стремящееся выйти из резонатора (рис. 2.19, *д* — *е*), может быть использовано как полезное выходное излучение лазера. К достоинствам неустойчивых резонаторов относятся: а) возможность использования больших объемов активного вещества, что связано с отсутствием фокусировки излучения (гауссова сжатия) к оси резонатора; б) возможность эффективной селекции поперечных типов колебаний; в) возможность использования только отражающей оптики (например, металлических зеркал) как для создания резонатора, так и для вывода излучения, а также простота управления выводимой из резонатора энергией и достижения оптимальных условий вывода излучения.

Недостатком неустойчивых резонаторов, ограничивающим их применение мощными лазерами, является необходимость применения активных сред с большим показателем усиления. Кроме того, поперечное сечение выходного пучка света в ближней зоне, как видно из рис. 2.19, *е*, имеет форму кольца (для круглых зеркал), что не всегда удобно. Однако вдали от выходного зеркала, в дальней зоне, темное пятно в центре исчезает.

Составные и дисперсионные резонаторы. Селекция продольных типов колебаний. Рассмотренные выше резонаторы не могут селективировать продольные типы колебаний, т. е. TEM_{0q} -моды, различающиеся по индексу q , поскольку дифракционные потери для них одинаковы. Для осуществления селекции продольных мод может быть использовано их отличие друг от друга по частоте [см. (2.26), (2.30)], для чего необходимо внутри резонатора ввести узкополосный дисперсионный (т. е. частотно зависимый) элемент. В качестве такого элемента могут быть использованы призмы, дифракционные решетки, специальные зеркала с частотно зависимым коэффициентом отражения, а также дополнительные резонаторы типа плоского резонатора (эталона Фабри — Перо) с малой базой.

Два или более связанных между собой резонатора (1 — 3 и 1 — 2) образуют составной резонатор. Его схема приведена на рис. 2.23, *а*. Зеркало 3 — полупрозрачное. Свойства такого резонатора аналогичны свойствам двух связанных кон-

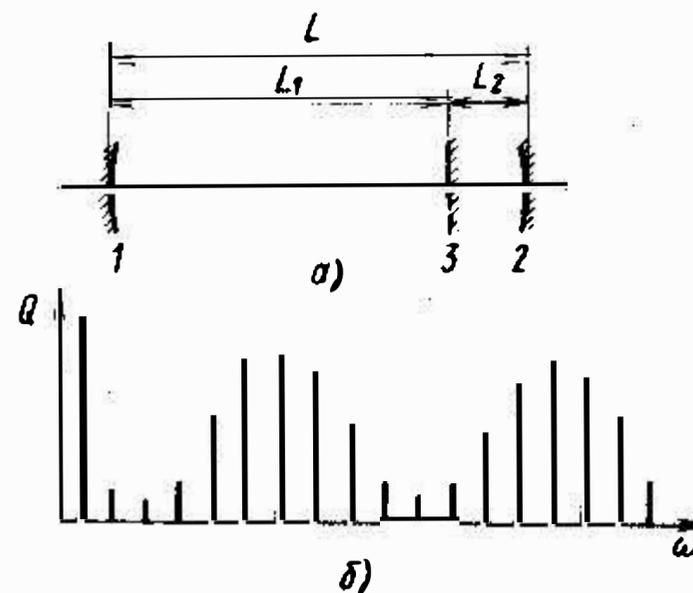


Рис. 2.23. Схема составного резонатора (а) и зависимость добротности его основных типов колебаний от частоты (б)

туров. Как видно из рис. 2.23, б, при соответствующем выборе величин L_2 , L_1 и коэффициента пропускания зеркала 3 составной резонатор способен селективировать аксиальные колебания (сравните с рис. 2.11).

Естественно, что проводить селекцию продольных колебаний в резонаторе необходимо лишь тогда, когда ширина спектральной линии спонтанного излучения $\Delta\omega_{\text{сп}}$ существенно больше расстояния между аксиальными модами (2.26). В противном случае такую селекцию будет осуществлять само активное вещество лазера.

В тех случаях, когда линия усиления достаточно широка ($\Delta\omega_{\text{спонт}} \gg \Delta\omega_e = 2\pi \frac{c}{2L}$), а частота настройки дисперсионного эле-

мента управляема, возможно перестраивать частоту лазерного излучения. Такие резонаторы, содержащие перестраиваемые дисперсионные элементы, называются *дисперсионными резонаторами*. Как правило, они используются в лазерах с перестраиваемой частотой излучения. Простейшим оптическим дисперсионным элементом является призма. Простейший призмный дисперсионный резонатор показан на рис. 2.24, а. Вследствие того что показатель преломления призмы n зависит от длины волны (для ее изготовления выбирают материалы с большой дисперсией $\frac{1}{n} \frac{dn}{d\lambda}$), волны с разными λ будут отклоняться на разные углы.

Меняя угол разворота призмы, можно осуществлять перестройку частоты.

Дисперсии призмы часто бывает недостаточно для получения одночастотной генерации с высокой степенью монохроматичности и для тонкой перестройки частоты лазерного излучения. Для этой цели применяют более сложные дисперсионные резонаторы (рис. 2.24, б). Дифракционная решетка 1 выполняет роль грубого дисперсионного элемента. Эталон Фабри — Перо 2 позволяет выделить одну продольную моду. Те-

лескопическая система 3, состоящая из одной или нескольких линз, служит для расширения пучка, выходящего из активного элемента, и уменьшения его угла расходимости, что необходимо для эффективной работы как дифракционной решетки, так и эталона. Грубая перестройка частоты осуществляется поворотом дифракционной решетки, а точная — поворотом эталона Фабри — Перо.

Кольцевой резонатор. Кольцевым называют открытый резонатор, зеркала которого обеспечивают распространение электромагнитных волн по замкнутому контуру (рис. 2.25). Система из четырех зеркал 1, 2, 3 и 4 образует замкнутый контур. В одно из плеч резонатора помещают активный элемент, например газоразрядную трубку газового лазера. В кольцевом резонаторе, как и в рассмотренных выше резонаторах, может существовать стоячая волна, образованная интерференцией двух волн, бегущих в противоположных направлениях. Если каким-либо образом устранить одну из бегущих волн, например сделав зеркало 4 полупрозрачным и поставив дополнительное зеркало 5, как показано на рис. 2.25, то в таком резонаторе можно осуществить режим бегущей волны. Иногда режим стоячей волны является энергетически менее выгодным, поскольку наиболее полное взаимодействие излучения с веществом здесь происходит лишь в пучностях светового поля. В бегущей волне поле в среднем более однородно, что позволяет полностью использовать активное вещество.

Кроме того, если вращать кольцевой резонатор вокруг оси, то длина пути для волн, распространяющихся по направлению и против направления вращения, будет различной. Это может быть использовано для измерения скорости вращения и построения лазерных гироскопов.

Резонатор с распределенной обратной связью. Это особый вид резонатора, в котором торцевые зеркала вообще отсутствуют, а положительная обратная связь обеспечивается рассеянием на периодических неоднородностях, образующих строго периодическую решетку. Важно, чтобы эта решетка образовывалась пространственными периодическими изменениями одного из параметров, от которого зависят условия распространения света. Это может быть показатель преломления, показатель усиления, толщина пленки или кристалла. Расстояние b между неоднородностями должно удовлетворять условию Вульфа — Брэгга:

$$b \sin \theta = m \lambda / 2,$$

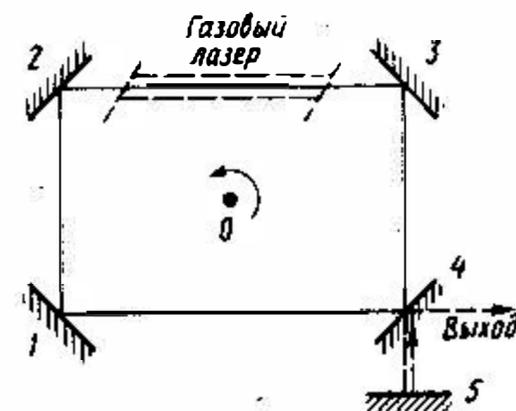


Рис. 2.25. Кольцевой резонатор

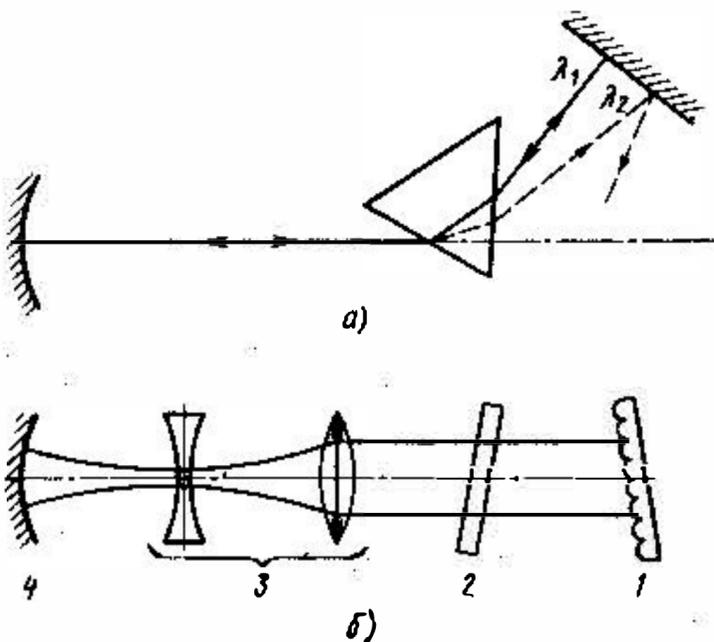


Рис. 2.24. Дисперсионные резонаторы: а) призмный, б) составной решетчатый: 1 — дифракционная решетка, 2 — эталон Фабри — Перо (двухплоский интерферометр), 3 — телескопическая система, 4 — зеркало резонатора

где b — период дифракционной решетки; θ — угол между нормалью к плоскости решетки и оптической осью резонатора; $m=1, 2, 3, \dots$ — порядок дифракции. В оптических резонаторах, реализованных таким образом, решетка является одновременно фильтром и отражателем, выполняя роль зеркал «обычных» резонаторов.

Распределенная обратная связь используется, в частности, в некоторых типах полупроводниковых лазеров. Она осуществляется следующим образом. Пусть электромагнитное излучение распространяется в тонком плоском диэлектрическом волноводе, заполненном активным веществом. Вблизи этого активного слоя на расстоянии l от него поместим плоскую дифракционную решетку с периодом b , как показано на рис. 2.26.

Как известно, электромагнитная волна, распространяющаяся в диэлектрическом волноводе, «просачивается» за его стенки на расстояние порядка длины волны; амплитуда волны вне волновода экспоненциально убывает в направлении, перпендикулярном плоскости волновода (в направлении x на рис. 2.26). Поэтому если расстояние l мало ($l \approx \lambda$), то дифракционная решетка будет влиять на волну, распространяющуюся внутри активного слоя. Распространяющаяся в волноводе мода будет «зацепляться» своим «хвостом» за поверхность, на которую нанесена дифракционная решетка. Если период b решетки составляет целое число длин полуволн света в волноводе, то брэгговское отражение волноводной моды от решетки изменит направление ее распространения на обратное, приводя тем самым к возникновению положительной обратной связи. Эта обратная связь будет равномерно распределена по всей длине z активного слоя, поэтому она называется распределенной обратной связью.

При определенных условиях с помощью той же дифракционной решетки можно осуществить вывод излучения из плоского диэлектрического волновода наружу. При этом излучение выводится с поверхности активного слоя под определенными углами, т. е. происходит распределенный съем энергии излучения.

Если период решетки b равен половине длины волны λ (в данном случае λ есть длина световой волны в активном слое с показателем преломления n), т. е. $b = \lambda/2n$, то волна рассеивается только назад, чем и создается обратная связь. Для решеток высшего по-

рядка с $m \geq 2$, кроме отражения вперед и назад, возможно рассеивание под дискретными углами к плоскости волновода. Так, для решетки с $m=2$ волна будет рассеиваться назад под углом $\theta = 180^\circ$, обеспечивая обратную связь во втором порядке дифракции, и перпендикулярно плоскости решетки ($\theta = 90^\circ$) в первом порядке. Для решетки с $m=3$ волна рассеивается назад в третьем порядке дифракции и под углами 60° и 120° в первом и втором порядке, как это схематически показано на рис. 2.26.

Таким образом, в активном слое волновода распространяются навстречу друг другу как минимум два колебания с одинаковой длиной волны. По мере того как одна из двух волн распространяется вдоль резонатора (вдоль оси z), за счет дифракции она получает энергию от волны, распространяющейся в противоположном направлении, что определяет взаимодействие этих волн и создает положительную обратную связь, распределенную по всей длине периодической структуры. Коэффициент потерь на излучение и обратную связь в такой системе можно изменять, например, изменяя расстояние l .

При наличии нескольких типов колебаний в резонаторе, т. е. нескольких волноводных мод, из-за их возможного взаимодействия общая картина излучения может значительно усложниться. С помощью системы, аналогичной изображенной на рис. 2.26, можно в обратном порядке осуществлять не вывод, а ввод излучения в тонкопленочный волновод через его боковую поверхность. Это часто используют в устройствах интегральной оптики.

§ 2.4. УСЛОВИЕ САМОВОЗБУЖДЕНИЯ И НАСЫЩЕНИЕ УСИЛЕНИЯ

Как было показано в § 2.1, необходимым условием для усиления электромагнитных колебаний в активном веществе является создание инверсии населенностей между его рабочими лазерными уровнями. Для возникновения генерации нужно, чтобы усиление за один проход в активном веществе превысило все потери, связанные прежде всего с частичным выводом излучения наружу, дифракционными потерями и неактивными потерями в рабочем элементе, т. е. необходимо выполнить условие самовозбуждения. Другими словами, генерация начнется тогда, когда инверсия населенностей между лазерными уровнями достигнет такого значения ΔN , при которой показатель усиления α_ω на частоте ω превысит показатель потерь β_ω :

$$\alpha_\omega = -k_\omega \geq \beta_\omega.$$

Показатель усиления, являющийся отрицательным показателем поглощения, связан с населенностью уровней соотношением (2.2), т. е.

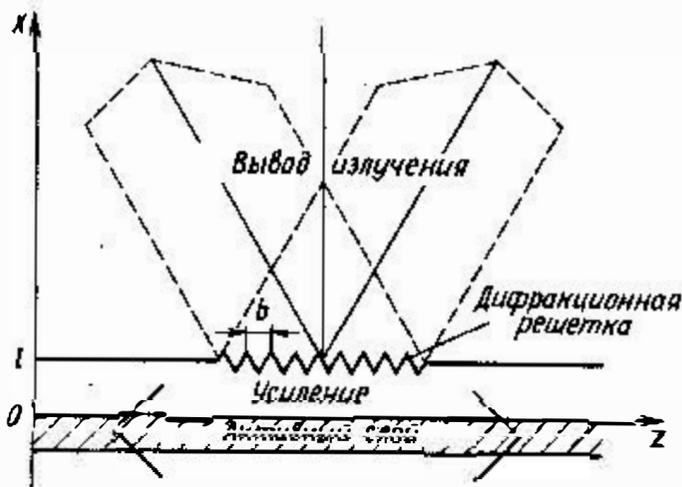


Рис. 2.26. Схема резонатора с распределенной обратной связью

частоте ω_{mn} мало. Тогда инверсию можно определить с помощью кинетических уравнений (2.16), предполагая независимость вероятности переходов $m \leftrightarrow n$ от интенсивности сигнала I_ω на рабочей частоте ω_{mn} . Такое приближение называется приближением слабого сигнала. Ему соответствуют сплошные линии изменения населенностей рабочих лазерных уровней на рис. 2.8, б и 2.9, б: в отсутствие генерации увеличение ρ_x приводит к увеличению инверсии ΔN и возрастанию усиления.

Пусть теперь на систему, в которой создана инверсия ΔN , начинает действовать мощное излучение на частоте лазерного перехода ω_{mn} вблизи максимума спектральной линии усиления. Под воздействием этого излучения между уровнями m и n начнутся интенсивные вынужденные переходы, стремящиеся выровнять населенности рабочих лазерных уровней. Ситуация по существу будет аналогична изображенной на рис. 2.6 для двухуровневой схемы с той разницей, что верхний и нижний уровень как бы поменялись местами, а вместо ρ_x выступает лазерное излучение. Это выравнивание населенностей рабочих уровней под действием генерируемого самой системой лазерного излучения приведет к ограничению (насыщению) показателя усиления. Процесс насыщения усиления в лазере, как и в генераторе любого другого типа, осуществляется за счет эффектов, нелинейных по интенсивности. В результате эффекта насыщения инверсия населенностей в лазере автоматически поддерживается на уровне $\Delta N_{\text{пор}}$, как показано пунктирными линиями на рис. 2.8, б, а интенсивность лазерного излучения достигает стационарного значения.

Для количественного описания насыщения необходимо определить изменение населенности рабочих уровней как функцию плотности энергии лазерного излучения при постоянной ρ_x аналогично тому, как это было сделано в § 2.2 для определения населенностей уровней как функции ρ_x . Проведя соответствующие преобразования, получим для случая $g_m = g_n$

$$\Delta N = N_m - N_n = \frac{\Delta N^0}{1 + 2\rho_{mn} B_{mn} \tau_m}, \quad (2.51)$$

где под $\rho_{mn} B_{mn}$ подразумевается интеграл (2.14), а τ_m — время релаксации верхнего лазерного уровня E_m в отсутствие генерации. Если его опустошение осуществляется только спонтанными переходами $m \rightarrow n$, то $\tau_m = 1/A_{mn}$. Величина ΔN^0 есть инверсная населенность при $\rho_{mn} = \rho_x = 0$, т. е. это инверсия, определенная в приближении слабого сигнала. Выражение (2.51) удобно записать как функцию интенсивности излучения в виде

$$\Delta N = \frac{\Delta N^0}{1 + I/I_S} \quad (2.52)$$

В случае, если кратности вырождения уровней m и n не равны, (2.52) перепишем в виде

$$\Delta N = \frac{N_m}{g_m} - \frac{N_n}{g_n} = \frac{\Delta N^0}{1 + [(g_m + g_n)/(2g_m)] I/I_S} \quad (2.52a)$$

В этих соотношениях I_S является эффективной характеристикой насыщения и называется *параметром насыщения*. Учитывая связь (2.47) показателя усиления с инверсией ΔN , получаем

$$\alpha_\omega = \frac{\alpha_\omega^0}{1 + I/I_S}, \quad (2.53)$$

где α_ω^0 называется *ненасыщенным показателем усиления* или *показателем усиления слабого сигнала*. Он определяет максимальное усиление, которое может быть достигнуто в данной системе при данном уровне накачки. Величина α_ω^0 определяется соотношением (2.47) при условии, что населенности N_m/g_m и N_n/g_n вычислены без учета вынужденных лазерных переходов. Параметр насыщения I_S определяется свойствами активного вещества. Формулы (2.8), (2.51) — (2.53) позволяют связать I_S с сечением оптического поглощения в максимуме спектральной линии σ_{max} и эффективным временем релаксации населенностей τ_m :

$$I_S \approx \hbar\omega/2\sigma_{\text{max}}\tau. \quad (2.54)$$

Очевидно, параметр насыщения I_S определяет ту интенсивность лазерного излучения, при которой скорость вызванных этим излучением вынужденных переходов равна скорости релаксационного распада населенности верхнего уровня $1/\tau$. При $I = I_S$ инверсия населенности ΔN и показатель усиления α_ω уменьшаются вдвое против своих максимальных значений при $I = 0$. Для видимой области спектра ($\tau \approx 1$ мс и $\sigma \approx 10^{-19}$ см²) определяем $I_S = 1$ кВт/см².

Характер насыщения усиления в лазере очень сильно зависит от характера уширения спектральной линии. Проведенное выше рассмотрение и формулы (2.51) — (2.54) справедливы лишь для однородно уширенной линии. Напомним, что при однородном уширении форма спектра поглощения, а следовательно, и усиления для каждой из активных частиц совпадает с общим контуром спектральной линии (см. § 2.6). Поэтому мощное монохроматическое излучение на частоте генерации ω равномерно перераспределяет частицы по энергетическим уровням, приводя к равномерному по спектру уменьшению показателя усиления. Контур спектральной линии при этом не изменяется и остается таким же, как и в отсутствие насыщения.

При неоднородном уширении во взаимодействии с монохроматической волной на частоте ω будут участвовать лишь те частицы, собственные частоты которых лежат вблизи этой частоты. Поэтому насыщение будет происходить не равномерно по спектру всего ансамбля частиц, а преимущественно вблизи частоты ω , образуя характерный «провал» в общем спектре. Этот эффект называется «частотным выжиганием дырки» на контуре линии усиления. В первом приближении насыщение показателя усиления (поглощения) для неоднородно уширенной линии может быть аппроксимировано выражением

$$\alpha_{\omega}^{\text{пол}} = \frac{\alpha_{\omega}^0}{(1 + I/I_S)^{1/2}} \quad (2.55)$$

Учет эффекта насыщения позволяет рассмотреть вопрос об усилении излучения лазерной активной средой и проанализировать работу лазера-усилителя. Ограничимся случаем однородно уширенной линией. Пусть активная среда за счет созданной в ней инверсии населенностей обладает ненасыщенным показателем усиления α_{ω}^0 . В реальной среде всегда присутствуют потери, которые будем характеризовать показателем нерезонансных потерь β . Приняв для простоты $g_m = g_n$, уравнение переноса излучения (1.169) в такой среде с учетом (2.53) запишется в виде

$$\frac{dI}{dz} = -\beta I + \frac{\alpha_{\omega}^0 I}{1 + I/I_S} \quad (2.56)$$

Коэффициенты β , α_{ω}^0 и I_S не зависят от интенсивности I . Это уравнение легко интегрируется, однако в общем виде его решение приводит к плохо обозреваемому трансцендентному уравнению, которое решают численно или графически. Для понимания сущности дела полезно рассмотреть два предельных случая.

1. Случай малого сигнала, когда $I \ll I_S$. В этом случае из (2.56) имеем

$$dI/dz = (\alpha_{\omega}^0 - \beta)I$$

и получаем экспоненциальный рост интенсивности с расстоянием, так что для усилителя с длиной активного элемента l имеем

$$I_{\text{вых}} = I_{\text{вх}} \exp[(\alpha_{\omega}^0 - \beta)l] \quad (2.57)$$

в полном согласии с законом Бугера — Ламберта (2.3).

2. Случай большой интенсивности сигнала, когда $I \gg I_S$. В этом случае (2.56) принимает вид

$$dI/dz = -\beta I + \alpha_{\omega}^0 I_S \quad (2.58)$$

В отсутствие нерезонансных потерь в активном веществе, т. е. при $\beta = 0$, из (2.58) получаем

$$I_{\text{вых}} = I_{\text{вх}} + \alpha_{\omega}^0 I_S l \quad (2.59)$$

При сильном насыщении ($I \gg I_S$) и отсутствии нерезонансных потерь экспоненциальный рост выходной интенсивности (2.57) сменяется линейным (2.59).

При $\beta \neq 0$ существует такая предельная интенсивность

$$I_{\text{max}} = \frac{\alpha_{\omega}^0}{\beta} I_S \quad (2.60)$$

при которой $dI/dz = 0$. Это означает, что дальнейшего усиления нет и лазерное излучение, интенсивность которого превышает значение (2.60), усиливаться не будет. Физически это означает такое состояние, при котором всё, что может излучить активное вещество с единицы длины в режиме полного насыщения, им же поглощается за счет нерезонансных потерь.

Из соотношения (2.60) вытекает важный вывод, что в мощном лазере — усилителе бегущей волны интенсивность выходного излучения ограничивается интенсивностью насыщения I_S , показателем нерезонансных потерь β и ненасыщенным показателем усиления α_{ω}^0 .

Проведенные выше рассуждения эффектов насыщения были выполнены для стационарных (квазистационарных) режимов работы. При импульсном нестационарном режиме эффект насыщения характеризуется энергией насыщения. При этом важную роль играют форма и длительность лазерных импульсов. Этот вопрос требует отдельного анализа.

Когда под воздействием накачки инверсия населенностей достигает пороговой величины, начинается генерация на одной моде, обладающей наивысшей добротностью и расположенной ближе всего к максимуму спектральной линии. При увеличении (иногда очень незначительном) интенсивности накачки условие самовозбуждения начинает выполняться и для других мод, обладающих меньшими добротностями или расположенных дальше от максимума спектральной линии. Многомодовый режим генерации обусловлен эффектами насыщения и наиболее легко возникает в активных средах с неоднородно уширенными спектральными линиями. Действительно, в неоднородно уширенной линии насыщение наступает не по всей линии усиления, а на частоте генерируемой моды («частотное выжигание дырок»). Благодаря ему при увеличении накачки коэффициент усиления для основной моды остается на прежнем уровне, в то время как для других мод он увеличивается, достигая порогового значения, и генерация начинается на этих модах, приводя к выжиганию

очередной «дырки». При дальнейшем увеличении интенсивности накачки в игру вступают все новые моды, так что их общее число может достигать нескольких десятков и даже сотен.

В активном элементе с однородно уширенной спектральной линией возникновение многомодовой генерации обусловлено другой причиной и связано с тем, что в активной среде, находящейся в резонаторе, каждой моде соответствует определенное пространственное распределение стоячей волны (см. рис. 2.10). В результате при генерации на какой-то определенной моде наиболее эффективное взаимодействие будет в областях, соответствующих пучностям стоячей волны. В этих же областях в первую очередь наступает насыщение усиления, в то время как в областях, соответствующих узлам стоячей волны, насыщения для данной волны нет. Для другой моды, отличающейся по частоте от первой, пространственное распределение другое и потому при увеличении интенсивности накачки вторая мода, первоначально имеющая более низкое усиление, может теперь достичь усиления, достаточного для начала генерации, поскольку в генерацию на этой моде дают вклад те области активной среды, которые не использовались при генерации первой моды. Этот эффект, обусловленный пространственно неоднородным насыщением инверсии населенностей внутри активной среды, называется *пространственным выжиганием дырок*. Естественно, что он может наблюдаться в активных материалах как с однородно, так и с неоднородно уширенными спектральными линиями.

Эффекты пространственного и частотного выжигания дыр приводят к тому, что большинство лазеров имеют тенденцию работать в многомодовом режиме. Для получения одномодовой или одночастотной генерации в лазерах, работающих в ультрафиолетовой, видимой и ближней ИК-областях, необходимо принимать меры, обеспечивающие селекцию типов колебаний, которые были рассмотрены в предыдущем параграфе. При многомодовой генерации возможно возникновение эффектов взаимодействия мод между собой, так что в этом случае их нельзя рассматривать как независимые.

§ 2.5. НЕСТАЦИОНАРНАЯ ГЕНЕРАЦИЯ, МОДУЛЯЦИЯ ДОБРОТНОСТИ И СИНХРОНИЗАЦИЯ МОД

Под нестационарным режимом работы лазера понимают такой импульсный режим, характерные времена которого короче времен релаксации населенностей рабочих уровней активной среды (см. § 2.1).

Для количественного описания процессов, происходящих при нестационарном режиме работы лазеров, необходимо решить соответствующие кинетические уравнения (2.10). Поскольку при

этом $dN_m/dt \neq 0$, эти уравнения будут нелинейными относительно переменных $N_m(t)$, и аналитическое решение получить в общем виде невозможно. Поэтому ограничимся качественным рассмотрением вопроса.

Пусть импульс накачки имеет ступенчатую форму (диаграмма 1 на рис. 2.27). В момент времени t_1 накачка включена, в момент времени t_4 выключена. После включения накачки населенность верхнего лазерного уровня будет возрастать и в момент времени t_2 достигнет пороговой населенности $N_{2\text{пор}}$, необходимой для развития генерации (диаграмма 2 на рис. 2.27). Дальнейшее возрастание $N_2(t)$ приводит к быстрому росту интенсивности поля на частоте генерации и соответственно к увеличению вероятности вынужденных переходов с верхнего лазерного уровня N_2 , что влечет за собой уменьшение населенности N_2 и инверсии населенностей. Это в свою очередь обуславливает падение усиления и уменьшение интенсивности поля излучения как внутри резонатора, так и на выходе. В результате в функции $N_2(t)$ появляется первая осцилляция (диаграмма 2), которой соответствует интенсивный всплеск мощности излучения на выходе генератора (диаграмма 3). Заметим, что момент времени t_3 начала генерации задержан относительно момента времени t_2 достижения пороговой инверсии, а момент времени t_2 — относительно начала накачки t_1 .

Математическое рассмотрение этого явления для генерации только в одной моде показывает, что временные зависимости $N_2(t)$, а следовательно, и $P_{\text{вых}}(t)$ представляют собой регулярную во времени последовательность световых импульсов (лазерных пиксов), амплитуда которых уменьшается приблизительно по экспоненциальному закону (диаграммы 2 и 3 на рис. 2.27). Такую

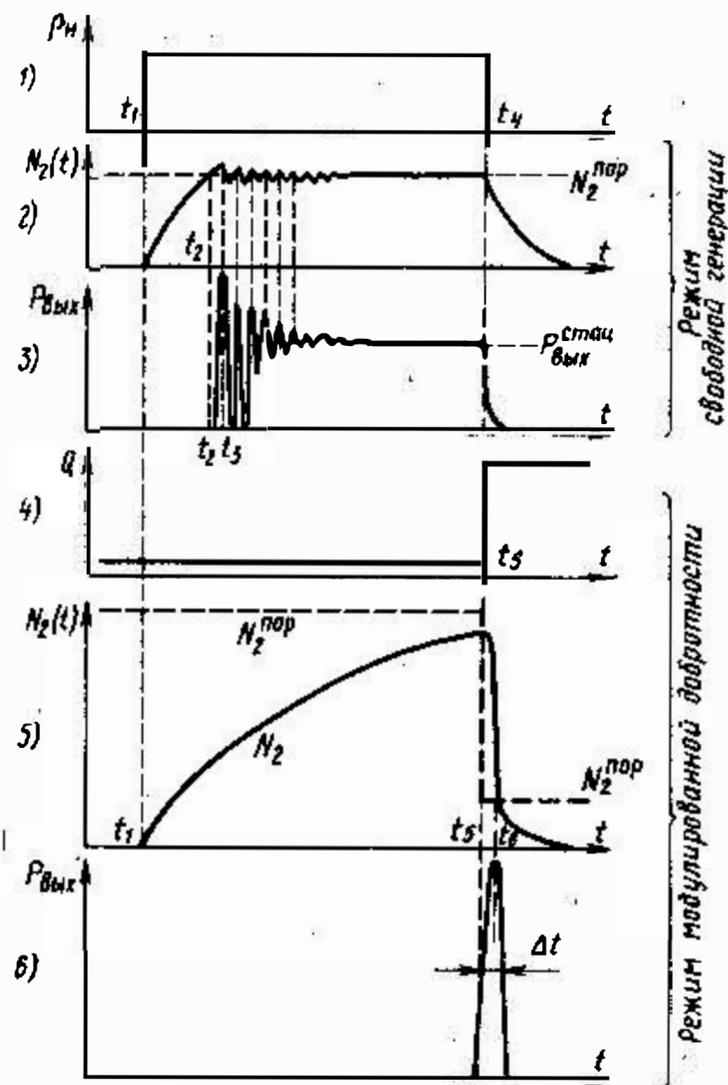


Рис. 2.27. Временные диаграммы развития лазерного импульса

генерацию называют *режимом регулярных пичков*. При больших t в отсутствие дополнительных источников флуктуаций режим регулярных пичков переходит в стационарный режим.

Основным физическим механизмом возникновения релаксационных колебаний является взаимодействие поля в резонаторе с активной средой. Эти колебания могут возникать не только в момент развития генерации при ступенчатом импульсе накачки, но и в случае, когда лазер, работающий в стационарном режиме, испытывает внезапное возмущение. Характерный период таких осцилляций превышает время жизни фотона в резонаторе $t_{\phi} = Q/\omega$ и время обхода резонатора. Обычно это время, приблизительно равное $2(t_2 - t_3)$, составляет 0,1...10 мкс.

Режим регулярных затухающих осцилляций может возникать лишь при одномодовой генерации и неизменности параметров лазера во времени. В реальных условиях, особенно для твердотельных лазеров и диэлектриках и для полупроводниковых лазеров, наблюдается режим *нерегулярных пичков*, интенсивность которых хаотически изменяется и не затухает во времени. Типичные временные зависимости $P_{\text{вых}}(t)$ для этого режима, называемого иногда *пичковым режимом генерации*, приведены на осциллограммах рис. 2.28, а, б. Такой режим может быть как в импульсных, так и в непрерывных лазерах.

Причины возникновения незатухающих хаотических флуктуаций в излучении лазера могут быть самыми разнообразными. Прежде всего они могут быть обусловлены эффектами пространственного выжигания дыр в активном веществе (см. § 2.4) и нестабильностью параметров лазера в реальных условиях, а также взаимодействием различных мод при многомодовой генерации.

Временные диаграммы 1—3 на рис. 2.27 соответствуют так называемому режиму свободной генерации лазера: добротность Q резонатора и соответственно пороговая населенность $N_{2\text{пор}}$ остаются неизменными во времени. Генерация

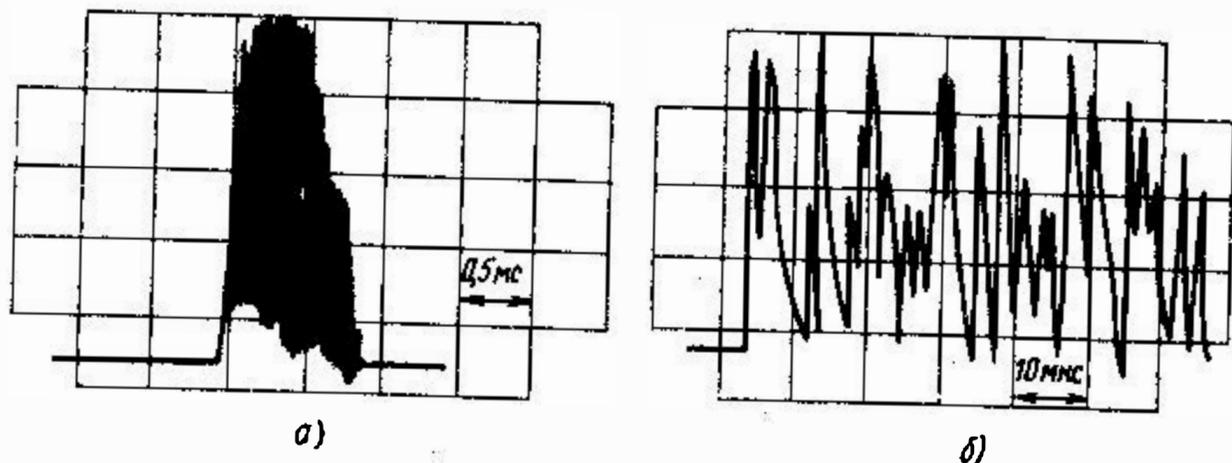


Рис. 2.28. Осциллограммы «пичковой» структуры лазерного излучения: а) масштаб 500 мкс/см; б) масштаб 10 мкс/см

в таком режиме работы начинается самопроизвольно в момент времени t_3 .

Представляет интерес импульсный режим работы генератора, при котором за сравнительно большой промежуток времени в активном веществе происходит накопление энергии накачки, создавая большую степень перенаселенности между лазерными уровнями, а затем в заданный момент времени накопленная энергия излучается в виде мощного светового импульса. Этот режим осуществляется благодаря модуляции добротности резонатора. Излучаемые световые импульсы имеют малую длительность и высокую импульсную мощность и называются *гигантскими импульсами*.

Принцип работы лазера в режиме *модулированной добротности* поясняется временными диаграммами 4—6 на рис. 2.27 и состоит в следующем. Поместим внутри лазерного резонатора затвор, вносящий большие потери и снижающий добротность резонатора. Потери вносятся на время действия накачки, с тем чтобы показатель усиления в активной среде возрос до очень большого значения, но генерация не начиналась. В этот промежуток времени активная среда накапливает энергию и инверсия достигает максимального значения. Если теперь открыть затвор, то усиление окажется намного выше порогового, что приведет к резкому возрастанию интенсивности вынужденного излучения и одновременному сбросу инверсии вследствие индуцированных переходов (см. диаграммы 4 и 5 на рис. 2.27). Запасенная в активном веществе энергия выделится в виде короткого (длительностью порядка нескольких наносекунд) и интенсивного импульса света (см. диаграмму 6 на рис. 2.27).

Предположим, что затвор открывается за время, времени развития лазерного импульса. Примем за $t=0$ момент открытия затвора, т. е. момент включения добротности. К этому моменту в активном веществе достигнута инверсия населенностей ΔN между лазерными уровнями E_m и E_n , значительно превышающая пороговую инверсию для генерации, т. е. $\Delta N > \Delta N_{\text{пор}}$. Для описания поведения системы в последующие моменты времени $t > 0$ необходимо решить кинетические уравнения с начальными условиями $\Delta N(0) = \Delta N_{\text{инт}}$ и $\rho_{\omega}(0) = \rho_{\omega}^{\text{инт}}$. Здесь ρ_{ω} — плотность энергии электромагнитного поля на частоте генерации ω . Будем считать, что генерация осуществляется на одной моде. При существенном превышении инверсии над пороговой показатель усиления α_{ω} велик, и генерация развивается настолько быстро, что всеми слагаемыми, описывающими изменение числа частиц на лазерных уровнях, кроме индуцированного испускания, можно пренебречь. Изменение во времени $\Delta N(t)$ будет определяться как

$$\frac{d\Delta N(t)}{dt} = \frac{d(N_m/g_m - N_n/g_n)}{dt} = -2B_{\omega}\rho_{\omega}(t)\Delta N(t). \quad (2.61)$$

Здесь $B_{\omega} \rho_{\omega}$ — вероятность вынужденных переходов в канале генерации. Множитель 2 означает, что появление каждого фотона сопровождается уменьшением инверсии на $\Delta N = -2$. В свою очередь плотность излучения на частоте рабочей моды ρ_{ω} будет определяться поступлением энергии в эту моду за счет процессов вынужденного испускания $B_{\omega} \Delta N \rho_{\omega}$ и потерями на излучение, равными Φ_{ω}/t_{ϕ} , где t_{ϕ} — время жизни фотона в резонаторе. Число фотонов Φ_{ω} связано с плотностью излучения ρ_{ω} соотношением $\Phi_{\omega} = \rho_{\omega}/\hbar\omega$, и кинетическое уравнение для числа фотонов или плотности электромагнитной энергии в моде записывается в виде

$$\frac{d\Phi_{\omega}}{dt} = \frac{d\rho_{\omega}}{\hbar\omega dt} = \left[B_{\omega} \Delta N(t) \hbar\omega - \frac{1}{t_{\phi}} \right] \Phi_{\omega}. \quad (2.62)$$

Результаты совместного решения уравнений (2.61), (2.62) с начальными условиями $\Delta N(0) = \Delta N_{нач}$ и $\Phi(0) = \Phi_{ин}$ представлены на рис. 2.29. Здесь по существу изображена в более крупном масштабе часть рис. 2.27. Подставляя в уравнение (2.62) $d\Phi_{\omega}/dt = 0$, нетрудно найти инверсию населенностей, соответствующую максимуму лазерного импульса

$$\Delta N = \frac{1}{B_{\omega} t_{\phi} \hbar\omega} = \frac{\omega}{Q B_{\omega} \hbar\omega}. \quad (2.63)$$

Сравнение с (2.50) показывает, что эта величина равна критической инверсии [знак равенства в (2.50)] для непрерывного режима работы лазера. Этот результат вполне объясним: до момента времени $t_{ин}$, когда $\Delta N > \Delta N_{пор}$, интенсивность лазерного импульса

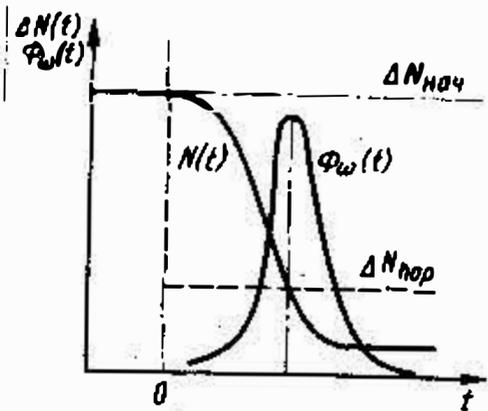


Рис. 2.29. Изменение инверсии $\Delta N(t)$ и числа фотонов $\Phi_{\omega}(t)$ в резонаторе в процессе генерации гигантского импульса при медленном включении добротности в момент времени $t=0$

нарастает, после чего, достигнув пика, импульс ослабляется. Форма лазерного импульса при этом асимметрична: время нарастания обычно несколько короче времени спада.

Максимальное число фотонов, запасенных внутри резонатора, равно $\Delta N_{ин}/2$. Если начальная инверсия намного превышает пороговую и потери в активном веществе малы [β в (2.58) равно нулю], то мощность лазерного излучения в максимуме

$$P_{max} \approx \frac{\hbar\omega}{2t_{\phi}} V \Delta N_{ин}, \quad (2.64)$$

где V — объем, занимаемый генерируемой модой в активном веществе.

Развитие генерации в рассмотренном виде будет происходить тогда, когда время включения добротности мало (меньше t_{ϕ}). В этом случае генерируется один моноимпульс. Если это условие не соблюдается, т. е. затвор открывается медленно ($\tau > 10^{-8}$ с), то может генерироваться не один, а несколько гигантских импульсов. При медленном включении добротности накопленная в активной среде энергия высвечивается в виде последовательных шагов.

Укажем типичные методы модуляции добротности лазера, подробнее рассматриваемые в гл. 10.

1. Одно из двух зеркал резонатора вращается вокруг оси. Потери в резонаторе будут очень высокими на протяжении всего цикла, за исключением короткого интервала времени, соответствующего параллельному расположению зеркал. Этот момент времени соответствует включению добротности. Часто вместо зеркала используют призму полного внутреннего отражения.

2. Внутри резонатора имеется специальный элемент — оптический модулятор, оптические свойства которого можно изменить с помощью внешних воздействий. Наиболее часто для этих целей используют электрооптические и акустооптические модуляторы, работающие на основе электрооптических и акустооптических эффектов в кристаллах (см. § 10.2).

3. Внутри резонатора помещают насыщающийся поглотитель, т. е. вещество, показатель поглощения которого уменьшается (насыщается) с ростом интенсивности излучения. Наиболее часто здесь применяют просветляющиеся красители. Эффект просветления определяется переходом поглощающих молекул красителя в возбужденное состояние и связанным с этим уменьшением показателя поглощения [см. формулу (2.2) и рис. 2.2].

Первый и второй методы модуляции добротности являются активными, а третий — пассивным. В последнем случае потери в резонаторе регулируются автоматически.

Рассмотрим теперь многомодовую генерацию. Пусть для определенности генерация осуществляется на $(2N+1)$ аксиальных модах, разделенных частотным интервалом (2.26): $\Delta\omega = \omega_{q+1} - \omega_q = \pi c/L$. Обозначим собственную частоту центральной моды ω_0 . Тогда полное электрическое поле $E(t)$ электромагнитной волны в некоторой произвольной точке, например на зеркале резонатора, можно записать в комплексной форме:

$$E(t) = \sum_{l=-N}^N E_l e^{i[(\omega_0 + l\Delta\omega)t + \phi_l]}, \quad (2.65)$$

где E_l и ϕ_l — амплитуда и фаза (при $t=0$) l -й моды.

Если не принимать специальных мер, то фазы отдельных мод будут случайными и полная мощность излучения будет равна сумме мощностей отдельных мод. Если фазы отдельных мод

каким-либо образом синхронизованы, то моды интерферируют и происходит явление, называемое *синхронизацией мод*.

Для простоты будем считать, что генерируется $(2N+1)$ аксиальных мод с равными амплитудами $E_i = E_0$ и все фазы равны нулю. Тогда выражение (2.65) переписывается в виде

$$E(t) = \sum_{i=-N}^N E_0 e^{i(\omega_0 + i\Delta\omega)t}, \quad (2.66)$$

что при суммировании дает

$$E(t) = A(t) e^{i\omega_0 t}, \quad (2.66)$$

где

$$A(t) = E_0 \frac{\sin[(2N+1)\Delta\omega t/2]}{\sin(\Delta\omega t/2)}. \quad (2.67)$$

Выражение (2.66) отображает гармоническую волну с частотой несущей ω_0 и амплитудой $A(t)$, промодулированной по закону (2.67). Выходная мощность лазера пропорциональна $|A(t)|^2$, т. е.

$$P_{\text{вых}}(t) \sim \frac{\sin^2[(2N+1)\Delta\omega t/2]}{\sin^2(\Delta\omega t/2)}. \quad (2.68)$$

На рис. 2.30 приведена временная зависимость выходной мощности, рассчитанная по соотношению (2.68), в случае генерации $2N+1=7$ мод с синхронизованными фазами и одинаковыми амплитудами. Отметим следующие важные свойства функции (2.68), являющиеся следствием интерференции мод.

1. Энергия излучается в виде последовательности коротких световых импульсов. Максимумам соответствуют моменты времени, когда знаменатель в (2.68) обращается в нуль. Два импульса разделены интервалом $\tau = 2\pi/\Delta\omega = 2L/c$. Это время, необходимое фотону для полного прохода резонатора.

2. Ширина импульса, определенная на полувысоте, приблизительно равна временному интервалу $\Delta\tau$ между вершиной импульса и ближайшим минимумом

$$\tau_n = \Delta\tau = \tau/(2N+1). \quad (2.69)$$

Число генерируемых мод $(2N+1)$ можно оценить как

отношение спектральной ширины линии усиления $\Delta\nu$ к межмодовому интервалу. Подставляя это отношение в (2.69), получаем выражение для длительности импульса

$$\tau_n \approx 1/\Delta\nu, \quad (2.70)$$

которое хорошо иллюстрирует соотношение неопределенностей $\delta E \delta t = \hbar \Delta\nu \Delta\tau \approx \hbar$.

Из (2.70) следует, что для получения очень коротких световых импульсов необходима большая ширина генерируемого спектра, т. е. активные материалы с большой шириной линии усиления. Для газовых лазеров низкого давления, как было показано в § 1.6, ширина спектральной линии $\Delta\nu \approx 10^9$ Гц. Поэтому в таких лазерах возможно получить импульсы длительностью порядка 1 нс. В твердотельных и жидкостных лазерах линии уширены и с их помощью можно генерировать световые импульсы длительностью порядка 1 пс = 10^{-12} с и менее.

3. В режиме синхронизации мод пиковая мощность в импульсе в $(2N+1)$ раз превышает сумму мощностей отдельных мод.

Синхронизации мод можно достигнуть путем модуляции усиления (или потерь) в лазере с частотой $\Omega = \Delta\omega = \pi c/L$, равной межмодовому интервалу. Ее можно осуществить с помощью как активного модулятора, управляемого внешним воздействием, так и соответствующей нелинейной оптической среды. Первый случай соответствует активной, второй — пассивной синхронизации мод.

Для того чтобы полнее представить, что означают приведенные цифры, укажем, что за время длительности импульса $\tau_n = 1$ пс свет проходит расстояние всего 0,3 мм. Каждый импульс может нести световую энергию порядка 1 Дж, т. е. импульсную мощность порядка 10^{12} Вт. Таким образом, в области пространства с размерами 0,3 мм сосредоточена мощность, равная 1 млн. МВт, и этот сгусток энергии движется со скоростью света.

§ 2.6. СВОЙСТВА ЛАЗЕРНОГО ИЗЛУЧЕНИЯ

Из предыдущего изложения следует, что основными свойствами лазерного излучения, принципиально отличающими его от излучения «классических» источников, являются: 1) монохроматичность, 2) пространственная и временная когерентность, 3) направленность, 4) высокая мощность и яркость. Рассмотрим их более подробно.

Монохроматичность и поляризация. Монохроматичность* характеризует степень концентрации излучения по спектру. Количественной характеристикой степени монохроматичности является ширина спектральной линии на уровне 0,5 от ее максимума

*От греч. *monos* — единственный, *chroma* — цвет.

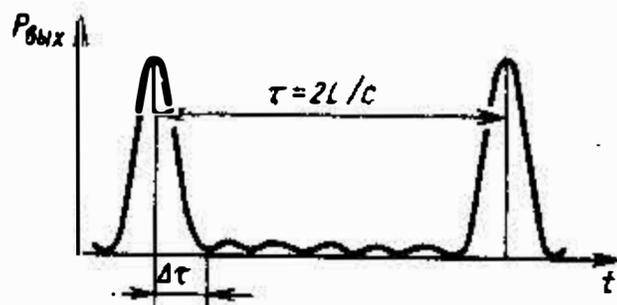


Рис. 2.30. Временная зависимость выходной мощности в случае генерации семи мод в режиме их синхронизации

или спектральный диапазон $\Delta\omega$ ($\Delta\lambda$), занимаемый группой линий. Более объективной характеристикой является не абсолютная, а относительная ширина спектра $\Delta\omega/\omega_0 = \Delta\lambda/\lambda_0$, где ω_0 и λ_0 — угловая частота и длина волны, соответствующие максимуму спектра. Величина $\Delta\omega/\omega_0$ называется *спектральной чистотой излучения*.

Ширина спектральной моды, выделяемой резонатором, определяется его добротностью:

$$\Delta\omega_c = \omega_0/Q = 1/t_\phi, \quad (2.71)$$

где t_ϕ — время жизни фотона в резонаторе.

В свою очередь величины Q и t_ϕ определяются потерями в резонаторе. Но резонатор лазера выполняет не только функции пассивного фильтра. Ширина спектральной линии, генерируемой лазером в одномодовом режиме, будет существенно меньше величины, определяемой соотношением (2.71) для пассивного резонатора. Если внутри резонатора находится активная среда, компенсирующая потери на частоте генерации, то для такой моды в режиме генерации $Q \rightarrow 0$ и $\Delta\omega \rightarrow 0$.

Теоретический предел ширины спектральной линии лазерного излучения определяется двумя факторами: 1) шумами, обусловленными тепловым излучением в резонаторе; 2) шумами, связанными со спонтанным излучением активного вещества. В оптическом диапазоне шумы за счет спонтанного излучения преобладают над тепловыми шумами. Если учитывать только шумы, вызванные спонтанными переходами, то окажется, что спектральная линия выходного лазерного излучения имеет лоренцеву формулу (1.148) с полушириной

$$\Delta\omega = 4\hbar\omega_0(\Delta\omega_0)^2/P, \quad (2.72)$$

где P — выходная мощность лазерного излучения (остальные обозначения те же).

В формуле (2.72) предполагается, что лазер генерирует в стационарном непрерывном режиме на одной моде в одночастотном режиме и его параметры не изменяются во времени. Величину $\Delta\omega$ можно рассматривать как теоретический предел ширины спектральной линии генерации.

Для лазера с выходной мощностью $P = 1$ мВт, излучающего в красной области спектра (гелий-неоновый лазер, $\lambda_0 = 0,63$ мкм, $\omega_0/2\pi = 4,76 \cdot 10^{14}$ Гц, $\hbar\omega_0 = 3,15 \cdot 10^{-19}$ Дж) и имеющего параметры резонатора, определенные в § 2.3 ($Q \approx 10^8$), из (2.72) получаем $\Delta\omega/2\pi \approx 0,2$ Гц, что соответствует спектральной чистоте $\Delta\omega/\omega_0 \approx 5 \cdot 10^{-16}$.

Для уяснения смысла такой спектральной чистоты определим требования, накладываемые на стабильность длины резонатора

для поддержания частоты в этих пределах. Из (2.30) для заданной моды находим

$$\Delta L/L = \Delta\omega/\omega_0,$$

т. е. при $L = 1$ м допустимое отклонение длины резонатора от заданной $\Delta L = 5 \cdot 10^{-16}$ м = $5 \cdot 10^{-7}$ нм. Если учесть, что типичные размеры атомов имеют порядок $0,1$ нм = 1 \AA , то, очевидно, стабилизировать длину резонатора в таких пределах весьма проблематично. Это означает, что в реальных условиях монохроматичность лазерного излучения определяется изменениями длины резонатора, вызванными тепловыми эффектами, вибрациями и т. д. В очень жестких условиях стабилизации в течение нескольких секунд было достигнуто $\Delta\omega/\omega_0 \approx 10^{-14}$. При обычных условиях, но с использованием методов стабилизации частоты можно получить спектральную чистоту порядка $10^{-12} \dots 10^{-13}$. В видимой области этому соответствует $\Delta\omega/2\pi \approx 50$ Гц.

Ни один из «классических» источников излучения не способен обеспечить такие параметры. Например, при выделении одной спектральной линии у ртутных или натриевых ламп в лучшем случае достигается отношение $\Delta\omega/\omega_0 \approx 10^{-6}$.

Отметим, что ширина спектра, генерируемого лазером, может быть существенно меньше «естественной» ширины спектральной линии, для которой $\Delta\nu_{\text{ест}} = \Delta\omega_{\text{ест}}/2\pi \approx 20$ МГц (см. § 1.6). На первый взгляд это может показаться парадоксальным, так как $\Delta\nu_{\text{ест}}$ определяется соотношением неопределенностей $\hbar\Delta\nu_{\text{ест}}\tau_{\text{ср}} \approx \hbar$, где $\tau_{\text{ср}}$ — время жизни атома в возбужденном состоянии, ограниченное спонтанными переходами. На самом деле в режиме генерации все активные атомы + резонатор выступают как единое целое, связанное полем излучения. В стационарном режиме благодаря непрерывному поступлению энергии извне (энергии накачки) время жизни этой системы в возбужденном состоянии может быть сколь угодно большим.

Все предыдущие рассуждения относились к одночастотной генерации в стационарном режиме. В большинстве случаев, если не принимать специальных мер по селекции типов колебаний, лазер будет генерировать в многомодовом режиме. В качестве примера на рис. 2.31 представлен спектр генерации на семи аксиальных ТЕМ₀₀-модах и семи неаксиальных ТЕМ₁₀-модах*.

При генерации в импульсном режиме спектр излучения для каждой моды будет уширен до $\Delta\nu_{\text{тис}} \approx 1/\tau_{\text{и}}$, где $\tau_{\text{и}}$ — длительность импульса. Поэтому в режиме генерации гигантских импульсов

*Часто одномодовым режимом генерирования лазерного излучения называют такой режим работы, при котором лазерное излучение содержит только продольные моды в пределах спектра частот данной линии спонтанного излучения. Режим работы лазера, при котором лазерное излучение содержит только одну продольную моду, называют *одночастотным режимом генерирования*.

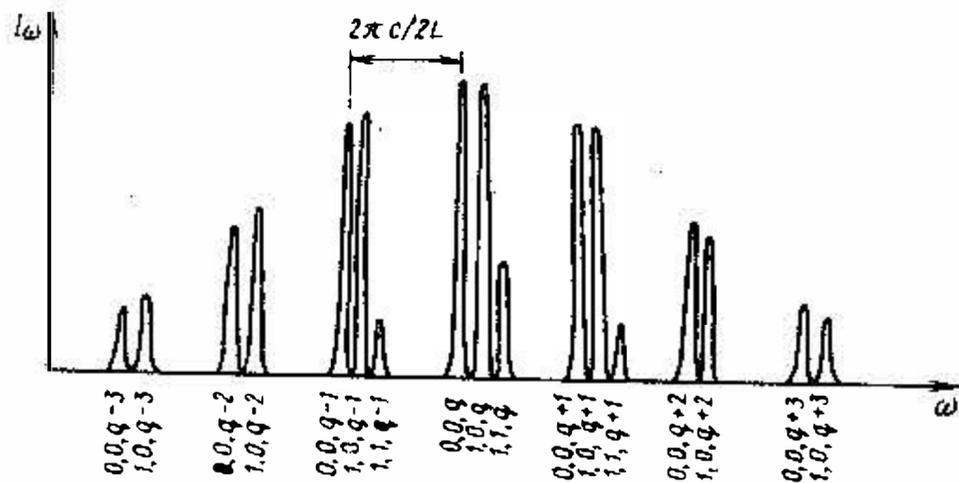


Рис. 2.31. Спектр излучения лазера в многомодовом режиме

при $\tau_p \approx 1$ нс даже при одномодовой генерации $\Delta\nu_{\text{ген}} \approx 1000$ МГц. В пиковом режиме генерации значение $\Delta\nu_{\text{ген}}$ для каждой моды будет определяться длительностью отдельных пиков.

Спектр лазерного излучения также может быть уширен вследствие неоднородности активного материала (например, за счет термических деформаций), что особенно характерно для твердотельных лазеров.

Остановимся теперь на вопросе о поляризации лазерного излучения. Поляризация, как правило, не фигурировала ранее при рассмотрении свойств резонаторов или процессов взаимодействия излучения с веществом. Это означает, что лазер, вообще говоря, может излучать неполяризованный свет. Такой свет, как известно, может быть разложен на две взаимно ортогональные поляризации. Поскольку они генерируются в одной и той же области одними и теми же частицами, то эти две поляризации будут «конкурировать» между собой. Часто это вредит стабильности работы лазера.

Обеспечить работу лазера на одной поляризации и получить на выходе плоскополяризованное излучение весьма просто. Для этого достаточно внутри резонатора ввести потери для одной из двух поляризаций. Например, достаточно поместить в резонатор прозрачную плоскопараллельную пластинку под углом Брюстера к его оси. Такие «окна Брюстера» практически всегда используются в газовых лазерах (см. гл. 7). В твердотельных лазерах на диэлектриках часто существует анизотропия оптических свойств самого активного вещества. Например, излучение рубинового лазера, как правило, поляризовано вследствие его двулучепреломления и несовпадения оптической оси кристалла с осью резонатора (см. гл. VIII).

Когерентность. Когерентность* характеризует согласованность (корреляцию) протекающих во времени и в пространстве

колебательных и волновых процессов (см. определение во введении). Математически степень когерентности описывается с помощью корреляционных функций, а экспериментально определяется по наблюдению интерференционной картины.

Наивысшей (полной) когерентностью обладает идеально монохроматическая линейно поляризованная волна, которая, как отмечалось в § 1.1, является абстракцией и в природе не существует. Наиболее близкими к такой волне характеристиками обладает лазерное излучение. Поэтому говорят, что оно является когерентным. Когерентность лазерного излучения является следствием согласованности процессов вынужденного испускания, лежащих в основе действия лазера.

Излучение «обычных», не лазерных источников определяется процессами спонтанного (самопроизвольного и несогласованного) испускания, поэтому такое излучение является некогерентным. Точнее, излучение классических источников обладает очень небольшой степенью когерентности: из сплошного спектра, даваемого, например, лампой накаливания, с помощью монохроматора можно выделить узкую спектральную линию, а с помощью диафрагмы — малую часть, излучаемую малым объемом источника. При определенных условиях с помощью подобного излучения можно наблюдать интерференционные эффекты, для чего оно должно быть, по крайней мере, частично когерентным.

Для электромагнитной волны можно определить два независимых понятия — пространственную и временную когерентность. Под пространственной когерентностью понимают корреляцию фаз электромагнитных волн, испущенных из двух различных точек источника в одинаковые моменты времени. Под временной когерентностью подразумевают корреляцию фаз электромагнитных волн, испущенных из одной и той же точки источника в различные моменты времени.

Степень взаимной когерентности можно экспериментально оценить по контрасту интерференционной картины χ :

$$\chi = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}, \quad (2.73)$$

где I_{min} и I_{max} — интенсивности в минимуме и максимуме интерференционных полос. Измерив интенсивности I_{max} и I_{min} вблизи выбранной точки экрана, можно определить функцию γ^1 , характеризующую степень взаимной когерентности первого порядка:

$$\gamma^1 = \frac{\chi}{2} \left(\frac{I_1 + I_2}{\sqrt{I_1 I_2}} \right). \quad (2.74)$$

Для наблюдения только пространственной когерентности в точках x_1 и x_2 на рис. 2.32 необходимо положить $\tau = 0$ ($t_1 = t_2$),

*От англ. coherency — согласованность, связность, связь.

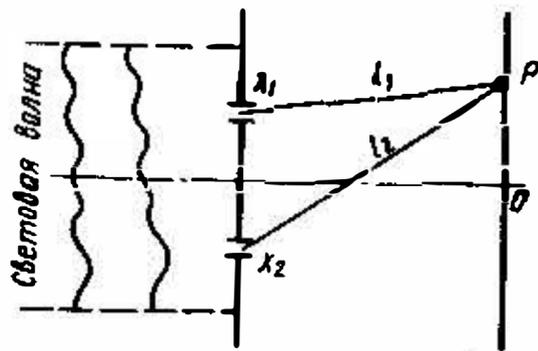


Рис. 2.32. Измерение степени когерентности электромагнитной волны с помощью интерферометра Юнга

т. е. проводить измерения вблизи точки O . Для наблюдения только временной когерентности отверстия x_1 и x_2 должны быть расположены сколь угодно близко (совпадать), но для двух интерферирующих волн должна быть обеспечена задержка во времени на τ , например путем разделения волны от отверстия x_1 на две части с помощью дополнительного полупрозрачного зеркала, как это осуществляется в интерферометре Майкельсона.

Если расположить источники оптического излучения в порядке уменьшения степени когерентности генерируемого ими излучения, то получится примерно такая последовательность: газовые лазеры — жидкостные лазеры — твердотельные лазеры на диэлектриках — полупроводниковые лазеры — газоразрядные лампы — светодиоды — лампы накаливания. При переходе от лазерных источников к нелазерным наблюдается резкий скачок. Поэтому можно считать, что классические (нелазерные) источники света генерируют некогерентное излучение, а лазерные — когерентное.

С когерентностью излучения связана «пятнистая», или «зернистая», картина лазерного света. Ее легко наблюдать визуально при рассмотрении лазерного излучения, рассеянного от экрана или стены. Наблюдаемый рассеянный свет состоит из хаотического скопления ярких и темных пятен, или «зерен». Это явление обусловлено интерференцией волн, рассеянных небольшими рассеивающими центрами, расположенными на поверхности экрана или стены. Чем выше степень когерентности падающего излучения, тем отчетливее проявляются отдельные «зерна» в пятнистой картине.

Направленность и яркость. Направленность определяет расходимость светового пучка в пространстве и характеризуется плоским или телесным углом, в котором распространяется большая часть излучения.

Тепловые источники, излучающие по законам абсолютно черного тела, дают ненаправленное излучение: поверхностная плотность излучаемой электромагнитной энергии не зависит от направления в пределах всего телесного угла 4π . От таких источников, подчиняющихся закону Ламберта, с помощью определенной системы диафрагм, линз и зеркал можно получить направленное излучение. Но эта направленность достигается за счет потерь большей части излучаемой энергии. Чем большую направ-

*В зарубежной литературе каждое зерно называют «стаклом», а саму картину — «стакло-картинкой».

вленность желательно получить от классического источника, тем меньшую долю световой энергии можно использовать.

Лазерное излучение по своей природе обладает высокой степенью направленности. Направленность излучения, генерируемого лазером, определяется свойствами резонатора (см. § 2.3). Расходимость лазерного пучка при генерации на одной аксиальной моде близка к предельно достижимой расходимости, которая ограничивается явлениями дифракции.

Дифракционная расходимость является следствием естественного расширения лазерного пучка по мере движения световых волн в пространстве. Она может быть пояснена с помощью принципа Гюйгенса — Френеля (рис. 2.35). Пусть на диафрагму диаметром d падает идеальная плоская электромагнитная волна с полной пространственной когерентностью. Волновой фронт для нее является плоским. После прохождения отверстия вследствие дифракции волновой фронт искажается и излучение приобретает дифракционную расходимость. Это поясняется рис. 2.33, а, на котором изображен волновой фронт $A'B'$, полученный из волнового фронта AB с помощью принципа Гюйгенса — Френеля. Если наблюдать световую картину за отверстием, например, поместив туда плоский экран, то можно увидеть известную дифракционную картину. Для круглого отверстия она будет представлять собой яркое центральное пятно, окруженное менее яркими концентрическими кольцами. Распределение интенсивности излучения для такой картины показано на рис. 2.33, б. Поскольку 84% энергии излучения, проходящего через отверстие, заключено в центральном пятне, мерой дифракции и мерой расходимости может служить угол θ_D , под которым из отверстия виден диаметр первого темного кольца. Вычисления дают, что при $d \gg \lambda$ угол расходимости

$$\theta_D = 2,44\lambda/d. \quad (2.75)$$

Очевидно, дифракционная расходимость лазерного излучения будет определяться соотношением (2.75) для пучка диаметром d , имеющего «жесткие» края. Для гауссова пучка, который получается для основной TEM₀₀-моды в резонаторе со сферическими зеркалами, распределение интенсивности в поперечном сечении

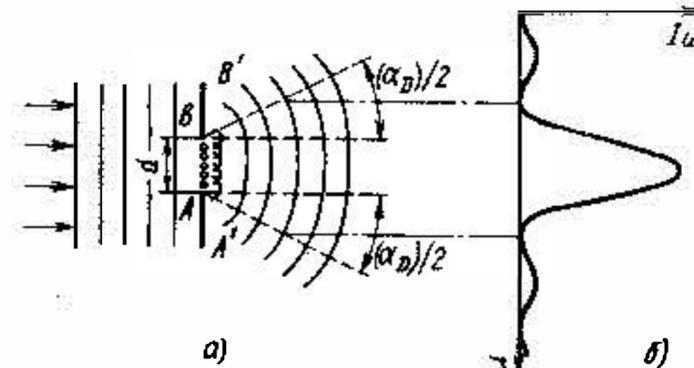


Рис. 2.33. Дифракционная расходимость плоского светового пучка (а) и распределение его интенсивности в поперечном направлении (б)

представляется гладкой экспоненциальной кривой (2.39a). Для такого пучка, как видно из (2.40), (2.41), на больших расстояниях радиус пятна r_{1k} увеличивается линейно с координатой z : $r_{1k} \approx \lambda z / (2\pi r_0)$, где r_0 — радиус в шейке пучка внутри резонатора при $z=0$, т. е. $r_0 = (\lambda L / 4\pi)^{1/2}$. Поэтому угол дифракционной расходимости гауссова пучка определяется выражением

$$\theta_D \approx \frac{2r_{1k}}{z} = \frac{\lambda}{\pi r_0} = \frac{2\lambda}{\pi d_0}, \quad (2.76)$$

где d_0 — диаметр пучка в наиболее узкой его части.

Область пространства вдоль оси лазерного пучка, расположенная на таком расстоянии от излучателя лазера, начиная с которого диаграмма направленности лазерного излучения остается постоянной, называется *дальней зоной лазерного излучения*.

Из сравнения выражений (2.75) и (2.76) видно, что при одинаковых диаметрах расходимость гауссова пучка меньше расходимости плоского пучка. Можно показать, что при прочих условиях гауссов пучок имеет минимально возможную дифракционную расходимость. В общем случае расходимость θ_D пространственно-когерентной волны можно записать в виде

$$\theta_D = \beta \lambda / d, \quad (2.77)$$

где β — числовой коэффициент порядка единицы; его точное значение зависит от распределения амплитуды поля, а также от способа, которым определены θ_D и d . Такой пучок называется *дифракционно-ограниченным*.

Соотношения (2.75) — (2.77) применимы только к пространственно-когерентной волне. У электромагнитной волны, не обладающей полной пространственной когерентностью, расходимость больше, чем у пространственно-когерентной волны, имеющей такое же распределение интенсивности. Это можно видеть из рис. 2.33, а: если волна не обладает пространственной когерентностью, то вторичные волны, испускаемые с поперечного сечения AB , не будут находиться в фазе и волновой фронт, образующийся в результате дифракции, будет иметь большую расходимость, чем θ_D в формуле (2.75).

В некоторых случаях, характерных для твердотельных лазеров на диэлектриках и инжекционных полупроводниковых лазеров, генерация происходит не равномерно по всему объему активного вещества, а в областях, образующих отдельные нити. Диаметр a этих нитей или «горячих точек» существенно меньше общего диаметра d светового пучка. Если отдельные пучки взаимно некоррелированы, то расходимость пучка в целом $\sim \beta \lambda / a$. Если бы такие пучки были коррелированы, то общий угол расходимости $\sim \beta \lambda / d$. Последнее эквивалентно множеству синхронно излучающих атомов.

Высокая направленность лазерного излучения определяет и его высокую яркость. Яркость источника электромагнитных волн есть мощность излучения, испускаемого с единицы поверхности в единичном телесном угле в направлении, перпендикулярном излучательной поверхности. Она имеет размерность ($\text{Вт}/\text{м}^2 \cdot \text{ср}$) и называется также *энергетической яркостью*.

Наряду с яркостью для описания энергетических характеристик лазерного излучения часто используют такие понятия, как *мощность излучения* (непрерывная и импульсная) и *энергия излучения*. Последняя характеризует энергию, переносимую световым потоком, и применима обычно к импульсным лазерам. Она определяет энергию электромагнитного поля в одном лазерном импульсе. Но энергетическая яркость лазерного излучения часто является более важной характеристикой, чем мощность, энергия или интенсивность излучения. Это связано с тем, что во многих практических случаях интерес представляет интенсивность, которую можно получить при фокусировке лазерного излучения линзой. А эта величина пропорциональна яркости пучка. Существенным является то, что хотя интенсивность пучка можно увеличить (например, за счет фокусировки), его яркость при этом остается без изменений.

Кроме энергетической яркости, являющейся объективной физической характеристикой электромагнитного излучения, вводится понятие *фотометрической яркости*. Фотометрическая яркость служит для оценки эффективности воздействия света на глаз человека. Эта эффективность воздействия определяется многими факторами и может быть различной у разных людей, зависеть от уровня освещенности (дневное зрение, ночное зрение), психологического состояния и т. д. Но наиболее сильной является ее зависимость от длины волны. Эта зависимость, полученная путем усреднения многочисленных данных, была утверждена Международной комиссией по освещению (МКО) и качестве стандартной кривой видности. Она приведена на рис. 2.34 и выражает чувствительность глаза к свету с различными длинами волн.

Таким образом, переход от энергетических величин к фотометрическим осуществ-

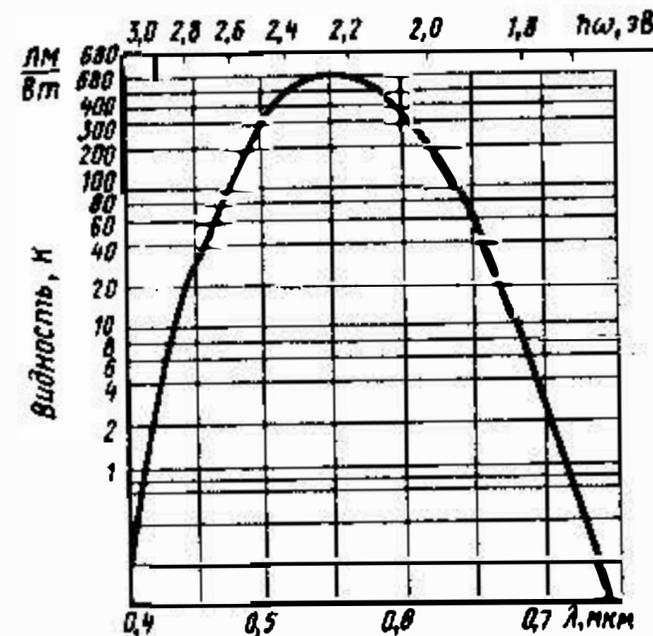


Рис. 2.34. Кривая видности для дневного зрения, определенная Международной комиссией по освещению

вляется через коэффициент K_λ , зависящий от длины волны. Этот коэффициент является световым (фотометрическим) эквивалентом потока излучения (энергетического) и называется *спектральной световой эффективностью монохроматического излучения* или *видностью*. Для данной длины волны фотометрическая яркость $V_{\nu,\lambda}$ связана с энергетической яркостью $V_{E,\lambda}$ соотношением

$$V_{\nu,\lambda} = K_\lambda V_{E,\lambda}. \quad (2.78)$$

Для немонахроматического света связь между V_ν и V_E , очевидно, будет определяться интегралом

$$V_\nu = \int V_{\nu,\lambda} d\lambda = \int K_\lambda V_{E,\lambda} d\lambda. \quad (2.79)$$

Аналогично полный световой (фотометрический) поток Φ_ν источника света связан со спектральной плотностью потока излучения соотношением

$$\Phi_\nu = \int K_\lambda \Phi_\lambda d\lambda.$$

Для нормального дневного зрения максимум функции видности приходится на длину волны $\lambda = 555$ нм (зеленый свет). На этой длине волны, как видно из рис. 2.34, мощность излучения 1 Вт эквивалентна световому потоку 680 лм, т. е. $K_{\lambda=555\text{ нм}} = K_{\text{max}} = 680$ лм/Вт. При $\lambda_{\text{min}} = 380$ нм и $\lambda_{\text{max}} = 780$ нм функция видности уменьшается почти до нуля. Эти длины волн приняты за фиолетовую и красную границы видимой области спектра.

Для оценки яркости лазерного излучения рассмотрим мало-мощный гелий-неоновый газовый лазер, генерирующий в красной области спектра на длине волны $0,63 \text{ мкм} = 630$ нм мощность 10^{-3} Вт. Диаметр пучка на выходном зеркале равен 3 мм. Угол расходимости определяется дифракционными эффектами, и из (2.75) получаем $\alpha_D = 2,44 \cdot 0,63 \cdot 10^{-6} / 10^{-3} \approx 10^{-3}$ рад. Соответствующий телесный угол пучка равен 10^{-6} ср. Выходной мощности 1 мВт на длине волны 630 нм соответствует $K_\lambda = 180$ лм/Вт, т. е. световой поток 0,18 лм. Разделив эту величину на площадь источника и телесный угол, получим значение яркости $V_\nu \approx 2,5 \cdot 10^{10}$ кд/м². Для того чтобы оценить эту величину, укажем, что средняя яркость солнечного диска за пределами земной атмосферы составляет $2 \cdot 10^9$ кд/м². Это означает, что яркость излучения, генерируемого даже маломощным лазером с выходной мощностью 1 мВт, превышает яркость солнечного излучения примерно в 10 раз. Если еще учесть, что солнечное излучение занимает широкий спектральный диапазон, а лазерное излучение обладает высокой степенью монохроматичности, то окажется, что спектральная яркость лазера V_λ будет на 10...12 порядков превышать спектральную яркость Солнца.

ГЛАВА 3

РАСПРОСТРАНЕНИЕ И ПРЕОБРАЗОВАНИЕ ЛАЗЕРНЫХ ПУЧКОВ

Для эффективного использования лазерного излучения необходимо знать его свойства, особенности распространения и возможности преобразования. Управление распространением лазерного пучка, т. е. его пространственное преобразование, является одной из важных задач, которую приходится решать на практике. В отличие от электронов, управление движением которых осуществляется с помощью внешних электрических и магнитных полей, фотоны непосредственно не реагируют на эти поля. Единственным способом остается использование отражения и преломления света на границах раздела или в общем случае явлений рефракции в неоднородных структурах.

Преобразования, связанные с изменением частоты (длины волны) лазерного излучения, основаны на использовании процессов нелинейного взаимодействия света с веществом, которые будут рассмотрены в гл. 5.

§ 3.1. РАСПРОСТРАНЕНИЕ ГАУССОВЫХ ПУЧКОВ

Напомним, что гауссовым пучком мы назвали такой пучок, распределение поля в поперечном сечении которого описывается функцией Гаусса (2.39а). Гауссовы пучки важны с практической точки зрения, поскольку для них наблюдается минимальная дифракционная расходимость (коэффициент β в (2.77) минимален) и такие пучки ближе к реальности, нежели плоские волны. Кроме того, гауссовы пучки весьма удобны для теоретического анализа, поскольку функция Гаусса — единственная функция, инвариантная по отношению к преобразованию Фурье (1.26), (1.27).

Распространение гауссова пучка в свободном пространстве мы начали обсуждать ранее в § 2.3, рассматривая свойства TEM₀₀-моды кофокального резонатора. Перепишем соотноше-

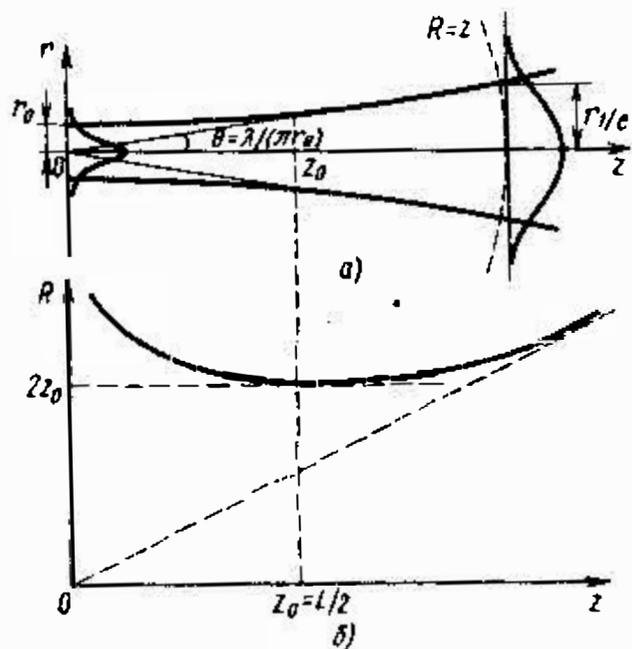


Рис. 3.1. Гауссов пучок в свободном пространстве.

а) зависимость ширины пучка $r_{1/e}$ от расстояния z и распределение интенсивностей в плоскостях $z=0$ и z ; б) зависимость радиуса кривизны сферического волнового фронта пучка R от расстояния. В точке $z=z_0$ радиус $R=R_{\min}=2z_0$, а $r_{1/e}=\sqrt{2}r_0$

(3.2). Основные характеристики гауссова пучка, распространяющегося в свободном пространстве, показаны на рис. 3.1. Важно, что вдоль всей траектории распространения пучок остается гауссовым, а волна в пучке имеет почти сферический волновой фронт с радиусом кривизны R^* . При больших z гиперболоид [см. формулу (3.2)], сечение которого показано на рис. 3.1, а, асимптотически переходит в конус с половинным углом при вершине

$$\theta = \arctg\left(\frac{\lambda}{\pi r_0}\right) \approx \frac{\lambda}{\pi r_0}, \quad (3.3)$$

который в § 2.6 определен как мера угловой расходимости. Отметим, что согласно теории дифракции волна, ограниченная в поперечном направлении апертурой радиуса r_0 , будет расходиться (дифрагировать) в дальней зоне при $z \gg \pi r_0^2/\lambda$ точно в соответствии с (3.3). Это означает, что любой пучок, ограниченный размером r_0 , не может иметь расходимость меньше (3.3) при его распространении в свободном пространстве.

Рассмотрим теперь прохождение гауссова пучка ТЕМ₀₀-моды

ния (2.40) — (2.42) для размера лазерного пятна $r_{1/e}$ и радиуса кривизны R поверхностей равных фаз в виде

$$r_{1/e}^2 = r_0^2 \left[1 + \left(\frac{\lambda z}{\pi r_0^2} \right)^2 \right], \quad (3.1)$$

$$R = z \left[1 + \left(\frac{\pi r_0^2}{\lambda z} \right)^2 \right]. \quad (3.2)$$

Отсюда следует, что гауссов пучок однозначно определен, если заданы: длина волны монохроматического излучения λ , минимальный размер пучка r_0 (радиус перетяжки) и его положение (плоскость $z=0$). Тогда размер пучка $r_{1/e}$ и его радиус кривизны в любой плоскости z определяются из (3.1),

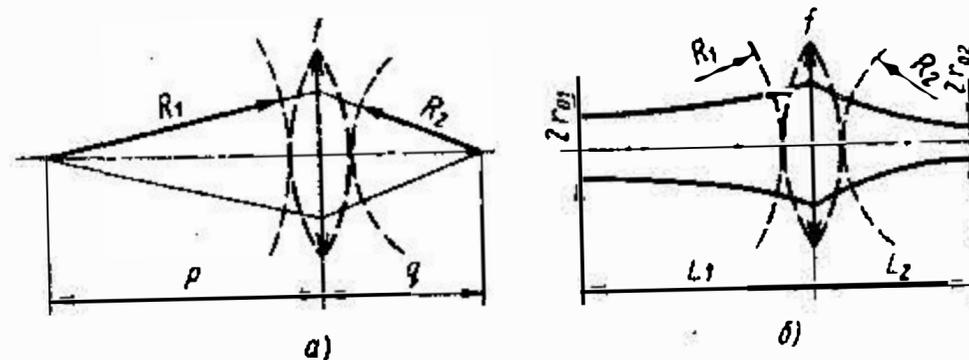


Рис. 3.2. Прохождение через тонкую линзу: а) сферической волны; б) гауссова пучка

через линзу или систему линз. Пусть мы имеем идеальную тонкую линзу. Согласно определению классической оптики, идеальная линза — это оптическая система, преобразующая одну сферическую волну в другую. В частном случае плоская волна может быть преобразована в сферическую или наоборот. Рассмотрим сначала прохождение через такую линзу идеальной сферической волны, испускаемой точечным источником P_1 . Эта волна фокусируется линзой в точку изображения P_2 , как показано на рис. 3.2, а. Из геометрической оптики хорошо известно соотношение

$$1/p + 1/q = 1/f, \quad (3.4)$$

где f — фокусное расстояние линзы (при $p \rightarrow \infty$, $q \rightarrow f$). Из рис. 3.2, а видно, что радиусы волновых фронтов R_1 и R_2 непосредственно до и после прохождения линзы составляют: $R_1 = p$; $R_2 = -q^*$. Отсюда имеем

$$1/R_1 - 1/R_2 = 1/f. \quad (3.5)$$

Из рис. 3.2, а нетрудно понять, что для осуществления функции преобразования одного сферического волнового фронта в другой линза должна выполнять фазовую задержку, которая зависит от расстояния до оси линзы. Обычная сферическая стеклянная линза осуществляет такую задержку за счет того, что ее толщина меняется по радиусу. А так как скорость света в материале, из которого изготовлена линза, в n раз меньше, чем в воздухе, то тем самым и достигается фазовая задержка, изменяющаяся по радиусу. Возможно реализовать и другой тип линзы, представляющий собой плоскопараллельную пластинку, изготовленную из материала, показатель преломления которого изменяется с координатой по заданному закону. Такие линзы называются *градиентными линзами*. Из рис. 3.2, а нетрудно заключить, что

* На самом деле фронты волн параболические, за исключением области плоскости $z=0$. Но для $z \gg r_0^2/\lambda$ отличия параболических поверхностей от сферических несущественны.

* Радиус кривизны принято считать положительным, если центр кривизны находится слева от волнового фронта, и отрицательным, если он находится справа от него, подразумевая направление волны слева направо.

для рассматриваемого случая фазовая задержка должна квадратично изменяться по мере удаления от центра линзы, т. е. при прохождении линзы волна должна приобретать фазовый множитель вида

$$e^{\pm ikr^2/2f}, \quad (3.6)$$

где знак «+» относится к собирающей, а знак «-» — к рассеивающей сферической линзе с фокусным расстоянием f .

Таким образом, сферическая линза преобразует радиус кривизны R_1 падающей волны в радиус кривизны R_2 выходящей волны. На рис. 3.2, а использовано приближение геометрической оптики и волновая природа света не учтена. Пусть теперь на ту же линзу падает гауссов пучок с размером пятна на поверхности линзы $r_1^{(1)} = r_1$ и радиусом кривизны волнового фронта R_1 , как показано на рис. 3.2, б. Очевидно, после прохождения идеальной тонкой линзы амплитудное распределение пучка на ее выходе будет таким же, как на входе, т. е. сохранится гауссово распределение с $r_1 = r_2$. Радиус кривизны R_2 выходящего гауссова пучка, показанного на рис. 3.2, б, будет также определяться формулой (3.5). Следовательно, мы знаем как амплитудное, так и фазовое распределение поля волны на выходе линзы. В обозначениях на рис. 3.2, б с учетом (3.1), (3.2) имеем

$$r_2^2 = r_1^2 = r_{01}^2 \left[1 + \left(\frac{\lambda L_1}{\pi r_{01}^2} \right)^2 \right], \quad (3.7)$$

$$R_2 = L_1 + \left[1 + \left(\frac{\pi r_{01}^2}{\lambda L_1} \right)^2 \right]. \quad (3.7a)$$

Зная размер пятна r_2 из (3.7) и радиус кривизны R_2 волнового фронта из (3.7a) и (3.5) непосредственно после линзы, можно вычислить соответствующие величины в любой точке пространства. В частности, размер пятна r_{02} перетяжки на выходе и место этой перетяжки L_2 можно найти из соотношений (3.5), (3.7). После несложных математических преобразований получаем

$$L_1 = f \pm (r_{01}/r_{02}) (f^2 - f_0^2)^{1/2}, \quad (3.8)$$

$$L_2 = f \pm (r_{02}/r_{01}) (f^2 - f_0^2)^{1/2}, \quad (3.8a)$$

откуда находим r_{02} и L_2 . Здесь

$$f_0 = \pi r_{01} r_{02} / \lambda. \quad (3.8b)$$

Приведенные соотношения полезны при рассмотрении распространения гауссовых пучков. Отметим, в частности, неочевидный

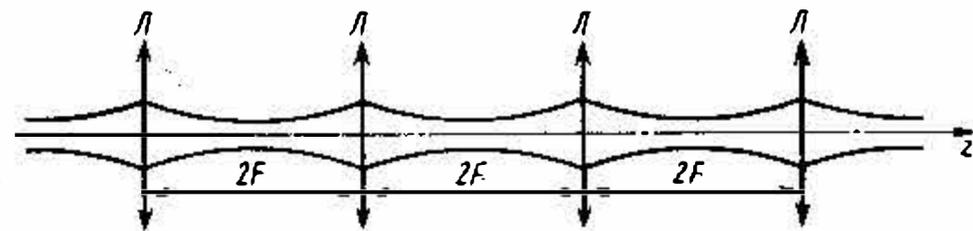


Рис. 3.3. Линзовый конфокальный волновод (световод)

результат: если местоположение первой перетяжки совпадает с фокальной плоскостью линзы ($L_1 = f$), то вторая перетяжка будет расположена также в фокальной плоскости ($L_2 = f$), что не совпадает с результатом, полученным в приближении геометрической оптики. В общем случае плоскости обеих перетяжек не связаны соотношением геометрической оптики, т. е. $1/L_1 + 1/L_2 \neq 1/f$.

Зная результат преобразования гауссова пучка одной линзой, можно рассмотреть егохождение через систему линз. Например, можно образовать линзовый волновод, как показано на рис. 3.3, который компенсирует естественную расходимость дискретно, можно предложить компенсировать её непрерывно. Для этого, как следует из предыдущих рассуждений, необходимо использовать среду распространения с изменяющимся показателем преломления $n = n(r, z)$. Из (3.6) ясно, что закон изменения показателя преломления должен быть квадратичным:

$$n(r, z) = n_0 - n_2/2 r^2. \quad (3.9)$$

Каждый элемент такой среды толщиной dz будет выполнять функцию линзы, компенсируя дифракционную расходимость пучка, как показано на рис. 3.3. Полная компенсация расходимости гауссова пучка будет достигнута при выполнении условия

$$n_0/n_2 = z_0^2 = (\pi r_0^2/\lambda)^2 \quad (3.10)$$

для пучка размером

$$r_0^2 = \lambda/\pi \sqrt{n_0/n_2}. \quad (3.11)$$

Следовательно, в среде, показатель преломления которой изменяется по закону (3.9), гауссов пучок с характерным размером, определяемым соотношением (3.11), будет распространяться не расходясь и иметь плоский фронт на всем протяжении.

До сих пор мы рассматривали распространение гауссова пучка, образованного идеальной монохроматической волной. Такой волны, как отмечалось в § 2.6, в природе не существует, хотя лазерное излучение может быть весьма близко к ней. В общем случае мы имеем дело с пакетом волн разных частот. Вследствие дисперсии показателя преломления каждая частота распростра-

вается с собственной фазовой скоростью $v_\phi = c/n(\omega)$. Поэтому при распространении импульсного сигнала в среде с дисперсией, т. е. в среде с $dn/d\omega \neq 0$, он будет расширяться и искажаться по форме.

Предположим, что вдоль оси z распространяется узкий волновой пакет с резонансной частотой ω_0 и $\Delta\omega \ll \omega_0$. Запишем поле при $z=0$ в виде разложения по монохроматическим волнам, т. е. представим его в виде интеграла Фурье (1.26):

$$E(0, t) = \int_{-\infty}^{\infty} E_0(\omega) e^{i\omega t} d\omega.$$

Так как каждая монохроматическая волна распространяется с собственной скоростью, то поле в точке z будет

$$E(z, t) = \int_{-\infty}^{\infty} E_0(\omega) e^{i(\omega t - k(\omega)z)} d\omega. \quad (3.12)$$

Разложим показатель экспоненты в ряд вблизи $\omega = \omega_0$ и ограничимся первыми двумя членами

$$i(\omega t - k(\omega)z) \approx i(\omega_0 t - k_0 z) + i(\omega - \omega_0) \left[t - \left(\frac{dk}{d\omega} \right)_{\omega_0} z \right].$$

Обозначим

$$v_{гр} = 1 / \left(\frac{dk}{d\omega} \right)_{\omega_0}. \quad (3.13)$$

Тогда соотношение (3.12) принимает вид

$$E(z, t) \approx e^{i(\omega_0 t - k_0 z)} g(t - z/v_{гр}),$$

где

$$g(t) = \int_{-\infty}^{\infty} E_0(\omega) e^{i(\omega - \omega_0)t} d\omega,$$

откуда видно, что величина $v_{гр}$, определяемая соотношением (3.13), представляет собой скорость распространения огибающей g поля $E(z, t)$ и называется групповой скоростью пакета волн. Групповая скорость $v_{гр} = d\omega/dk$ связана с фазовой скоростью $v_\phi = \omega/k$ соотношением

$$v_{гр} = \frac{d\omega}{dk} \approx v_\phi - \lambda \frac{dv_\phi}{d\lambda} = \frac{v_\phi}{1 - (\omega/v) \frac{dv}{d\omega}}, \quad (3.14)$$

справедливым лишь при малых $\Delta\omega/\omega_0$. Групповая скорость $v_{гр}$ (3.14) в приближении геометрической оптики есть скорость

распространения светового луча и соответствует реальной физической скорости распространения светового пучка.

В среде, в которой отсутствует дисперсия (что бывает крайне редко), $\frac{dv}{d\omega} = \frac{dv}{d\lambda} = 0$ и $v_{гр} = v_\phi$. В области нормальной дисперсии

$$\frac{dv}{d\omega} < 0, \quad \frac{dv}{d\lambda} > 0 \text{ и } v_{гр} < v_\phi,$$

а в области аномальной дисперсии (см. § 1.8)

$$\frac{dv}{d\omega} > 0; \quad \frac{dv}{d\lambda} < 0 \text{ и } v_{гр} > v_\phi.$$

§ 3.2. ОПТИЧЕСКИЕ СВОЙСТВА АТМОСФЕРЫ

Рассмотренные в предыдущем параграфе законы распространения гауссова пучка относятся к идеальной однородной среде, в которой отсутствуют поглощение, рассеяние и дисперсия. В реальных средах присутствие этих факторов приводит к ослаблению интенсивности и искажению формы пучка.

Оптические свойства атмосферы в значительной степени зависят от её состояния: влажности, давления, температуры, загрязненности. Они изменяются с высотой и географическим положением. Наличие ветра, завихрений (турбулентностей), восходящих и нисходящих потоков воздуха приводит к тому, что свойства атмосферы постоянно меняются как во времени, так и в пространстве. Показатель преломления воздуха невелик ($n \leq 1,01$) и определяется молекулярной рефракцией. Однако его изменения на уровне $10^{-4} \dots 10^{-3}$, вызванных указанными выше причинами, вполне достаточно для искажения фронта электромагнитной волны и, как следствие, искажения формы и направления распространения светового пучка в атмосфере. Указанные причины вызывают такие явления в природе, как миражи, мерцание звезд, способность видеть за горизонт и т. п. Как правило, эти факторы начинают сказываться на больших расстояниях.

Одной из важных характеристик атмосферы является её оптическое пропускание (прозрачность), которое характеризует способность атмосферы пропускать оптическое излучение заданного спектрального диапазона. Коэффициент пропускания T атмосферы, как любой другой однородной среды в состоянии термодинамического равновесия, в силу (1.170) экспоненциально убывает с расстоянием R

$$T = I/I_0 = \exp(-\beta R), \quad (3.15)$$

где β — показатель ослабления. Ослабление оптического излучения в атмосфере определяется не только его поглощением, но и рассеянием. Рассеяние осуществляется молекулами воздуха и аэрозолями. Поэтому

$$\beta = \beta_a + \beta_p + \beta_m \quad (3.16)$$

где $\beta_a = k_\omega$ — показатель оптического поглощения; β_p — рассеяние молекулами воздуха (как указывалось в § 1.7, такое рассеяние называется рэлеевским) и β_m — рассеяние аэрозолями. Величины β_a и β_p зависят от типа и концентрации молекул, а β_m — от концентрации и размеров частиц аэрозоля. Очевидно, что показатель ослабления β изменяется с высотой, зависит от атмосферных условий и местоположения источника или объекта. Поэтому соотношение (3.15) справедливо лишь для небольших расстояний R , в пределах которых свойства атмосферы можно считать неизменными. Иногда соотношение (3.15) в силу его удобства применяют и для больших L , вводя эффективный показатель ослабления β .

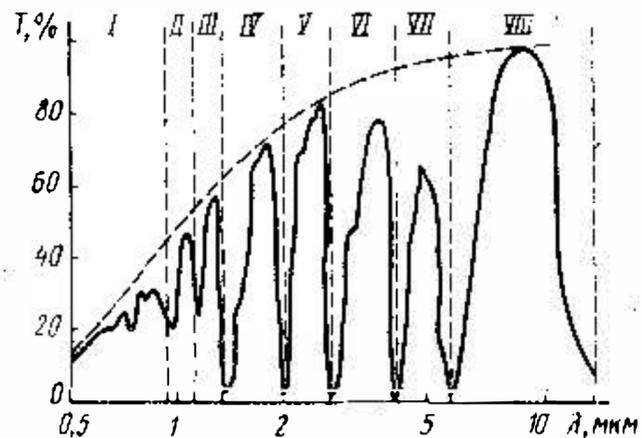


Рис. 3.4. Спектр оптического пропускания слоя атмосферы толщиной 1 км на уровне моря. Обозначены восемь (I—VIII) атмосферных окон пропускания

молекул в многообразном переходе между их колебательно-вращательными состояниями (см. § 1.2). Между полосами поглощения расположены области прозрачности, называемые «атмосферными окнами», где потери определяются лишь процессами рассеяния. Эти области отмечены на рис. 3.4. Очевидно, для передачи излучения через атмосферу с минимальными потерями необходимо выбирать лазер, длина волны которого расположена в окне прозрачности. Обычно это длины волн вблизи 0,7; 1,06; 1,6; 3,9 и 10 мкм. Лазеры, работающие на длинах волн, совпадающих с областью поглощения определенных молекул, могут быть использованы для их количественного контроля в атмосфере.

Вода довольно интенсивно поглощает ИК-излучение, так что в спектре оптического пропускания морской воды имеется одно довольно узкое окно прозрачности в синне-зеленой области спектра от 0,46 до 0,56 мкм. Показатель ослабления β зависит от многих факторов и может изменяться от $\sim 0,01$ до 10 м^{-1} и более.

§ 3.3. ОТРАЖЕНИЕ, ПРЕЛОМЛЕНИЕ И РЕФРАКЦИЯ СВЕТА

Процессы отражения и преломления (рефракции) света при прохождении его через неоднородные среды и границы раздела лежат в основе пространственных преобразователей светового пучка.

Напомним, что *рефракция** света в широком смысле означает преломление световых волн. В узком смысле слова под рефракцией понимают искривление луча в среде с непрерывно изменяющимися характеристиками. В той области частот, где поглощением можно пренебречь, распространение света в среде полностью определяется вещественной частью комплексного показателя преломления \hat{n} , т. е. показателем преломления n , называемым иногда индексом рефракции. Если известен закон изменения n в пространстве, то в приближении геометрической оптики траектория светового луча легко определяется с помощью принципа Ферма (§ 1.1). Например, фокусирующее действие оптической линзы основано на преломлении световых лучей при

Оптическое пропускание атмосферы различно в различных спектральных диапазонах. В коротковолновой УФ-области пропускание ограничено поглощением молекулами озона O_3 , рэлеевским (молекулярным) рассеянием, для которого $\beta_r \sim (1/\lambda)^4$, а также рассеянием аэрозолями и пылевидными частицами. На рис. 3.4 показан типичный спектр пропускания атмосферы. Сильное поглощение вблизи длин волн 0,94; 1,13; 1,38; 1,90; 2,7; 4,3 и 6,0 мкм обусловлено парами воды (H_2O), углекислого газа (CO_2) и озона (O_3). Каждая из полос поглощения имеет сложное строение и состоит из множества узких линий, определяемых сложным строением энергетического спектра

прохождении границы раздела двух сред с различными значениями показателя преломления. Эту функцию могут осуществлять в упоминавшиеся в § 3.1 градиентные линзы. Создавая определенный профиль показателя преломления $n(z, y, z)$, можно локализовать световой пучок или осуществить какое-либо другое его пространственное преобразование. Исключительно важное значение в оптоэлектронике и интегральной оптике играют диэлектрические волноводы (световоды), где пучок света распространяется, испытывая многократное отражение и рефракцию.

Рассмотрим прохождение света через границу раздела двух непоглощающих (диэлектрических) сред. Пусть эта граница будет плоской (зеркальной). При падении на такую границу плоской монохроматической волны часть ее отразится, а часть пройдет в другую среду. Амплитуды, фазы и направления распространения отраженной и преломленной (прошедшей) волн определяются согласованием волновых полей в соответствии с граничными условиями. Требование непрерывности фазы приводит к равенству тангенциальных (параллельных границе раздела) составляющих волновых векторов падающей (индекс 1), отраженной (индекс R) и преломленной (индекс 2) волн:

$$(\mathbf{k}_1)_t = (\mathbf{k}_R)_t = (\mathbf{k}_2)_t = \mathbf{k}_t,$$

при этом $|\mathbf{k}| = \omega/v = 2\pi/\lambda = 2\pi n/\lambda_0$, где λ — длина волны света в веществе. В случае изотропных сред нормальные составляющие для отраженной и падающей волн равны по величине и противоположны по направлению: $(\mathbf{k}_R)_n = -(\mathbf{k}_1)_n$. Это допускает лучевую трактовку прохождения границы раздела двух изотропных сред в приближении геометрической оптики, как показано на рис. 3.5, а. При падении пучка света с интенсивностью I_1 из изотропной среды с показателем преломления n_1 на плоскую границу раздела с изотропной однородной средой с показателем преломления n_2 часть его I_R отражается, а часть $I_T = I_2$ проходит (преломляется). При этом:

1) угол отражения θ_R равен углу падения θ_1 :

$$\theta_R = \theta_1; \quad (3.17)$$

2) частота электромагнитных колебаний не меняется и произведение $n \sin \theta$ одинаково как для падающего (индекс 1), так и для преломленного (индекс 2) лучей:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2,$$

т. е. выполняется закон преломления Снелля (Снеллауса)

$$\sin \theta_2 / \sin \theta_1 = v_2 / v_1 = n_1 / n_2 = n_{12} = 1/n_{21}, \quad (3.18)$$

где n_{12} и n_{21} есть относительные показатели преломления;

*От лат. *refractio* — преломление.

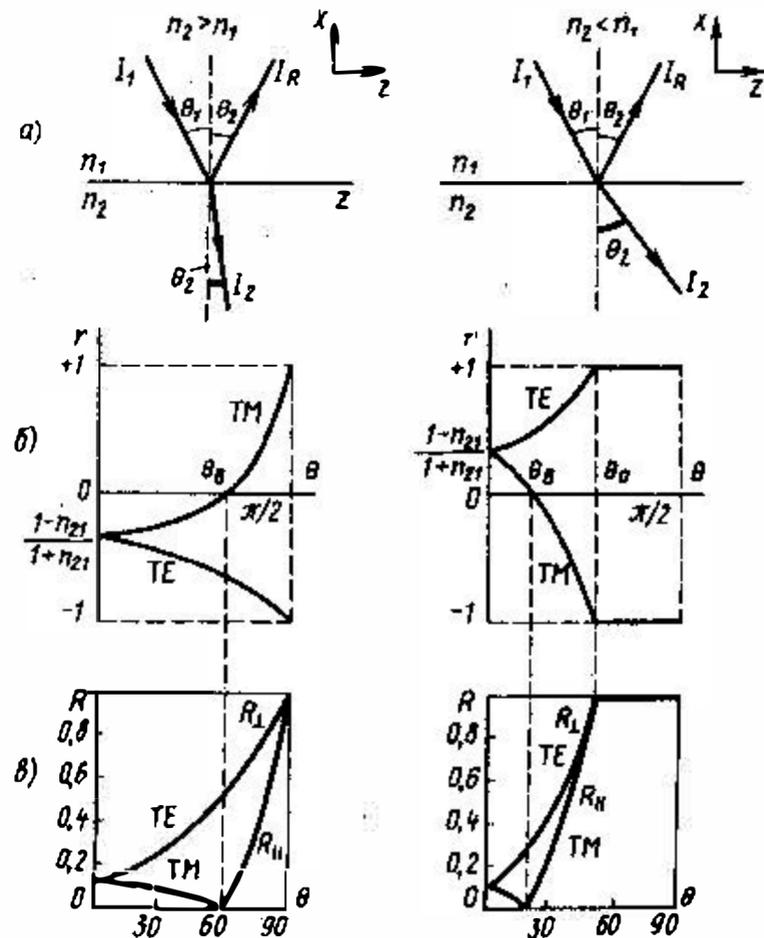


Рис. 3.5. Прохождение границы раздела двух изотропных сред: левая колонка — из оптически менее плотной в оптически более плотную среду ($n_1 < n_2$); правая — из оптически более плотной в оптически менее плотную среду ($n_1 > n_2$)

3) падающий, отраженный и преломленный лучи лежат в одной плоскости с нормалью к границе раздела (для изотропных сред);

4) интенсивность как отраженного I_R , так и прошедшего I_T света зависит как от его поляризации, так и от угла падения θ_1 . Эти зависимости устанавливаются соотношениями Френеля.

Напомним, что электромагнитное поле можно представить линейной комбинацией двух типов волн, для которых: 1) вектор напряженности электрического поля E перпендикулярен плоскости падения, т. е. лежит в плоскости границы раздела; эта волна обозначается ТЕ и называется поперечной электрической волной (или волной с перпендикулярной поляризацией) и 2) вектор напряженности магнитного поля H перпендикулярен плоскости падения; эта волна обозначается ТМ и называется поперечной магнитной волной (или волной с параллельной поляризацией).

Такое разделение удобно, поскольку для заданной симметрии задачи система уравнений Максвелла (1.3) разделяется на две

независимые системы трех уравнений, каждая из которых содержит для волны ТЕ — E_y , H_x и H_z , а для волны ТМ — H_y , E_x и E_z .

Решения задачи о прохождении электромагнитной волны через рассматриваемую границу раздела хорошо известны. Коэффициенты отражения r и пропускания t (по амплитуде) зависят от угла падения θ_1 и поляризации и даются формулами Френеля: для волны ТЕ

$$r_{TE} = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} = -\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)}, \quad (3.19)$$

$$t_{TE} = \frac{2n_1 \cos \theta_1}{n_1 \cos \theta_1 + n_2 \cos \theta_2};$$

для волны ТМ

$$r_{TM} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} = \frac{\text{tg}(\theta_1 - \theta_2)}{\text{tg}(\theta_1 + \theta_2)}, \quad (3.20)$$

$$t_{TM} = \frac{2n_1 \cos \theta_1}{n_2 \cos \theta_1 + n_1 \cos \theta_2}.$$

Коэффициенты отражения по интенсивности получаются из (3.19), (3.20):

$$R_1 = |r_{TE}|^2 = \frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)}, \quad (3.21)$$

$$R_1 = |r_{TM}|^2 = \frac{\text{tg}^2(\theta_1 - \theta_2)}{\text{tg}^2(\theta_1 + \theta_2)}. \quad (3.22)$$

При нормальном падении луча ($\theta_1 = 0$) понятие плоскости падения теряет смысл и для любой поляризации

$$R = \frac{I_R}{I_1} = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2} = \left(\frac{1 - n_{21}}{1 + n_{21}} \right)^2. \quad (3.23)$$

Зависимости (3.19) — (3.22) для коэффициентов отражения r и R как функции угла падения приведены на рис. 3.5, б, в. Из анализа этих зависимостей можно сделать следующие выводы.

1. Для волны ТМ при $\theta_1 + \theta_2 = \pi/2$ величина $R_1 = 0$. Угол

$$\theta_1 = \theta_B = \text{arctg}(n_2/n_1), \quad (3.24)$$

при котором коэффициент отражения R_1 равен нулю, называется *углом Брюстера*.

2. При отражении от поверхности оптически менее плотной среды ($n_2 < n_1$, $n_{21} < 1$ — правая колонка рис. 3.5) существует предельный угол падения θ_c

$$\theta_1 = \theta_c = \arcsin n_{21} = \arcsin n_2/n_1, \quad (3.25)$$

для которого $R=1$. Для углов падения $\theta_1 > \theta_c$ при любой поляризации коэффициент отражения по интенсивности равен единице, а коэффициент пропускания — нулю: электромагнитная волна испытывает полное отражение от границы раздела. Поэтому θ_c называется *углом полного внутреннего отражения*.

3. Так как θ_1 и θ_2 вещественны и вещественны функции (3.19), (3.20), то, следовательно, фаза каждой компоненты отраженной и прошедшей волн либо равна фазе соответствующей компоненты падающей волны, либо отличается от нее на π (если r или t отрицательны). Так как t_{TE} и t_{TM} положительны при любых θ_1 , то фаза прошедшей волны всегда равна фазе падающей. Фаза отраженной волны будет зависеть от относительных значений θ_1 и θ_2 . Если $n_2 > n_1$, то $\theta_2 < \theta_1$ и из (3.19) получаем $r_{TE} < 0$, т. е. фазы отраженной и падающей волн для ТЕ-поляризации отличаются друг от друга на π . Для ТМ-волны из (3.20) получаем, что фаза изменится на 180° для случая $\theta_1 + \theta_2 > \pi/2$, когда знаменатель в (3.20) станет отрицательным. Аналогичное рассмотрение нетрудно провести для случая $n_2 < n_1$, когда вторая среда оптически менее плотная, чем первая.

Вернемся теперь к эффекту полного внутреннего отражения. Более полное рассмотрение вопроса о прохождении электромагнитной волны через границу раздела показывает следующее. Если луч падает на поверхность раздела под углом, большим критического угла полного внутреннего отражения: $\theta_1 > \theta_c$, то он отражается не точно в месте падения, а проникает на некоторое расстояние во вторую среду с $n_2 < n_1$ и снова выходит, но на расстоянии D вдоль поверхности раздела под углом, равным углу отражения, как показано на рис. 3.6, а. При этом амплитуда прошедшей волны экспоненциально затухает по мере удаления от границы раздела с показателем затухания

$$\alpha = \frac{2\pi}{\lambda_1} \sqrt{\sin^2 \theta_1 - (n_2/n_1)^2}. \quad (3.26)$$

Следовательно, при углах падения, при которых $\sin \theta_1 > n_2/n_1$, прошедшая (преломленная) волна представляет собой поверхностную волну, амплитуда которой экспоненциально затухает по нормали к поверхности раздела. Расстояние D на рис. 3.6 связано с α соотношением

$$D = 0,52 n_2 2\pi/\alpha. \quad (3.27)$$

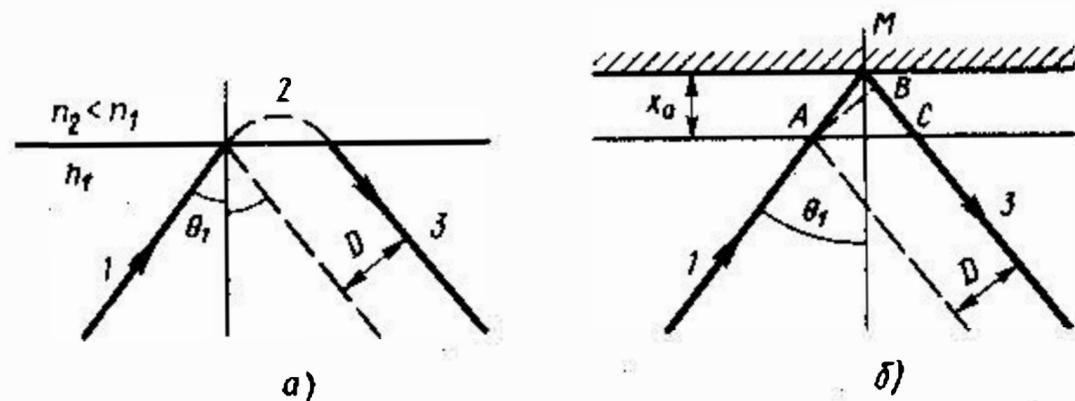


Рис. 3.6. Прохождение светового луча при эффекте полного внутреннего отражения (а) и эквивалентное ему отражение от металлической поверхности (б):

1 — падающий луч; 2 — поверхностная волна; 3 — отраженный луч

Из рис. 3.6, а видно, что при полном внутреннем отражении отраженная волна приобретает фазовый сдвиг $e^{-i\Phi}$, зависящий от угла падения. Эта зависимость может быть получена из (3.26), (3.27):

$$\Phi \approx 2 \arctg \left[\frac{\sqrt{\sin^2 \theta_1 - n_{21}^2}}{\cos \theta_1} \right]. \quad (3.28)$$

Можно показать, что полное внутреннее отражение плоской волны на границе раздела двух диэлектрических сред сопровождается теми же эффектами, что и отражение от идеальной металлической плоскости, смещенной на некоторое расстояние x_0 от поверхности раздела рассматриваемых сред (рис. 3.6, б). Важно, что x_0 зависит от угла падения. Из рис. 3.6, б нетрудно найти:

$$D = 2x_0 \sin \theta_1. \quad (3.29)$$

Фазовый сдвиг $\Delta\phi$, связанный с разностью хода $\delta = AM + MB = 2x_0 \cos \theta_1$ (рис. 3.6, б), равен

$$\Delta\phi = kn_1 \delta = 2x_0 kn_1 \cos \theta_1 = 2x_0 \alpha.$$

К этому сдвигу необходимо добавить скачок фазы, равный π , который появляется при отражении от металлической поверхности. В результате получаем соотношение $\Delta\phi = 2\alpha x_0 + \pi$, которое полезно сравнить с (3.28).

§ 3.4. ПЛАНАРНЫЙ ДИЭЛЕКТРИЧЕСКИЙ ВОЛНОВОД (СВЕТОВОД)

Эффект полного внутреннего отражения, рассмотренный в предыдущем параграфе, широко используется для направленной передачи света по различного рода диэлектрическим волноводам, называемым также световодами.

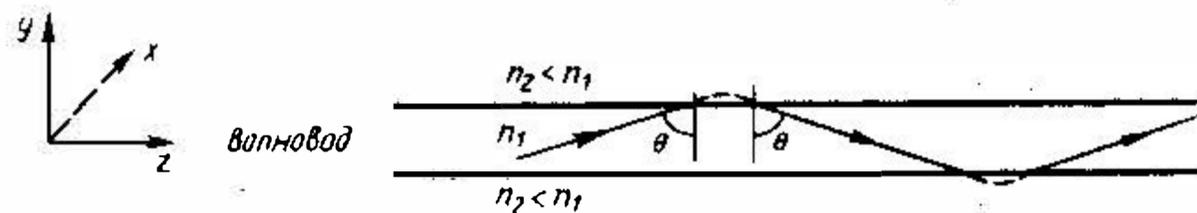


Рис. 3.7. Распространение луча в плоском планарном волноводе

Рассмотрим распространение света по простейшему планарному волноводу, представляющему собой оптически прозрачный диэлектрический слой толщиной d с показателем преломления $n_1 = \sqrt{\epsilon_1}$, расположенный между двумя диэлектрическими полупространствами с показателями преломления $n_2 = \sqrt{\epsilon_2}$ так, что $n_2 < n_1$. Распространение светового луча в таком волноводе схематически показано на рис. 3.7. Световая волна испытывает многократное полное внутреннее отражение и распространяется по зигзагообразному пути, поэтому ее иногда называют *зигзагообразной волной*. Для количественного описания электромагнитной волны, распространяющейся в структуре типа изображенной на рис. 3.7, необходимо решить уравнения Максвелла с соответствующими граничными условиями. Такая задача решается в курсах электродинамики при рассмотрении распространения электромагнитных волн СВЧ-диапазона в соответствующих волноводах. Результаты решения показывают следующее.

Распространяясь по оптическому волноводу и многократно отражаясь от границ раздела, плоская волна накладывается сама на себя. Результирующая волна в волноводе представляет собой сумму полей первоначальной и отраженных волн. Ее интенсивность будет нарастать только в том случае, если основная и отраженная волны складываются в фазе. Это условие приводит к резонансному эффекту. Он означает, что в оптических волноводах (как и в металлических СВЧ-волноводах) могут распространяться лишь волны, у которых распределение электромагнитного поля по сечению волновода имеет вполне определенную структуру. Волна, распределение поля в которой удовлетворяет условию резонанса, называется *волноводной модой*. Каждой из разрешенных для данного волновода мод в геометрическом приближении может быть сопоставлен определенный дискретный угол распространения и эффективный показатель преломления n_{eff} , определяющий скорость движения волноводной моды вдоль оптической оси.

Поясним сказанное с помощью рис. 3.8. Пусть в световоде под углом φ к его оси распространяется плоская волна с вектором распространения (волновым вектором) $\mathbf{k} = k_0 n_1$, совпадающим с направлением оптического луча. Коэффициент распространения света вдоль оси z равен

$$\beta = k_z = k_0 n_1 \cos \varphi, \quad (3.30)$$

а вдоль оси x

$$\chi = k_x = k_0 n_1 \sin \varphi. \quad (3.31)$$

Световая волна в волноводе испытывает многократные отражения. Собственная волноводная мода соответствует замкнутому в сердцевине свету. Для нее коэффициент распространения определяется из условия формирования двух стоячих плоских волн e^{ix} и e^{-ix} , а потому χ принимает дискретные собственные значения. В силу (3.30), (3.31) коэффициент распространения β в направлении z также может принимать только дискретные значения. Следовательно, углы φ и $\theta = \pi/2 - \varphi$ дискретны. Ситуация здесь аналогична той, которая рассматривалась в § 2.3 для плоского резонатора. Однако в нашем случае дополнительно надо учесть фазовый сдвиг (3.28), образующийся при полном внутреннем отражении.

Найдем собственные значения угла φ . Для образования стоячей волны должны интерферировать между собой такие бегущие волны, которые в точке с одной и той же координатой x имеют одинаковые фазы. Это означает, что, когда волна дважды отразилась от границ волновода и вернулась к тому же значению x , ее фаза должна измениться на величину, кратную 2π , т. е.

$$2\pi[2dn_1 \sin \varphi / \lambda_0] - 2\Phi = 2\pi m, \quad (3.32)$$

где Φ определяется (3.28), $m = 0, 1, 2, \dots$ и λ_0 — длина волны в вакууме.

Разрешенные значения $\varphi = \varphi_m$ и их дискретность определяются через m , и из (3.32) имеем

$$\sin \varphi_m = \frac{(m\pi + \Phi)\lambda_0}{2\pi d n_1}. \quad (3.33)$$

Существует максимальное значение m , вытекающее из условия полного внутреннего отражения. Максимальный угол

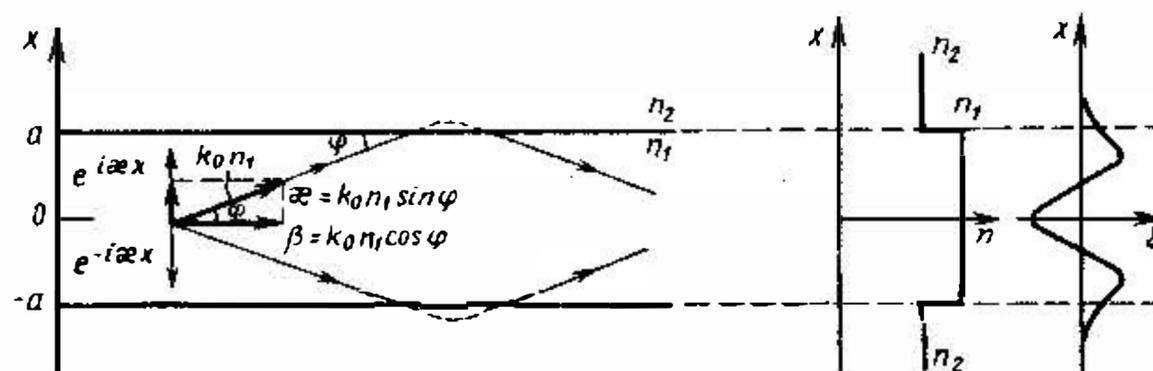


Рис. 3.8. Образование волноводных мод в планарном световоде

$\varphi_c = \pi/2 - \theta_c$, при котором происходит полное отражение, определяется из (3.25) и рис. 3.8 как

$$\varphi_c = \arccos n_2/n_1 = \arcsin \sqrt{(n_1^2 - n_2^2)/n_1^2} = \arcsin \sqrt{2\Delta}. \quad (3.34)$$

Параметр Δ называется удельной разницей показателей преломления и определяется через показатели преломления волновода и оболочек

$$\Delta = (n_1^2 - n_2^2)/2n_1^2 \approx n_1 - n_2/n_1. \quad (3.35)$$

Из (3.33), (3.34), (3.35) находим

$$m \leq \frac{2dn_1}{\lambda_0} \sqrt{2\Delta} - \frac{\Phi}{\pi} = \frac{2V}{\pi} - \frac{\Phi}{\pi}, \quad (3.35a)$$

где $V = \frac{\pi d}{\lambda_0} n_1 \sqrt{2\Delta}$.

Каждому значению m от $m=0$ до $m=m_{\max}$ отвечает своя волноводная мода, распространяющаяся в световоде с минимальными потерями. Каждой моде соответствует свое распределение поля. В качестве примера на рис. 3.9 показаны распределения амплитуд поля для трех волноводных мод с индексами $m=0, 1, 2$ и соответствующие им геометрические приближения. Эти волны могут переносить электромагнитную энергию по световоду на большие расстояния. Заметим, что часть поля (а следовательно, часть полной энергии) ведомой волны просачивается за стенки волновода, т. е. переносятся вытекающими волнами. Скорость распространения волноводной моды вдоль оси z различна для разных мод:

$$v = c/n_{\text{эф}} = c/n_m,$$

где эффективный показатель преломления $n_{\text{эф}} = n_m$ для m^{II} -моды связан с ее постоянной распространения

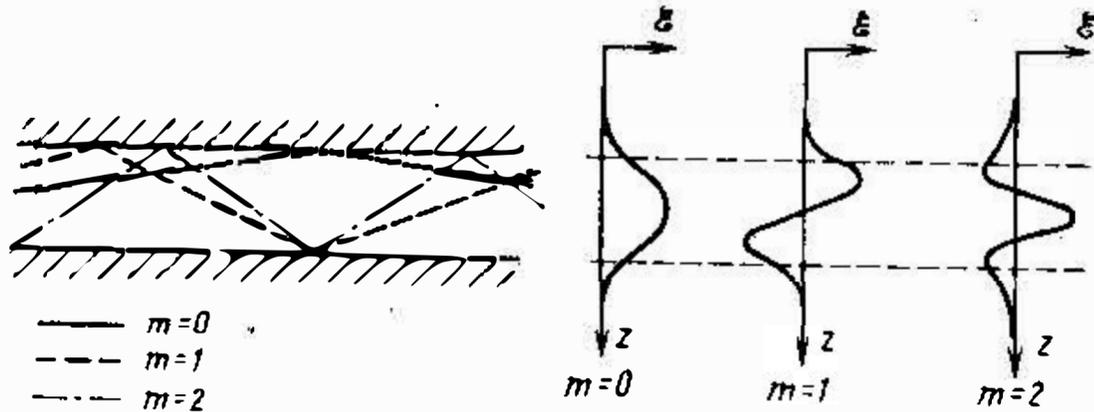


Рис. 3.9. Волноводные моды в плоском оптическом волноводе: справа — распределение электрического поля для наименьших мод $m=0, 1$ и 2 ; слева — условное геометрическое изображение хода лучей

$$n_m = c/v = \frac{\beta_m}{k} = \beta_m c/\omega. \quad (3.36)$$

Постоянная распространения β_m для моды, распространяющейся под углом φ_m к оси волновода, находится из (3.30). Отметим, что для проявления волноводных свойств для мод с малыми φ_m иногда достаточно очень незначительных изменений показателя преломления ($\Delta n \leq 10^{-3}$). Это изменение не обязательно должно быть скачкообразным, оно может быть и плавным.

§ 3.5. РАСПРОСТРАНЕНИЕ СВЕТА В АНИЗОТРОПНЫХ СРЕДАХ И ЭЛЕМЕНТЫ КРИСТАЛЛООПТИКИ

Рассмотрим распространение света в кристаллах. Так как период кристаллической решетки ($\sim 0,5$ нм) во много раз меньше длины световой волны, то кристалл можно рассматривать как однородную, но анизотропную среду. Сначала ограничимся рассмотрением той области частот, в которой отсутствует поглощение, т. е. кристалл прозрачен и ведет себя как диэлектрик.

Уравнения Максвелла и материальные уравнения в виде (1.4) были записаны ранее для изотропной среды, где поляризация вещества не зависела от направления вектора электрического поля E , и потому диэлектрическую проницаемость $\epsilon = \epsilon_0 \epsilon_r$ можно было считать скалярной величиной. Поляризация кристалла в поле световой волны в общем случае может зависеть от направления вектора E , а потому связь между вектором электрической индукции D и напряженностью электрического поля E должна быть записана в виде, позволяющем учесть анизотропию, а именно соотношением, в котором каждая компонента вектора D связана линейно с компонентами E :

$$\left. \begin{aligned} D_x &= \epsilon_0 \epsilon_{xx} E_x + \epsilon_0 \epsilon_{xy} E_y + \epsilon_0 \epsilon_{xz} E_z \\ D_y &= \epsilon_0 \epsilon_{xy} E_x + \epsilon_0 \epsilon_{yy} E_y + \epsilon_0 \epsilon_{yz} E_z \\ D_z &= \epsilon_0 \epsilon_{xz} E_x + \epsilon_0 \epsilon_{yz} E_y + \epsilon_0 \epsilon_{zz} E_z \end{aligned} \right\} \quad (3.37)$$

Девять величин ϵ_{ij} являются постоянными среды (на самом деле они зависят от частоты и температуры) и составляют тензор диэлектрической проницаемости. Вектор D равен произведению этого тензора на вектор E . В анизотропной среде эти векторы неколлинеарны.

Соотношения (3.37) можно записать в более компактной форме:

$$D_i = \epsilon_0 \sum_j \epsilon_{ij} E_j$$

или

$$D_i = \epsilon_0 \epsilon_{ij} E_j$$

В последнем выражении использована формальная тензорная запись, в которой знак суммирования обычно опускают, а указанием на суммирование по всем значениям j служит двукратное появление этого индекса в произведении.

В кристаллооптике доказывается, что из закона сохранения энергии следует симметричность компонент тензора диэлектрической проницаемости, т. е. $\epsilon_{ij} = \epsilon_{ji}$. Там же доказывается, что для любого кристалла можно найти три таких главных направления x , y и z , для которых справедливы соотношения

$$D_x = \epsilon_0 \epsilon_x E_x; \quad D_y = \epsilon_0 \epsilon_y E_y; \quad D_z = \epsilon_0 \epsilon_z E_z.$$

Другими словами, выбором системы координат матрица ϵ_r может быть диагонализирована и приведена к виду

$$\epsilon_r = \begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix}. \quad (3.38)$$

Система координат, в которой тензор диагонализирован, называется *главной системой координат*, а её оси — *главными диэлектрическими осями*. Для кристаллических тел главная система координат обычно совпадает с кристаллографической. Симметрия кристаллов накладывает некоторые ограничения на диагональные компоненты ϵ_x , ϵ_y и ϵ_z , которые называются *главными значениями тензора ϵ_r* . В кубических кристаллах $\epsilon_x = \epsilon_y = \epsilon_z$. Это означает, что в таких кристаллах $D = \epsilon_0 \epsilon E$, где ϵ — скаляр, и по своим оптическим свойствам они изотропны, т. е. эквивалентны аморфному телу. В кристаллах с гексагональной, тригональной и тетрагональной сингониями $\epsilon_x = \epsilon_y \neq \epsilon_z$. В кристаллах, имеющих более низкую симметрию, $\epsilon_x \neq \epsilon_y \neq \epsilon_z$. Их оптические свойства *анизотропны*.

Соответствующие главным значениям тензора диэлектрической проницаемости показатели преломления $n_x = \sqrt{\epsilon_x}$, $n_y = \sqrt{\epsilon_y}$ и $n_z = \sqrt{\epsilon_z}$ называются *главными показателями преломления* (в кристаллооптике обычно рассматривают диэлектрические кристаллы, для которых $\sigma = 0$ и ϵ вещественна).

В общем случае плотность электрической энергии в электромагнитной волне

$$\rho_E = ED/2. \quad (3.39)$$

В главной системе координат

$$2\rho_E = \epsilon_x \epsilon_0 E_x^2 + \epsilon_y \epsilon_0 E_y^2 + \epsilon_z \epsilon_0 E_z^2 = \frac{D_x^2}{\epsilon_0 \epsilon_x} + \frac{D_y^2}{\epsilon_0 \epsilon_y} + \frac{D_z^2}{\epsilon_0 \epsilon_z}, \quad (3.40)$$

откуда следует, что поверхность постоянной плотности энергии ρ_E в пространстве D_x , D_y , D_z имеет форму эллипсоида. Действительно, если обозначить вектор $\frac{1}{\sqrt{2\rho_E \epsilon_0}} D$ через a , то уравнение

для поверхности постоянной плотности энергии (3.40) можно переписать

$$a_x^2/n_x^2 + a_y^2/n_y^2 + a_z^2/n_z^2 = 1. \quad (3.41)$$

Это есть уравнение эллипсоида с главными осями вдоль направлений x , y , z . Длины полуосей эллипсоида равны соответственно n_x , n_y и n_z . Этот эллипсоид определяет поверхность постоянной плотности энергии и называется *эллипсоидом показателя преломления* или *оптической индикатрисой*. Его использование удобно при анализе оптических свойств анизотропных кристаллов.

В изотропных средах фазовая скорость волны, определяемая волновым вектором k , и групповая (лучевая) скорость, определяемая вектором Пойнтинга S , совпадают. В анизотропных кристаллах этого нет. Чтобы показать это, предположим, что в кристалле распространяется плоская монохроматическая волна, характеризуемая фазовым множителем $e^{i\omega(t - n/cr)}$, где n — единичный вектор нормали к волновому фронту, или к плоскости постоянной фазы, а n — показатель преломления (он веществен, так как потери отсутствуют). Волновой вектор этой волны $k = \frac{\omega n}{c}$, а её фазовая скорость

$$V_\phi = \frac{c}{n}. \quad (3.42)$$

Найдем для такой волны вектор Пойнтинга, характеризующий направление распространения потока энергии. При этом учтем, что для плоской монохроматической волны дифференцирование по времени d/dt сводится к умножению на $i\omega$, а действие оператора Δ (дифференцирование по координатам) формально можно заменить векторным умножением на $-\frac{i\omega n}{c}$. Благодаря этому уравнения Максвелла $\text{rot} H = \partial D/\partial t$ и $\text{rot} E = -\partial B/\partial t$ можно преобразовать:

$$D = -\frac{n}{c} [n \times H], \quad H = \frac{n}{\mu_0 c} [n \times S].$$

Эти уравнения показывают, что вектор D перпендикулярен H и оба вектора перпендикулярны n . Вместе с тем вектор Пойнтинга $S = [E \times H]$ перпендикулярен векторам E и H . Фазовая

скорость света v_ϕ направлена по \mathbf{k} , а групповая скорость $v_{гр}$ — по \mathbf{S} . Связь между величинами $|v_{гр}|$ и $|v_\phi|$ определяется углом β между векторами \mathbf{D} и \mathbf{E} :

$$|v_{гр}| = |v_\phi| / \cos \beta. \quad (3.43)$$

В оптически изотропных средах (например, в кубических кристаллах, для которых $\epsilon_x = \epsilon_y = \epsilon_z = \epsilon$) векторы \mathbf{D} и \mathbf{E} коллинеарны и фазовая скорость совпадает с групповой, т. е. $v_\phi = v_{гр}$. В анизотропных средах в общем случае направление потока энергии неколлинеарно направлению распространения фазы, т. е. $v_\phi \neq v_{гр}$. Этот случай поясняет рис. 3.10.

Подчеркнем, что в анизотропных кристаллах понятие «показатель преломления» не имеет такого простого отношения к закону преломления, как в изотропных средах. Расчеты, основанные на решениях уравнений Максвелла, показывают, что в общем случае в анизотропной среде в одном и том же направлении могут распространяться две плоские электромагнитные волны, плоскости поляризации которых взаимно перпендикулярны. Показатели преломления и скорости распространения этих волн различны. Такое явление называется *двулучепреломлением*. Определить два разрешенных направления поляризации и соответствующие им показатели преломления можно с помощью оптической индикатрисы. Для этого через центр эллипсоида проводят плоскость, перпендикулярную направлению распространения. Линия пересечения этой плоскости с эллипсоидом образует эллипс, две оси которого параллельны двум искомым направлениям поляризации, а длины полуосей равны показателям преломления. Отметим, что под направлением распространения здесь подразумевается направление нормали, а не потока энергии к волновому фронту. В кристаллах кубической симметрии оптическая индикатриса представляет собой сферу. Любое её сечение есть окружность, поэтому оптические свойства таких кристаллов не зависят от направления, т. е. являются оптически изотропными. В кристаллах с $n_x = n_y \neq n_z$ такое сечение (окружность) может быть проведено только одно. В кристаллах с $n_x \neq n_y \neq n_z$ таких сечений может быть два. Направления, перпендикулярные этим сечениям, называются *оптическими осями* кристалла (они совпадают с главными диэлектрическими осями), а сами кристаллы — соответственно *одноосными* и *двуосными*. Свет, распространяющийся вдоль оптической оси кристалла, не испытывает двулучепреломления.

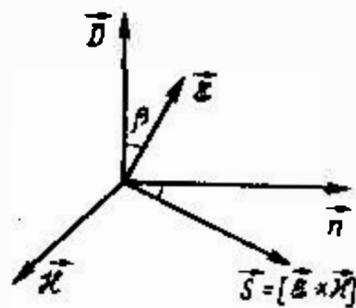


Рис. 3.10. Относительная ориентация векторов \mathbf{D} , \mathbf{E} , \mathbf{k} и \mathbf{S} в анизотропном кристалле (векторы \mathbf{D} , \mathbf{E} , \mathbf{k} и \mathbf{S} лежат в одной плоскости)

В одноосных кристаллах уравнение эллипсоида показателей преломления (3.41) упрощается:

$$a_x^2/n_0^2 + a_y^2/n_0^2 + a_z^2/n_e^2 = 1. \quad (3.44)$$

Здесь оптическая ось направлена по z и совпадает с осью симметрии кристалла. Величина n_0 называется *показателем преломления для обыкновенного луча*, а n_e — *главным показателем преломления для необыкновенного луча*. При $n_e < n_0$ кристалл называют *оптически отрицательным*, а при $n_e > n_0$ — *оптически положительным*.

Рассмотрим прохождение света через границу раздела, образованную анизотропной средой, аналогично тому, как это было сделано в § 3.3 для изотропных сред. Пусть одна из сред (среда n_2 на рис. 3.5), образующих границу раздела, будет обладать оптической анизотропией. Поскольку показатель преломления анизотропной среды зависит от направления электрического поля световой волны, т. е. от направления её распространения и поляризации, то простые закономерности, рассмотренные выше для границы раздела изотропных сред, полностью не выполняются. В частности, при вхождении в анизотропную среду под произвольным углом световой луч будет раздваиваться на две компоненты, соответствующие различным поляризациям, т. е. будет наблюдаться двойное лучепреломление. Это раздвоение будет наблюдаться даже при нормальном падении луча на границу раздела.

Для определенности рассмотрим прохождение волны через плоскопараллельную пластину одноосного кристалла, находящуюся на воздухе ($n_1 \approx 1$) при нормальном падении пучка света на ее переднюю грань. Оптическая ось кристалла составляет угол ϕ с плоскостью раздела, как показано на рис. 3.11. Как указывалось выше, в анизотропной среде в произвольном направлении распространяются две линейно поляризованные волны. Это оз-

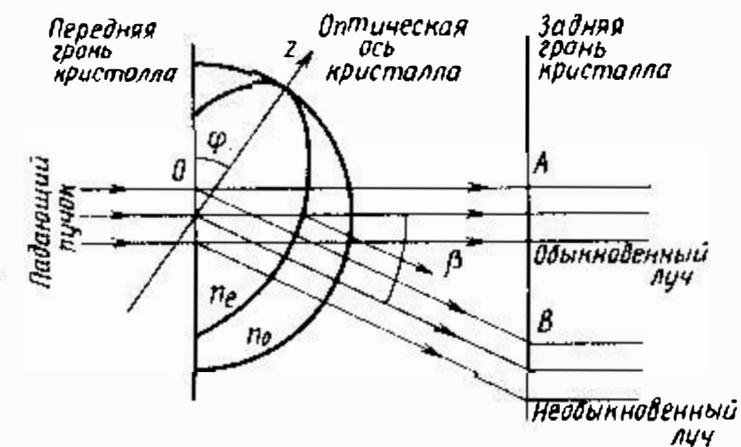


Рис. 3.11. Двойное лучепреломление в одноосном кристалле

начает, что имеется две поверхности волновых векторов. В основном кристалле одна из этих поверхностей — сфера, и соответствующая ей волна имеет сферический фронт. Это обыкновенная волна, и для нее кристалл является как бы изотропной средой. Необыкновенная волна в одноосном кристалле имеет поверхность волновых векторов в виде эллипсоида. Показатель преломления для этой волны зависит от направления ее распространения. Поэтому даже при перпендикулярном падении на границу раздела необыкновенная волна будет отклоняться, как показано на рис. 3.11: оптический путь OB , равный произведению геометрического пути на показатель преломления, для необыкновенной волны короче, чем оптический путь OA , и тем самым выполняется принцип Ферма. Для нахождения направления распространения необыкновенного луча поступаем следующим образом. Зная направление волнового вектора k падающей волны, находим n , проводя из центра O прямую до пересечения с поверхностью показателей преломления (см. рис. 3.11). В точке пересечения падающего луча с поверхностью n_e восстанавливаем перпендикуляр, который указывает направление луча S_e . Для обыкновенной волны векторы k_0 и S_0 параллельны друг другу. Для необыкновенной волны векторы k_e и S_e образуют угол β , который называется углом двойного лучепреломления.

В результате при прохождении границы раздела в одноосном кристалле падающий луч раздваивается, как показано на рис. 3.11. Первый луч (обыкновенный) ведет себя аналогично преломленным волнам в изотропной среде. Вторым луч (необыкновенный) аномален. В общем случае он лежит не в плоскости падения.

Явление двойного лучепреломления широко используется для получения и анализа поляризованного света, для управления поляризационными характеристиками лазерного излучения (§ 3.6), а также для выполнения условий фазового синхронизма в приборах нелинейной оптики (гл. 5).

§ 3.6. ПРЕОБРАЗОВАНИЕ ЛАЗЕРНЫХ ПУЧКОВ

Можно выделить следующие преобразования лазерного излучения: а) пространственное; б) амплитудное; в) поляризационное; г) фазовое; д) частотное и е) временное. Часто все эти преобразования оказываются взаимосвязанными, но иногда их можно осуществить по отдельности. Рассмотрим каждое из них.

Пространственное преобразование. Это преобразование, связанное с изменением пространственного распределения оптического излучения. Оно осуществляется в процессе распространения лазерного пучка либо в свободном пространстве, либо через соответствующую оптическую систему. Законы, по которым происходит преобразование гауссова пучка, были рассмотрены нами

в § 3.1. В качестве иллюстрации применения полученных там результатов обратимся к важному для практики случаю фокусирования лазерного излучения на большом расстоянии от выходного зеркала резонатора, т. е. на большом удалении от перегретки каустики исходного гауссова пучка при $z \gg L/2$ (см. рис. 3.1) или $L_1 \gg \frac{\pi}{\lambda} r_{01}^2$ (см. рис. 3.2, б). Пусть фокусировка осуществляется

короткофокусной линзой с фокусным расстоянием $f \ll L_1$, но $f > f_0$ (см. рис. 3.3, в). Считаем, что диаметр линзы D существенно превышает размеры освещенной области r_1 и r_2 из (3.7а). При этих условиях из (3.7а) получаем уменьшение радиуса пятна в месте перегретки

$$(r_{02}/r_{01}) \approx \sqrt{f^2 - f_0^2}/L_1 \approx f/L_1, \quad (3.45)$$

а из (3.8б)

$$L_2 \approx f. \quad (3.46)$$

Это означает, что при большом удалении от перегретки каустики тонкая положительная линза с $f \ll L_1$ концентрирует излучение гауссова пучка в своей фокальной области, увеличивая его интенсивность в $(L_1/f)^2$ раз.

Полученный результат не является неожиданным, поскольку при больших z фокусируемая волна близка к плоской и, следовательно, она фокусируется в точке, по определению называемой точкой фокуса.

Из (3.45) формально следует, что при $L_1 \rightarrow \infty$ в фокусе линзы получается пятно бесконечно малого радиуса r_{02} . Однако такая экстраполяция недопустима по двум причинам. Во-первых, при $z = L_1 \rightarrow \infty$ из (3.7а) следует, что диаметр освещенного пятна на линзе становится бесконечно большим ($r_1 \rightarrow \infty$), что противоречит исходному предположению о том, что линза не диафрагирует пучок, т. е. не нарушает гауссова распределение. Во-вторых, минимальный размер пятна r_{02} , в который можно сфокусировать оптическое излучение, ограничен дифракционными явлениями. Действительно, угол дифракционной расходимости гауссова пучка (2.87) не может существенно превышать стерadian, откуда имеем для идеального когерентного гауссова пучка

$$r_{02} \geq \lambda/\pi.$$

Поскольку размер пятна на линзе $r_1 = \lambda L_1/(\pi r_{01})$, то очевидно, что формулы (3.45), (3.46) справедливы при условии

$$f > r_1.$$

В совокупности с условием $f \ll L_1$ это означает, что наилучшая пространственная концентрация энергии основной TEM_{00} -моды

лазерного излучения возможна при фокусировании излучения тонкой линзой, расположенной на большом расстоянии от выходного зеркала лазера. При этом фокусное расстояние линзы должно быть меньше этого расстояния, но больше размера пятна на линзе, т. е.

$$L_1 > f > r_2.$$

Нетрудно убедиться, что эти условия легко выполнимы.

Соотношение (3.7а) дает возможность определить размер фокальной области в продольном направлении (вдоль оси распространения z), где наиболее сильно сконцентрирована энергия излучения. Определяя этот размер как расстояние, где плотность энергии излучения уменьшается в два раза относительно точки максимальной концентрации $z=f$, имеем из (3.7а):

$$\Delta z = 2\pi r_0^2 / \lambda. \quad (3.47)$$

При фокусировке излучения с $\lambda=1$ мкм в пятно радиусом ~ 10 мкм плотность потока энергии почти постоянна в почти цилиндрической области длиной 1200 мкм. При этом в фокусе волновой фронт является плоским.

Аналогичным образом возможно рассмотреть и другие пространственные преобразования гауссова пучка. Например, можно провести его расширение с помощью процедуры, обратной рассмотренной выше. Такое расширение бывает необходимо для уменьшения диаграммы направленности пучка. Отметим, что при изменении диаметра пучка в месте перетяжки с помощью оптической системы происходит изменение как его интенсивности, так и его расходимости. При этом яркость остается неизменной (если пренебречь потерями в оптической системе). В связи с этим уместно напомнить оптическую теорему, согласно которой яркость изображения, сформированного любой оптической системой, не может быть выше яркости исходного источника при условии, что и источник и изображение находятся в среде с одним и тем же показателем преломления.

Одним из важных в практическом отношении видов пространственного преобразования лазерного пучка наряду с фокусировкой является изменение его направления распространения. Оно может осуществляться как с помощью обычных оптических элементов типа зеркал, призм, оптических клиньев, уголкового отражателей и т. д., так и с помощью специальных устройств, называемых дефлекторами и предназначенных для быстрого углового перемещения луча аналогично тому, как это осуществляется в электронно-лучевой телевизионной трубке. Эти устройства будут рассмотрены в гл. 10.

Амплитудное преобразование. Можно выделить два вида амплитудного преобразования светового пучка: а) амплитудную мо-

дуляцию, т. е. изменение по заданному закону амплитуды (интенсивности) пучка во времени и в пространстве и б) усиление, т. е. увеличение его интенсивности. Первый вид преобразования осуществляется с помощью оптических модуляторов и рассматривается в гл. 10. Второй вид амплитудного преобразования осуществляется при пропускании пучка через лазерный усилитель или цепочку усилителей.

Рассмотрим, до какого значения может быть увеличена амплитуда лазерного излучения. При малых интенсивностях лазерный усилитель обладает линейными характеристиками, так что интенсивность выходного сигнала пропорциональна интенсивности на входе (2.57). При больших интенсивностях на процессы амплитудного преобразования, как указывалось в § 2.4, существенное влияние начинают оказывать эффекты насыщения, так что при $I \gg I_s$ предельная интенсивность, до которой может быть усилен лазерный пучок, определяется соотношением (2.60). Однако довольно часто в усилителях, обладающих малыми нерезонансными потерями β , теоретический предел (2.60) реализовать невозможно, поскольку максимальное значение светового потока ограничивается другими явлениями, в частности — лучевой (оптической) стойкостью активного элемента. Для твердотельных лазеров на диэлектриках эта величина будет приведена в § 8.3.

Отметим, что ненасыщенный коэффициент усиления усилителя, определяемый при $\beta \leq \alpha$ через (2.57) как

$$K = I_{\text{вых}} / I_{\text{вх}} = \exp(\alpha_0^0 l), \quad (3.48)$$

нельзя делать слишком большим. Иначе могут возникнуть два нежелательных эффекта: паразитные осцилляции и усиленное спонтанное излучение. В практически реализуемых случаях при оптимальном выборе l величина $(\alpha_0^0 l)_{\text{max}}$ достигает значений 3...5.

Поляризационные преобразования. Если лазерный пучок не поляризован, то выделить из него составляющую с заданной поляризацией возможно с помощью поляризатора. Поляризатором называют оптическое устройство, которое преобразует проходящий через него естественный свет в поляризованный. В зависимости от типа создаваемой поляризации поляризатор называют линейным, циркулярным или эллиптическим. Обычно применяют линейные поляризаторы.

Действие поляризатора состоит в том, что он, разделяя первоначальный пучок на две ортогональные по поляризации компоненты, пропускает одну из них и поглощает или отклоняет другую. Для этой цели могут быть использованы оптические явления: дихроизм, двойное лучепреломление, отражение и рассеяние. В видимом и ближнем ИК-диапазонах часто применяют

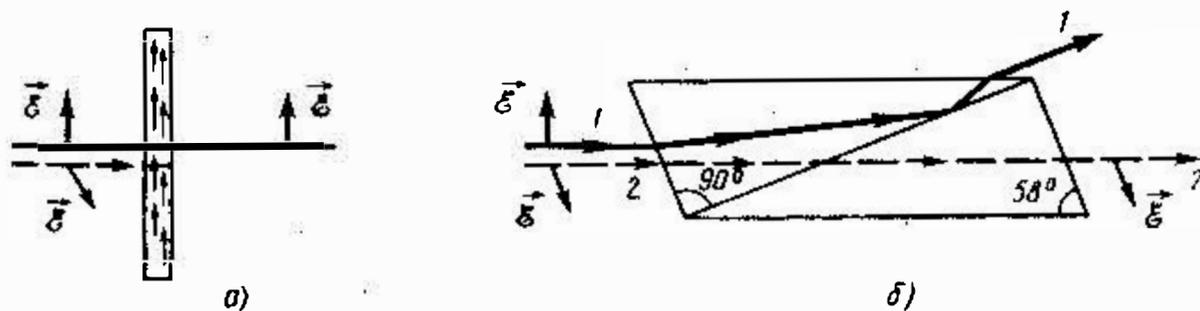


Рис. 3.12. Поляризаторы:
а) пленочный поляризатор; б) призма Николя

дихроичные поляризаторы в виде поляроидных пленок, называемых просто *поляроидами* (рис. 3.12, а). Их действие основано на различии в показателях поглощения для света, поляризованного в различных направлениях. Таким свойством, известным под названием *дихроизм*, обладают некоторые анизотропные кристаллы, а также пленки органических полимеров с молекулами в виде длинных цепочек, почти полностью выстроившихся параллельно друг другу в результате растяжения или какой-либо другой обработки. В качестве такого полимера применяют, например, синтетический поливиниловый спирт, содержащий йод. Поляризующими центрами являются йодвиниловые комплексы, ориентированные в пленке путем её вытягивания. Существуют и другие типы поляроидов, работающих в различных спектральных диапазонах от 0,4 до 2 мкм и более. Хороший поляроид пропускает около 80% света, поляризованного в рабочем направлении, и менее 1% света в перпендикулярном направлении. Спектральный диапазон работы поляроида ограничен областью, где дихроизм используемого материала максимален.

Действие призмных поляризаторов основано на использовании двулучепреломления в оптически прозрачных анизотропных кристаллах. Различие в показателях преломления обыкновенного и необыкновенного лучей (см. § 3.5) в таких материалах позволяет при соблюдении некоторых условий полностью разделить их. Типичный двупреломляющий поляризатор состоит из двух склеенных вместе призм из исландского шпата. Комбинация призм подбирается таким образом, чтобы один из лучей (например, обыкновенный) испытал полное внутреннее отражение на границе раздела, а второй (например, необыкновенный) оказался пропущенным, как это показано на рис. 3.12, б. Такое устройство называют *призмой Николя* или просто *ником**

Наилучшим образом поляризующие устройства работают в параллельном световом пучке. В сходящемся или расходящемся пучке степень поляризации уменьшается. Рабочую *апертуру*

поляризатора определяют как максимально допустимый угол раствора светового конуса, при котором выходящий из поляризатора свет остается линейно поляризованным. Для призмы Николя этот угол составляет около 24°. Луч, падающий вне этого конуса, т. е. под углом более 12° к оси призмы Николя, не будет испытывать полное отражение на границе склейки.

Кроме призмы Николя существует целый ряд других поляризаторов, принцип действия которых основан на использовании двойного лучепреломления в кристаллах, например, призмы Глана — Томпсона, призмы Фуко и т. п. Все они работают, как правило, в видимом и ближнем ИК-диапазонах, что ограничено прозрачностью используемых кристаллов. В средней и далекой ИК-областях спектра используют отражательные поляризаторы. Их принцип действия основан на различии в коэффициентах отражения для лучей, поляризованных в плоскости падения и перпендикулярно этой плоскости. Наиболее сильно это различие наблюдается при угле падения, равном углу Брюстера [см. рис. 3.5, соотношения Френеля (3.19), (3.20), (3.24)]. Пропуская свет через стопу пластин, расположенных под углом Брюстера к падающему лучу, на выходе получим поляризованный свет, вектор E которого параллелен плоскости падения.

Степенью поляризации излучения называют величину

$$V = I_a / (I_a + I_b), \quad (3.49)$$

где I_a — интенсивность полностью поляризованной компоненты, а I_b — интенсивность неполяризованной компоненты, не имеющей постоянного соотношения фаз с компонентой I_a . Компонента I_a может быть линейно, циркулярно или эллиптически поляризована.

Надо сказать, что излучение, генерируемое большинством лазеров (но не отраженное или рассеянное различными предметами), является линейно поляризованным. Причины того обсуждались в § 2.6. Поэтому рассмотренные выше поляризационные устройства в большинстве случаев используются как анализаторы и фильтры.

Если мы имеем какую-либо одну форму поляризации (например, линейно поляризованный свет), то её с высокой эффективностью можно преобразовать в любую другую форму (например, в циркулярно поляризованный свет). Это достигается с помощью так называемых фазовых или волновых пластинок.

Фазовая, или волновая, пластинка является оптическим элементом, который, не меняя интенсивности и степени поляризации монохроматического поляризованного пучка света, расщепляет его на компоненты, сдвигает фазу одной из них относительно другой и затем вновь соединяет эти компоненты в единый пучок. Фазовый сдвиг одной компоненты относительно другой обозна-

*По имени Уильяма Николя, который в 1828 г. разрезал кристалл исландского шпата (см. рис. 3.12, б), склеил обе его части канадским бальзамом и создал высококачественный поляризатор.

чают через δ . Наиболее распространены линейные фазовые пластинки со сдвигом фаз 90 и 180° . Они называются соответственно четвертьволновыми и полуволновыми пластинками, поскольку разность хода Δ между соответствующими волновыми фронтами двух компонент составляет $\lambda/4$ и $\lambda/2$. Естественно, что речь идет об относительном, а не абсолютном изменении фазы. Абсолютное изменение фазы, вызванное введением пластинки внутрь светового пучка, обычно в сотни раз превышает относительный сдвиг фаз δ .

На практике чаще всего используют двупреломляющие или отражательные фазовые пластинки. К первым относятся пластинки из кристаллического кварца или исландского шпата, вырезанные параллельно оптической оси, а также пластинки слюды, пленки ориентированного целлофана или поливинилового спирта, стеклянные пластинки с остаточным натяжением и т. п. Для создания сдвига по фазе может быть использован эффект Фарадея (см. гл. 10). При работе в микроволновом радиодиапазоне в качестве фазовых пластинок используются решетки из металлических проволок, стержней или пластин.

Состояние поляризации прошедшего через фазовую пластинку света зависит как от разности хода (т. е. величины фазового сдвига δ), так и от углов между плоскостью поляризации света и главными направлениями в пластинке. В качестве примера укажем, что если линейно поляризованный свет падает на четвертьволновую пластинку, то в общем случае выходящий свет будет эллиптически поляризован. Если плоскость поляризации падающего луча делит пополам угол между главными направлениями (см. § 3.5), то свет, прошедший через четвертьволновую пластинку, будет поляризован по кругу, т. е. циркулярно поляризован. Таким образом, фазовая пластинка в четверть волны при выполнении определенных ориентационных соотношений преобразует линейно поляризованный свет в циркулярно поляризованный.

Фазовое преобразование. Этот вид преобразования имеет смысл выполнять для лазерного излучения, обладающего высокой степенью когерентности. Он может осуществляться или как фазовая задержка (фазовый сдвиг), или как фазовая модуляция. Задержка по фазе возникает, если изменяется оптическая длина пути света. Поэтому для реализации фазовой задержки достаточно ввести в световой пучок соответствующий оптический элемент, например плоскопараллельную пластинку. Толщина пластинки и ее показатель преломления на заданной длине волны будут определять фазовую задержку.

Для реализации фазовой модуляции необходимо управлять показателем преломления или геометрическими размерами элемента, осуществляющего фазовую задержку. Это может быть достигнуто путем использования рассмотренных в § 10.1 различ-

ного рода электрооптических, магнитооптических и акустооптических эффектов в твердых телах.

Частотное преобразование. Оно осуществляет изменение частоты (длины волны) лазерного излучения. Можно выделить два вида таких преобразований. В первом случае когерентное лазерное излучение одной частоты преобразуется в когерентное излучение другой частоты. Это основной вид частотных лазерных преобразований. Он основан на использовании нелинейных оптических явлений, рассмотренных в гл. 5. Здесь может быть осуществлено как дискретное преобразование частоты, основанное на генерации гармоник (§ 5.2) и использовании вынужденного комбинационного рассеяния, так и ее плавная перестройка в параметрических генераторах света (§ 5.3). Отметим, что КПД таких преобразований может быть достаточно высок, но для их реализации, как правило, требуется лазерное излучение довольно большой интенсивности. Эти преобразования обладают исключительно малой инерционностью.

Второй вид частотных преобразований основан на использовании явлений люминесценции неорганических и органических веществ, в том числе обычной люминесценции под воздействием внешнего возбуждающего излучения, эффектов гашения и стимуляции видимой люминесценции инфракрасным излучением, различного рода каскадных и многоступенчатых процессов и т. п. Для этого вида преобразований когерентность преобразуемого излучения не играет никакой роли. Как правило, такие преобразования используются для визуализации инфракрасного или ультрафиолетового излучения. Их инерционность обычно велика. Примером преобразования инфракрасного излучения в видимое может служить рассмотренный в § 9.1 арсенид-галлиевый светодиод с преобразователем (см. рис. 9.5).

Временное преобразование. Это преобразование изменяет временные характеристики лазерного пучка, как правило, длительность и форму выходного лазерного импульса. Оно осуществляется с помощью нелинейных оптических элементов. Например, пропуская световой импульс через оптический затвор, пропускание которого нелинейно зависит от интенсивности света, получим изменение формы этого импульса. В качестве такого нелинейного элемента может быть использован, например, пассивный фототропный затвор.

Временное преобразование лазерного импульса можно осуществлять теми же методами, которые применяются для амплитудной модуляции.

с использованием адиабатического и одноэлектронного приближений, сущность которых аналогична адиабатическому и одноэлектронному приближениям для многоэлектронных атомов (см. § 1.2).

Основные особенности энергетического спектра идеального кристалла обусловлены трансляционной симметрией кристаллической решетки. Электрон движется в поле $V(\mathbf{r})$, обладающем пространственной периодичностью решетки:

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}), \quad (4.1)$$

где вектор \mathbf{R} определяется через базисные векторы \mathbf{a}_1 , \mathbf{a}_2 и \mathbf{a}_3 , которые образуют элементарную ячейку кристалла:

$$\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3 \quad (4.2)$$

(n_1 , n_2 и n_3 — любые целые числа).

В отличие от атомов и молекул потенциал (4.1) не обладает ни сферической, ни аксиальной симметрией. Поэтому квантовые числа n , l , m , L , S , J и т. д. в кристалле теряют физический смысл. В то же время условие (4.1) позволяет по-новому провести классификацию энергетических состояний в кристаллах.

Из трансляционной симметрии потенциала $V(\mathbf{r})$ вытекает важнейшее свойство волновой функции, описывающей движение электронов в кристалле. Это свойство определяется теоремой Блоха, которая гласит: собственные состояния Ψ одноэлектронного гамильтониана с потенциалом (4.1) могут быть выбраны в виде произведения экспоненциальной функции $e^{i\mathbf{k}\mathbf{r}}$ (плоской волны) на функцию, периодическую с периодом решетки:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \quad (4.3)$$

где

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}).$$

Это означает, что волновая функция электрона, движущегося в периодическом поле, представляет собой модулированную плоскую волну и состояние электрона определено по всему кристаллу. Функции вида (4.3) называют блоховскими функциями, а описываемые ими электроны — блоховскими электронами. Вектор \mathbf{k} в (4.3) называется волновым вектором. Он определяет состояние системы и, следовательно, является квантовым числом для рассматриваемой задачи.

При образовании кристалла из N атомов каждый из атомарных уровней энергии будет расплываться и смешаться вследствие эффекта Штарка, что приведет к образованию соответствующих полос или зон энергии. Число независимых значений волнового вектора \mathbf{k} в пределах одной зоны равно числу элементарных ячеек в объеме кристалла. Расстояние между этими состояниями $\Delta k = 2\pi/Na$ мало

ГЛАВА 4 ОПТИЧЕСКИЕ И ФОТОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

Оптические явления в полупроводниках составляют ту основу, на которой базируется работа как отдельных элементов, так и целых оптоэлектронных и интегрально-оптических систем. В настоящей главе рассматриваются процессы взаимодействия оптического излучения с однородными полупроводниками и возникающие вследствие этого разнообразные оптические явления, которые используются или могут быть использованы в приборах оптической электроники. Оптические и фотоэлектрические явления, возникающие в неоднородных структурах типа p - n -переходов и квантово-размерных слоев, будут рассмотрены в гл. 9, 10 и 11.

§ 4.1. ЭНЕРГЕТИЧЕСКИЕ СОСТОЯНИЯ В ПОЛУПРОВОДНИКОВЫХ КРИСТАЛЛАХ

Процессы поглощения и испускания света в полупроводниках лежат в основе действия большинства приборов полупроводниковой оптоэлектроники. Как и в атомах и молекулах, эти процессы обусловлены квантовыми переходами между разрешенными электронными состояниями. Но в отличие от изолированных атомов и молекул собственные энергетические состояния электрона в кристалле не привязаны к какому-то конкретному узлу (не локализованы), а определены по кристаллу в целом. Настоящая книга рассчитана на читателей, знакомых с физикой полупроводников, и потому в этом параграфе мы лишь напомним основные особенности собственных и примесных состояний в полупроводниках.

— Определение энергетического спектра кристалла как многоядерной и многоэлектронной системы исходя лишь из «первых принципов» не представляется реальным, так как задача исключительно сложна. Поэтому прибегают к различным упрощениям. Подавляющее большинство задач физики твердого тела решают

(стремятся к нулю при $N \rightarrow \infty$), и для кристаллов обычных размеров можно считать, что волновой вектор как квантовое число изменяется квазинепрерывно. Этому изменению соответствует квазинепрерывный спектр состояний электрона в каждой из зон.

При заданном квантовом числе k одноэлектронное уравнение Шредингера имеет не одно, а бесконечное семейство решений с собственными дискретными значениями, которые можно классифицировать с помощью второго квантового числа n , присутствующего в индексах волновых функций (4.3). Это квантовое число не имеет глубокого физического смысла и по существу означает номер зоны n .

Из теоремы Блоха (4.3) видно, что состояния, описываемые волновыми векторами k и k' , отличающимися друг от друга на вектор обратной решетки q :

$$k' = k + q,$$

полностью эквивалентны. Действительно, для любого вектора обратной решетки iqR $e^{iqR} = 1$, поэтому, если соотношение (4.3) справедливо для k' , то оно должно быть справедливым также для k . При заданном n собственные значения энергии электрона $E_{n,k}$ и его собственные волновые функции $\Psi_{n,k}(r)$ представляют собой периодические функции k в пространстве обратной решетки:

$$\begin{aligned} \Psi_{n,k+q}(r) &= \Psi_{n,k}(r), \\ E_{n,k+q} &= E_{n,k}. \end{aligned} \quad (4.4)$$

Это обстоятельство является следствием наличия у электрона волновых свойств. Электронная волна (волна де Бройля), распространяющаяся в кристалле, испытывает дифракцию и интерференцию, а кристаллическая решетка служит для этой волны объемной дифракционной решеткой.

Для полного описания всей совокупности энергетических состояний кристалла область значений k достаточно ограничить одной элементарной ячейкой обратной решетки, т. е. ограничиться рассмотрением одной зоны Бриллюэна. Так как волновой вектор k может принимать квазинепрерывный ряд значений, то энергетические состояния электрона в поле, обладающем трансляционной симметрией (4.1), необходимо характеризовать семейством непрерывных функций $E_{n,k}$ [или $E_n(k)$]. Эти функции определяют зонную структуру кристалла. Совокупность всех электронных уровней, описываемых функцией $E_n(k)$ при заданном n , называют энергетической зоной.

Волновое число k является основным квантовым числом, характеризующим состояние электрона в кристалле. Поэтому зависимость $E(k)$ имеет фундаментальное значение для физики твердого тела. В частности, скорость движения электрона в кристалле можно отождествить с групповой скоростью волнового пакета

$$v = \frac{d\omega}{dk} = \frac{dE}{\hbar dk}$$

Пусть на электрон, описываемый блоховской функцией (4.3), действует сила F . Тогда изменение энергии за время dt будет равно $Fvdt$. С другой стороны, $dE = (dE/dk)dk = \hbar v dk$, откуда $F = \hbar dk/dt$. Это равенство можно рассматривать как аналог закона Ньютона, если величину $\hbar k$ определять как аналог импульса. По этой причине величину $\hbar k$ называют квазиимпульсом, подчеркивая

тем самым, что волновой вектор k в задаче о движении частицы в периодическом поле решетки играет такую же роль, как волновой вектор k свободного электрона. Продолжая эту аналогию, нетрудно показать, что величина

$$m^* = \frac{\hbar^2}{d^2E/dk^2} \quad (4.5)$$

играет роль массы и называется эффективной массой. Так как в кристалле энергия зависит не только от значения волнового вектора k , но и от его направления, то в общем случае эффективная масса является тензорной величиной.

Расчет зонной структуры кристалла, т. е. нахождение зависимостей $E(k)$, представляет в общем случае довольно сложную задачу. Будем интересоваться лишь результатами этих расчетов. В качестве примера на рис. 4.1 приведены части энергетических диаграмм $E(k)$ германия и арсенида галлия. Обычно зонную структуру изображают в виде зависимостей для различных характерных направлений k . Эти направления совпадают с направлениями наивысшей симметрии в зоне Бриллюэна, на которые приходится экстремумы $E(k)$. Для полупроводников, кристаллизующихся в структуры алмаза и сфалерита, такими характерными направлениями являются $\langle 100 \rangle$, $\langle 111 \rangle$ и $\langle 110 \rangle$. Все точки, лежащие на границе зоны Бриллюэна, принято обозначать прописными буквами латинского алфавита, а внутри зоны — прописными буквами греческого алфавита: точки $(100) — X$, $(111) — L$ и $(110) — K$ на границе зоны при $k = 2\pi/a$ и $[100] — \Delta$, $[111] — \Lambda$ и $[110] — \Sigma$ внутри зоны. Центр зоны Бриллюэна, т. е. состояние $k=0$, обозначают Γ . Нижние индексы у букв характеризуют симметрию состояния и означают неприводимые представления, к которым принадлежит волновая функция в данной

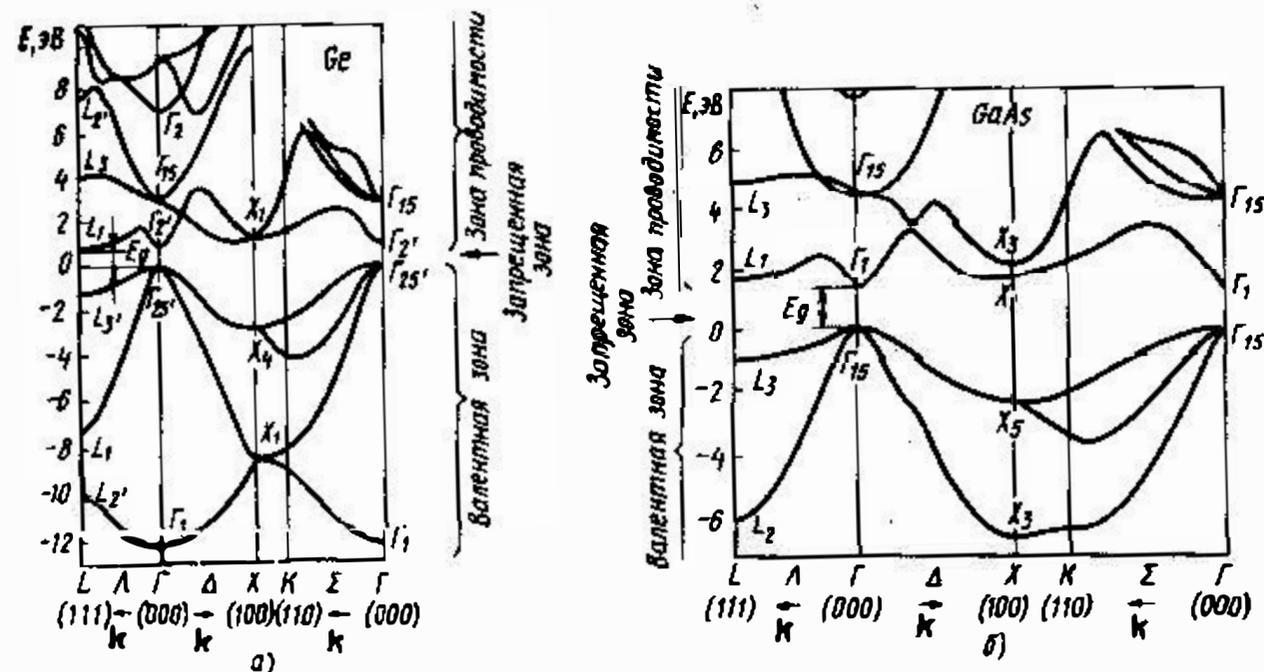


Рис. 4.1. Зонная структура германия (а) и арсенида галлия (б)

точке зоны Бриллюэна. Чтобы отличить состояния, принадлежащие валентной зоне и зоне проводимости, их помечают верхним индексом v или c соответственно.

Как видно из рис. 4.1, валентная зона и зона проводимости имеют сложное строение и состоят из нескольких подзон. Область энергий, в которой нет разрешенных электронных состояний при любом k , составляет запрещенную зону. Для кристаллов германия эта область находится от 0 до 0,7 эВ, а для арсенида галлия — от 0 до $\sim 1,5$ эВ (за начало отсчета энергии принят потолок валентной зоны). Разрешенные состояния, примыкающие к этой области, и представляют наибольший интерес, так как именно в этих состояниях могут находиться свободные носители заряда, способные участвовать в процессах электропроводности (электроны вблизи абсолютного минимума зоны проводимости и дырки вблизи максимума валентной зоны). По этой причине зонную структуру кристаллов часто представляют в упрощенном виде (рис. 4.2). Энергетические зазоры, отсчитанные от потолка валентной зоны до одного из минимумов зоны проводимости, обозначают символом E_g с соответствующим индексом сверху, как показано на рис. 4.2. Ширина запрещенной зоны соответствует минимальному энергетическому зазору. Для Ge $E_g = E_g^L = L_1^c - \Gamma_{25}^v$, а для арсенида галлия $E_g = E_g^F = \Gamma_1^c - \Gamma_{25}^v$.

Как видно из рис. 4.1 и 4.2, наинизший (абсолютный) минимум зоны проводимости германия (L_1^c — минимум) расположен на границе зоны Бриллюэна, в то время как максимум валентной зоны — в центре зоны, в точке Γ . В арсениде галлия абсолютный минимум зоны проводимости и максимум валентной зоны в k -пространстве совпадают: оба находятся в центре зоны Бриллюэна. Полупроводники, у которых максимум валентной зоны и минимум зоны проводимости расположены в одной и той же точке k -пространства, называют полупроводниками с прямой структурой зон или прямозонными. Полупроводники, у которых максимум валентной зоны и минимум зоны проводимости расположены в различных точках k -пространства, называют полупроводниками с непрямой структурой зон или непрямозонными. К первым относятся GaAs, InSb, CdS и др., ко вторым — Ge, Si, GaP, SiC и др. Соответствующие энергетические зазоры называют прямыми (например, E_g^L) и непрямыми (например, E_g^X или E_g^F).

Помимо традиционного материала полупроводниковой электроники — кремния — в оптоэлектронике наиболее широкое применение находят полупроводниковые соединения элементов III и V групп Периодической системы Д. И. Менделеева (полупроводники $A^{III}B^V$). Рассмотрим особенности энергетического спектра этих материалов. В полупроводниках группы $A^{III}B^V$ на-

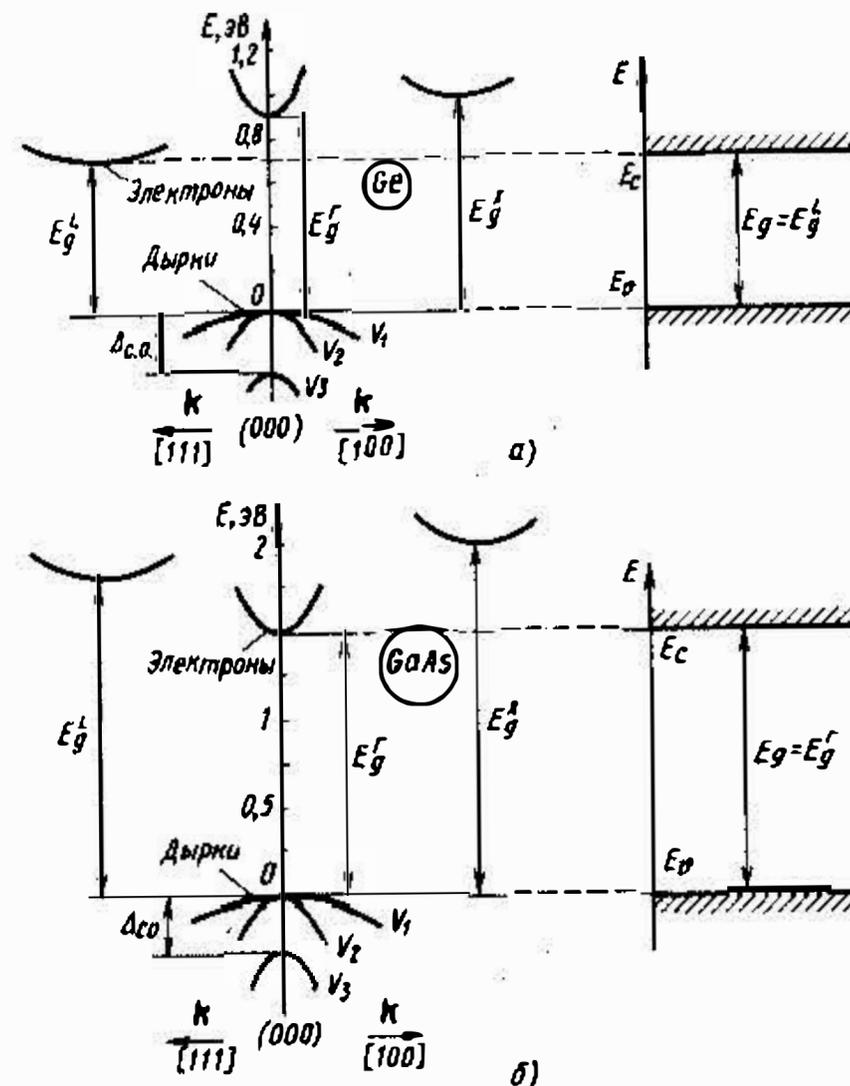


Рис. 4.2. Упрощенные энергетические диаграммы для электронных состояний в кристаллах германия (а) и арсенида галлия (б)

блюдается изменение их основных свойств по мере изменения суммы атомных номеров элементов, составляющих данное соединение. С увеличением $\Sigma z = z_{III} + z_V$ происходит монотонное уменьшение соответствующих энергетических зазоров, как показано на рис. 4.3. Скорость изменения «прямого» энергетического зазора E_g^L заметно больше, чем «непрямого» E_g^X . Это приводит к тому, что соединения $A^{III}B^V$ с $\Sigma z \geq 56$ обладают прямозонной, а соединения с $\Sigma z \leq 56$ — непрямозонной структурой, как схематически показано на рис. 4.3. Исключение составляют нитриды GaN и AlN, кристаллизующиеся в структуру вюрцита. Остальные полупроводники группы $A^{III}B^V$ кристаллизуются в кубическую структуру цинковой обманки (сфалерит).

Полупроводники группы $A^{III}B^V$ обладают сходной структурой валентной зоны. Она состоит из трех подзон: зоны тяжелых дырок V_1 , зоны легких дырок V_2 и зоны V_3 , отщепленной спин-орбитальным взаимодействием. Эффективные массы дырок

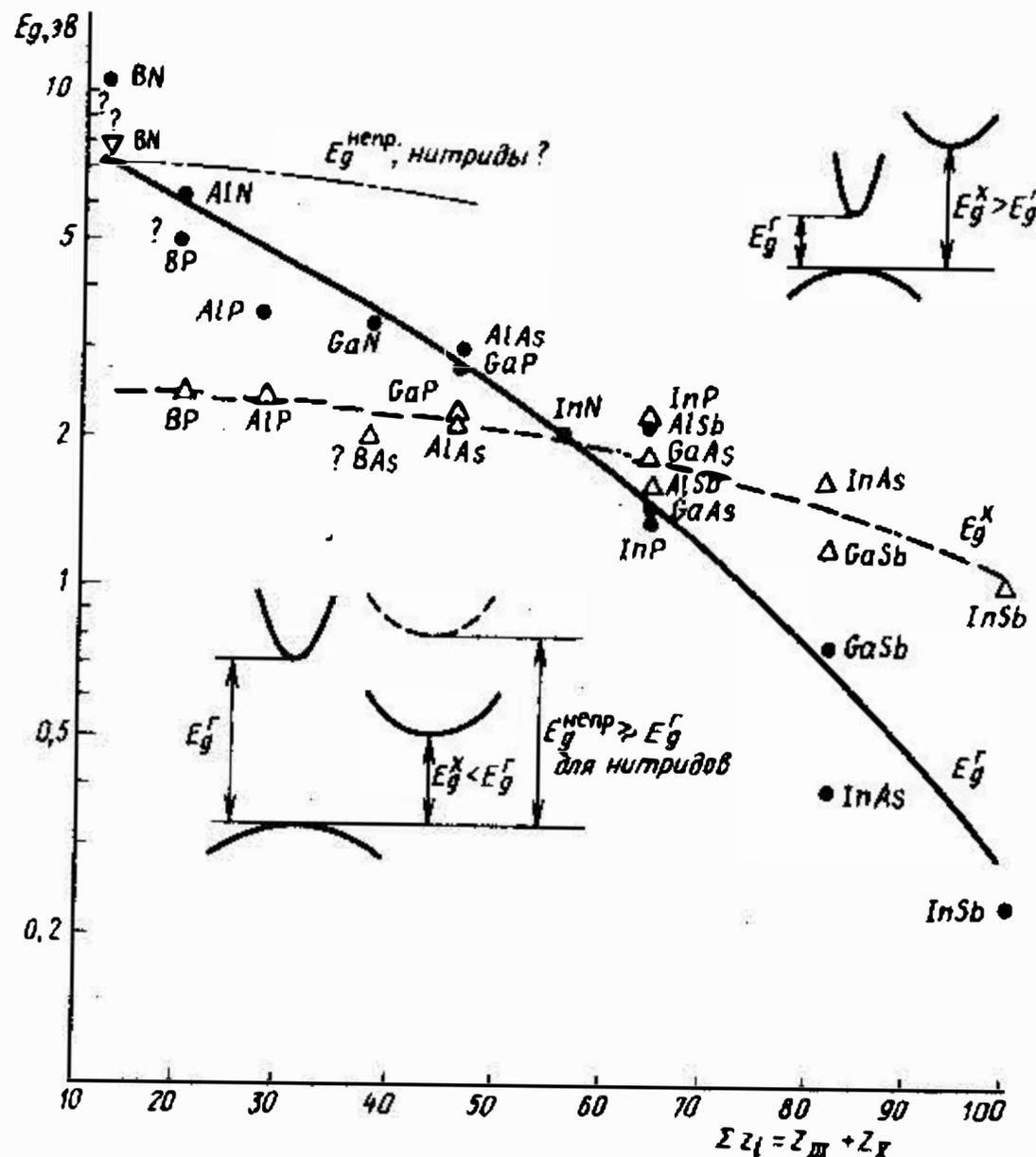


Рис. 4.3. Общие закономерности в изменении «прямого» (E_g^Γ , сплошная линия) и «непрямого» (E_g^X , пунктировая линия) энергетических зазоров в полупроводниках $A^{III}B^{IV}$

изменяются сравнительно слабо. Например, для тяжелых дырок отношение m_p^*/m изменяется от 0,45 для InSb до 0,9 для GaP.

Эффективная масса в Γ -минимуме зоны проводимости изотропна и монотонно возрастает с увеличением прямого энергетического зазора, так что хорошо выполняется предсказываемое теорией соотношение

$$\frac{1}{m_n^\Gamma} = \frac{1}{m_0} \left(1 + \frac{E_p}{E_g^\Gamma} \right), \quad (4.6)$$

где E_p определяется через матричный элемент оператора импульса $P_{cv}(0)$ и для большинства полупроводников с тетраэдрической координацией атомов $E_p \approx 20$ эВ.

Закон дисперсии, т. е. зависимость $E(k)$, вблизи Γ -минимума, изотропен. Отличие $E(k)$ от простой квадратичной зависимости

$$E_n^\Gamma = \frac{\hbar^2 k^2}{2m_n^\Gamma}, \quad (4.7)$$

при $m_n^\Gamma = \text{const}$ сравнительно невелико и проявляется лишь в узкозонных материалах типа InSb.

Для «непрямых» X - и L -минимумов закон дисперсии анизотропен. Изоэнергетические поверхности представляют собой эллипсоиды вращения, как у Ge и Si, что учитывается введением поперечной m_\perp и продольной m_\parallel эффективных масс. Средняя эффективная масса проводимости для электронов, находящихся в этих минимумах, определяется соотношением

$$3/m_n = 1/m_{n\parallel} + 2/m_{n\perp}. \quad (4.8)$$

Плотность состояний $\rho(E)$, т. е. число электронных состояний в единичном интервале энергий в единице объема кристалла, является функцией энергии. Для зон с параболическим законом дисперсии (4.7) имеем

$$\rho_c(E) = \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} (E - E_c)^{1/2}. \quad (4.9)$$

Аналогично для валентной зоны

$$\rho_v(E) = \frac{1}{2\pi^2} \left(\frac{2m_p}{\hbar^2} \right)^{3/2} (E - E_v)^{1/2}. \quad (4.9a)$$

Для анизотропного закона дисперсии в X - и L -долинах вместо m_n в (4.9) должна стоять величина m_c , называемая эффективной массой плотности состояний. Для эллипсоидальных поверхностей постоянной энергии

$$(m_c)^{3/2} = N_c (m_\perp^2 m_\parallel)^{1/2}, \quad (4.10)$$

где N_c — число эквивалентных минимумов по различным симметричным кристаллографическим направлениям (число долин).

Наряду с полупроводниками $A^{III}B^{IV}$ в оптической электронике применяются и другие группы бинарных полупроводниковых соединений. Среди них важны халькогениды (т. е. сульфиды, селениды и теллуриды) элементов II и IV групп, а именно — теллуриды кадмия и ртути (CdTe и HgTe) и халькогениды свинца — PbS, PbSe и PbTe. Все эти материалы являются прямозонными, но в отличие от полупроводников $A^{III}B^{IV}$ максимум валентной зоны и абсолютный минимум зоны проводимости халькогенидов свинца и олова расположены не в центре зоны Бриллюэна.

лозна, т. е. не при $k=0$, а в направлении (III). Температурный коэффициент ширины запрещенной зоны $\partial E_g/\partial T$ у них положительный, а барический коэффициент $\partial E_g/\partial P$ — отрицательный.

Халькогениды свинца кристаллизуются в гранцентрированной кубической решетке типа структуры каменной соли. Для них характерно весьма заметное отклонение от стехиометрии, что усложняет технологию изготовления приборов на их основе. Избыток атомов свинца приводит к увеличению концентрации свободных электронов, а избыток халькогена дает свободные дырки. Для этих кристаллов, обладающих большой долей ионной связи, характерны весьма большие значения статической диэлектрической проницаемости.

В идеальных кристаллах кроме собственных электронных состояний существуют энергетические состояния, связанные с колебательными движениями ядер, образующих кристаллическую решетку. Как и в молекулах, этот вид движения квантован. Наличие трансляционной симметрии в расположении атомов позволяет классифицировать колебательные состояния с помощью того же квантового числа k , которое использовалось для классификации электронных состояний. Решение уравнения для колебательного движения ядер в трехмерной решетке приводит к ограниченному числу колебательных ветвей в противоположность бесконечному числу электронных энергетических зон. Общее число различных колебательных ветвей, называемых также *типами колебаний* или *модами*, равно утроенному числу атомов в примитивной ячейке кристалла.

В периодических атомных структурах коллективные движения представляют собой колебания, распространяющиеся по кристаллу в виде упругих волн. Эти колебания квантованы. Квант колебаний кристаллической решетки называется *фононом*.

Отметим, что каждой волне в квантовой механике можно сопоставить совокупность квазичастиц, например квантованием колебаниям электромагнитного поля с частотой ω — совокупность фотонов, которым приписывается энергия $\hbar\omega$ и волновой вектор k . Электронную волну, распространяющуюся по кристаллу, также можно рассматривать как движение частицы (точнее, квазичастицы), которой приписывают некоторую эффективную массу m_n , заряд $-e$, спин $1/2$ и волновой вектор k . Аналогично фонон можно рассматривать как квазичастицу с энергией $\hbar\Omega$, волновым вектором k и нулевым спином, движущуюся в кристалле. Колебательная энергия в волне данного типа определяется аналогично (1.57):

$$E_{\text{фонон}} = \hbar\Omega_k (q + 1/2).$$

Здесь $\hbar\Omega_k$ — энергия фонона с волновым вектором k ; q — число фононов.

Полная колебательная энергия кристалла определяется суммированием по всем значениям k и всем колебательным ветвям.

До сих пор рассматривались энергетические состояния идеальных кристаллов. Как известно, примеси определяют ряд важнейших свойств веществ, особенно полупроводников. Примесные

состояния в кристаллах условно можно разделить на слабо связанные («мелкие») и сильно связанные («глубокие») состояния. Энергия связи мелкого центра существенно меньше ширины запрещенной зоны, а для глубокого она сравнима с шириной запрещенной зоны.

Мелкие примесные состояния, называемые также *водородоподобными состояниями*, хорошо описываются с помощью *приближения эффективной массы*. В этом приближении возмущающий потенциал, создаваемый примесным атомом в кристалле, считается кулоновским и записывается аналогично (1.36), а действие кристалла учитывается введением эффективной диэлектрической проницаемости ϵ_r :

$$V(r) = -\frac{e^2}{4\pi\epsilon_0\epsilon_r r}. \quad (4.11)$$

Если связь примесного электрона (дырки) с донорным (акцепторным) атомом слабая, то волновая функция примесного состояния будет «размазана» по кристаллу, и большая часть плотности вероятности $|\Psi(r)|^2$ окажется вне центральной ячейки, содержащей примесный атом. При этом разумно представить электрон (дырку) связанным с донором (акцептором) силами кулоновского притяжения вследствие избыточного заряда примеси и медленно движущимся вокруг примесного центра в области неискаженного кристалла. Движение такого связанного электрона (дырки) будет в значительной мере определяться свойствами зоны проводимости (валентной зоны) вблизи экстремума. Это означает, что описание связанного примесного состояния основывается на блоховских функциях для соответствующей зоны, с которой связано данное примесное состояние, для ограниченной области k в окрестности границы зоны. Уравнение, описывающее движение слабосвязанного примесного электрона в кристалле, будет аналогично уравнению (1.38) с той разницей, что вместо массы свободного электрона m будет эффективная масса m_n :

$$-\frac{\hbar^2}{2m_n} \nabla^2 \Psi(r) - \frac{e^2}{4\pi\epsilon_0\epsilon_r r} \Psi(r) = E \Psi(r). \quad (4.12)$$

Решения этого уравнения аналогичны решениям уравнения для атома водорода. В случае изотропной эффективной массы спектр разрешенных значений энергии, например для водородоподобного донора, будет аналогичен (1.39), где вместо массы свободного электрона m будет фигурировать масса m_n , а вместо диэлектрической проницаемости вакуума — диэлектрическая проницаемость среды:

$$E_n = -\frac{m_n e^4}{8(\epsilon_0 \epsilon_r \hbar)^2} \frac{1}{n^2} = -\frac{R_\infty}{\epsilon_r^2} \left(\frac{m_n}{m}\right) \frac{1}{n^2} = -\frac{E_d}{n^2}. \quad (4.13)$$

Здесь R_∞ — постоянная Ридберга, а n — главное квантовое число.

За начало отсчета энергии в (4.13) аналогично (1.39) принято состояние свободного электрона. В данном случае это будет дно зоны проводимости. Состояние с $n=1$ называется основным, а состояния с $n=2, 3, \dots$ — возбужденными. По аналогии с атомом водорода вводят энергию ионизации донора

$$E_d = \frac{R_\infty}{\epsilon_r^2} \left(\frac{m_n}{m}\right).$$

Найдем энергию ионизации водородоподобного донора, например, для арсенида галлия. При $\epsilon_r = 13$, $m_n^r = 0,07m$ имеем $E_d \approx 5,6$ мэВ. Для акцепторных состояний энергия отсчитывается от потолка валентной зоны. Аналогично для энергии ионизации водородоподобного акцептора в GaAs при $m_p \approx 0,5m$ получаем $E_d \approx 40$ мэВ.

По аналогии с атомом водорода вводят понятие *эффективного боровского радиуса* для доноров:

$$a_d = \frac{\hbar^2 \epsilon_0 \epsilon_r}{\pi m_n e^2} = \epsilon_r \left(\frac{m}{m_n}\right) a_B, \quad (4.14)$$

где a_B — боровский радиус для основного состояния атома водорода, называемый также радиусом первой боровской орбиты: $a_B = \hbar^2 \epsilon_0 / \pi m e^2 \approx 0,053$ нм.

Оценка по формуле (4.14) дает для доноров в арсениде галлия $a_d \approx 10$ нм. Это значение существенно превышает размеры элементарной ячейки кристалла (период идентичности арсенида галлия $a = 0,54$ нм), что дает возможность рассматривать донорные состояния, основываясь на блоховских функциях зоны проводимости для ограниченной области значений волнового вектора вблизи границы зоны. Это означает, что волновую функцию донорного состояния $\Psi_d(\mathbf{r})$ разумно выразить через блоховские состояния $\Psi_k(\mathbf{r})$ идеального (невозмущенного) кристалла:

$$\Psi_d(\mathbf{r}) = \sum_{\mathbf{k}} C(\mathbf{k}) \Psi_{\mathbf{k}}(\mathbf{r}). \quad (4.15)$$

Коэффициенты $C(\mathbf{k})$ указывают вклад волновой функции $\Psi(\mathbf{r})$ с заданным значением волнового вектора \mathbf{k} в формирование донорной волновой функции $\Psi_d(\mathbf{r})$, т. е. эти коэффициенты имеют смысл волновой функции донора в \mathbf{k} -представлении.

Для электрона, слабо связанного с донором, можно предположить, что в (4.15) наибольший вклад в формирование функции $\Psi_d(\mathbf{r})$ дадут ближайшие состояния, и со значительной амплитудой под знак суммы войдут лишь члены с \mathbf{k} вблизи экстремума зоны. Действительно, если (4.15) подставить в (4.12) и провести преобразование Фурье, то с точностью до постоянного множителя получим

$$C(\mathbf{k}) \sim \frac{1}{[(k a_d)^2 + 1]^2}. \quad (4.16)$$

Отсюда видно, что амплитуды $C(\mathbf{k})$ существенны только в области $k \lesssim a_d^{-1}$, т. е. волновая функция электрона на водородоподобном доноре локализована в области \mathbf{k} -пространства с размерами порядка $1/a_d$. Это позволяет представить донорные состояния на диаграмме $E(\mathbf{k})$ так, как показано на рис. 4.4. Для сравнения там же изображен спектр энергетических уровней донора на диаграмме $E(x)$. Соотношение

$$\Delta k \approx 1/a_d \quad (4.17)$$

характеризует «размазанность» волновой функции примесного состояния в \mathbf{k} -пространстве и по существу является следствием соотношения неопределенностей Гейзенберга. Действительно, «неопределенность координаты» для электрона, связанного с донором, $\Delta x \approx a_d$, а «неопределенность импульса» $\Delta p \approx \hbar \Delta k$, следовательно, $\Delta x \Delta p = a_d \hbar \Delta k = \hbar$, откуда получаем соотношение (4.17).

Из (4.16), (4.17) следует, что чем сильнее электрон (дырка) локализован вблизи примесного центра в обычном простран-

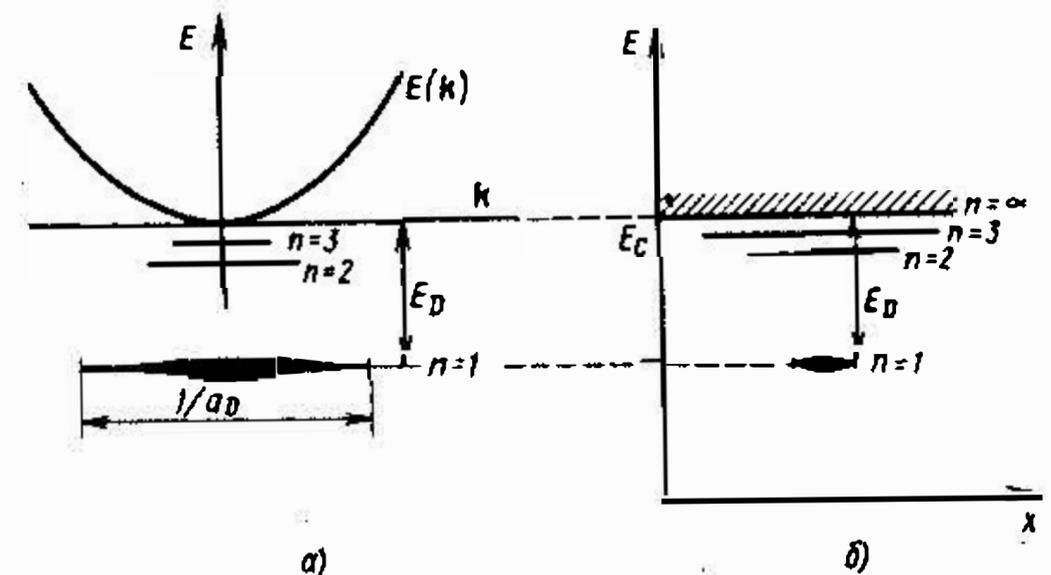


Рис. 4.4. Состояния водородоподобного донора на диаграмме $E(\mathbf{k})$, т. е. в \mathbf{k} -пространстве (а) и в геометрическом r -пространстве (б)

стве, тем в большей степени он «размазан» в k -пространстве.

В многодолинных полупроводниках типа германия и кремния волновая функция для донорного состояния должна строиться из комбинаций компонент, соответствующих каждому эквивалентному минимуму зоны проводимости. В таких полупроводниках необходимо учитывать анизотропию эффективной массы. Это приводит к усложнению структуры примесных уровней по сравнению с простой водородоподобной моделью. Сложное строение валентной зоны усложняет также энергетический спектр акцепторных состояний.

Кроме мелких примесей, описываемых водородоподобной моделью, в твердых телах возможно существование дефектов (в том числе примесей), возмущающий потенциал которых существенно отличается от кулоновского потенциала (4.11) и локализован вблизи дефекта в пределах одной элементарной ячейки. Указанные примеси создают в запрещенной зоне глубокие уровни энергии.

Примесные состояния, создаваемые элементами с незастроенными внутренними оболочками (например, $3d$ - и $4f$ -оболочками), могут образовывать в твердых телах систему энергетических уровней, схожих с уровнями свободных атомов или ионов (см. гл. 8).

Примеси, свободные носители заряда, дефекты решетки искажают энергетический спектр идеального кристалла. Эти искажения проявляются не только в примесной, но и в собственной области спектра. В первую очередь они будут заметны вблизи пороговых значений энергий, т. е. вблизи экстремумов зон E_c и E_v .

Проследим за образованием таких хвостов при легировании полупроводника простыми донорами или акцепторами. При слабом легировании примеси не взаимодействуют друг с другом и создают в запрещенной зоне локализованные состояния, возникающие в результате отщепления их от зоны проводимости или валентной зоны (рис. 4.4). С увеличением степени легирования среднее расстояние R между примесными атомами уменьшается, они начинают взаимодействовать друг с другом, их уровни перекрываются, приводя к появлению зоны примесных состояний. Когда концентрация примеси N такова, что среднее расстояние $R \sim (N)^{-1/3}$ между атомами примерно равно боровскому радиусу a_d примесного состояния (4.15), то примесная зона сливается с разрешенной зоной E_c или E_v и энергия ионизации примеси становится равной нулю — возникает примесная проводимость металлического типа. Корректное рассмотрение такой ситуации показывает, что такой переход происходит, когда отношение среднего расстояния между атомами примеси к боровскому радиусу водородоподобной примеси a_d примерно равно трем:

$$R_{\text{ср}}/a_d \approx 3.$$

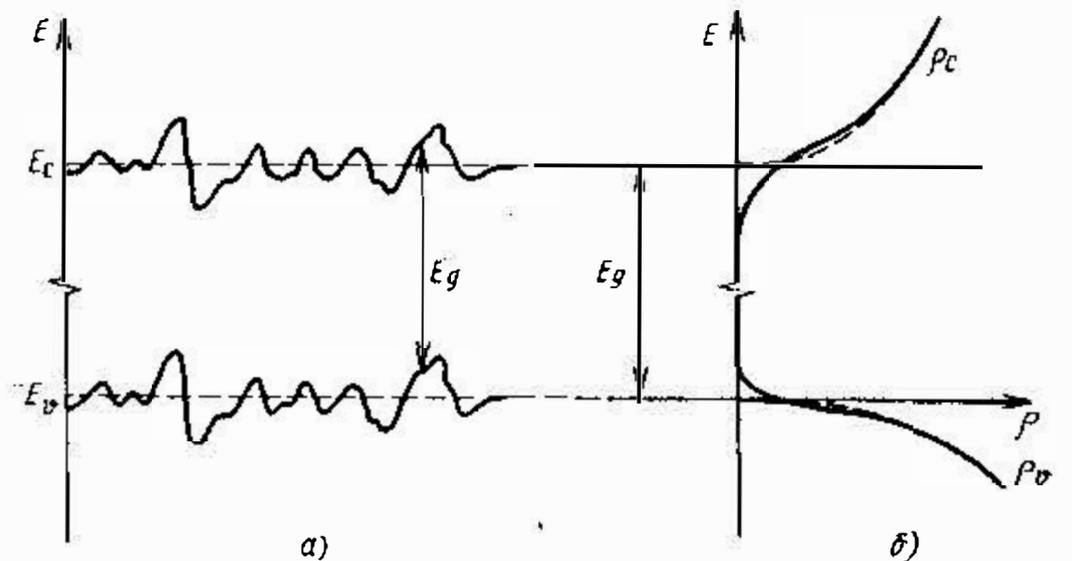


Рис. 4.5. Образование «хвостов» плотности состояний при сильном легировании полупроводников (с учетом компенсации): а — схематическое изображение границ зон, искривленных потенциалом флуктуирующий; б — «хвосты» плотности состояний (пунктиром показана плотность состояний в нелегированном кристалле)

При случайном распределении примесей $R_{\text{ср}} = \left(\frac{3}{4N}\right)^{1/3}$, где N — концентрация примеси.

Случайное распределение ионизованных примесей в кристалле приводит к появлению в них флуктуирующих электрических полей, которые вызывают флуктуационные искривления зон. Случайный характер этих флуктуаций приводит к появлению «хвостов» состояний, плотность которых экспоненциально убывает вглубь запрещенной зоны, как это схематически показано на рис. 4.5.

При гауссовом законе распределения флуктуаций плотность состояний в зоне проводимости полупроводника n -типа имеет вид

$$\rho_c(E) = (2\sqrt{2} V_{\text{ф.в.}})^{1/2} (m_n^{3/2}/\pi^2 \hbar^3) y[(E - E_c)/\sqrt{2} V_{\text{ф.в.}}], \quad (4.18)$$

где $V_{\text{ф.в.}}$ — среднеквадратичная флуктуация потенциальной энергии, а E_c — край номинальной (невозмущенной) параболической зоны проводимости с эффективной массой m_n . Безразмерная функция $y[(E - E_c)/\sqrt{2} V_{\text{ф.в.}}]$ определяется выражением

$$y(x) = \pi^{-1/2} \int_{-\infty}^x (x-z)^{1/2} \exp(-z^2) dz.$$

Его анализ показывает, что при высоких энергиях ρ_c ведет себя как $(E - E_c)^{1/2}$, т. е. как в простой невозмущенной зоне,

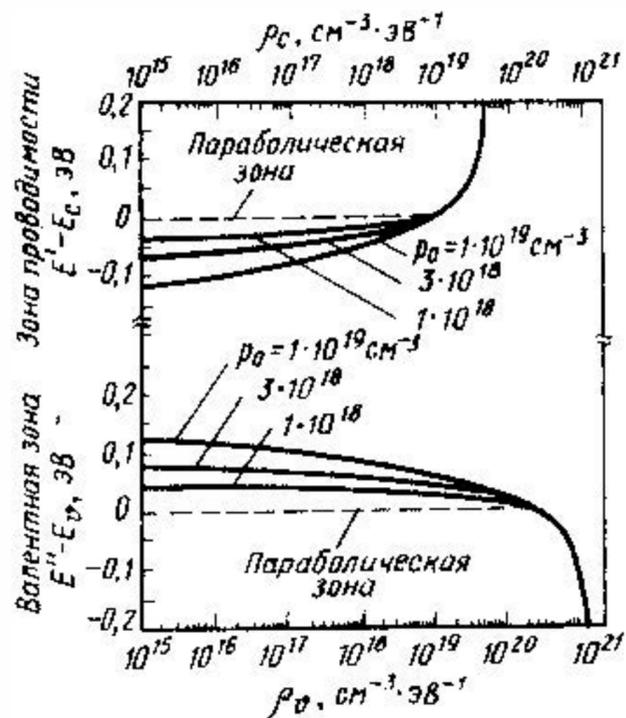


Рис. 4.6. Плотность состояний в хвостах зон GaAs, рассчитанная по модели Кейна для разных концентраций дырок при комнатной температуре

§ 4.2. ЭЛЕКТРОННЫЕ СОСТОЯНИЯ В ТВЕРДЫХ РАСТВОРАХ И НЕУПОРЯДОЧЕННЫХ СИСТЕМАХ

Большое значение для оптической электроники и лазерной техники имеют материалы, свойства которых можно плавно изменять в заданных пределах. К таким материалам относятся твердые растворы и в первую очередь полупроводниковые твердые растворы замещения. Идеальным твердым раствором замещения считается кристалл с идеальной решеткой, по соответствующим узлам которой статистически независимо распределены различные атомы составляющих твердый раствор компонентов. Как правило, используют *изовалентные* твердые растворы замещения, в которых замещаемый и замещающий атомы находятся в одном столбце Периодической системы элементов, а образующие твердый раствор компоненты принадлежат к одной группе соединений. Типичными примерами являются твердые растворы германия и кремния ($\text{Si}_x\text{Ge}_{1-x}$), твердые растворы между соединениями $\text{A}^{\text{III}}\text{B}^{\text{V}}$, между соединениями $\text{A}^{\text{IV}}\text{B}^{\text{VI}}$ и др. Твердые растворы между различными группами материалов, например $\text{A}^{\text{III}}\text{B}^{\text{V}}$ — $\text{A}^{\text{IV}}\text{B}^{\text{VI}}$, называются *гетеровалентными*.

Принципиально важной особенностью твердых растворов является возможность путем изменения их состава плавно и в широких пределах изменять основные параметры материала, в том числе: Параметры решетки, ширину запрещенной зоны, показате-

ль преломления, диэлектрическую проницаемость, теплопроводность и т. п. Это позволяет получать материал с точно заданным набором основных характеристик, что особенно важно при создании приборов оптоэлектроники, интегральной оптики, приборов на основе гетеропереходов и варизонных структур.

Плотность состояний в валентной зоне будет определяться аналогичными соотношениями с заменой m_n на m_p и изменением знака аргумента «квадрат». В качестве примера на рис. 4.6 в полулогарифмическом масштабе изображены рассчитанные для арсенида галлия плотности состояний в хвостах зон при разных концентрациях дырок. Пунктиром показаны плотности состояний в идеальном кристалле GaAs.

Плотность состояний в валентной зоне будет определяться аналогичными соотношениями с заменой m_n на m_p и изменением знака аргумента «квадрат». В качестве примера на рис. 4.6 в полулогарифмическом масштабе изображены рассчитанные для арсенида галлия плотности состояний в хвостах зон при разных концентрациях дырок. Пунктиром показаны плотности состояний в идеальном кристалле GaAs.

тель преломления, диэлектрическую проницаемость, теплопроводность и т. п. Это позволяет получать материал с точно заданным набором основных характеристик, что особенно важно при создании приборов оптоэлектроники, интегральной оптики, приборов на основе гетеропереходов и варизонных структур.

Период идентичности идеального твердого раствора замещения A_xB_{1-x} , образованного соединениями A и B, линейно изменится от a_A к a_B (закон Вегарда). Каждый из составляющих его атомов располагается в своем узле решетки. Поэтому в идеальном твердом растворе трансляционная симметрия в геометрии решетки сохраняется. Но статистически независимое распределение компонентов (или атомов) A и B в твердом растворе приводит к нарушению трансляционной симметрии в его потенциальной энергии, как это проиллюстрировано на рис. 4.7. Нарушение условия (4.1) приводит, строго говоря, к несправедливости теоремы Блоха и невозможности рассматривать волновой вектор k как хорошее квантовое число для электрона. Однако в изовалентных твердых растворах замещения это нарушение сравнительно невелико. Средняя длина пробега L носителя заряда существенно больше межатомных расстояний a , так что $kL \gg 1$. Поэтому нарушение трансляционной симметрии потенциальной энергии, вызванное неупорядоченным распределением атомов в узлах кристаллической решетки твердого раствора, можно рассматривать как малое возмущение, т. е. представить $V(r)$ в виде

$$V(r) = V_0(r) + V_1(r), \quad (4.19)$$

где $V_0(r)$ — периодический потенциал, обладающий трансляционной симметрией; $V_1(r)$ — непериодическая часть потенциала, обусловленная статистически беспорядочным распределением атомов твердого раствора типа A_xB_{1-x} или $\text{A}_x\text{B}_{1-x}\text{C}$ по узлам подрешетки R_{AB} .

Часто $V_1(r) \ll V_0(r)$ и влияние непериодической части можно рассматривать как малое возмущение. В первом приближении этим малым возмущением пренебрегают и твердый раствор рас-

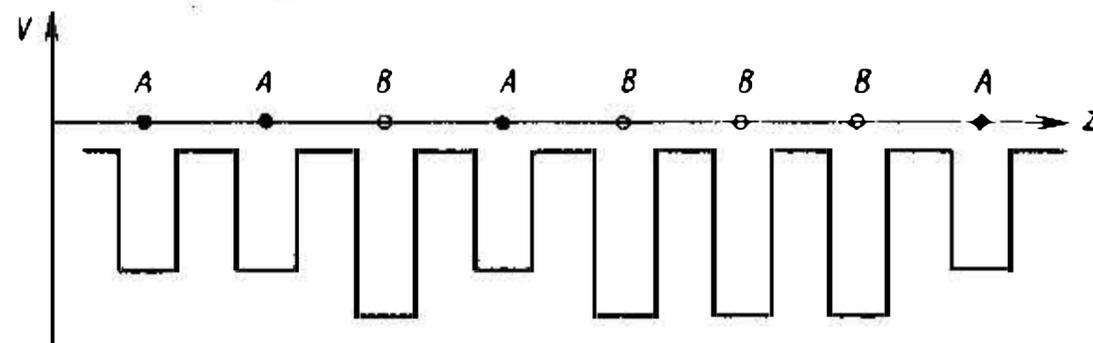


Рис. 4.7. Схематическое изображение изменения потенциальной энергии с координатой z в твердом растворе A_xB_{1-x}

смаатривают как идеальный кристалл с параметрами решетки, вытекающими из закона Вегарда, и со свойствами, определяемыми некоторыми «псевдоатомами», располагающимися строго упорядоченно в узлах решетки. Свойства этих псевдоатомов в свою очередь зависят от состава твердого раствора и планируются, изменяются от A к B с весовыми множителями x и $(1-x)$ соответственно, в твердом растворе A_xB_{1-x} . Такое приближение называют «*приближением виртуального кристалла*». В этом приближении правомерно пользоваться теми же понятиями зонной теории, что и для идеальных кристаллов, в частности: «*зонная структура*», «*волновой вектор*», «*правила отбора*», «*ширина запрещенной зоны*» и т. п.

Непериодическая, флуктуирующая часть потенциала рассматривается как малое возмущение к «виртуальному кристаллу» и с его учетом проводят методами теории возмущений аналогично тому, как это делается, например, для кристаллов с дефектами. Флуктуация V_1 тем больше, чем больше отличаются по своим свойствам образующие твердый раствор атомы. Основные значения здесь имеют различия в геометрических размерах (ковалентный, ионный радиусы) и электрохимических свойств замещаемых и замещающих атомов. Как правило, если атомные радиусы отличаются больше чем на 15%, то непрерывный ряд твердых растворов замещения вообще не образуется.

Для оптоэлектроники и интегральной оптики важны твердые растворы между полупроводниками группы $A^{III}B^V$. Их образование возможно при замещении как в анионной, так и в катионной подрешетках или в обеих подрешетках одновременно. Среди них могут быть трехкомпонентные (например, $Al_xGa_{1-x}As$, $Ga_xIn_{1-x}As$; $GaAs_{1-x}P_x$), четырехкомпонентные ($Ga_xIn_{1-x}P_yAs_{1-y}$, $[Al_xGa_{1-x}]_yIn_{1-y}P$) и более сложные системы. На рис. 4.8 приведены

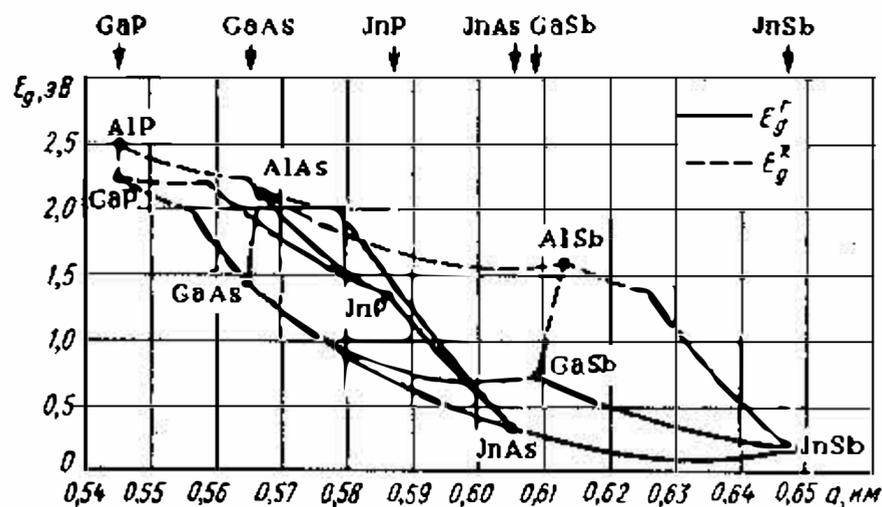


Рис. 4.8. Ширина запрещенной зоны в период решетки для некоторых твердых растворов полупроводников группы $A^{III}B^V$

диаграмма значений ширины запрещенной зоны и периоды идентичности для бинарных полупроводников $A^{III}B^V$ и их твердых растворов. Из трехкомпонентных растворов наиболее близки к идеальной системе типа $GaAs - AlAs$ ($Al_xGa_{1-x}As$), поскольку ковалентные радиусы атомов Al и Ga отличаются менее чем на 0,5%, а их электроотрицательности близки.

Параметры зонной структуры плавно изменяются с изменением состава твердого раствора. Поскольку валентная зона в полупроводниках $A^{III}B^V$ имеет сходную структуру, основные изменения в энергетическом спектре их твердых растворов определяются различием в структуре зоны проводимости составляющих компонентов. Наиболее важным здесь является характер изменения соответствующих энергетических зазоров E_g^Gamma , E_g^X и E_g^L в зависимости от состава твердого раствора. Эти зависимости отклоняются от линейных и для заданной точки зоны Бриллюэна могут быть представлены квадратичной функцией состава x в виде

$$E_g(x) = E_g(0) + [E_g(1) - E_g(0)]x - cx(1-x). \quad (4.20)$$

Коэффициент нелинейности c положительный и тем больше, чем больше отличаются по своим свойствам атомы замещаемых и замещающих компонент. Если при изменении состава x не происходит смены абсолютного минимума зоны проводимости, то характер зонной структуры не меняется. Например, твердые растворы $Ga_xIn_{1-x}As$ и $InAs_xSb_{1-x}$ во всем диапазоне изменения $0 \leq x \leq 1$ обладают прямой структурой энергетических зон. Если при изменении состава происходит смена абсолютного минимума зоны проводимости, то в зависимости $E_g(x)$ наблюдается излом.

На рис. 4.9 представлены изменения энергетических зазоров в трехкомпонентных твердых растворах, часто применяемых в качестве основных рабочих материалов оптоэлектронных устройств. Эти зависимости тесно связаны с общими закономерностями в изменении E_g (см. рис. 4.3). Для кристаллов $GaAs_{1-x}P_x$ с $x < 0,46$ Γ -минимум зоны проводимости расположен ниже X - и L -минимумов, и эти твердые растворы обладают прямозонной структурой. Твердые растворы $GaAs_{1-x}P_x$ при $x > 0,46$ обладают непрямозонной структурой. Аналогично кристаллы $Ga_xAl_{1-x}As$ при $x < 0,36$ обладают непрямозонной, а при $x > 0,36$ — прямозонной структурой. Состав твердого раствора, при котором происходит переход от прямой к непрямой структуре зон, называется *переходным*.

Кроме трехкомпонентных систем большое значение для оптоэлектроники имеют четырехкомпонентные твердые растворы типа $A_xB_{1-x}C_yD_{1-y}$, типичным представителем которых является арсенид — фосфид галлия — индия: $Ga_xIn_{1-x}P_yAs_{1-y}$. Одновременное замещение двух компонент приводит к появлению до-

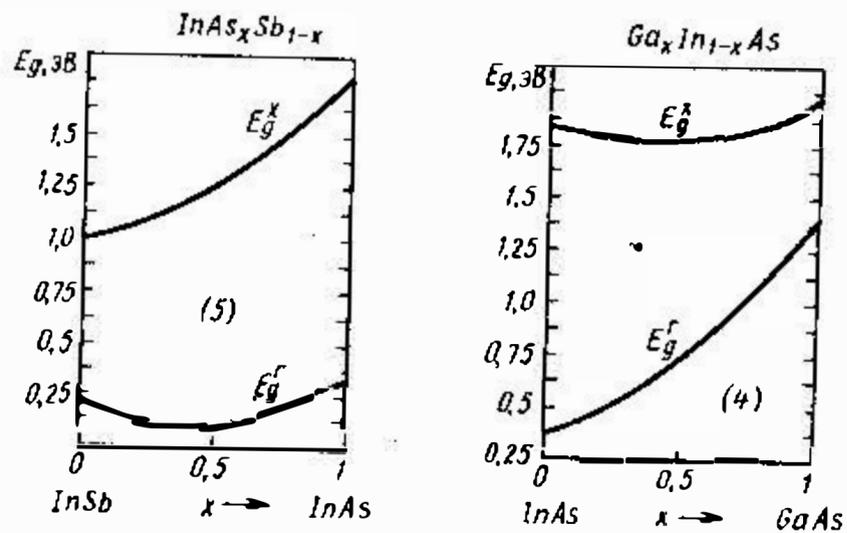
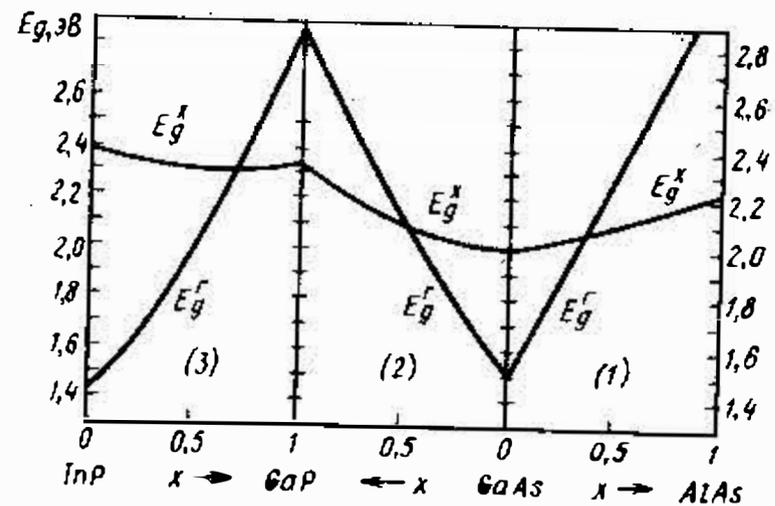


Рис. 4.9. Изменение прямого E_{Γ}^{Γ} и непрямого E_{Γ}^X энергетических зазоров в трехкомпонентных твердых растворах

$\text{Al}_x\text{Ga}_{1-x}\text{As}$ (1); $\text{GaAs}_{1-x}\text{P}_x$ (2); $\text{Ga}_x\text{In}_{1-x}\text{P}$ (3);
 $\text{Ga}_x\text{In}_{1-x}\text{As}$ (4); $\text{InAs}_x\text{Sb}_{1-x}$ (5) при $T=300\text{ K}$

полнительной степени свободы и к возможности в определенных пределах независимо управлять шириной запрещенной зоны и периодом кристаллической решетки такого материала. Это обстоятельство является исключительно важным для создания гетероструктур с идеальной контактной границей. Появляется возможность создавать так называемый *изопериодный* (называемый так же *изорешеточный*) ряд твердых растворов, где подбором соотношений x и y обеспечивается изменение ширины запрещенной зоны при постоянстве периода идентичности кристаллической решетки.

В качестве исходных компонентов твердого раствора $A_xB_{1-x}C_yD_{1-y}$ можно рассматривать четыре бинарных соединения: 1 — AC ; 2 — BC ; 3 — AD и 4 — BD . При выполнении закона Вегарда период кристаллической решетки аддитивно зависит от

молярной концентрации каждой из компонент в твердом растворе, т. е.

$$a(x, y) = xa_1 + (1-x)a_2 + x(1-y)a_3 + (1-x)(1-y)a_4, \quad (4.21)$$

где $a_1 \dots a_4$ — периоды решеток соответствующих бинарных соединений.

Изменение энергетических зазоров с составом четырехкомпонентных твердых растворов следует тем же закономерностям, что и трехкомпонентных типа $A_xB_{1-x}C$. Зависимости $E_g(x, y)$ для каждой из точек зоны Бриллюэна могут быть определены, если известны данные по составляющим твердый раствор бинарным и тройным системам, используя интерполяционное соотношение, вытекающее из (4.20):

$$E(xy) = E_4 + (E_3 - E_4)x + (E_2 - E_4)y + (E_1 + E_4 - E_2 - E_3)xy - [c_{24} + (c_{13} - c_{24})x]y(1-y) - [c_{34} + (c_{12} - c_{34})y]x(1-x), \quad (4.22)$$

где E_n — энергетический зазор в заданной точке зоны Бриллюэна бинарного соединения, а c_{mn} — коэффициенты нелинейности для трехкомпонентного твердого раствора, образованного бинарными соединениями m и n . Зависимости $E(xy) = E_g(xy)$ могут быть изображены графически аналогично тому, как это показано на рис. 4.9, однако они будут трехмерными, как это изображено на рис. 4.10. Поэтому изменение ширины запрещенной зоны в зависимости от состава твердых растворов типа $A_xB_{1-x}C_yD_{1-y}$ изоб-

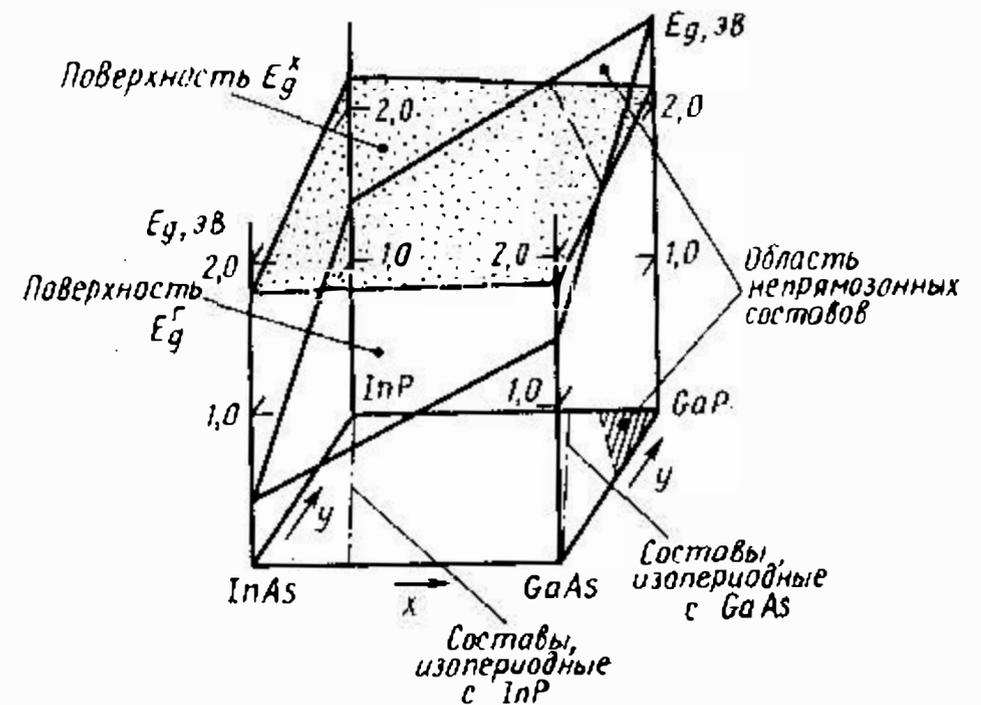


Рис. 4.10. Пространственное изображение изменений прямого E_{Γ}^{Γ} и непрямого E_{Γ}^X энергетических зазоров в четырехкомпонентном твердом растворе $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ при $T=300\text{ K}$

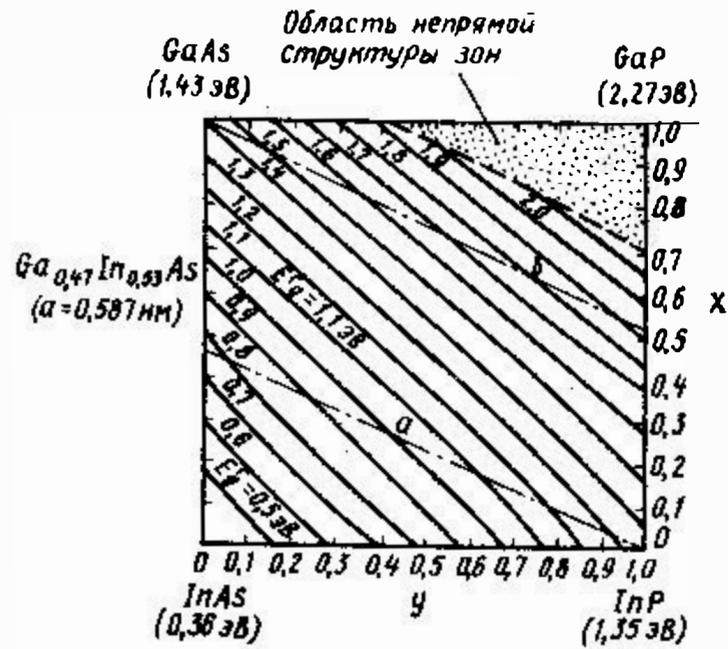


Рис. 4.11. Плоскость составов ($x-y$) для четырехкомпонентного твердого раствора $Ga_xIn_{1-x}P_yAs_{1-y}$ при $T=300$ К. Координаты x и y показывают состав в каждой точке. Сплошные линии получены проекцией линий $E_g^x(x,y) = \text{const}$ (рис. 4.10) и указывают изопериодические составы с длиной E_g^x . Штрих-пунктирные линии соответствуют составам твердых растворов, изопериодным InP (а) и $GaAs$ (б).

Как указывалось выше, одним из достоинств многокомпонентных систем является возможность получения изопериодного ряда твердых растворов. На рис. 4.11 пунктирными линиями показаны составы, изопериодные InP ($a=0,587$ нм) и $GaAs$ ($a=0,565$ нм). Эти зависимости получаются из (4.21), (4.22) и для системы $Ga_xIn_{1-x}P_yAs_{1-y}$ соотношение между x и y , удовлетворяющее условию изопериодичности с InP при $T=300$ К, имеет вид

$$y = \frac{0,1896 - 0,40519x}{0,1896 + 0,0127x}$$

Аналогичное соотношение для составов, изопериодных с $GaAs$:

$$y = \frac{0,40519(1-x)}{0,1896 + 0,0127x}$$

Наряду с твердыми растворами на основе соединений $A^{\text{III}}B^{\text{V}}$ в оптоэлектронике применяются узкозонные материалы на основе халькогенидов свинца, олова и ртути, а именно — твердые растворы в системах $PbTe - SnTe$ ($Pb_xSn_{1-x}Te$) и $CdTe - HgTe$ ($Cd_xHg_{1-x}Te$). Особенностью этих материалов является возможность получения сколь угодно малой ширины запрещенной зоны,

ражают в виде проекции на плоскость xy линий $E_g = \text{const}$, т. е. линий пересечения поверхности $E(x,y)$ с плоскостями $E = \text{const}$. В качестве примера такие зависимости показаны для $Ga_xIn_{1-x}P_yAs_{1-y}$ на рис. 4.11. Точка с координатами xy на плоскости составов определяет молярную долю бинарных псевдокомпонентов в твердом растворе. Область составов, где $E_g^x < E_g^y$, на рис. 4.11 заштрихована. Она соответствует твердым растворам с не прямой структурой зон. Основное поле диаграммы рис. 4.10 и 4.11 занято прямозонными материалами.

Как указывалось выше,

Вплоть до бесшелевого состояния $E_g = 0$. Эта возможность связана с особенностями строения энергетических зон составляющих твердый раствор бинарных соединений. Так, симметрия зоны проводимости $PbTe$ соответствует симметрии валентной зоны $SnTe$, и наоборот. Поэтому по мере добавления $SnTe$ к $PbTe$ зоны сближаются до тех пор, пока при $x=0,6$ не наступит их перекрытия, после чего при дальнейшем увеличении концентрации $SnTe$ зоны как бы меняются местами и начинают снова расходиться. Этот эффект называется инверсией зон. Он проиллюстрирован на рис. 4.12.

Наличие флюктуирующей части потенциала $V_i(r)$ в твердом растворе (4.19) должно приводить к «размазыванию» и уширению его энергетического спектра. Предположим, что мы имеем идеальный твердый раствор A_xB_{1-x} , состав которого x сохраняется постоянным по всему объему вещества. Но этот состав может быть постоянным только в среднем, т. е. в макроскопическом объеме. В каждой точке этого объема вследствие статистически неупорядоченного распределения компонент A и B состав будет флюктуировать вблизи среднего значения x . Эти микроскопические флюктуации состава тем более заметны, чем меньше область локализации, т. е. чем меньше выделенный объем.

Допустим, что в определенных узлах решетки могут находиться либо атомы сорта A , либо атомы сорта B . Пусть концентрация таких узлов есть N . Из них в достаточно большом объеме V атомами A заняты xNV , а атомами B — $(1-x)NV$ узлов. Оценим флюктуационное изменение Δx доли узлов A в некотором объеме радиуса R , считая R большим по сравнению со средним расстоянием между атомами A . Среднее число атомов A в этом объеме есть xNR^3 . При случайном характере заполнения узлов атомами A и B из теории вероятности имеем величину флюктуаций:

$$\Delta x = \frac{(x(1-x)NR^3)^{1/2}}{NR^3}$$

В объеме $R^3 \gg a^3$, ограниченном десятками и сотнями периодов идентичности a и включающем $10^3 \dots 10^6$ элементарных ячеек кристалла, усреднение состава происходит весьма эффективно. Напротив, в объеме, ограниченном несколькими элементарными

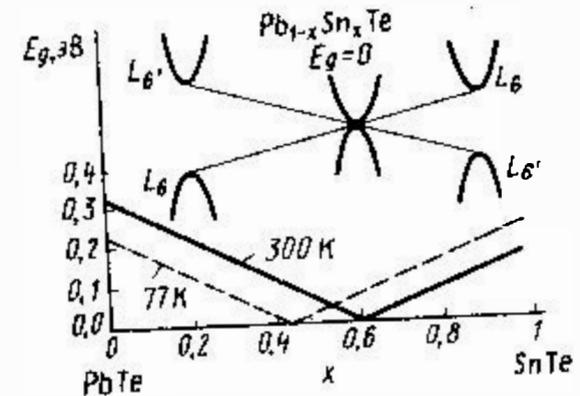


Рис. 4.12. Изменение зонной структуры с изменением состава твердого раствора $Pb_{1-x}Sn_xTe$ (эффект инверсии зон)

ячейками, флуктуации могут быть достаточно сильны. По этой причине наиболее чувствительны к принципиально существующим в твердом растворе микроскопическим флуктуациям состава будут состояния с малыми радиусами локализации (глубокие примеси, связанные экситоны и т. п.).

Для примера рассмотрим размытие дна зоны проводимости. Энергия, соответствующая дну зоны E_c , есть функция состава $E_c(x)$. При малых флуктуациях Δx она может быть представлена в виде

$$E_c(x, \Delta x) = E_c(x) + \alpha \Delta x,$$

где $E_c(x)$ — энергия дна зоны проводимости в приближении виртуального кристалла при $\Delta x \rightarrow 0$, а $\alpha = (dE_c(x)/dx)_{\Delta x \rightarrow 0}$. Таким образом, в рассматриваемом объеме возникает флуктуация потенциала с глубиной (высотой) ямы (горба)

$$\Delta V = \alpha \Delta x = \alpha \frac{[x(1-x)NR^3]^{1/2}}{NR^3}$$

Если ΔV и R удовлетворяют неравенству

$$\Delta V \gg \hbar^2/m_n R^2, \quad (4.23)$$

где m_n — эффективная масса электрона в зоне, для которой рассматриваются флуктуации, то такая яма содержит электронный уровень, близкий к ΔV . При обратном неравенстве (4.23) яма вообще не создает уровня. Вероятностный характер флуктуаций Δx и ΔV приводит к тому, что край дна зоны проводимости размывается так, что плотность состояний $g(E)$ убывает по экспоненциальному закону вглубь запрещенной зоны:

$$\ln g(E)/g(0) = -\sqrt{E/E_0}. \quad (4.24)$$

Здесь E — энергия, отсчитываемая вниз от дна зоны проводимости $E_c(x)$. Энергия E_0 является мерой размытия края зоны твердого раствора. Детальный расчет дает для E_0 :

$$E_0 = \frac{1}{178} \frac{\alpha^4 x^2 (1-x)^2 m_n^2}{N^6 N^2}. \quad (4.25)$$

Этот результат справедлив при $E \gg E_0$.

Аналогичная ситуация будет наблюдаться и для валентной зоны. В этом случае в (4.25) $\alpha = dE_v/dx$, а вместо m_n должна стоять эффективная масса тяжелой дырки.

Таким образом, микроскопические флуктуации состава и потенциала в твердом растворе должны вызывать неоднородное уширение энергетического спектра электронных состояний. Флю-

ктуации энергетического положения дна зоны проводимости и потолка валентной зоны приводят к появлению хвоста плотности состояний аналогично тому, как это наблюдается в сильно легированных полупроводниках. Примесные состояния будут испытывать сдвиг по энергии и расщепление, приводящие к неоднородному уширению их энергетических уровней. Это уширение тем больше, чем меньше боровский радиус состояния.

Оценки, проведенные по (4.25), показывают, что уширение $\delta E \approx E_0$ для собственных состояний в твердых растворах полупроводников $A^{\text{III}}B^{\text{V}}$ не превышает долей мэВ и в большинстве случаев ею можно пренебречь. Для твердых растворов $A^{\text{II}}B^{\text{VI}}$ типа ZnS-CdS E_0 достигает нескольких мэВ, а для узкозонных материалов пренебрежимо мала. Для мелких (водородоподобных) примесей с боровскими радиусами порядка 10 им и более уширение не превышает 1 мэВ, в то время как для неводородоподобных глубоких центров вследствие малости R эта величина может составлять сотни мэВ и её необходимо учитывать.

Наряду с кристаллами в оптической и квантовой электронике используются аморфные материалы, в том числе аморфные полупроводники. Среди них наибольшее значение имеют аморфный гидрогенизированный кремний и халькогенидные стекла. Эти материалы относятся к неупорядоченным системам с отсутствием дальнего порядка в расположении атомов. Непериодическая часть потенциала у них настолько велика, что вызванные ею флуктуации дна зоны проводимости E_c и потолка валентной

зоны E_v имеют порядок ширины запрещенной зоны. Электроны в зоне проводимости и дырки в валентной зоне разбиваются на «жабли», скапливающиеся в пространственно разделенных ямах потенциального рельефа, где образуются локализованные состояния. Электропроводность таких материалов при низких температурах носит прыжковый характер и при $T=0$ такие состояния вообще не участвуют в проводимости. При высоких температурах их электропроводность обусловлена тепловыми забросами электронов из локализованных флуктуациями потенциала состояний в область делокализованных состояний. Энергии E_c и E_v , разделяющие в неупорядоченных системах локализованные и делокализованные состояния, играют роль границ разрешенных и запрещенных зон и их называют *порогами подвижности*, как это показано на рис. 4.13. Если уровень Ферми лежит внутри полосы локализации, то эле-

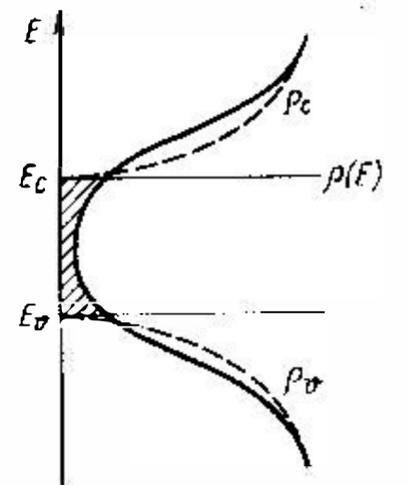


Рис. 4.13. Плотность электронных состояний в неупорядоченном аморфном полупроводнике. Область локализованных состояний заштрихована. Штриховыми линиями показана плотность состояний в эквивалентном кристалле

ктропроводность аморфного полупроводника экспоненциально возрастает с температурой. Если E_F лежит вне области локализации, то материал обладает металлической проводимостью.

Отметим, что по современным представлениям порог подвижности существует лишь в трехмерных неупорядоченных системах. В одномерных и двумерных неупорядоченных системах состояния локализованы при всех энергиях.

§ 4.3. ОПТИЧЕСКИЕ ПЕРЕХОДЫ В ПОЛУПРОВОДНИКАХ

В отличие от отдельных атомов и молекул энергетические состояния в твердых телах определяются по всему кристаллу и описываются волновыми функциями типа блоховских функций (4.3). Это накладывает особенности на процессы взаимодействия таких систем с электромагнитным излучением. Напомним, что под оптическими переходами понимают квантовые переходы, совершаемые под действием электромагнитного излучения оптического диапазона.

Рассмотрим квантовый переход из состояния j в состояние f при поглощении фотона $\hbar\omega$ и определим, при каких условиях такой переход возможен. Диаграмма взаимодействия для процесса поглощения фотона имеет вид

$$E_j \xrightarrow{\hbar\omega_k} E_f = V_{k_f k_j}^{e-\text{фотон}} \quad (4.26)$$

Здесь $V_{k_f k_j}^{e-\text{фотон}}$ — матричный элемент электрон-фотонного взаимодействия для состояний с волновыми векторами k_j и k_f .

Рассмотрим условия, при которых переход (4.26) возможен, т. е. возможен непосредственный оптический переход $k_j \rightarrow k_f$ при взаимодействии только с фотоном. Как и ранее (см. § 1.3), матричный элемент (4.26)

$$V_{k_f k_j}^{e-\text{фотон}} = \int \Psi_{k_f} V \Psi_{k_j} d\tau. \quad (4.27)$$

Возмущение V в нашем случае можно записать в виде плоской монохроматической волны:

$$V = A e^{i\mathbf{k}_{\text{фот}} \cdot \mathbf{r}} = A e^{i \frac{2\pi}{\lambda} (\mathbf{n} \cdot \mathbf{r})}, \quad (4.28)$$

где $\mathbf{k}_{\text{фот}}$ — волновой вектор фотона, а \mathbf{n} — единичный вектор в направлении распространения волны. Временной множитель $e^{-i\omega t}$ входит в коэффициент A .

Волновые функции электронов в кристалле, как обычно, записываем в виде блоховских функций (4.3), т. е.

$$\Psi_{k_j} = e^{i\mathbf{k}_j \cdot \mathbf{r}} u_{k_j}(\mathbf{r}), \quad \Psi_{k_f} = e^{i\mathbf{k}_f \cdot \mathbf{r}} u_{k_f}(\mathbf{r}), \quad (4.29)$$

где $u_{k_j}(\mathbf{r})$ и $u_{k_f}(\mathbf{r})$ — периодические функции с периодом решетки.

Подставляя (4.28) и (4.29) в (4.27), нетрудно убедиться, что подынтегральная функция в матричном элементе будет содержать множитель $\exp[i(-\mathbf{k}_f + \mathbf{k}_j + \mathbf{k}_{\text{фот}}) \cdot \mathbf{r}]$, который быстро осциллирует в пространстве, за исключением случая $\mathbf{k}_j + \mathbf{k}_{\text{фот}} - \mathbf{k}_f = 0$. Поэтому при интегрировании по \mathbf{r} при всех значениях k_j и k_f , не удовлетворяющих этому равенству, величина $V_{k_f k_j}^{e-\text{фот}}$ обращается в нуль. Отсюда вытекает правило отбора по волновому вектору при непосредственных оптических переходах $j \rightarrow f$:

$$\mathbf{k}_j + \mathbf{k}_{\text{фот}} = \mathbf{k}_f, \quad (4.30)$$

где $\mathbf{k}_{\text{фот}}$ — волновой вектор фотона ($\mathbf{k}_{\text{фот}} = \frac{2\pi}{\lambda} \mathbf{n}$).

Соотношение (4.30) аналогично закону сохранения импульса в классической физике. Действительно, если обе части выражения (4.30) умножить на \hbar , то получим правило отбора по квазиимпульсу:

$$\mathbf{p}_j + \mathbf{p}_{\text{фот}} = \mathbf{p}_f. \quad (4.30a)$$

Подчеркнем, что $\mathbf{p}_j = \hbar \mathbf{k}_j$ и $\mathbf{p}_f = \hbar \mathbf{k}_f$ не есть импульсы частиц в твердом теле, но по аналогии называются квазиимпульсами. Соотношения (4.30), (4.30a) выражают условия, при которых матричный элемент $V_{k_f k_j}^{e-\text{фот}}$ будет отличен от нуля. Для осуществления реального перехода, как отмечалось в § 1.3 [см. соотношение (1.87)], необходимо выполнение закона сохранения энергии. В нашем случае для непосредственного перехода $j \rightarrow f$ с поглощением фотона $\hbar\omega$ это приводит к условию

$$E_j + \hbar\omega = E_f.$$

Это условие легко получается из (4.27), если вместо волновых функций $\Psi_{\mathbf{k}}$ записать волновые функции $\phi_{\mathbf{k}}$, зависящие от време-

ни: $\phi_{k_j} = e^{i\frac{E_j}{\hbar} t} \Psi_{k_j}(\mathbf{r})$, и учесть временной множитель в (4.29).

Очевидно, приведенные рассуждения справедливы не только для процессов поглощения, но и для процессов испускания фотона. Таким образом, для осуществления непосредственно оптического перехода как с испусканием, так и с поглощением фотона в твердых телах необходимо выполнение двух законов сохранения — энергии и волнового вектора:

$$E_j \pm \hbar\omega = E_f, \quad (4.31)$$

$$\mathbf{k}_j \pm \mathbf{k}_{\text{фот}} = \mathbf{k}_f. \quad (4.32)$$

Знак плюс соответствует поглощению, а знак минус — испусканию. Поскольку волновой вектор k в кристаллах определяется с точностью до вектора обратной решетки q , соотношение (4.30) справедливо с точностью до $2\pi/a$.

В видимой и инфракрасной областях спектра волновой вектор фотона $2\pi/\lambda$ значительно меньше волнового вектора электрона у границ зоны Бриллюэна $2\pi/a$. Кроме того, импульс фотона с $\lambda \approx 1$ мкм соответствует изменению энергии электрона при параболическом законе дисперсии и $m^* \approx m$ всего на $7 \cdot 10^{-7}$ эВ. Поэтому величинами $k_{\text{фот}}$ и $p_{\text{фот}}$ в соотношениях (4.30), (4.31) можно пренебречь. Тогда в соответствии с законом сохранения (4.32) при взаимодействии только с фотоном могут осуществляться только переходы без изменений волнового вектора:

$$k_j = k_f \quad (4.33)$$

На диаграмме $E(k)$ такой переход изображается вертикальной линией и называется *прямым* или *вертикальным переходом*.

Таким образом, непосредственные оптические переходы с поглощением и испусканием фотона возможны лишь между состояниями с одинаковыми волновыми векторами, т. е. расположенными в одной и той же точке k -пространства. Типичным примером такого перехода является межзонный переход из k_j -состояния валентной зоны в k_f -состояние зоны проводимости (рис. 4.14, а). Диаграмма взаимодействия для подобного перехода может быть аналогичной диаграмме (4.26), но в таком виде она изображает процесс развития на языке электронных переходов, который является довольно неудобным, когда зоны целиком заполнены электронами. В рассматриваемом случае в результате поглощения фотона $\hbar\omega$ в зоне проводимости появляется свободный электрон, а в валентной зоне — дырка, что в явном виде не отражено на диаграмме (4.26). При таком описании трудно про-

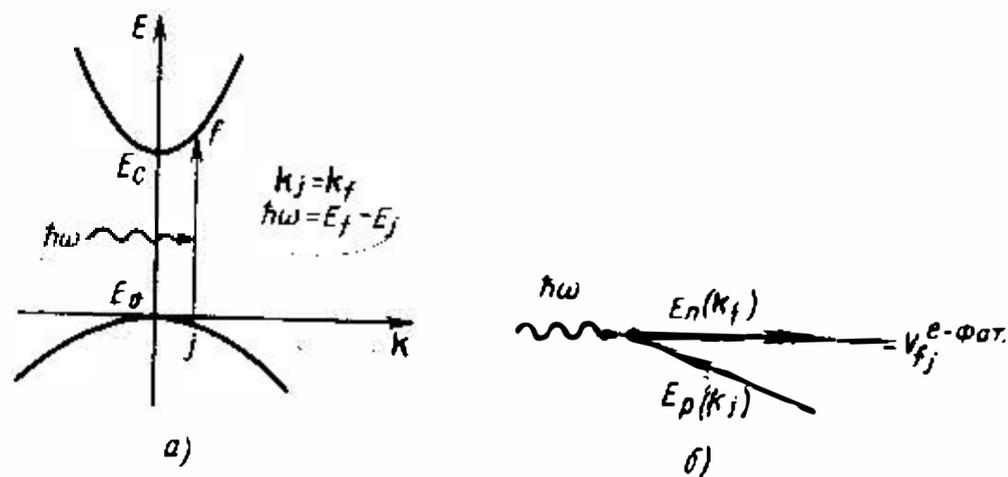


Рис. 4.14. Прямой оптический переход из j -состояния валентной зоны в f -состояние зоны проводимости при поглощении фотона $\hbar\omega$ (а) и соответствующая диаграмма взаимодействия (б)

следить за тем, что происходит с образующейся в процессе поглощения дыркой, поскольку для этого нужно рассмотреть движение всех валентных электронов в отсутствие одного из них. По этой причине графическое изображение процесса модернизируем, введя явное изображение дырки в валентной зоне.

Энергию дырки, как обычно принято, будем отсчитывать вниз от потолка валентной зоны E_v (рис. 4.14, а). Тогда волновая функция для состояния с волновым вектором k в валентной зоне может быть записана аналогично (1.67):

$$\begin{aligned} \varphi_{v,k} &= \Psi_{v,k}(r) e^{-\frac{i}{\hbar} [E_v(0) - E_v(k)] t} \\ &= \Psi_{v,k}(r) e^{-\frac{i}{\hbar} [E_v(k) - E_v(0)] (-t)} = \Psi_{v,k}(r) e^{-\frac{i}{\hbar} E_p(k) t} \end{aligned} \quad (4.34)$$

где $E_p(k)$ — энергия дырки, отсчитываемая вниз от потолка валентной зоны.

Для электрона в зоне проводимости энергия по-прежнему отсчитывается вверх, поэтому волновая функция имеет вид

$$\varphi_{c,k} = \Psi_{c,k} e^{-\frac{i}{\hbar} E_c(k) t} \quad (4.35)$$

Функции $\Psi_{v,k}$ и $\Psi_{c,k}$ являются блоховскими функциями (4.29). Из сравнения временных множителей в (4.34) и (4.35) видно, что на временной диаграмме движению дырки при увеличении времени должна соответствовать линия, направленная в сторону отрицательных времен.

С учетом сказанного процесс поглощения фотона, сопровождающийся межзонным переходом электрона из валентной зоны в зону проводимости, можно изобразить диаграммой взаимодействия рис. 4.14, б. До того как фотон поглотился, валентная зона была полностью заполнена, а зона проводимости — пустая. Следовательно, до акта взаимодействия имелась только одна фотонная линия. В момент взаимодействия фотонная линия исчезает (фотон поглощается), но появляются две новые линии, соответствующие рождению электрона в зоне проводимости и дырки в валентной зоне, т. е. можно наглядно проследить за более сложными процессами взаимодействия.

Фотон практически не изменяет волновой вектор электрона или дырки, поэтому оптический переход в первом приближении теории возмущений возможен лишь между состояниями с одинаковыми значениями волнового вектора. Если j - и f -состояния принадлежат, например, одной долине (одной подзоне) зоны

проводимости, или валентной зоны, или различным зонам, экстремумы которых расположены в разных точках зоны Бриллюэна. Тогда непосредственный оптический переход $j \leftrightarrow f$ между этими состояниями будет невозможен, поскольку для них $V_{ff} = 0$.

Как указывалось в § 1.3, равенство $V_{ff} = 0$ не означает, что переход $j \leftrightarrow f$ вообще невозможен. Он невозможен лишь в первом приближении теории возмущений, но может быть возможен во втором приближении при переходе через промежуточное виртуальное состояние. В этом случае необходимо рассмотреть второе слагаемое в выражении (1.81) для матричного элемента перехода, т. е. процесс, описываемый диаграммой (1.89). Так как взаимодействие только с фотоном практически не изменяет волнового вектора электрона, то для осуществления перехода $j \leftrightarrow f$ при $|\mathbf{k}_j - \mathbf{k}_f| > \mathbf{k}_{\text{фот}}$ требуется дополнительный процесс, приводящий к рассеянию волнового вектора.

Закон сохранения волнового вектора (4.32) вытекает из условия трансляционной симметрии кристалла, поскольку волновые функции электронов представлены в виде блоховских функций (4.29). Различные нарушения трансляционной симметрии, определяемые дефектами, будут выполнять роль рассеивающих центров и приводить к возможности «нарушения» закона сохранения (4.32).

Таким образом, для реализации оптического перехода из j в f -состояние с существенно различными значениями волновых векторов \mathbf{k}_j и \mathbf{k}_f необходим дополнительный процесс рассеяния с участием какого-либо рассеивающего центра или квазичастицы, взаимодействие с которой способно изменить волновой вектор электрона (или дырки). Такими квазичастицами и рассеивающими центрами могут быть кванты колебаний кристаллической решетки (фононы), свободные носители заряда (электроны и дырки), примесные атомы, границы раздела и т. д.

Рассмотрим в качестве примера оптический переход $j \rightarrow f$, определяющий процесс поглощения фотона $\hbar\omega$ с рассеянием за счет взаимодействия с фононом $\hbar\Omega_{\mathbf{k}}$. Этот процесс разрешен во втором приближении теории возмущений, и ему соответствует следующая диаграмма взаимодействия:

$$\frac{V_{jl} V_{lf}}{(E_l - E_j - \hbar\omega)} \quad (4.36)$$

На этой диаграмме фононная ветвь изображена пунктирной линией и для определенности показано, что рассеяние осуществляется путем испускания фонона $\hbar\Omega$ с волновым вектором \mathbf{k} (рассеяние может осуществляться также за счет поглощения фонона). Матричный элемент $V^{\text{эфон}}$ учитывает электрон-фононное взаимо-

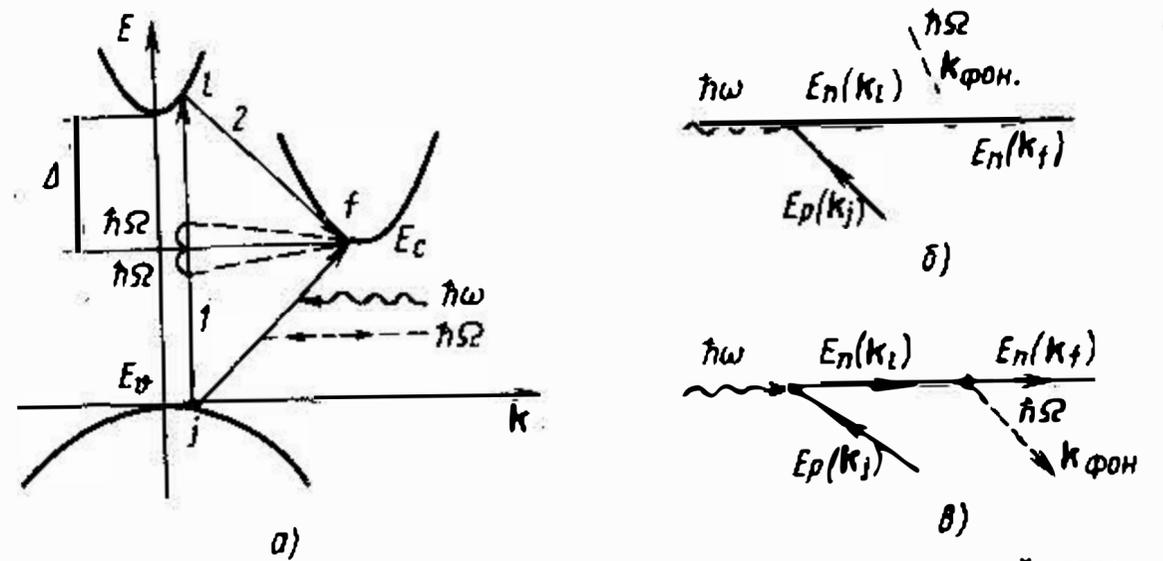


Рис. 4.15. Непрямой оптический переход из j -состояния валентной зоны в f -состояние зоны проводимости при поглощении фотона $\hbar\omega$ и рассеянии на фононе:

a — на диаграмме $E(\mathbf{k})$;

$b, в$ — диаграммы взаимодействия для процессов поглощения фотона при одновремениом поглощении (b) и испускании ($в$) фонона с энергией $\hbar\Omega$ и волновым вектором $\mathbf{k}_{\text{фон}} = \mathbf{k}_f - \mathbf{k}_j$. Состояние l виртуальное

действие, т. е. является матричным элементом рассеяния. Состояние E_l есть виртуальное состояние.

Если начальное j -состояние принадлежит валентной зоне, а конечное f -состояние — непрямому экстремуму, то процесс поглощения фотона $\hbar\omega$ будет изображен, как на рис. 4.15. Под воздействием электромагнитного возмущения электрон переходит из j -состояния в промежуточное, виртуальное l -состояние. Таким состоянием, например, может являться более высоко лежащее состояние зоны проводимости (переход 1 на рис. 4.15, a). Время δt пребывания электрона в этом промежуточном состоянии чрезвычайно мало, так что из соотношения неопределенности $\delta E \delta t \gtrsim \hbar$ неопределенность в энергии δE может быть достаточно велика, при переходе в виртуальное состояние $j \rightarrow l$ закон сохранения энергии не соблюдается. Находясь в виртуальном l -состоянии, электрон взаимодействует с фононом (переход 2 на рис. 4.15, a), в результате чего электрон переходит в конечное f -состояние. Полный переход $j \rightarrow f$ возможен, если соблюдаются законы сохранения энергии и волнового вектора. Но поскольку фонон обладает энергией $\hbar\Omega$ и волновым вектором $\mathbf{k}_{\text{фон}}$, соответствующие законы сохранения при оптических переходах с участием фононов, как видно из рис. 4.15, можно переписать в виде

$$E_j + \hbar\omega \pm \hbar\Omega_{\text{фон}} = E_f, \quad (4.37)$$

$$\mathbf{k}_j \pm \mathbf{k}_{\text{фон}} = \mathbf{k}_f. \quad (4.38)$$

Знаки в этих выражениях соответствуют поглощению и испусканию фонона. Им отвечают диаграммы взаимодействия изображенные на рис. 4.15, б, в. Оптические переходы между состояниями с различными значениями волнового вектора называются *непрямыми* (рис. 4.15). Они разрешены лишь во втором приближении теории возмущений, поэтому вероятность таких переходов существенно меньше, чем прямых.

Заметим, что в матричных элементах для непрямых переходов стоит энергетический множитель $(\hbar\omega - E_f)^{-1}$. Поэтому, чем ближе виртуальное состояние расположено по энергии к конечному f -состоянию, тем больше будет вероятность непрямого перехода.

§ 4.4. ОПТИЧЕСКОЕ ПОГЛОЩЕНИЕ В КРИСТАЛЛАХ

Поглощение электромагнитного излучения твердым телом может осуществляться различными путями. В силу адиабатического приближения все процессы, приводящие к оптическому поглощению, можно разбить на две группы: 1) процессы, в результате которых энергия электромагнитного поля передается электронной подсистеме, т. е. поглощение фотона сопровождается переводом электрона в более высокое энергетическое состояние; 2) процессы, в результате которых энергия поля непосредственно передается решетке, т. е. поглощение фотона сопровождается возбуждением одного или нескольких фононов. Этим двум случаям в идеальном кристалле соответствует со-

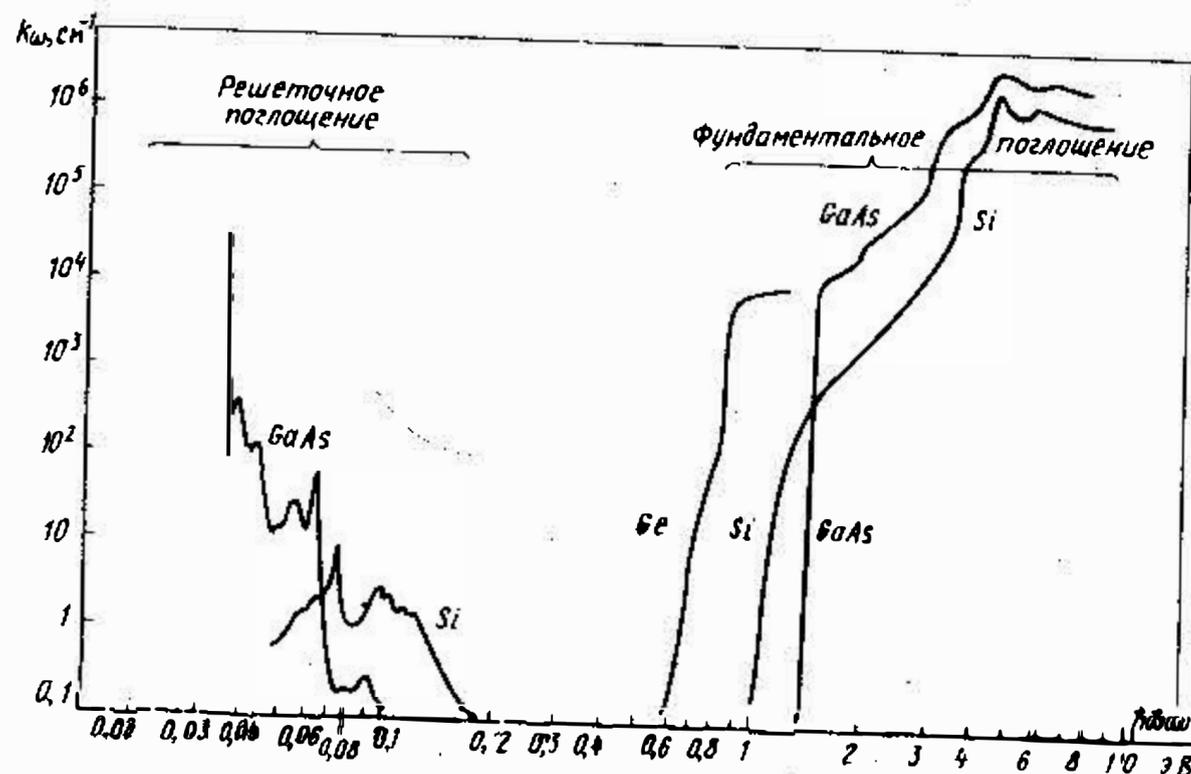


Рис. 4.16. Спектры собственного поглощения нелегированных кристаллов арсенида галлия, кремния и германия при комнатной температуре

бственное межзонное фундаментальное поглощение и собственное решеточное поглощение. В качестве примера на рис. 4.16 представлены спектры собственного поглощения нелегированных кристаллов кремния (Si), германия (Ge) и арсенида галлия (GaAs) в широком диапазоне энергий фотонов (длин волн) и значений показателя поглощения.

Наличие примесей и свободных носителей заряда в реальных кристаллах приводит к появлению дополнительного примесного поглощения и поглощения свободными носителями заряда. Поэтому различают следующие механизмы оптического поглощения в твердых телах: 1) фундаментальное поглощение; 2) решеточное поглощение; 3) примесное поглощение; 4) поглощение свободными носителями заряда. Электронные переходы, ответственные за указанные процессы поглощения, изображены на рис. 4.17. Рассмотрим особенности перечисленных механизмов поглощения.

Фундаментальное (собственное межзонное) поглощение. Фундаментальное поглощение определяется межзонными переходами электронов из заполненной зоны в свободную. В зависимости от энергетического расстояния между этими зонами такое поглощение наблюдается от инфракрасной до рентгеновской области спектра, включая видимую и ультрафиолетовую области. Так как в твердом теле имеется множество заполненных и свободных энергетических зон, каждая из которых в свою очередь состоит из нескольких подзон, то в спектрах фундаментального поглощения наблюдается ряд широких полос с собственной структурой, что определяется сложным строением каждой из зон.

Минимальная энергия фотонов, при которой начинается собственное поглощение в идеальных неметаллических кристаллах, определяется минимальным энергетическим зазором между полностью заполненной (валентной) зоной и свободной зоной (зоной проводимости), т. е. шириной запрещенной зоны E_g полупроводника или диэлектрика. Область вблизи $\hbar\omega \approx E_g$ называется *краем собственного (фундаментального) поглощения*. Обычно именно эта область представляет наибольший интерес, поскольку она дает богатую информацию об энергетических состояниях электронов вблизи потолка валентной зоны и вблизи абсолютного миниму-

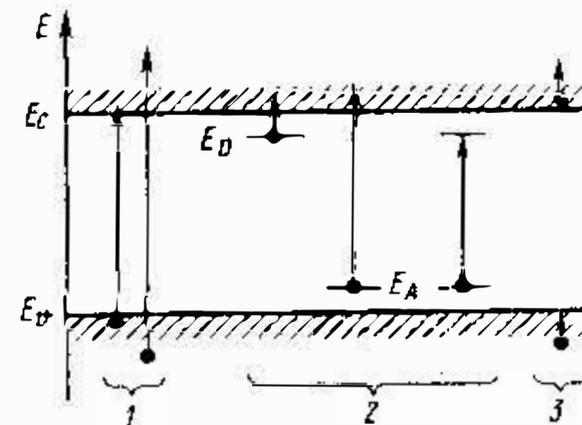


Рис. 4.17. Схематическое изображение электронных переходов, ответственных за процессы оптического поглощения в полупроводниках: 1 — фундаментальное поглощение; 2 — примесное поглощение; 3 — поглощение свободными носителями заряда

ма зоны проводимости, т. е. в наиболее важных областях энергетического спектра кристалла.

Форма края собственного поглощения определяется прежде всего особенностями структуры энергетических зон материала. В кристаллах с прямой структурой энергетических зон (см. § 4.1) край собственного поглощения определяется прямыми оптическими переходами из j -состояния валентной зоны в f -состояние зоны проводимости, как изображено на рис. 4.14. В кристаллах с непрямой структурой зон край собственного поглощения определяется непрямыми оптическими переходами $j \rightarrow f$ (см. рис. 4.15).

Рассмотрим сначала край собственного поглощения полупроводниковых кристаллов с прямой структурой энергетических зон. Показатель поглощения k_ω пропорционален вероятности перехода w_{ff} и плотности начальных (занятых) и конечных (свободных) электронных состояний, разделенных энергией $\hbar\omega$, для которых выполняются правила отбора (4.31), (4.32). Число таких пар состояний, приходящихся на единичный интервал энергий и единицу объема кристалла, называется комбинированной плотностью состояний и обозначается $N(\hbar\omega)$. То есть $N(\hbar\omega) = N(E_f - E_j)$ — плотность состояний, разделенных энергией $(E_f - E_j)$. Тогда

$$k_\omega \sim w_{ff} N(\hbar\omega). \quad (4.39)$$

Учитывая, что энергия свободного электрона отсчитывается от дна зоны проводимости E_c , а энергия дырки — от потолка валентной зоны E_v , можно записать

$$E_j = \hbar\omega - E_f - E_g.$$

Для параболических зон

$$E_f = \hbar^2 k^2 / (2m_n), \quad E_j = \hbar^2 k^2 / (2m_p),$$

откуда

$$\hbar\omega - E_g = \hbar^2 k^2 / (2m_r). \quad (4.40)$$

Здесь m_r — приведенная эффективная масса:

$$1/m_r = 1/m_n + 1/m_p.$$

Так как для прямых межзонных переходов должно выполняться условие (4.33), то комбинированную плотность состояний при этом рассчитывают аналогично обычной плотности состояний для зоны проводимости или валентной зоны. Объем k -пространства между сферами, разделенными по энергии от $\hbar\omega$ до $\hbar\omega + d(\hbar\omega)$, равен $4\pi k^2 dk$. На одно состояние в k -пространстве

приходится объем $(2\pi)^3/V$, где V — объем кристалла. Принимая для удобства $V=1$ и считая, что в каждом состоянии может находиться два электрона, получаем с учетом (4.40) число состояний в интервале энергий от $\hbar\omega$ до $\hbar\omega + d(\hbar\omega)$:

$$N(\hbar\omega) d(\hbar\omega) = \frac{8\pi k^2 dk}{(2\pi)^3} \frac{(2m_r)^{3/2}}{2\pi^2 \hbar^3} (\hbar\omega - E_g)^{1/2} d(\hbar\omega). \quad (4.41)$$

Вероятность перехода w_{ff} определяется квадратом матричного элемента $|V_{ff}|^2$. Ее зависимость от волнового вектора k (или от энергии $\hbar\omega$) определяется видом волновых функций u_{kf} и u_{kj} в (4.29).

Для большинства полупроводниковых материалов симметрия волновых функций u_{kf} и u_{kj} такова, что матричный элемент V_{ff} слабо зависит от волнового вектора. Такие переходы называются разрешенными по симметрии. В этом случае можно положить $V_{ff} = V_{cv}(k=0) = \text{const}$. Тогда из (4.39), (4.41) получаем спектральную зависимость показателя поглощения для прямых разрешенных переходов:

$$k_\omega = A(\hbar\omega - E_g)^{1/2}. \quad (4.42)$$

В коэффициент пропорциональности A входят квадрат матричного элемента $|V_{cv}|^2$, приведенная эффективная масса $m_r^{3/2}$ и другие константы. Оценка коэффициента пропорциональности может быть проведена с помощью соотношения

$$A \approx \frac{e^2 (2m_r)^{3/2}}{nch^2 m \epsilon_0} f_{cv}. \quad (4.43)$$

Коэффициент f_{cv} , называемый силой осциллятора, имеет порядок около единицы. Для полупроводников группы $A^{III}B^V$

$$f_{cv} = \frac{3|V_{cv}(0)|^2}{2 m \hbar \omega} \approx \frac{1}{3} \left(\frac{m}{m_n^r} - 1 \right). \quad (4.44)$$

Для оценки коэффициента A примем $m_p = m$, $m_n = 0,25m$, $n=4$, что дает $A \approx 10^4 \text{ см}^{-1} \cdot \text{эВ}^{-1/2}$. Из (4.42) следует, что при $\hbar\omega = E_g$, $k_\omega = 0$. В случае $\hbar\omega - E_g = 0,01 \text{ эВ}$ $k_\omega \approx 10^3 \text{ см}^{-1}$, т. е. собственное поглощение для прямых разрешенных переходов очень быстро возрастает с увеличением энергии фотонов при $\hbar\omega \gtrsim E_g$.

В реальном полупроводнике, содержащем примеси, свободные носители заряда или другие дефекты, край собственного поглощения будет отличаться от (4.42) в связи с возникновением хвостов плотности состояний и откликом функций ρ_c и ρ_v вблизи E_c и E_v от (4.9) (см. рис. 4.5 и 4.6). Это в свою очередь приведет

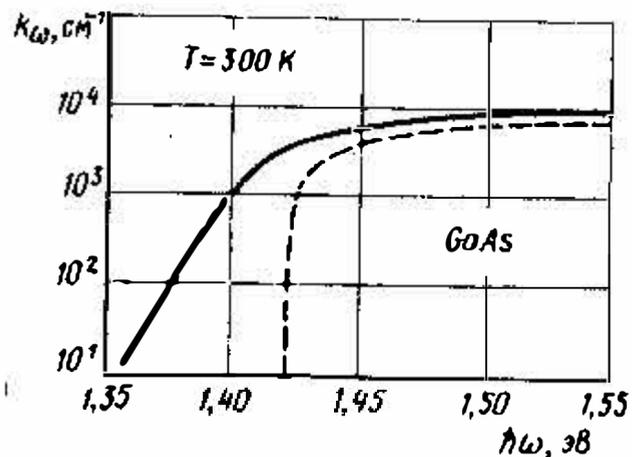


Рис. 4.18. Влияние хвостов плотности состояний на край поглощения арсенида галлия.

Пунктирной кривой изображен спектр собственного поглощения GaAs, рассчитанный по (4.72) в предположении параболических зон. Сплошная кривая — экспериментальные данные для GaAs, легированного мелкими акцепторами до $p = 1,2 \cdot 10^{18} \text{ см}^{-3}$ ($T = 300 \text{ К}$)

к возникновению характерных экспоненциальных хвостов на краю собственной поглощения прямозонного полупроводника. В качестве примера на рис. 4.18 сравниваются рассчитанные по (4.42) и измеренные экспериментально для легированного до $p = 1,2 \cdot 10^{18} \text{ см}^{-3}$ арсенида галлия спектры поглощения вблизи края фундаментальной полосы. Обратите внимание на взаимное соответствие данных на рис. 4.18 и рис. 4.6.

В области энергий фотонов, заметно превышающих ширину запрещенной зоны, особенности спектра со-

бственного поглощения определяются в основном особенностями функции $N(\hbar\omega)$ в (4.39). В свою очередь комбинированная плотность состояний $N(\hbar\omega)$ как функция энергии $\hbar\omega = E_f - E_j$ при $k_f = k_j$ определяется особенностями структуры энергетических зон материала. Очевидно, максимум функции $N(\hbar\omega)$ должен наблюдаться в той области энергий, где дисперсионные кривые $E(k)$ для начальных состояний E_j (валентная зона) и конечных состояний E_f (зона проводимости) идут параллельно друг другу. В этой области $E_f - E_j \approx \text{const}$. Из рис. 4.1 можно видеть, что этому условию могут отвечать переходы вблизи состояний $L_3^V - L_1^C$ и $\Delta_3^V - \Delta_1^C$ ($X_3^V - X_1^C$).

Современные методы расчета зонной структуры полупроводников позволяют достаточно надежно определить общий вид спектра $N(\hbar\omega)$ для конкретных материалов и по нему найти мнимую часть комплексной диэлектрической проницаемости или комплексного показателя преломления, связанные с показателем поглощения k_ω соотношениями (1.168), (1.177). Для большинства полупроводников с тетраэдрической координацией атомов, кристаллизующихся в структуры алмаза или сфалерита, в спектре ϵ_2 наблюдаются характерные максимумы при энергиях E_1 и E_2 , как показано на рис. 4.19. Междузонные оптические переходы, формирующие эти особенности, указаны стрелками на вставке рис. 4.19. Максимумам ϵ_2 в силу (1.209) соответствуют максимумы в спектре отражения при $\hbar\omega > E_g$. Обратим внимание, что край собственного поглощения, обусловленный прямыми переходами, на рис. 4.19 расположен при энергии $\hbar\omega = E_0$. Характерные значе-

ния энергий для GaAs: $E_1 = 2,9 \text{ эВ}$ и $E_2 = 5,0 \text{ эВ}$, а $E_0 = 1,425 \text{ эВ}$. В группе полупроводников $A^{III}B^V$, как, впрочем, и в других группах материалов, энергии E_0 , E_1 и E_2 сдвигаются с изменением ΣZ_i аналогично тому, как это показано на рис. 4.3.

Вернемся снова к краю собственного поглощения, но для полупроводников с непрямой структурой энергетических зон (Ge, Si, GaP, SiC и др.). Оптические переходы, определяющие край поглощения в таких материалах, будут непрямыми, как изображено на рис. 4.15, а законы сохранения будут определяться взаимодействием с фононами согласно соотношениям (4.37) и (4.38).

Для определенности рассмотрим процесс непрямого оптического поглощения фотона $\hbar\omega$ с одновременным поглощением фонона $\hbar\Omega$ при переходе через ближайшее виртуальное состояние l в зоне проводимости. Диаграмма взаимодействия для такого перехода изображена на рис. 4.15, в. Допустим, что оптический переход $j \rightarrow l$ в виртуальное состояние разрешен. Тогда, как и ранее, оптический матричный элемент $V_{ij}^{0\text{фот}}$ можно считать не зависящим от волнового вектора, т. е. положить $V_{ij}^{0\text{фот}}(\mathbf{k}) = V_{ij}(0) = \text{const}$. При этом показатель поглощения будет пропорционален эффективной плотности состояний, разделенных энергией $\hbar\omega + \hbar\Omega$, и вероятности взаимодействия с фононами. Поскольку рассматривается процесс, протекающий с поглощением фононов, вероятность взаимодействия будет пропорциональна числу фононов N_Ω . Фононы подчиняются статистике Бозе — Эйнштейна, поэтому

$$N_\Omega = 1 / (e^{\hbar\Omega/kT} - 1). \quad (4.45)$$

При непрямых оптических переходах (в отличие от прямых) возможны переходы из любого занятого j -состояния валентной зоны в любое свободное f -состояние зоны проводимости, лишь бы выполнялся закон сохранения энергии (4.37). Тогда комбинированная плотность состояний $N(\hbar\omega)$ при непрямых переходах будет определяться интегралом по всем возможным парам состояний, разделенных энергией $\hbar\omega + \hbar\Omega$, от произведения плот-

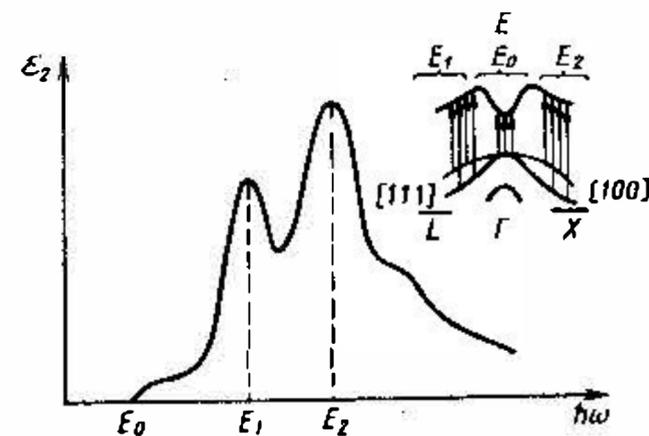


Рис. 4.19. Спектр $\epsilon_2(\hbar\omega)$ в области собственного поглощения, характерный для полупроводников с тетраэдрической координацией атомов. На вставке на энергетической диаграмме вертикальными стрелками показаны прямые переходы, формирующие характеристические особенности спектра $\epsilon_2(\hbar\omega)$

ностей начальных $\rho_v(E_j)$ и конечных $\rho_c(E_f)$ состояний. Для зон с параболическим законом дисперсии из (4.9) имеем

$$\begin{aligned} \rho_v(E_j) &= \frac{1}{2\pi^2} (2m_p/\hbar^2)^{3/2} E_j^{1/2}, \\ \rho_c(E_f) &= \frac{N_c}{2\pi^2} (2m_n/\hbar^2)^{3/2} E_f^{1/2}. \end{aligned} \quad (4.46)$$

В отличие от (4.37) энергии E_j и E_f в (4.46) отсчитываются от разных точек. Поэтому вместо соотношения (4.37) можно написать

$$\hbar\omega = E_f + E_g + E_j - \hbar\Omega_k. \quad (4.47)$$

Тогда

$$N(\hbar\omega) = \int_0^\infty dE_f \int_0^\infty \rho_c \rho_v \delta(E_f + E_g + E_j - \hbar\omega - \hbar\Omega_k) dE_j \approx \int_0^{\hbar\omega - E_g + \hbar\Omega_k} \rho_c \rho_v dE_f.$$

Подставляя (4.46) и интегрируя, находим

$$N(\hbar\omega) = \frac{N}{\pi^3} \left(\frac{2}{\hbar^2}\right)^3 (m_n m_p)^{3/2} (\hbar\omega - E_g + \hbar\Omega_k)^2. \quad (4.48)$$

Таким образом, спектральная зависимость показателя поглощения для непрямых оптических переходов с поглощением фонона (индекс a) имеет вид

$$k_\omega^{(a)} = \frac{B}{\frac{\hbar\Omega}{e^{kT} - 1}} (\hbar\omega - E_g + \hbar\Omega)^2. \quad (4.49)$$

В коэффициент пропорциональности B входят произведение квадратов оптического матричного элемента $V_{ij}^{c \rightarrow v}$ и матричного элемента рассеяния $V_{ij}^{v \rightarrow c}$, эффективные массы плотности состояний, энергетический множитель $1/(\hbar\omega - E_g)^2$ и другие константы.

Аналогичные рассуждения можно провести и для переходов с испусканием фононов. При этом знак перед $\hbar\Omega_k$ в формулах (4.47) — (4.49) изменится на противоположный. Вероятность эмиссии фонона с учетом его спонтанного испускания пропорциональна $N_\Omega + 1$. Поэтому показатель поглощения для переходов с испусканием фононов (индекс e)

$$k_\omega^{(e)} = \frac{B}{1 - e^{-\hbar\Omega/kT}} (\hbar\omega - E_g - \hbar\Omega_k)^2. \quad (4.50)$$

Из выражений (4.49) и (4.50) видно, что для непрямых переходов поглощение фотона с одновременным испусканием фонона будет начинаться при энергии $\hbar\omega_e = E_g + \hbar\Omega_k$, а с поглощением фонона — при энергии $\hbar\omega_a = E_g - \hbar\Omega_k$. Эти энергии называются пороговыми энергиями для непрямых переходов. Суммарный показатель поглощения

$$k_\omega = k_\omega^{(a)} + k_\omega^{(e)}$$

Как видно из (4.49) и (4.50), спектр собственного поглощения вблизи края основной полосы при непрямых электронных переходах в координатах $\sqrt{k_\omega}(\hbar\omega)$ будет представлять собой две прямые, пересекающие ось абсцисс в точках $E_g + \hbar\Omega$ и $E_g - \hbar\Omega$ (рис. 4.20). Непрямая ширина запрещенной зоны E_g лежит посредине между этими точками. В соответствии со статистикой распределения фононов (4.45) процессы с их поглощением будут заметно проявляться при сравнительно высоких температурах (обычно выше 70 К). При низких температурах возможны лишь процессы с испусканием фононов, поэтому $k_\omega^{(e)} \rightarrow 0$ при $T \rightarrow 0$.

На рис. 4.20 учтено взаимодействие только с одним типом фононов. На самом деле необходимо учесть все ветви колебаний: поперечную и продольную акустическую, а также оптические. При их учете показатель поглощения k_ω для непрямых переходов можно представить в виде суммы

$$k_\omega = \sum_i (k_\omega^{(a)} + k_\omega^{(e)}), \quad (4.51)$$

где индекс i относится к данному типу фононов, а индексы e и a — к компонентам с испусканием и поглощением фононов соответственно.

Поскольку непрямые переходы разрешены лишь во втором порядке теории возмущений, то их интенсивность существенно меньше, чем прямых, а край собственного поглощения в не-

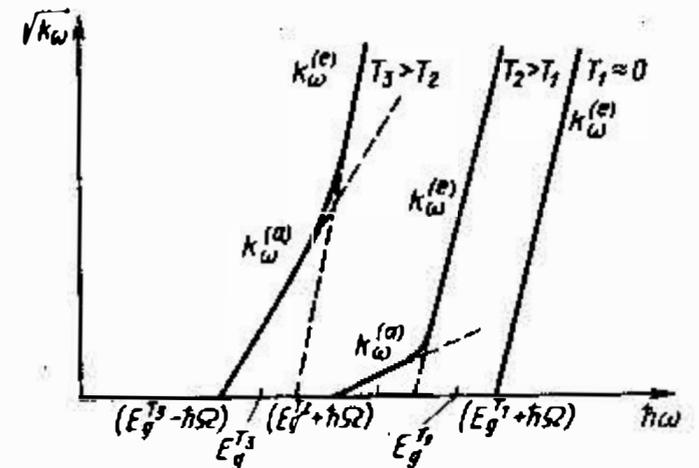


Рис. 4.20. Край собственного поглощения при непрямых электронных переходах в координатах $\sqrt{k_\omega}(\hbar\omega)$ при различных температурах

прямозонных материалах менее резкий, чем в прямозонных (см. рис. 4.16).

В процессе собственного поглощения фотона в одной и той же точке кристалла появляются неравновесные носители заряда противоположного знака — электрон в зоне проводимости и дырка в валентной зоне (переход 1 на рис. 4.17). Очевидно, между ними существует взаимодействие, которым мы пренебрегали. Это взаимодействие, осуществляемое преимущественно за счет сил кулоновского притяжения, приводит к тому, что рожденные в результате межзонного поглощения фотона электрон и дырку следует рассматривать не независимо, а как связанную электроно-дырочную пару. Эта пара взаимодействующего электрона и дырки ведет себя как единая квазичастица, получающая название *экситон**. Ситуация здесь по существу аналогична рассмотренной в § 1.2 для атома водорода с той разницей, что взаимодействие между положительно и отрицательно заряженными частицами будет осуществляться не в свободном пространстве, а в кристалле. Как и атому водорода, экситону можно приписать свою кинетическую энергию, связанную с движением центра масс, и внутреннюю энергию, обусловленную кулоновским взаимодействием электрона и дырки. Но в отличие от атома водорода экситон является возбужденной, неравновесной квазичастицей (отсюда происходит его название). Он может исчезнуть либо в результате теплового «довозбуждения», т. е. термической диссоциации, сопровождающейся возникновением свободных электрона и дырки, либо вследствие *аннигиляции* с испусканием фотона, либо за счет передачи своей энергии решетке, т. е. фононам.

В полупроводниках характерные размеры экситона обычно превышают размеры элементарной ячейки кристалла. Такие экситоны называются экситонами большого радиуса или экситонами Ванье — Мотга. Для них действие кристалла может быть учтено простой заменой m и M в (1.36) — (1.40) на эффективные массы m_e и m_p и введением вместо ϵ_0 диэлектрической проницаемости ϵ_r , учитывающей ослабление кулоновского взаимодействия в кристалле. Приведением к центру масс волновую функцию экситона можно представить в виде

$$\Phi_{\mathbf{K},n} = \frac{1}{\sqrt{V}} e^{i\mathbf{K}\mathbf{R}} \Psi_n(\mathbf{r}), \quad (4.52)$$

где $\mathbf{K} = \mathbf{k}_e + \mathbf{k}_p$ — экситонный волновой вектор; \mathbf{R} — координаты центра масс; V — объем кристалла; $\mathbf{r} = \mathbf{r}_e - \mathbf{r}_p$.

*От англ. excite — возбуждать.

Плоская волна, как и в случае атома водорода, отображает движение центра масс, а волновая функция $\Psi_n(\mathbf{r})$ описывает состояние внутреннего движения. Она удовлетворяет уравнению

$$\left(-\frac{\hbar^2}{2m_r} \nabla_r^2 - \frac{e^2}{4\pi\epsilon_r\epsilon_0 r} \right) \Psi_n(\mathbf{r}) = E_n \Psi_n(\mathbf{r}), \quad (4.53)$$

аналогичному уравнению (1.38) для атома водорода. Поэтому внутренняя энергия экситона E_n будет определяться аналогично (1.39):

$$E_n = -\frac{m_r e^4}{8(\epsilon_r \epsilon_0 \hbar)^2 n^2} = -\frac{R_{ex}}{n^2}, \quad (4.54)$$

где R_{ex} — энергия ионизации экситона (экситонный ридберг), n — главное квантовое число ($n = 1, 2, 3, \dots$). Отметим, что уравнение (4.53) записано для изотропных параболических зон и справедливо для экситонов большого радиуса. Для экситонов малого радиуса, называемых экситонами Френнеля, приближение эффективной массы и уравнение (4.53) несправедливы.

Из сравнения (4.54) с (1.39) получаем, что экситонный ридберг связан с постоянной Ридберга R_∞ соотношением

$$R_{ex} = \frac{(m_r/m)}{e^2} R_\infty. \quad (4.55)$$

Это означает, что для полупроводника с $m_p \approx 0,1m$ и $\epsilon_r = 10$ энергия ионизации экситона в 1000 раз меньше энергии ионизации атома водорода и составляет ~ 13 мэВ.

Из уравнения (4.53) видно, что экситонные волновые функции $\Psi_n(\mathbf{r})$ аналогичны атомным волновым функциям для атома водорода, но вместо боровского радиуса $a_B = \frac{\hbar^2 \epsilon_0}{\pi m e^2} \approx 0,053$ им необходимо ввести экситонный радиус

$$a_{ex} = \frac{\hbar^2 \epsilon_r \epsilon_0}{\pi m_r e^2} = \epsilon_r \left(\frac{m}{m_r} \right) a_B. \quad (4.56)$$

Здесь картина по существу аналогична рассмотренной в § 4.1 для водородоподобных примесей [ср. выражения (4.13) и (4.54); (4.14) и (4.56)], за исключением того, что экситон как квазичастица может свободно перемещаться по кристаллу.

Полная энергия экситона E состоит из его кинетической энер-

гн $\frac{\hbar^2 K^2}{2M}$, определяемой движением центра масс, и внутренней энергии ($E_g + E_n$), поэтому можно записать:

$$E(K) = \frac{\hbar^2 K^2}{2M} + E_g + E_n, \quad (4.57)$$

где K — волновой вектор экситона; $M = m_n + m_p$, E_g — ширина запрещенной зоны, а спектр E_n определяется соотношением (4.54). Из (4.57) следует, что энергетический спектр экситонов содержит серию дискретных параболических зон ниже E_g , которые сливаются в континуум при более высоких энергиях, как показано на рис. 4.21, а. Каждая из зависимостей $E(K)$ с заданным n образует экситонную зону. Точка 0 есть энергия невозбужденного кристалла. Минимальная энергия $E_{gx} \approx E_g - E_{ex}$, необходимая для создания экситона, называется экситонной шириной запрещенной зоны.

Экситонные эффекты и экситонные состояния не могут быть изображены на обычных энергетических диаграммах типа рис. 4.14 и 4.17, поскольку эти диаграммы справедливы в одноэлектронном приближении, а экситон состоит как минимум из двух частиц. Поэтому рис. 4.21, а существенно отличается от рис. 4.14. В частности, по оси абсцисс на рис. 4.21, а отложен волновой вектор экситона K , а не электрона или дырки в отдельности. Общим для рис. 4.21, а и 4.14 является начало отсчета энергий, за которое принято состояние невозбужденного кристалла.

Непосредственный оптический переход в экситонное состоя-

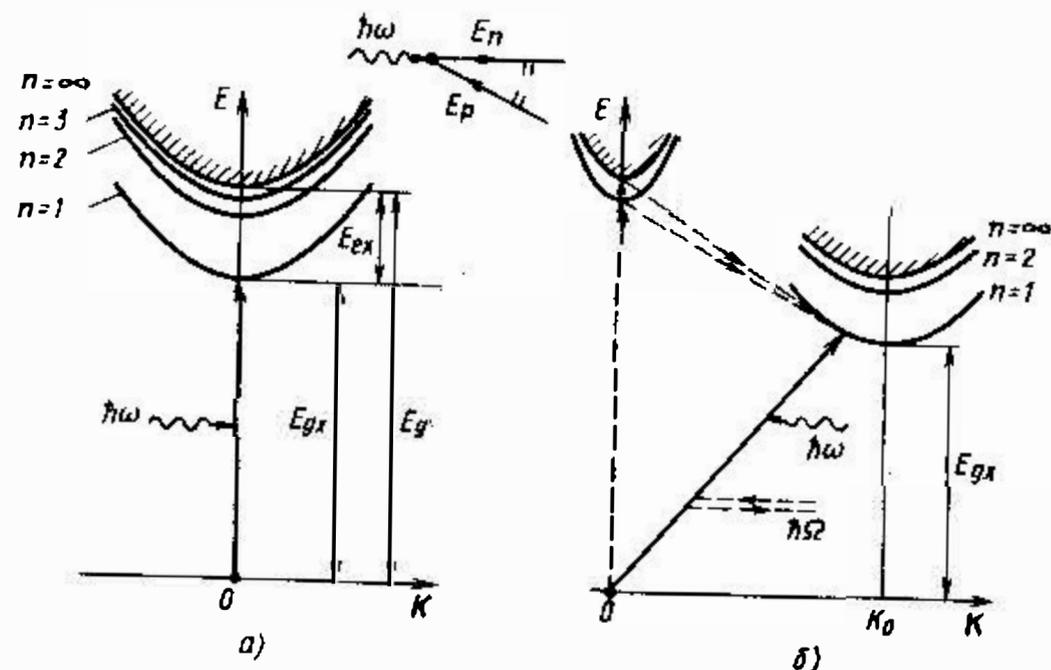


Рис. 4.21. Прямые (а) и непрямые (б) оптические переходы в экситонные состояния в кристаллах

ние при поглощении фотона $\hbar\omega$ будет возможен при соблюдении законов сохранения энергии и квазиимпульса:

$$\hbar\omega = \frac{\hbar^2 K^2}{2M} + E_g + E_n;$$

$$\hbar k_{\text{фот}} = \hbar K.$$

Поскольку $\hbar k_{\text{фот}} = 0$ и начальному состоянию отвечает точка $K=0$, непосредственные оптические переходы возможны только в состоянии $K=0$. Поэтому спектр собственного поглощения кристалла с учетом экситонных эффектов должен состоять из серии линий при

$$\hbar\omega = E_g - (E_{ex}/n^2), \quad (4.58)$$

которые переходят в непрерывный спектр при $\hbar\omega > E_g$. Как и в атоме водорода, интенсивность экситонных линий поглощения уменьшается с увеличением квантового числа n пропорционально $(1/n)^3$. В то же время при больших n число экситонных линий, приходящихся на единичный интервал энергий, возрастет пропорционально кубу квантового числа n .

Каждая из экситонных линий уширена вследствие конечности времени жизни экситона и наличия дефектов. Поэтому в реальных условиях в спектре поглощения обычно проявляется одна экситонная линия с $n=1$, а линии с $n \geq 2$ накладываются друг на друга, переходя в сплошной спектр (4.12), как это показано на рис. 4.22 для арсенида галлия. Заметим, что экситонные эффекты влияют на величину и форму края собственного поглощения не только при низких, но и при комнатной температуре.

Экситонные эффекты влияют также на форму края поглощения в полупроводниках с непрямой структурой зон. Диаграмма, иллюстрирующая экситонное поглощение при не прямых переходах, представлена на

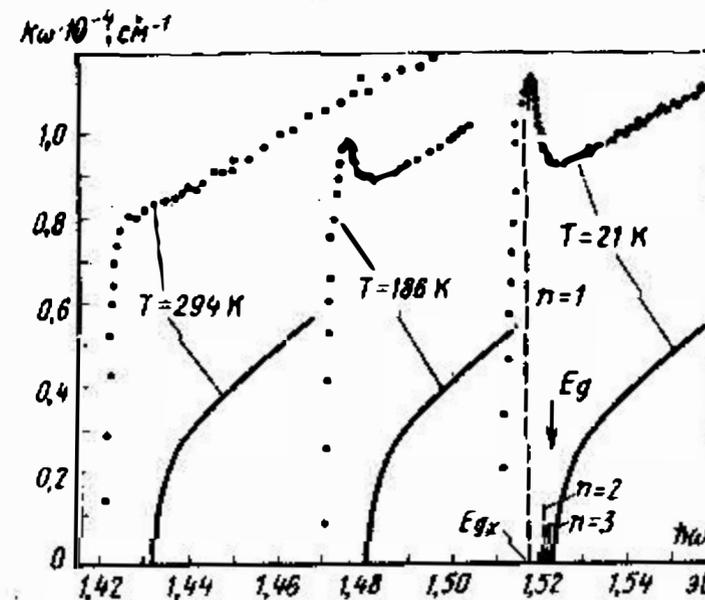


Рис. 4.22. Край собственного поглощения арсенида галлия при различных температурах. Точками обозначены экспериментальные данные, сплошными линиями — расчетные по соотношениям (4.42), (4.43) и (4.44). Штриховые линии показывают положение и относительную интенсивность линий экситонного поглощения

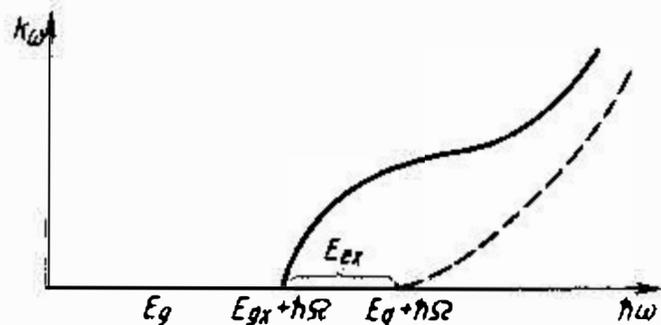


Рис. 4.23. Спектр собственного поглощения при непрямы переходов с учетом (сплошная кривая) и без учета (штриховая кривая) экситонных эффектов

бое состояние вблизи неё. Поэтому спектральная зависимость показателя поглощения от каждого экситонного состояния с заданным n будет определяться изменением плотности состояний в экситонной зоне, которая изменяется аналогично плотности состояний (4.9), если вместо эффективной массы m_n подставить массу $M = m_n + m_p$:

$$\rho_{ex}(E) = \frac{1}{2\pi^2} \left(\frac{2M}{\hbar^2} \right)^{3/2} E^{1/2} \quad (4.59)$$

Влияние экситонного эффекта на спектр поглощения при непрямом переходе с испусканием фонона $\hbar\Omega$ показано на рис. 4.23.

В полупроводниковых твердых растворах с изменением их состава край собственного поглощения будет сдвигаться по энергии в соответствии с изменением ширины запрещенной зоны. В качестве примера на рис. 4.24 приведены экспериментальные данные по краю собственного поглощения кристаллов $GaAs_{1-x}P_x$, изменение энергетических зазоров в которых приведено на рис. 4.9. В области составов $x \leq x_c = 0,46$ край поглощения определяется прямыми, а в области $x > x_c$ — непрямыми переходами в экситонные состояния. Из рис. 4.24 видно, что форма края собственного поглощения в твердых растворах с прямой структурой зон аналогична таковой для эквивалентного кристалла. Определяющую роль в формировании края поглощения вносят экситонные эффекты. Уширение (4.25) экситонных линий, вызванное наличием флуктуирующей части $V_1(r)$ кристаллического потенциала (4.19), невелико. Вообще фундаментальные оптические свойства полупроводниковых твердых растворов, определяемые прямыми вертикальными электронными переходами, довольно хорошо описываются в приближении виртуального кристалла (см. § 4.2). Непериодическая добавка $V_1(r)$ в (4.19) не оказывает существенного влияния на эти переходы. Это справедливо для всех изовалентных твердых растворов замещения.

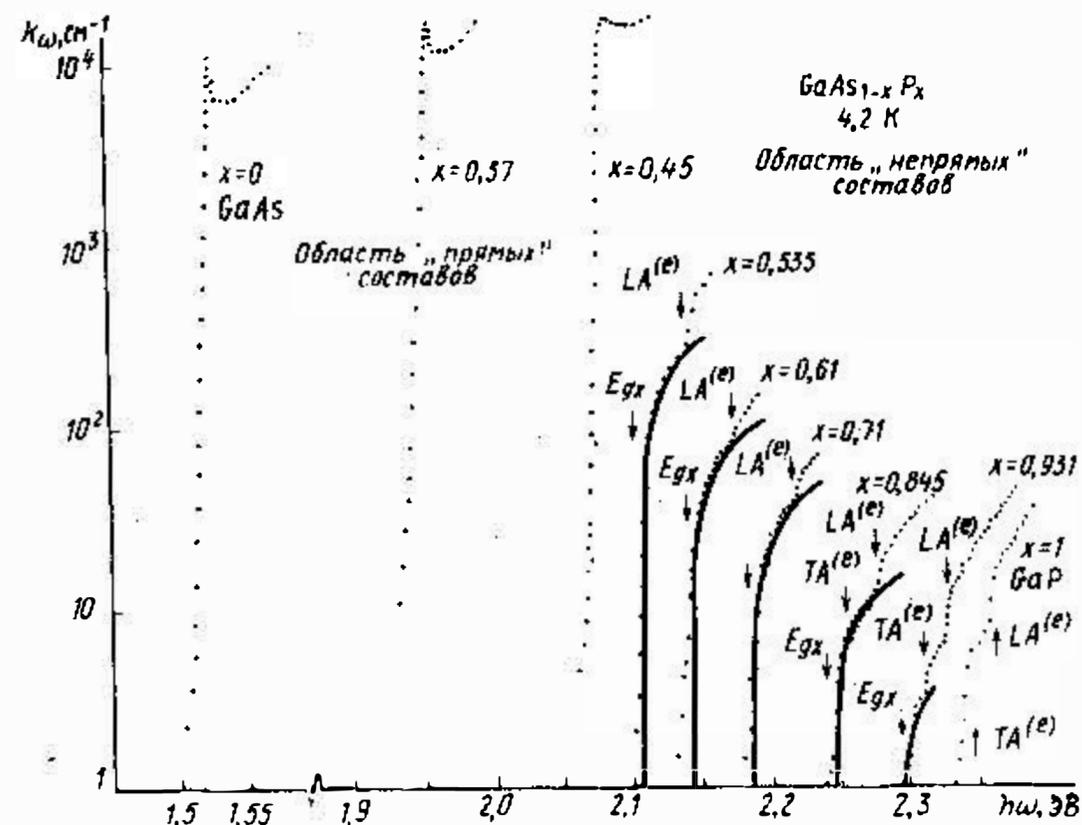


Рис. 4.24. Край собственного поглощения твердых растворов $GaAs_{1-x}P_x$ при низкой температуре для различных составов «х». Область составов с $x < 0,46$ соответствует прямой структуре энергетических зон, и край поглощения определяется прямыми переходами в экситонные состояния. Область составов с $x > 0,46$ соответствует непрямой структуре энергетических зон, и край поглощения определяется непрямыми переходами в экситонные состояния. Безфонная компонента обозначена штриховыми линиями

На оптические свойства, определяемые непрямыми переходами, непериодическая добавка $V_1(r)$ в (4.19), наоборот, может оказывать сильное влияние. Прежде всего нарушение трансляционной симметрии гамильтониана в твердых растворах может вызвать «нарушение» закона сохранения волнового вектора (4.32) при оптических переходах (напомним, что соотношения (4.30) и (4.32) вытекают из условия идеальной трансляционной симметрии потенциальной энергии кристалла). Это приводит к появлению в твердых растворах так называемых *непрямых бесфононных переходов*, т. е. оптических переходов без участия фононов между состояниями с различными значениями волнового вектора.

Непрямой бесфононный оптический переход рассматривается как переход через виртуальное состояние, в качестве которого может выступать наименьшее реальное состояние при $\mathbf{k} = 0$, а рассеяние осуществляется на флуктуирующем потенциале, вызванном статистически неупорядоченным распределением компонентов твердого раствора по узлам соответствующей подрешетки. При этом вместо концентрации рассеивающих центров необходи-

мо ввести коэффициент, учитывающий степень разупорядочения твердого раствора A_xB_{1-x} . В отсутствие упорядочения этот коэффициент равен $x(1-x)$.

Возможность не прямых переходов без участия фононов в твердых растворах обуславливает появление бесфононной компоненты в спектре. Её пороговая энергия равна не прямой ширине запрещенной зоны E_g для переходов зона — зона и экситонной ширине запрещенной зоны $E_{gx} = E_g - E_{ex}$ для переходов в экситонные состояния. Показатель поглощения для таких переходов пропорционален $x(1-x)$ и зависит как от параметров зонной структуры материала, так и от различия свойств замещающего и замещаемого атомов в твердых растворах замещения.

В силу указанных причин спектр собственного поглощения твердого раствора $GaAs_{1-x}P_x$ в области «непрямых» составов $0,46 < x \leq 1$ обладает следующими особенностями. В «чистом» GaP наблюдаются переходы в экситонные состояния с испусканием поперечных акустических (ТА) и продольных акустических (ЛА) фононов (см. рис. 4.24). Переходы с поглощением фононов не проявляются, так как температура мала. Бесфононная компонента в GaP отсутствует. По мере образования твердого раствора $GaAs_{1-x}P_x$ и увеличения $(1-x)$ появляется и возрастает по интенсивности бесфононная компонента с пороговой энергией, в точности равной экситонной ширине запрещенной зоны E_{gx} (на рис. 4.24 эта компонента обозначена сплошной линией). Быстрый рост $k_{\omega}^{6\phi}$ при увеличении содержания арсенида галлия в $GaAs_{1-x}P_x$ обусловлен двумя причинами: 1) увеличивается коэффициент разупорядочения $x(1-x)$; 2) уменьшается энергетический зазор $\Delta \approx E_{\Gamma_1^c} - E_{X_1^v}$ (см. рис. 4.9).

Кроме увеличения интенсивности бесфононной компоненты, как видно из рис. 4.24, при увеличении $(1-x)$ происходит быстрое (резонансное) увеличение интенсивности ЛА-компоненты, что также определяется уменьшением энергетического зазора Δ между виртуальным Γ_1^c и конечным реальным X_1^v состояниями в $GaAs_{1-x}P_x$.

Эффекты, связанные с увеличением вероятности не прямых оптических переходов вследствие уменьшения энергетического зазора между виртуальным и реальным состояниями в полупроводниках, называются *эффектами зонной структуры*. Они могут быть использованы для увеличения эффективности люминесценции в полупроводниках с не прямой структурой энергетических зон.

Отметим, что для ТА-компоненты в спектре собственного поглощения $GaAs_{1-x}P_x$ эффект зонной структуры не проявляется, поскольку оптические переходы через наименьшее Γ_1^c -состояние в X_1^v -минимум зоны проводимости с участием поперечных аку-

стических (ТА) фононов запрещены правилами отбора по симметрии.

В неупорядоченных системах типа стекол или аморфного кремния край поглощения «размыт» по энергии в связи с «размытостью» функции плотности состояний (рис. 4.13) так, что определить границу локализованных и делокализованных состояний по спектру поглощения здесь бывает затруднительно.

Решеточное поглощение. Оптическое поглощение на колебаниях решетки происходит в результате взаимодействия электромагнитного поля световой волны с движущимися зарядами узлов решетки. Энергия фотона $\hbar\omega$ передается решетке, увеличивая её колебания. Другими словами, при решеточном поглощении уничтожение фотона приводит к рождению фонона. Энергия электронов при этом в *adiaбатическом* приближении остается неизменной.

Законы сохранения энергии и волнового вектора при взаимодействии фотона с N фононами записываются в виде

$$\hbar\omega = \sum_{i=1}^N \pm \hbar\Omega_i(\mathbf{k}_{i\text{фон}}), \quad (4.60)$$

$$0 \approx \mathbf{k}_{\text{фот}} = \sum_{i=1}^N \pm \mathbf{k}_{i\text{фон}}. \quad (4.61)$$

Знак плюс соответствует испусканию, а минус — поглощению i -го фонона. При взаимодействии с одним фононом (однофононное поглощение) имеем

$$\hbar\omega = \hbar\Omega(0), \quad (4.62)$$

$$\mathbf{k}_{\text{фот}} = \mathbf{k}_{\text{фон}} \approx 0.$$

Это означает, что в процессе однофононного поглощения могут участвовать только оптические фононы. Точнее, фотон может поглощаться при возбуждении лишь поперечных оптических колебаний решетки (ГО-фононов), поскольку они создают в полярных кристаллах электрический дипольный момент, взаимодействующий с полем световой волны. По этой причине спектры однофононного решеточного поглощения и отражения в полярных кристаллах очень схожи с соответствующими спектрами гармонического осциллятора с собственной частотой $\omega_0 = \Omega_0$ и затуханием γ (ср. рис. 1.22 и 4.25). Показатель поглощения в максимуме может достигать очень больших значений ($k_{\omega} > 10^3 \text{ см}^{-1}$), чему соответствует коэффициент отражения, близкий к 100%. Эта узкая область спектра в инфракрасном диапазоне, где наблюдается интенсивное поглощение и резкий максимум коэффициента отражения, называется *областью ста-*

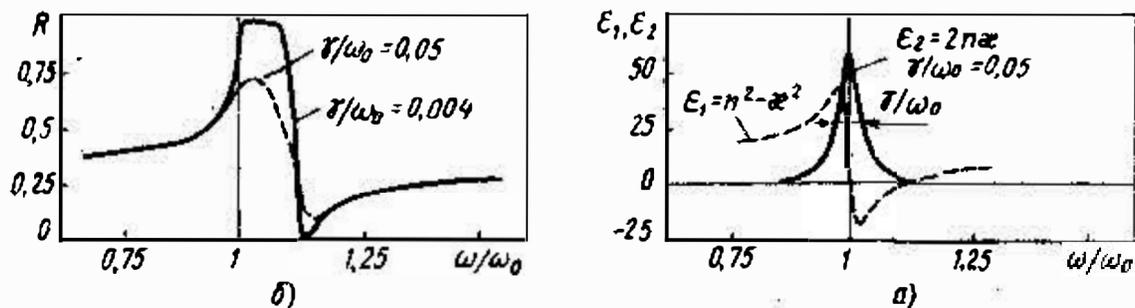


Рис. 4.25. Действительная ϵ_1 и мнимая ϵ_2 части комплексной диэлектрической проницаемости как функции частоты для модели одного классического осциллятора (а) и спектральная зависимость коэффициента отражения (б). ω_0 — собственная (резонансная) частота осциллятора; γ — параметр затухания

точных лучей. В кристаллах с ионной связью коэффициент отражения R в этой области близок к 100%. В кристаллах с чистой ковалентной связью (например, в Si и Ge) однофононное поглощение вообще отсутствует, поскольку колебания атомов не приводят к возникновению электрического дипольного момента.

Кроме однофононного возможно двухфононное поглощение, когда энергия фотона передается решетке, возбуждая одновременно два фонона $\hbar\Omega_1$ и $\hbar\Omega_2$ так, чтобы выполнялись правила отбора:

$$\hbar\omega = \hbar\Omega_1 + \hbar\Omega_2, \quad (4.63)$$

$$k_{\text{фон}1} = k_{\text{фон}2}.$$

Учет законов сохранения (4.63) совместно с законами дисперсии фононов приведет к тому, что спектр двухфононного решеточного поглощения будет представлять собой комбинацию из сравнительно узких линий поглощения в инфракрасной области спектра. Вероятность двухфононных процессов существенно меньше, чем однофононных, поэтому показатель двухфононного поглощения сравнительно невелик ($5-500 \text{ см}^{-1}$).

Процессы решеточного поглощения с участием комбинаций из трех и более фононов имеют еще меньшую вероятность и проявляются крайне слабо.

Примесное поглощение. Процессы оптического поглощения с участием примесей весьма разнообразны. Они могут быть обусловлены переходами носителей заряда из локализованного примесного состояния в одну из разрешенных зон либо в другое локализованное состояние. В первом случае процессы поглощения сопровождаются генерацией свободных носителей заряда одного знака, и спектр поглощения имеет вид сравнительно широкой полосы. Во втором случае свободные носители заряда не возникают, и спектр поглощения имеет вид узких линий. Возможны также оптическое поглощение с образованием экситонов, связанных на примесных центрах (связанных экситонов),

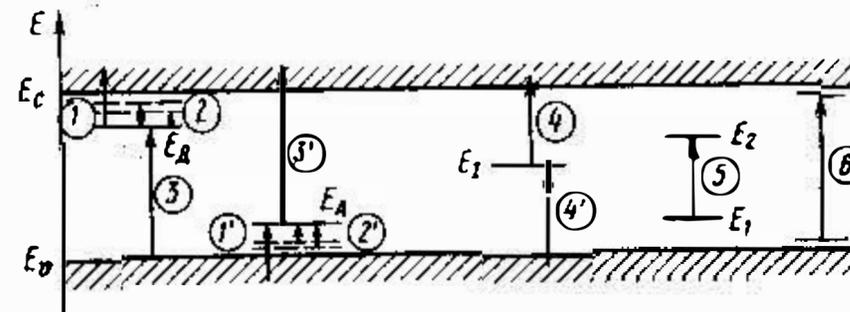


Рис. 4.26. Электронные переходы, ответственные за примесное поглощение в кристаллах: 1 и 1' — фотоионизация; 2 и 2' — фотовозбуждение; 3 и 3' — фотонейтрализация мелких водородоподобных примесей; 4 и 4' — фотоионизация и фотонейтрализация глубокого центра; 5 — внутривещные переходы; 6 — фотогенерация связанного экситона

и поглощение, обусловленное возбуждением локализованных колебаний примесных атомов.

Таким образом, можно выделить следующие процессы, приводящие к примесному поглощению в твердых телах: 1) фотоионизацию примесей; 2) фотовозбуждение примесей; 3) фотонейтрализацию примесей; 4) фотогенерацию связанных экситонов; 5) возбуждение локализованных колебаний примесных атомов. Электронные переходы, обуславливающие примесное поглощение, соответствующее пунктам (1 — 4), схематически изображены на энергетической диаграмме рис. 4.26. Рассмотрим сущность каждого из этих процессов.

1. **Фотоионизация мелких примесей.** Этот процесс обусловлен переходом электрона (дырки) из связанного состояния в свободное при поглощении фотона $\hbar\omega$ (переходы 1 и 1' на рис. 4.26). Для осуществления такого перехода энергия фотона должна быть больше энергии связи электрона (дырки) с донором (акцептором), т. е. больше E_D или E_A . До взаимодействия с фотоном носитель заряда был локализован на примесном центре и примесный атом был нейтральным. Процесс взаимодействия с фотоном сопровождается переходом носителя заряда в свободную зону, т. е. ионизацией примеси. Спектр поглощения для такого процесса будет определяться как электронной структурой примесного состояния, так и структурой свободной зоны, в которую осуществляется переход.

Мелкие примесные состояния, как указывалось в § 4.1, хорошо описываются в приближении эффективной массы с помощью водородоподобной модели. Из выражения (4.16) видно, что волновая функция электрона на водородоподобной примеси локализована в области k -пространства с характерными размерами порядка обратной величины боровского радиуса. На диаграмме $E(k)$ процесс фотоионизации мелкого водородоподобного доио-

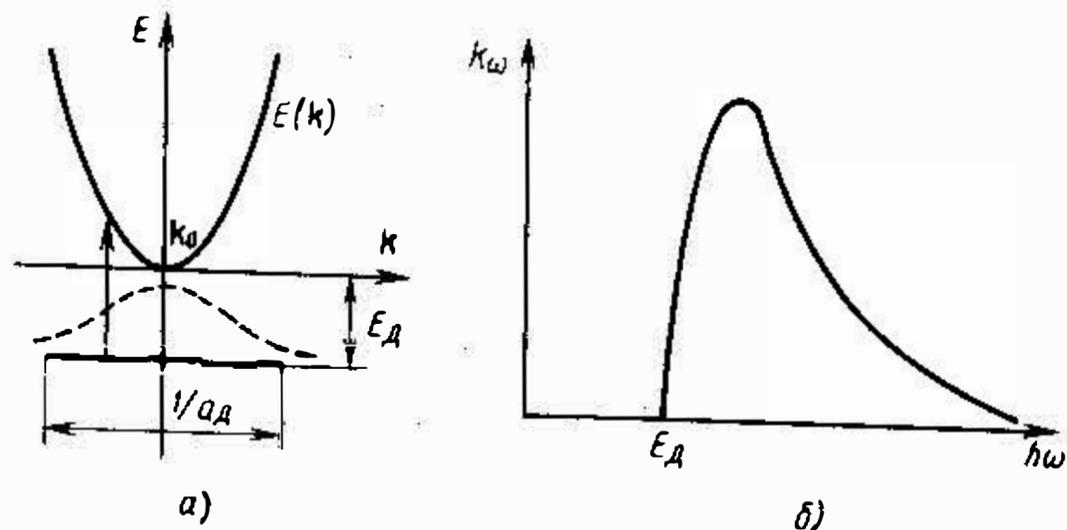


Рис. 4.27. Оптическое поглощение, обусловленное фотоионизацией мелкого водородоподобного донора:
 а) электронный переход на диаграмме $E(k)$; б) спектр поглощения

ра изображен так, как показано на рис. 4.27, а. Минимальная энергия, при которой возможна фотоионизация (пороговая энергия), равна E_d . При $h\omega = E_d$ переход будет осуществляться в точку $k = k_0$, соответствующую минимальной энергии свободного электрона. Так как плотность разрешенных состояний в зоне проводимости возрастает пропорционально корню квадратному из энергии [см. (4.9)], то при увеличении $h\omega$ будет наблюдаться резкий рост показателя поглощения (рис. 4.27, б). При $h\omega > E_d$ плотность конечных состояний в разрешенной зоне будет монотонно возрастать, а плотность начальных состояний — убывать в соответствии с формулой (4.16) (пунктирная кривая на рис. 4.27, а). Поэтому в спектре поглощения $k_\omega(h\omega)$ будет наблюдаться характерный максимум и спад, как видно из рис. 4.17, б. Для простых водородоподобных примесей спектр поглощения, вызванный процессами фотоионизации, определяется формулой

$$k_\omega = A \frac{(\hbar\omega/E_d - 1)^{3/2}}{(\hbar\omega/E_d)^5}, \quad (4.64)$$

где коэффициент пропорциональности A связан с концентрацией доноров N_d , энергией ионизации E_d и эффективной массой электронов в зоне m_n соотношением

$$A = \frac{128\pi e^2 \hbar}{3\pi m_n E_d} N_d.$$

Максимум функции (4.64), т. е. максимум спектра фотоионизации водородоподобной примеси, приходится на энергию $h\omega = \frac{10}{7} E_d$.

Аналогично происходят процессы фотоионизации мелких акцепторов.

2. Фотовозбуждение мелких примесей. Этот процесс обусловлен переходом электрона (дырки) из одного связанного состояния с $n=1$ (основного) в другое (возбужденное) связанное состояние с $n=2, 3, 4, \dots$ при поглощении фотона (переходы 2 и 2' на рис. 4.26). Поскольку такие переходы осуществяются между дискретными уровнями энергии, спектр фотовозбуждения будет состоять из серии дискретных узких линий поглощения при энергии фотонов, несколько меньшей энергии ионизации.

В качестве примера на рис. 4.28 представлены экспериментальные данные по спектру примесного поглощения в кристаллах кремния, легированного водородоподобным акцептором — бором ($E_A = 44,3$ мэВ). При $h\omega < E_A$ наблюдаемая серия узких дискретных линий обусловлена переходами в возбужденные состояния примеси, т. е. процессами фотовозбуждения дырок, а широкая полоса при $h\omega > E_A$ соответствует фотоионизации атомов бора.

Процессы фотовозбуждения и фотоионизации мелких примесей в полупроводниках будут проявляться лишь при низких температурах, когда $kT \ll E_d(E_A)$ и примеси термически не ионизованы. Поскольку энергии ионизации мелких примесей в полупроводниках составляют 5—100 мэВ, спектры их фотоионизации и фотовозбуждения расположены в инфракрасной области при $\lambda = 10 \dots 200$ мкм.

3. Фотонейтрализация мелких примесей. Этот процесс вызван переходами электронов из валентной зоны на уровни ионизованных доноров вблизи дна зоны проводимости или с уровней ионизованных акцепторов в зону проводимости (переходы 3 и 3' на рис. 4.26). Такие переходы могут наблюдаться лишь тогда, когда примесный центр ионизован, например, за счет одновременного легирования донорами и акцепторами (за счет компенсации). Процессам фотонейтрализации доноров и акцепторов в полупроводниках соответствует появление дополнительных полос поглощения при энергиях фотонов $h\omega \geq [E_s - E_d(E_A)]$. Так как для мелких примесей $E_d(E_A) \ll E_s$, то спектры фотонейтрализации обычно сливаются со спектрами собственного поглощения и проявляются лишь в виде дополнительной «ступеньки» на краю основной полосы поглощения. Если энергия ионизации примеси мала, а её концентрация такова, что образуется примесная зона,

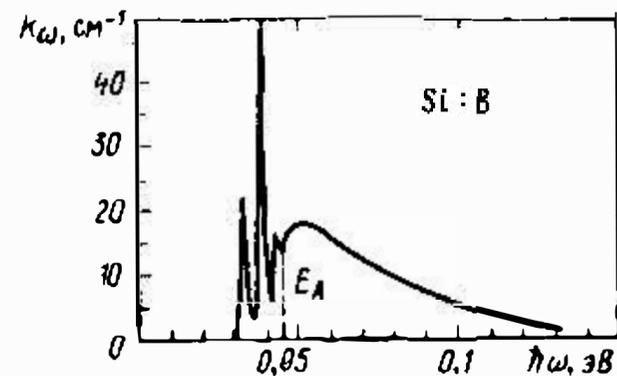


Рис. 4.28. Спектр примесного поглощения акцептора бора в кремнии при $T = 4,2$ К

сливающаяся с зоной разрешенных значений энергий, то полоса фотонейтрализации примеси сливается с собственным поглощением и неотделима от неё. Такие переходы в образующие «хвосты» плотности состояний (см. рис. 4.6) часто называют «квазиимежзонными» переходами.

4. **Фотоионизация и фотонейтрализация глубоких примесей**
 Так же как фотоионизация и фотонейтрализация мелких примесей, она определяется переходами типа примесь — зона или зона — примесь (переходы 4 и 4' на рис. 4.26). Но в отличие от мелких примесей возмущающий потенциал для глубоких центров локализован вблизи примеси. Боревский радиус для такого центра мал и волновая функция «размазана» в k -пространстве. Примесное поглощение, связанное с глубокими неводородоподобными примесями, проявляется обычно в виде широких бесструктурных полос, простирающихся от энергий фотонов, соответствующих энергии ионизации E_I^{opt} (или нейтрализации $E_c - E_I^{opt}$) примеси до края собственного поглощения. Форма спектра поглощения для процесса фотоионизации глубокого центра с энергией ионизации E_I^{opt} может быть аппроксимирована выражением

$$k_{\omega} = B \frac{(\hbar\omega/E_I^{opt} - 1)^{1/2}}{(\hbar\omega/E_I^{opt})^3}, \quad (4.65)$$

где E_I^{opt} — оптическая энергия ионизации центра, а коэффициент B пропорционален концентрации центров. Формула (4.65) получена для некулоновского (короткодействующего) потенциала, локализованного в пределах одной центральной ячейки.

Глубокие примеси из-за сильной пространственной локализации заряда обычно сильно взаимодействуют с колебаниями решетки и с локальными колебаниями центра. Поэтому при рассмотрении оптических переходов с участием таких примесей необходимо учитывать электрон-фононное взаимодействие, которое может заметно влиять на форму спектра поглощения, особенно в области длинноволнового края. На рис. 4.29 сплошной линией показан спектр поглощения, рассчитанный по формуле (4.65), т. е. без учета электрон-фононного взаимодействия. Пунктирные кривые соответствуют данным, полученным

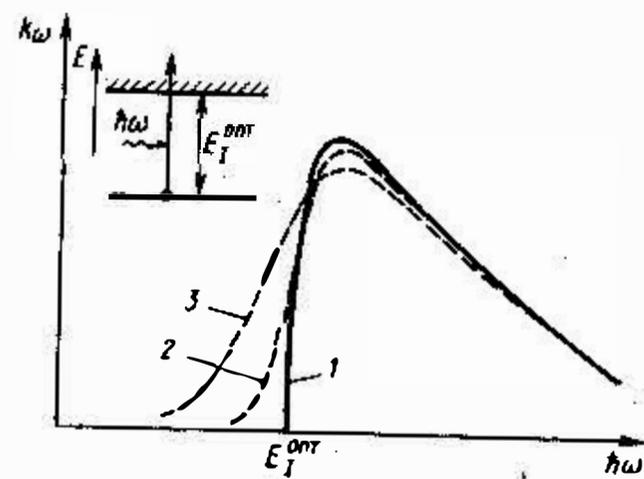


Рис. 4.29. Спектры примесного поглощения, обусловленные фотонейтрализацией глубокого центра без учета (кривая 1) и с учетом (кривые 2, 3) электрон-фононного взаимодействия.

E_I^{opt} — оптическая энергия ионизации центра

с учетом этого взаимодействия. Учет электрон-фононного взаимодействия приводит к тому, что вместо резкого длинноволнового края, описываемого соотношением (4.65) будет наблюдаться характерный экспоненциальный спад («хвост») показателя поглощения, который будет зависеть от температуры и от константы электрон-фононного взаимодействия.

Для глубоких центров, сильно взаимодействующих с колебаниями решетки, наблюдается различие в оптической и термической энергиях активации (см. § 4.6) Оптическая энергия ионизации центра определяется по энергии, соответствующей спаду показателя поглощения в длинноволновой области до уровня 0,5 от максимума.

5. **Внутрицентровые переходы.** Эти процессы определяют примесное поглощение в том случае, когда, поглощая фотон, примесный центр переходит из одного разрешенного состояния E_1 в другое состояние E_2 , при этом носитель заряда остается связанным с центром. Они аналогичны процессам фотовозбуждения мелких водородоподобных центров в полупроводниках, но осуществляются при переходах между глубокими состояниями и не образуют водородоподобную серию в спектре.

Типичным примером примесей, для которых наблюдается сильное оптическое поглощение, обусловленное внутрицентровыми переходами, являются элементы с незастроенными внутренними оболочками (3d-группа железа, 4f-лантаноиды и 5f-актиноиды). Эти оболочки (особенно 4f и 5f) довольно хорошо экранированы 4s-, 5s-, 5p- и 6s-электронами от внешних воздействий. Поэтому при помещении в различного рода кристаллические и аморфные матрицы такие атомы сохраняют в основном свойства свободных ионов. Электронные переходы между энергетическими состояниями внутренних 3d- и 4f-оболочек и будут являться внутрицентровыми переходами. Соответствующие линии поглощения могут быть уширены вследствие расщепления и уширения энергетических уровней примесного центра под воздействием внутрикристаллического поля решетки.

6. **Поглощение на связанных экситонах.** Этот процесс определяется генерацией связанного на дефекте экситона под действием фотона и схематически изображен переходом б на рис. 4.26.

При определенных условиях неравновесные электрон и дырка в кристаллах могут образовывать не только экситон, свободно перемещающийся по кристаллу и называемый свободным, но и локализованный на дефекте экситон, являющийся связанным. Энергию связи связанного экситона отсчитывают от энергии свободного экситона, т. е. от экситонной ширины запрещенной зоны. Для образования устойчивого состояния энергия связанного экситона должна быть меньше энергии свободного экситона. Экситон связывается, попадая в потенциальную яму дефекта. Так как связанный экситон локализован, то в спектре поглощения он

будет проявляться в виде очень узких линий, расположенных обычно вблизи запрещенной зоны.

В большинстве случаев связанные экситоны образуются на нейтральных центрах, хотя при определенных условиях их образование возможно и на заряженных дефектах. Особенно эффективно связанные экситоны образуются на так называемых *изоэлектронных ловушках*. Изоэлектронной примесью является любая примесь, изоэлектронно замещающая основной атом решетки. Например, в германии элементы IV группы (C, Si, Sn) будут являться изоэлектронными примесями замещения. Однако не любая изоэлектронная примесь способна связать экситон. Изоэлектронные примеси замещения образуют связанные состояния только тогда, когда примесный и замещаемый атом существенно различаются как по электроотрицательности, так и по ковалентным радиусам. При этом примесный центр образует потенциальную яму, в которую может быть захвачен электрон или дырка. После того как носитель одного знака будет локализован, изоэлектронный центр приобретает заряд (в отличие от обычных доноров и акцепторов) и затем довольно легко захватывает носитель противоположного знака — образуется экситон. Подробнее оптические свойства связанных экситонов будут рассмотрены в § 4.6.

7. *Поглощение на локальных колебаниях примесей.* Этот процесс определяется возбуждением колебаний примесного атома, связанного упругими силами с основными атомами решетки. Частота таких колебаний отличается от частот колебаний идеальной решетки, поэтому в спектре инфракрасного поглощения кристалла, содержащего примеси, могут появиться новые резонансные линии, соответствующие возбуждению локализованных колебаний. Типичным примером здесь является кислород в кремнии. Легирование полупроводникового кремния кислородом обуславливает проявление характерных пиков поглощения в области энергий 64 и 141 мэВ.

Поскольку возбуждение локализованных колебаний не приводит к изменению электронной энергии кристалла, на диаграмме рис. 4.26 соответствующие переходы изобразить невозможно.

Поглощение свободными носителями заряда. Если в кристалле имеются свободные носители заряда (электроны в зоне проводимости или дырки в валентной зоне), то их можно считать образующими электронный газ и поглощение электромагнитной энергии такими свободными носителями рассматривать, воспользовавшись в первом приближении классическими соотношениями для электронного газа. Согласно теории Друде, описывающей колебания электронов в металле под воздействием периодического электрического поля, проводимость σ на частоте

ω определяется временем релаксации $\tau(E)$ для электрона с энергией E :

$$\sigma = \frac{Ne^2}{m^*} \left\langle \frac{\tau(E)}{1 + \omega^2 \tau^2} \right\rangle, \quad (4.66)$$

где N — концентрация носителей заряда, а m^* — их эффективная масса. Усреднение в (4.66) проводится по энергии E .

Показатель поглощения k_ω связан с главным показателем поглощения χ и удельной электрической проводимостью σ соотношениями (1.168) и (1.177). Из этих соотношений находим

$$k_\omega = \frac{\sigma}{c\epsilon_0 n} = \frac{Ne^2}{m^* c\epsilon_0 n} \left\langle \frac{\tau}{1 + \omega^2 \tau^2} \right\rangle. \quad (4.67)$$

При $\langle \tau \rangle \omega \gg 1$ имеем

$$k_\omega = \frac{Ne^2}{m^* c\epsilon_0 n \omega^2} \left\langle \frac{1}{\tau} \right\rangle. \quad (4.67a)$$

Если $\langle 1/\tau \rangle$ не зависит от энергии E носителей заряда в зоне, то из (4.67a) получаем $k_\omega \sim \lambda^2$, т. е. показатель поглощения на свободных носителях заряда монотонно возрастает при увеличении длины волны.

Более точное рассмотрение вопроса о поглощении света свободными носителями заряда в кристаллах может быть проведено на основе теории квантовых переходов, аналогично тому, как это было сделано для собственного поглощения. Для поглощения фотона $\hbar\omega$ свободный носитель заряда должен перейти из одного разрешенного состояния j в другое состояние f . В пределах одной долины такой переход может быть только непрямым, как показано на рис. 4.30. Это означает, что для его осуществления необходимо рассеяние квазимпульса, т. е. участие в процессе поглощения третьей частицы, роль которой могут выполнять фононы или примеси. Совершенно свободные носители заряда в идеальной кристаллической решетке не могли бы поглощать электромагнитную энергию. По этой причине показатель поглощения на свободных носителях заряда зависит от механизма рассеяния и определяется тем, происходит ли рассеяние на акустических, оптических колебаниях решетки (фононах) или примесях.

В общем случае спектральная зависимость показателя поглощения на свободных носителях заряда может быть аппроксимирована степенной функцией

$$k_\omega \sim N\lambda^r, \quad (4.68)$$

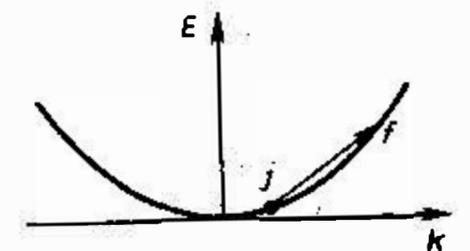


Рис. 4.30. Электронный переход, ответственный за оптическое поглощение свободными носителями заряда

§ 4.5. РЕФРАКЦИЯ СВЕТА В ПОЛУПРОВОДНИКАХ

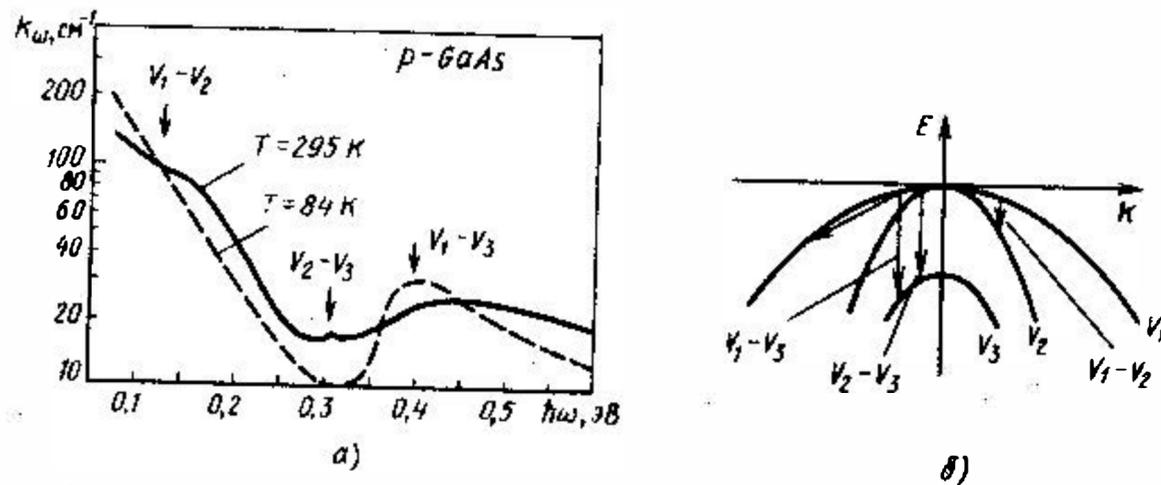


Рис. 4.31. Оптическое поглощение свободными дырками в p -GaAs: а) спектры поглощения при $T=295$ К (сплошная линия) и $T=84$ К (пунктирная линия) образца арсенида галлия, легированного цинком до $p=2,7 \cdot 10^{17}$ см $^{-3}$; б) переходы, ответственные за поглощение

где N — концентрация свободных носителей заряда, а λ — длина волны электромагнитного излучения. Показатель степени r зависит от механизма рассеяния и в первом приближении при рассеянии на акустических фононах $r \approx 1,5$, при рассеянии на оптических фононах $r \approx 2$ и при рассеянии на ионизованных примесях $r \approx 3,5$.

Как указывалось в § 4.1, каждая из энергетических зон в полупроводниках (зона проводимости и валентная зона) имеет сложную структуру и состоит из нескольких подзон. Под действием фотонов свободные носители заряда могут переходить из одной подзоны в другую (при выполнении законов сохранения энергии и волнового вектора). В зависимости от относительного расположения этих подзон на диаграмме $E(k)$ такие переходы могут быть как прямыми, так и непрямыми. Электронные переходы между подзонами по существу аналогичны межзонным переходам из валентной зоны в зону проводимости, за исключением того, что степень заселенности начального состояния, а следовательно, и показатель поглощения будут зависеть от уровня легирования материала и температуры. Оптическое поглощение, определяемое переходами свободных носителей заряда между подзонами одной зоны, будет накладываться на монотонный спектр поглощения (4.68), определяемый переходами носителей заряда в пределах одной подзоны.

В качестве примера на рис. 4.31, а изображен спектр поглощения кристаллов арсенида галлия p -типа, содержащих $2,7 \cdot 10^{17}$ см $^{-3}$ свободных дырок. Стрелками указана интерпретация соответствующих особенностей в спектре. На диаграмме $E(k)$ (рис. 4.31, б) показаны переходы, ответственные за поглощение. В отличие от аналогичных диаграмм рис. 4.27 и 4.30 здесь изображены не электронные, а дырочные переходы.

Современные методы полупроводниковой технологии позволяют создавать эпитаксиальные слои и структуры с заданным профилем показателя преломления. Вне конкуренции здесь стоят твердые растворы, где путем изменения их состава по координате можно создавать заданное распределение показателя преломления и тем самым управлять распространением световых пучков в материале (см. § 3.3 и 3.4). Часто это имеет решающее значение для создания различного рода оптоэлектронных и интегрально-оптических элементов и устройств.

Определим спектральную зависимость показателя преломления света в полупроводниках. Связь между вещественной и мнимой частями ϵ^* и n^* устанавливается фундаментальными соотношениями Крамерса — Кронига (1.178), (1.179), которые служат основой для описания спектральной зависимости показателя преломления $n(\omega)$ или $n(\lambda)$. В газах, где наблюдаются узкие линии поглощения, интегрирование (1.179) приводит к известной формуле Зельмейера (1.181в). В твердых телах наличие зон разрешенных значений энергий дает широкие области собственного поглощения. Современные методы расчета зонной структуры полупроводников позволяют определить вид спектра $\epsilon_2(\hbar\omega)$ для конкретного материала и по нему с помощью (1.179) найти показатель преломления и его дисперсию $dn/d\lambda$. Такая процедура требует громоздких вычислений и не позволяет представить $n(\hbar\omega)$ в виде простых аналитических выражений. Хотя такие расчеты «из первых принципов» представляют значительный интерес, их точность весьма ограничена. Поэтому в подавляющем большинстве случаев для нахождения $n(\lambda)$ и $n(\hbar\omega)$ используют полуэмпирические методы.

Как указывалось в § 4.4, общий вид спектра $\epsilon_2(\omega)$ для большинства полупроводниковых материалов подобен изображенному на рис. 4.19. Этот реальный спектр удобно аппроксимировать двумя эквивалентными осцилляторами с собственными энергиями, отвечающими положениям наиболее важных максимумов ϵ_2 при E_1 и E_2 , а вклад в дисперсию $n(\hbar\omega)$ краевых оптических переходов вблизи $\hbar\omega = E_g^r$ учесть в виде равномерного распределения осцилляторов с $\epsilon_2 = A = \text{const}$ от $\hbar\omega = E_0 = E_g^r$ до $\hbar\omega = E_1$, как это показано на рис. 4.32. Подставляя $\epsilon_2(\hbar\omega)$ в виде, изображенном на

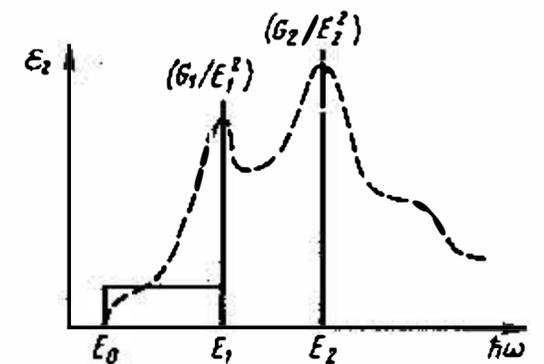


Рис. 4.32. Модельные представления $\epsilon_2(\hbar\omega)$ для расчета спектральной зависимости показателя преломления полупроводников

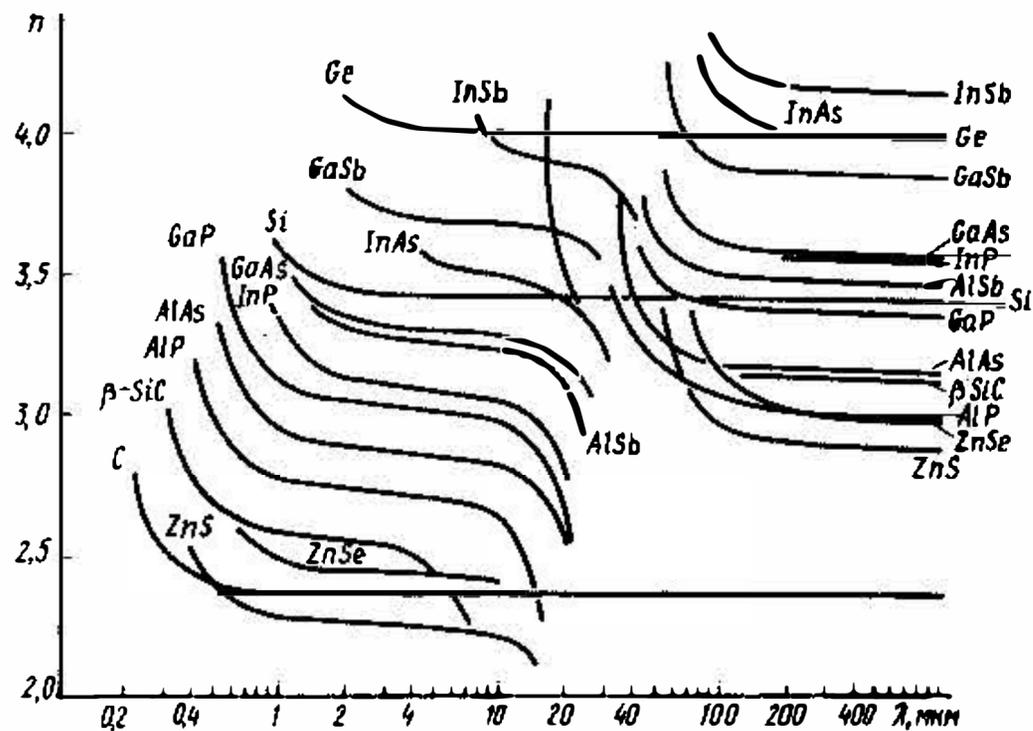


Рис. 4.33. Спектральные зависимости показателя преломления света полупроводников при $T=300$ К

рис. 4.32, в соотношение (1.179), получаем дисперсионное соотношение для $n(\hbar\omega)$:

$$n^2(\hbar\omega) = 1 + \frac{A}{\pi} \ln \frac{E_1^2 - (\hbar\omega)^2}{E_0^2 - (\hbar\omega)^2} + \frac{G_1}{E_1^2 - (\hbar\omega)^2} + \frac{G_2}{E_2^2 - (\hbar\omega)^2} \quad (4.69)$$

Входящие в это выражение значения параметров для ряда полупроводников приведены в табл. 4.1.

В выражении (4.69) учтены только межзонные электронные переходы, т. е. электронная поляризуемость вещества в поле световой волны. При низких частотах в ИК-диапазоне в полупроводниках, обладающих долей ионной связи, необходимо учитывать вклад поляризуемости решетки, т. е. вклад решеточного поглощения. Поскольку однофононное поглощение, как указывалось в § 4.4, хорошо описывается одноосцилляторным приближением при $E_{\text{то}} = \hbar\Omega_{\text{то}}$, то его вклад в показатель преломления легко учитывается добавлением в правую часть выражения (4.69) еще одного слагаемого

$$\frac{G_{\text{то}}}{E_{\text{то}}^2 - (\hbar\omega)^2} \quad (4.70)$$

Параметр $G_{\text{то}}$ пропорционален силе эквивалентного осциллятора при частоте поперечного оптического фонона $\Omega_{\text{то}}$. Экстраполяция (4.69) к $\hbar\omega \rightarrow 0$ без учета (4.70) дает высокочастотную (оптическую) диэлектрическую проницаемость материала ϵ_{∞} , а с учетом

(4.70) — низкочастотную (статическую) диэлектрическую проницаемость ϵ_0 . Дисперсионный параметр $G_{\text{то}}$ определяет разницу этих величин:

$$\Delta\epsilon = \epsilon_0 - \epsilon_{\infty} = \frac{G_{\text{то}}}{(\hbar\Omega_{\text{то}})^2} \quad (4.71)$$

Значения этих величин приведены в табл. 4.1, а на рис. 4.33 представлены спектральные зависимости показателя преломления ряда полупроводниковых материалов.

Используя рассмотренные выше приближения, нетрудно определить спектральные зависимости $n(\hbar\omega)$ в твердых растворах. Поскольку для прямых разрешенных переходов хорошо выполняется приближение виртуального кристалла (§ 4.2), то дисперсионное выражение (4.69) остается справедливым и для твердых растворов замещения. Зависимости характеристических энергий E_0 , E_1 и E_2 от состава твердого раствора x определяются соотношениями (4.20), (4.22). Для зависимостей $A(x)$, $G_1(x)$ и $G_2(x)$ справедливы линейные приближения. На рис. 4.34 представлены спектральные зависимости показателя преломления для широко применяемых в оптоэлектронике твердых растворов $\text{Al}_x\text{Ga}_{1-x}\text{As}$ и $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$.

Соотношение (4.69) позволяет найти температурный коэффициент dn/dT , который в полупроводниках определяется в основном температурным смещением E_0 , E_1 и E_2 . Для арсенида галлия эта величина составляет $dn/dT \approx 4,5 \cdot 10^{-5} \text{ К}^{-1}$.

Из (4.69) следует, что в полупроводниковых материалах, принадлежащих одному классу (например, $\text{A}^{\text{III}}\text{B}^{\text{V}}$; $\text{A}^{\text{IV}}\text{B}^{\text{VI}}$ и т. п.), уменьшение E_0 , E_1 и E_2 приводит к увеличению n . Поэтому показатель преломления, как правило, увеличивается с уменьшением ширины запрещенной зоны полупроводника, в чем нетрудно убедиться из рис. 4.33 и 4.34. Это обстоятельство оказывается весьма существенным для работы полупроводниковых лазеров на гетероструктурах (гл. 9).

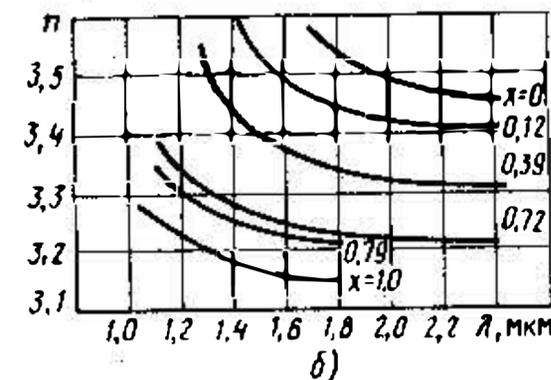
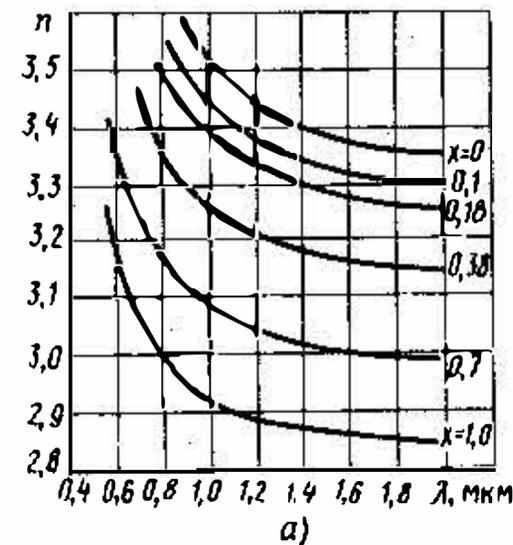


Рис. 4.34. Спектральные зависимости показателя преломления света в твердых растворах $\text{Al}_x\text{Ga}_{1-x}\text{As}$ (а) и $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$ (б) при $T=300$ К

Численные значения параметров в осцилляторных моделях
[соотношения (4.99) — (4.101)] для расчета дисперсии $n(\lambda\omega)$ в полупроводниках при $T=293$ К

Материал	Период решетки, $\mu\text{м}$	E_0 , эВ	E_{11} , эВ	E_{12} , эВ	E_{21} , эВ	E_{22} , эВ	τ_{10} , нэВ	A	G_1 , эВ ²	G_2 , эВ ²	$G_{10} 10^3$, эВ ²	ϵ_{∞}	ϵ_0
Ge	0,56575	0,805	2,1	4,4	—	—	—	0,780	22,2887	183,331	—	—	16,00
Si	0,54307	—	3,4	4,3	—	—	—	—	23,7537	159,275	—	—	11,67
C (алмаз)	0,35668	—	7,4	12,6	—	—	—	—	36,9961	630,929	—	—	5,64
β -SiC	0,4359	—	6,0	9,7	98,68	—	—	—	69,840	342,74	3,148	—	6,58
AlSb	0,6135	2,22	2,8	4,4	39,53	—	—	0,470	38,548	89,852	0,226	10,65	9,82
AlAs	0,56622	3,03	3,8	5,3	45,1	—	—	0,402	25,0559	150,838	0,364	8,16	9,95
AlP	0,5451	3,5	5,0	6,0	54,47	—	—	0,374	35,2995	176,714	0,636	7,41	9,55
GaSb	0,60954	0,8	2,0	4,2	28,58	—	—	0,783	13,187	156,05	0,0964	13,60	14,78
GaAs	0,56531	1,435	2,9	5,0	33,33	—	—	0,584	30,0432	151,197	0,205	10,88	12,73
GaP	0,54505	2,78	3,7	5,3	45,24	—	—	0,420	31,4388	160,537	0,449	9,09	11,28
InSb	0,64787	0,185	2,0	4,2	22,34	—	—	1,627	25,301	97,024	0,0969	15,29	17,23
InAs	0,60583	0,360	2,5	4,7	27,14	—	—	1,166	14,6475	167,261	0,255	12,35	15,81
InP	0,58687	1,345	2,8	5,1	37,65	—	—	0,604	26,0399	128,707	0,439	9,55	12,65
ZnS	0,5406	3,60	5,3	7,0	34	—	—	0,369	21,5334	158,8582	0,370	5,10	8,30
ZnSe	0,56676	2,68	4,5	6,3	25,62	—	—	0,428	31,9545	127,156	0,196	5,92	8,91
ZnTe	0,6101	2,28	3,7	5,5	21,94	—	—	0,464	28,259	123,0536	0,114	7,28	9,64
CdTe	0,6477	1,50	3,2	5,5	17,40	—	—	0,572	28,5519	94,3804	0,0902	7,18	10,16

§ 4.6. ЛЮМИНЕСЦЕНЦИЯ

Основные понятия. По определению С. И. Вавилова, *люминесценция* * есть излучение, избыточное над тепловым при данной температуре и обладающее длительностью, значительно превышающей период световых колебаний. Для наблюдения люминесценции вещество принципиально необходимо вывести из состояния термодинамического равновесия, т. е. возбудить. Люминесценция возникает вследствие квантовых переходов из верхнего возбужденного энергетического состояния в нижнее с испусканием электромагнитного излучения, т. е. определяется процессами, обратными поглощению света. В это же время понятие люминесценции относится не к отдельным атомам или молекулам, а к веществу в целом.

При люминесценции каждый из актов возбуждения и излучения разделен промежуточными процессами, что приводит к конечным значениям длительности послесвечения. Люминесценция, обладающая малой длительностью послесвечения (10^{-8} с и менее)**, называется *флуоресценцией*, а большой длительностью — *фосфоресценцией*. Времена послесвечения при фосфоресценции могут достигать нескольких часов.

Для большинства твердых тел люминесцентные свойства выражены очень слабо. Легирование вещества специальными примесями с целью повышения эффективности люминесценции называется активацией, а сами примеси — активаторами люминесценции. Активированные диэлектрики, обладающие высокой эффективностью люминесценции в видимой области спектра, называются *кристаллофосфорами* или *люминофорами*. Элементарными процессами, обуславливающими люминесценцию в таких веществах, обычно являются внутрицентровые оптические переходы, т. е. электронные переходы между энергетическими уровнями, принадлежащими иону (или атому) активатора. Такую люминесценцию иногда называют *мономолекулярной*.

В полупроводниках люминесценция обычно обусловлена рекомбинацией неравновесных носителей заряда, и поэтому её часто отождествляют с *рекомбинационным излучением*. Не любой из актов рекомбинации дает вклад в люминесценцию. Рекомбинация называется *излучательной*, если избыточная энергия рекомбинирующих частиц выделяется в виде кванта электромагнитной энергии — фотона. Если энергия рекомбинирующих частиц в конечном итоге передается решетке, приводя к её разогреву (выде-

*От лат. *lumen* (*luminis*) — свет и *essent* — слабое действие.

**Интервал времени 10^{-8} с выбран потому, что он соответствует времени жизни атома в возбужденном состоянии для переходов, разрешенных в дипольном приближении (см. § 1.5).

ляется в виде квантов колебаний решетки — фононов), то такая рекомбинация называется *безызлучательной*.

В зависимости от способа возбуждения люминесценции различают: фотолюминесценцию (возбуждение светом); катодолюминесценцию (возбуждение быстрыми электронами, предварительно ускоренными в электронной пушке); радиолюминесценцию и рентгенолюминесценцию (возбуждение радиоактивным и рентгеновским излучением); хемилюминесценцию (возбуждение при химических реакциях); биолюминесценцию, как особый вид хемилюминесценции у живых организмов, связанный с участием ферментов (ферментная хемилюминесценция); электролюминесценцию (возбуждение электрическим полем). Способы возбуждения электролюминесценции в свою очередь могут быть различными, а именно: возбуждение путем инжекции неосновных носителей заряда через гомо- или гетеропереход; возбуждение ударной ионизацией в сильном электрическом поле; возбуждение при туннелировании.

Отметим, что процессы излучательной рекомбинации в полупроводниках очень чувствительны к наличию в материале дефектов (примесей, дислокаций, вакансий и т. д.). По этой причине люминесцентные методы часто используют в качестве методов неразрушающего контроля кристаллов, эпитаксиальных слоев, готовых приборов и т. д.

Рассмотрим процессы, определяющие люминесценцию полупроводников. Их особенностью является то, что возбужденные носители заряда могут быть как свободными, т. е. находиться в разрешенных зонах, так и связанными, т. е. локализованными на дефектах (в том числе на примесях).

После того как в полупроводнике тем или иным путем созданы неравновесные носители заряда, в большинстве случаев довольно быстро (за 10^{-11} — 10^{-13} с) устанавливается *квазиравновесие* отдельно в валентной зоне и в зоне проводимости. Это справедливо тогда, когда время жизни неравновесных носителей заряда в зонах больше времени внутризонной релаксации. Неравновесные носители за очень короткий промежуток времени приобретают температуру, равную температуре решетки, и большую часть времени свободного существования в зоне проводят в состояниях, не отличающихся от равновесных.

Таким образом, в первом приближении можно считать, что генерация неравновесных носителей приводит лишь к изменению концентрации свободных носителей заряда, в то время как распределение этих носителей по энергии в зонах и средняя кинетическая энергия частиц (в отсутствие вырождения) остаются неизменными. При этом распределение неравновесных носителей заряда по энергии в каждой из зон можно характеризовать, как и для равновесного состояния, функцией Ферми — Дирака, однако с ~~несколько иными значениями входящих в нее параметров.~~

Напомним, что статистика Ферми — Дирака справедлива лишь для термодинамически равновесных систем. Для неравновесных систем с нормальным распределением частиц по состояниям в каждой из зон используют функцию Ферми — Дирака, в которой энергия Ферми заменяется величиной F^* , называемой *квазиуровнем Ферми*.

В отличие от равновесного состояния положения квазиуровней Ферми для электронов в зоне проводимости и дырок в валентной зоне различны и обозначаются F_n^* и F_p^* соответственно. Тогда распределение носителей заряда и их концентрации в каждой из зон можно выразить соотношениями

$$f_n = \frac{1}{\exp\left(\frac{E_n - F_n^*}{kT}\right) + 1}; \quad n = N_c \frac{1}{\exp\left(\frac{E_c - F_n^*}{kT}\right) + 1}, \quad (4.72)$$

$$f_p = \frac{1}{\exp\left(\frac{F_p^* - E_p}{kT}\right) + 1}; \quad p = N_v \frac{1}{\exp\left(\frac{F_p^* - E_v}{kT}\right) + 1}.$$

Здесь E_n , E_p — энергии электрона и дырки в зонах, отсчитанные от дна зоны проводимости E_c и потолка валентной зоны E_v соответственно; T — температура решетки (в данном случае это не совсем очевидно); F_n^* , F_p^* — энергетические положения квазиуровней Ферми; N_c , N_v — эффективные плотности состояний в зоне проводимости и в валентной зоне.

Физический смысл введенных квазиуровней Ферми для электронов и дырок поясняет рис. 4.35. Естественно, что положения квазиуровней Ферми F_n^* и F_p^* зависят от интенсивности возбуждения. В отсутствие возбуждения квазиуровни Ферми для электронов и дырок совпадают и их положение определяется обычным уровнем Ферми. С увеличением интенсивности возбуждения возрастает концентрация неравновесных носителей заряда в каждой из зон, а следовательно, расстояние между квазиуровнями Ферми. Если распределение носителей заряда в каждой из зон невырожденно, то вместо распределения Ферми, как и для равновесных состояний, можно записать распределение Больцмана. Тогда

$$n = N_c \exp\left(-\frac{E_c - F_n^*}{kT}\right); \quad p = N_v \exp\left(-\frac{F_p^* - E_v}{kT}\right) \quad (4.73)$$

и произведение неравновесных концентраций

$$np = N_c N_v \exp\left(\frac{E_g}{kT}\right) \exp\left(\frac{F_n^* - F_p^*}{kT}\right). \quad (4.74)$$

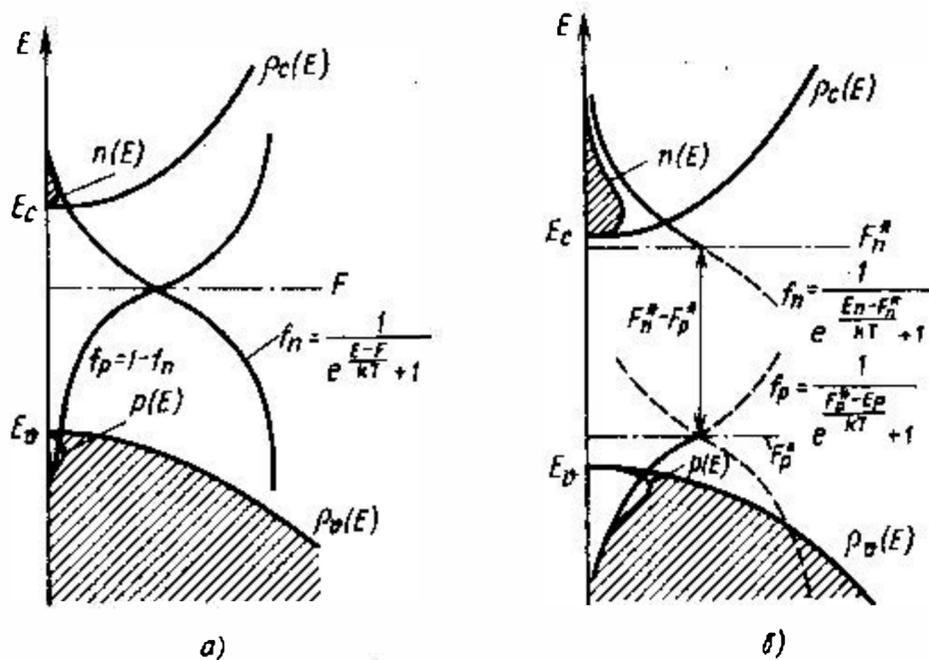


Рис. 4.35. Распределение носителей заряда по энергии:
 а — в равновесном состоянии; б — в неравновесном состоянии при
 ———— влиянии возбуждения

Расстояние между квазиуровнями Ферми определяется уровнем возбуждения и, как нетрудно получить из (4.74),

$$F_n^* - F_p^* = kT \ln(np/n_i^2).$$

Строго говоря, квазиуровень Ферми характеризует заполнение энергетических уровней, лежащих только между ним и границей соответствующей зоны. Заполнение уровней, расположенных между квазиуровнями Ферми, устанавливается процессами рекомбинации.

Условия инверсии и усиления в полупроводниках. Люминесцентное излучение в твердых телах может быть обусловлено как спонтанными, так и вынужденными оптическими переходами при взаимодействии с фотонами. При обычных условиях преобладают спонтанные переходы, определяющие спонтанную, некогерентную люминесценцию. Рассмотрим условия, при которых в полупроводниках возможна вынужденная, когерентная люминесценция, т. е. усиление рекомбинационного излучения и достижение лазерного эффекта.

Для того чтобы процессы вынужденного испускания преобладали над процессами поглощения фотонов, необходимо, как известно, создать инверсию населенностей между двумя энергетическими уровнями E_m и E_n . Это условие, записанное в виде $N_m > N_n$ при $E_m > E_n$, было выведено в § 2.1 для невзаимодействующих или слабо взаимодействующих атомов или молекул, когда каждому из активных центров можно приписать «свои» энергетические уровни, а каждый электрон рассматривать связанным

с конкретным атомом или молекулой, т. е. локализованным в пространстве. В полупроводниках состояния электронов в зоне проводимости и в валентной зоне описываются не локализованными волновыми функциями, а блоховскими функциями (4.3), «размазанными» в пространстве. Соответственно каждый из электронов в кристалле нельзя считать принадлежащим отдельному атому. Поэтому обычное условие инверсии для полупроводников неприемлемо.

Условие усиления электромагнитной волны в полупроводнике можно получить, рассматривая полное число переходов с поглощением и испусканием фотонов так же, как это было сделано в § 2.1. Для определенности возьмем собственный полупроводник и допустим, что процессы взаимодействия с фотонами определяются только переходами зона — зона. Тогда число переходов $\Delta N_{\text{погл}}$ с поглощением фотона будет пропорционально: вероятности перехода из валентной зоны в зону проводимости w_{vc} ; комбинированной плотности состояний $N(\hbar\omega)$; вероятности того, что состояние в валентной зоне занято электроном (f_v), а в зоне проводимости свободно ($1 - f_n$), и спектральной плотности излучения $\rho(\hbar\omega)$ на частоте ω . При наличии возбуждения и установлении квазиравновесия в каждой из зон вероятности заполнения состояний определяются функциями Ферми (4.72), при этом вероятность того, что состояние в валентной зоне будет занято электроном, $f_v = (1 - f_p)$. Следовательно, число поглощенных фотонов в единице объема

$$\Delta N_{\text{погл}} = w_{vc} N(\hbar\omega) (1 - f_p) (1 - f_n) \rho(\hbar\omega).$$

Число индуцированных переходов с испусканием фотонов будет пропорционально вероятности перехода w_{cv} , вероятности того, что состояние в зоне проводимости занято (f_n), а в валентной зоне свободно (f_p), т. е.

$$\Delta N_{\text{инд}} = w_{cv} N(\hbar\omega) f_p f_n \rho(\hbar\omega).$$

Тогда условие превышения числа испущенных фотонов над поглощенными за счет переходов зона — зона можно переписать в виде $\Delta N_{\text{инд}} > \Delta N_{\text{погл}}$.

Вероятности вынужденных переходов с поглощением w_{vc} и испусканием w_{cv} фотонов на основании принципа детального равновесия равны, т. е. $w_{vc} = w_{cv}$ (аналогично равенству коэффициентов Эйнштейна B_{mn} и B_{nm} для вынужденных переходов). Тогда из последнего соотношения имеем

$$f_n f_p > (1 - f_p) (1 - f_n)$$

$$f_n > (1 - f_p),$$

откуда с учетом (4.72) получаем условие

$$F_n^* - F_p^* > E_n - E_p. \quad (4.75)$$

Поскольку минимальное энергетическое расстояние между E_n и E_p равно ширине запрещенной зоны E_g , это неравенство можно записать в виде

$$F_n^* - F_p^* > E_g. \quad (4.76)$$

Соотношение (4.76) по существу определяет условие инверсии населенностей в собственном полупроводнике для переходов зона — зона. Его физической смысл очевиден: для того чтобы обеспечить преобладание усиления за счет процессов вынужденного испускания в полупроводнике над процессами собственного поглощения, необходимо создать такие избыточные неравновесные концентрации носителей заряда в зоне проводимости и в валентной зоне, при которых расстояние между квазиуровнями Ферми будет превышать ширину запрещенной зоны полупроводника. Это означает, что один или оба квазиуровня Ферми F_n^* и F_p^* должны находиться внутри разрешенных зон, т. е. возбуждение должно быть настолько интенсивным, чтобы создать вырождение в зоне проводимости и в валентной зоне.

В общем случае, применяя аналогичные рассуждения, нетрудно получить условие инверсии для любых типов переходов с испусканием фотона

$$F_n^* - F_p^* > \hbar\omega. \quad (4.76a)$$

Условие (4.76a) является необходимым, но не достаточным для получения усиления или генерации в системе в целом. Как и ранее (см. § 2.4), достаточным условием является превышение активного усиления над всеми потерями. В полупроводниках с непрямой структурой энергетических зон выполнение этого условия затруднено. Это объясняется прежде всего неактивными процессами поглощения на свободных носителях заряда. Указанные процессы разрешены только во втором приближении теории возмущений. Для полупроводников с непрямой структурой зон переходы зона — зона также возможны лишь во втором приближении теории возмущений, т. е. имеют тот же порядок вероятности, что и не прямые переходы с участием свободных носителей заряда. Увеличение интенсивности накачки не приводит к повышению общего коэффициента усиления, поскольку одновременно с ростом инверсии будет возрастать коэффициент потерь, который пропорционален концентрации свободных носителей заряда. По этой причине получить лазерный эффект на межзонных переходах в полупроводниках с непрямой структурой зон очень трудно, а в ряде случаев принципиально невозможно.

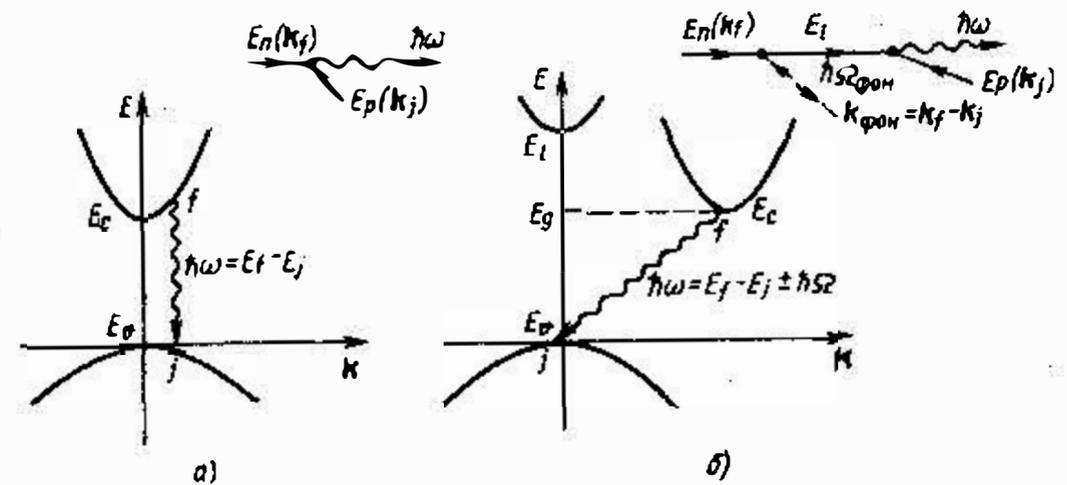


Рис. 4.36. Прямые (а) и не прямые (б) межзонные излучательные переходы

Для полупроводников с прямой структурой зон переходы зона — зона разрешены в первом порядке теории возмущений, и их вероятность существенно превышает вероятность переходов, определяющих паразитное поглощение на свободных носителях зарядов. В связи с этим для получения когерентной люминесценции и лазерного эффекта необходимо применять прямозонные полупроводники.

Связь между оптическим поглощением и люминесценцией. Основные закономерности, присущие квантовым переходам при взаимодействии с оптическим излучением, рассмотренные в § 2.3 и 4.3, справедливы и для излучательной рекомбинации в полупроводниках. Поскольку процессы люминесценции и поглощения определяются односторонними, но противоположно направленными квантовыми переходами, то между ними должна существовать количественная связь. Проследим эту связь на примере собственной излучательной рекомбинации зона — зона. Очевидно, она непосредственно связана с собственным поглощением полупроводника. По существу оба процесса обусловлены одинаковыми, но противоположно направленными электронными переходами. Как и в случае оптического поглощения, межзонные излучательные переходы могут быть прямыми (рис. 4.36, а) и не прямыми (рис. 4.36, б) в зависимости от расположения экстремумов соответствующих зон в k -пространстве. При не прямых переходах, как и ранее, для выполнения законов сохранения часть квазимпульса должна быть передана какой-либо третьей частице (фонону, дефекту решетки и т. д.). При этом ей передается и часть энергии.

Полное число межзонных излучательных переходов Φ в единицу времени в единице объема пропорционально произведению концентрации свободных электронов n и дырок p :

$$\Phi = Vnp. \quad (4.77)$$

Постоянную B можно вычислить из принципа детального равновесия по результатам измерения показателя поглощения k_{ω} , используя теорию Ван — Русбрека — Шокли. Проведение таких расчетов позволяет определить скорость излучательной рекомбинации в предположении, что квантовый выход равен единице, а следовательно, оценить так называемое излучательное время жизни $\tau_{\text{изл}}$, т. е. время жизни, ограничиваемое только излучательными переходами. Рассмотрим для простоты невырожденный материал.

В условиях термодинамического равновесия число актов излучательной рекомбинации с испусканием фотонов в частотном интервале $\omega \dots \omega + d\omega$ равно числу электронно-дырочных пар, генерируемых термодинамически равновесным излучением той же частоты. Скорость генерации пар

$$G(\omega)d\omega = w(\omega) \frac{\rho(\omega)}{\hbar\omega} d\omega, \quad (4.78)$$

где $w(\omega)$ — вероятность поглощения фотона с частотой ω в единицу времени, а $\rho(\omega)/\hbar\omega$ — число фотонов с частотой ω в единице объема в единичном интервале частот.

Средняя длина свободного пробега фотона равна $1/k_{\omega}$. Тогда среднее время жизни фотона $\frac{1}{k_{\omega}} \frac{n}{c}$, т. е.

$$w(\omega) = k_{\omega} c/n.$$

При термодинамическом равновесии для любого интервала скорость генерации равна скорости рекомбинации. Обозначим скорость рекомбинации в условиях равновесия через $R_0(\omega)$. Тогда

$$R_0(\omega)d\omega = w \frac{\rho(\omega)}{\hbar\omega} d\omega = k_{\omega} \frac{c \rho(\omega)}{n \hbar\omega} d\omega. \quad (4.79)$$

Плотность электромагнитной энергии $\rho(\omega)$ в условиях термодинамического равновесия определяется функцией Планка (1.107). Подставляя (1.107) в (4.79) и учитывая, что скорость света в веществе равна c/n , получаем выражение для скорости рекомбинации:

$$R_0(\omega) = \frac{\omega^2 n^2}{\pi^2 c^2} \frac{k_{\omega}}{\hbar\omega} \frac{1}{e^{y} - 1}, \quad (4.80)$$

с помощью которого можно, зная спектральную зависимость показателя поглощения k_{ω} , найти соответствующий спектр люминесценции $R(\omega)$.

Интегрируя по всем частотам, определяем полное число актов рекомбинации в единице объема в единицу времени, т. е. полную скорость рекомбинации:

$$R_0 = \int_0^{\infty} R_0(\omega) d\omega = \frac{k^3 T^3}{\pi^2 c^2 \hbar^3} \int_0^{\infty} \frac{k_{\omega} n^2 y^2}{e^y - 1} dy,$$

где $y = \hbar\omega/kT$.

В термодинамически неравновесном, но стационарном состоянии, когда равновесие устанавливается в каждой из зон отдельно, а также в отсутствие вырождения в каждой из зон скорость излучательной рекомбинации

$$R = \Phi = \frac{np}{n_i^2} R_0, \quad (4.81)$$

или с учетом (4.74)

$$R = R_0 \exp\left(\frac{F_n - F_p}{kT}\right), \quad (4.81a)$$

где $n_i^2 = n_0 p_0$ — квадрат собственной концентрации носителей заряда в материале при заданной температуре T в условиях термодинамического равновесия.

Отсюда получаем постоянную B в соотношении (4.77):

$$B = R_0/n_i^2.$$

Характерные значения B при комнатной температуре для некоторых полупроводников с непрямыми (Ge, Si, GaP) и прямыми (GaAs, GaSb, InP, InSb, PbS) межзонами переходами приведены в табл. 4.2. Естественно, что в случае прямых переходов вероятность излучательной рекомбинации на три-четыре порядка выше, чем в случае не прямых.

Таблица 4.2

Величины, характеризующие скорость межзонной излучательной рекомбинации в полупроводниках при комнатной температуре

Переходы	Вещество	E_g , эВ	n_i , см ⁻³	B , см ³ с ⁻¹	$\tau_{\text{изл, макс}}$, с	$\tau_{\text{изл, макс}} \text{ при } \rho(n) = 10^{16} \text{ см}^{-3}$
Непрямые	Ge	0,665	$2,3 \cdot 10^{13}$	$5 \cdot 10^{-14}$	0,4	2000
	Si	1,086	$8 \cdot 10^9$	$2 \cdot 10^{-15}$	$3 \cdot 10^4$	50000
	GaP	2,27	2	$5 \cdot 10^{-14}$	$5 \cdot 10^{12}$	2000
Прямые	GaAs	1,428	$1 \cdot 10^6$	$7 \cdot 10^{-10}$	700	0,15
	InP	1,351	$7 \cdot 10^4$	$1 \cdot 10^{-9}$	70	0,1
	GaSb	0,70	$9 \cdot 10^{11}$	$2 \cdot 10^{-10}$	0,003	0,5
	PbS	0,37	$3 \cdot 10^{15}$	$2 \cdot 10^{-11}$	$6 \cdot 10^{-3}$	5
	InSb	0,18	$2 \cdot 10^{16}$	$4 \cdot 10^{-11}$	$6 \cdot 10^{-7}$	2,5

Зная n_i и V , можно вычислить время затухания малых возмущений концентрации носителей заряда при наличии только излучательной рекомбинации зона — зона. Предположим, что равновесные концентрации носителей n_0 и p_0 изменились на δn и δp соответственно. Тогда скорость рекомбинации

$$\Phi = R_0 + \delta R = R_0 \frac{(n_0 + \delta n)(p_0 + \delta p)}{n_0 p_0}$$

Отбрасывая член $\delta n \delta p / n_0 p_0$ второго порядка малости, получаем:

$$\delta R / R_0 = \delta n / n_0 + \delta p / p_0 \quad (4.82)$$

В собственном полупроводнике $n_0 = p_0 = n_i$, $\delta n = \delta p$ и время жизни носителей заряда по отношению к излучательной рекомбинации зона — зона максимально:

$$\tau_{\text{жизн}} \text{ макс} = \frac{\delta n}{\delta R} = \frac{\delta p}{\delta R} = \frac{n_0 p_0}{R(n_0 + p_0)} = \frac{1}{n_i B} \quad (4.83)$$

Вычисленные по (4.83) для различных полупроводников значения $\tau_{\text{жизн}} \text{ макс}$ приведены в табл. 4.2. Как видно из этих данных, $\tau_{\text{жизн}} \text{ макс}$ очень сильно зависит от ширины запрещенной зоны.

Рассмотрим теперь легированный полупроводник для определенности p -типа. Тогда $p_0 \gg n_0$ и из (4.82) имеем $\delta R / R_0 \approx \delta n / n_0$. Излучательное время жизни, определяемое переходами зона — зона:

$$\tau_{\text{жизн}} = \frac{\delta n}{\delta R} \approx \frac{n_0}{R_0} = \frac{n_i^2}{p_0 R_0} = \frac{1}{p_0 B} = \frac{2n_i}{p_0} \tau_{\text{жизн}} \text{ макс} \quad (4.84)$$

Соответствующие значения $\tau_{\text{жизн}}$, вычисленные для $p_0 = 10^{16} \text{ см}^{-3}$, также представлены в табл. 4.2. Если концентрация основных носителей заряда поддерживается постоянной, то $\tau_{\text{жизн}}$ сравнительно слабо зависит от E_g . При этом в полупроводниках с непрямой структурой зон $\tau_{\text{жизн}}$ при прочих равных условиях примерно на четыре порядка больше, чем в полупроводниках с прямой структурой зон.

Таким образом, соотношение (4.80) с учетом (4.81) позволяет по известному спектру поглощения k_ω определить спектр люминесценции Φ_ω :

$$\Phi_\omega = A (\hbar\omega)^2 \frac{k_\omega}{e^{kT} - 1} \quad (4.85)$$

Это соотношение, как и (4.80), справедливо до тех пор, пока функции распределения носителей заряда в зонах с точностью до постоянного множителя не отличаются от равновесных. Коэффициент пропорциональности A слабо зависит от энергии фотона $\hbar\omega$ и его можно считать постоянным для данного типа переходов.

Для прямых разрешенных переходов спектр поглощения зона — зона определяется соотношением (4.42), откуда с помощью (4.85) спектр собственной люминесценции

$$\begin{aligned} \Phi_\omega &= A' (\hbar\omega)^2 (\hbar\omega - E_g)^{1/2} \exp\left(-\frac{\hbar\omega - E_g}{kT}\right) \approx \\ &\approx \Phi_0 \left(\frac{\hbar\omega - E_g}{E_g}\right)^{1/2} \exp\left(-\frac{\hbar\omega - E_g}{kT}\right). \end{aligned} \quad (4.86)$$

Здесь учтено, что обычно $\hbar\omega \gg kT$, что дает возможность пренебречь единицей в знаменателе (4.85), а в коэффициент A' для удобства вычислений включен множитель $\exp(-E_g/kT)$.

Из (4.86) видно, что максимум спектра люминесценции зона — зона расположен при энергии $\hbar\omega = (E_g + kT/2)$, а ширина спектра пропорциональна kT . В качестве примера на рис. 4.37 приведены спектры люминесценции нелегированного антимонида индия. Сплошной кривой представлены экспериментальные данные, пунктирной — расчет по (4.86).

Спектр собственного излучения при непрямых межзонных переходах, определяемых диаграммами на рис. 4.36, б, получается из соотношений (4.49) — (4.51) и (4.85) в виде

$$\Phi_\omega \approx \Phi_\pm \left(\frac{\hbar\omega - E_g \pm \hbar\Omega}{E_g}\right)^2 \exp\left(-\frac{\hbar\omega - E_g \pm \hbar\Omega}{kT}\right), \quad (4.87)$$

где знаки «+» и «-» соответствуют процессам испускания фотона с одновременным испусканием (знак «+») и поглощением (знак «-») фонона $\hbar\Omega$.

Механизмы излучательной рекомбинации в полупроводниках. Процессы, приводящие к рекомбинационному излучению в полу-

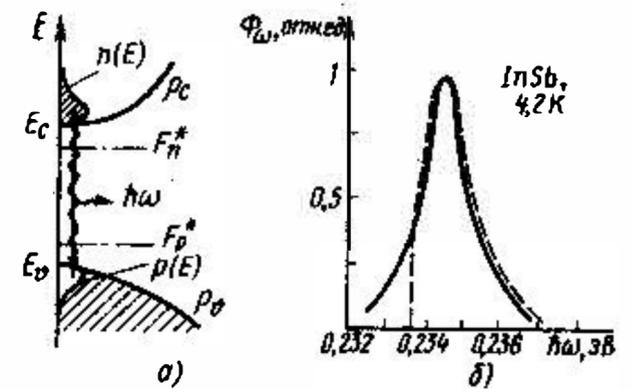


Рис. 4.37. Излучательная рекомбинация, обусловленная межзонными электронными переходами: а) энергетическая диаграмма; б) спектр излучательной рекомбинации InSb при $T = 4,2 \text{ К}$. Сплошной кривой — эксперимент, пунктирная — теория (расчет по формуле (4.86))

проводниках, могут быть довольно многочисленными. Они начинаются с возникновения неравновесных свободных носителей заряда одного или двух типов (электронов и дырок) под воздействием возбуждения. Непосредственно процессу рекомбинации могут предшествовать диффузия носителей заряда, ускорение их полем, захват ловушками, образование свободных экситонов, диффузия свободных экситонов и захват их дефектом. Следовательно, рекомбинационные процессы обычно не сводятся к какому-либо одному простому виду, и люминесценция в полупроводниках может быть обусловлена различными по своей природе электронными переходами. Различают следующие виды рекомбинации: 1) излучательную рекомбинацию зона — зона, т. е. излучательную рекомбинацию свободного электрона со свободной дыркой; 2) излучательную рекомбинацию свободных экситонов; 3) излучательную рекомбинацию связанных экситонов; 4) излучательную рекомбинацию зона — примесь (или примесь — зона) — рекомбинацию свободных носителей заряда с носителями противоположного знака, захваченными примесными центрами; 5) межпримесную излучательную рекомбинацию — рекомбинацию носителей заряда противоположного знака, захваченных двумя различными, пространственно разделенными центрами; 6) внутрицентровые излучательные переходы, т. е. электронные переходы между двумя энергетическими уровнями, принадлежащими одному и тому же центру; 7) внутризонные излучательные переходы между подзонами одной и той же зоны, включающие «горячие» носители заряда.

Соответствующие электронные переходы схематически изображены на рис. 4.38. Механизмы 1 и 2 излучательной реком-

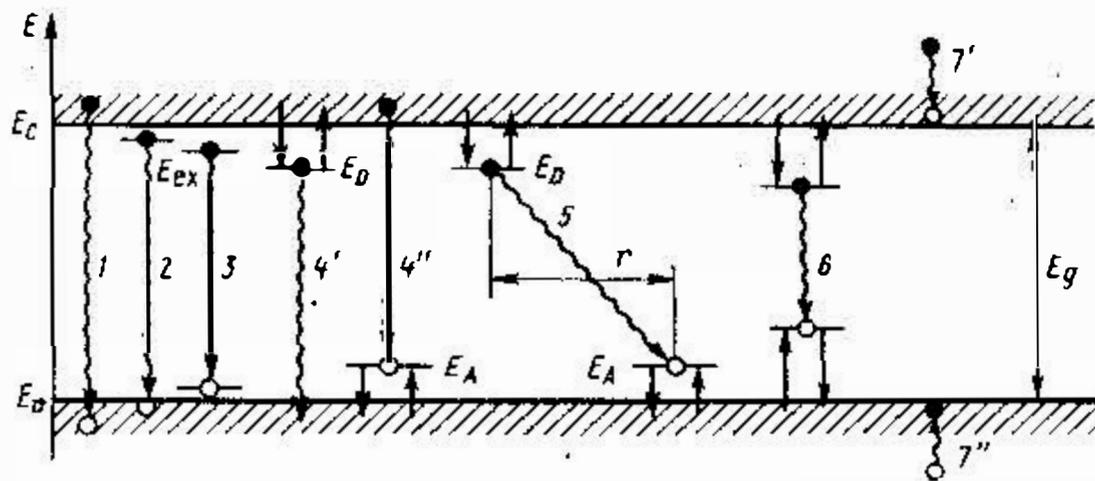


Рис. 4.38. Схематическое изображение электронных переходов, ответственных за различные механизмы излучательной рекомбинации в полупроводниках:

1 — зона-зона; 2, 3 — рекомбинация свободных и связанных экситонов; 4 — рекомбинация свободного носителя со связанным на доноре (акцепторе); 5 — межпримесная излучательная рекомбинация; 6 — внутрицентральный переход; 7 — внутризонные переходы с участием горячих носителей.

бинации являются собственными, а механизмы 3, 4, 5 и 6 — примесными. Поскольку процессы испускания и поглощения фотона определяются аналогичными электронными переходами, каждому из процессов излучательной рекомбинации можно сопоставить такой же процесс оптического поглощения, в чем легко убедиться, сравнивая рис. 4.38 и 4.17. Отметим особенности указанных механизмов люминесценции.

1. Собственная излучательная рекомбинация зона — зона была рассмотрена для нелегированных полупроводников. При увеличении степени легирования мелкими донорами или акцепторами в прямозонных полупроводниках начинают играть существенную роль излучательные переходы в хвосты плотности состояний (см. рис. 4.5). Такие процессы называются «квазимежзонными» переходами.

2. Излучательная рекомбинация свободных экситонов является процессом, обратным экситонному поглощению. Из принципа детального равновесия по спектрам экситонного поглощения с помощью (4.85) можно вычислить спектр и вероятность излучательной рекомбинации через экситонные состояния, которая оказывается выше вероятности переходов зона — зона. Физически это объясняется тем, что электрон и дырка в экситоне связаны и «подготовлены» к рекомбинации. Однако следует помнить, что экситон может не только рекомбинировать, но и диссоциировать с образованием свободных носителей заряда. Вероятность диссоциации также велика, особенно при высоких температурах вследствие малости энергии связи экситона E_{ex} . Отметим, что на рис. 4.38 экситонный переход 2 изображен условно.

При очень сильном возбуждении, когда концентрация экситонов значительна (высока плотность «экситонного газа»), при некоторых условиях возможно образование так называемых экситонных капель и экситонных молекул — биэкситонов, которые также проявляются в рекомбинационном излучении.

3. Излучательная рекомбинация связанных экситонов. Излучательная рекомбинация связанных экситонов возникает в случае, когда пара электрон — дырка перед рекомбинацией связывается на примесном центре (или каком-либо другом дефекте), а последующий излучательный переход происходит как переход центра из возбужденного состояния в равновесное. Такой механизм излучательной рекомбинации может быть достаточно эффективным. Он особенно важен в полупроводниках с непрямой структурой энергетических зон, поскольку вероятность излучательной рекомбинации через такой центр существенно больше вероятности межзонных переходов.

Как отмечалось в § 4.4, связанные экситоны наиболее активно образуются на изоэлектронных ловушках. В зависимости от того, в какой последовательности происходит захват носителей

заряда изоэлектронным центром при образовании связанного экситона, различают *изоэлектронные доноры* и *изоэлектронные акцепторы*. Если изоэлектронный центр имеет меньший атомный номер, чем замещаемый основной атом решетки, то возникающие вследствие разностей их электроотрицательностей короткодействующие силы будут притягивать электрон. После электрона захватывается дырка. Такой центр называется изоэлектронным акцептором (так как дырка движется в кулоновском поле). Если центр захватывает сначала дырку, а затем электрон, то он называется изоэлектронным донором.

Потенциал, создаваемый изоэлектронным центром, является довольно короткодействующим. Поэтому, несмотря на возможно малую энергию связи, этот центр обладает свойствами глубокой, неводородоподобной примеси, что значительно облегчает условия рекомбинации в полупроводниках с непрямой структурой зон.

Типичным примером изоэлектронной ловушки может служить атом азота в фосфиде галлия — полупроводнике с непрямой структурой энергетических зон. Энергия связи электрона на нейтральной изоэлектронной ловушке атома N (атом азота в GaP замещает атом фосфора и является изоэлектронным акцептором) обусловлена короткодействующими центральными силами и составляет $\sim 0,01$ эВ. Волновая функция электрона в значительной степени деформирована и локализована вблизи атома N, поэтому она имеет намного большую протяженность в k -пространстве, чем, например, волновая функция электрона, локализованного на мелком, водородоподобном донорном центре (кулоновское взаимодействие). Это облегчает условия рекомбинации. При взаимодействии с такой ловушкой большую вероятность приобретает бесфононный переход, когда закон сохранения квазиимпульса (4.30) выполняется за счет взаимодействия не с квантами колебаний решетки — фононами, а с примесным атомом, т. е. с тем же дефектом, с которым связан экситон. Спектр излучательной рекомбинации связанных экситонов существенно уже спектра свободных экситонов, так как связанный экситон локализован в пространстве и его кинетическая энергия равна нулю. При низких температурах излучение связанных экситонов проявляется обычно в виде серии очень узких линий (рис. 4.39, а). С повышением температуры проявляется температурное уширение и отдельные линии сливаются.

Бесфононная линия излучения (*A*-линия на рис. 4.39, а) расположена резонансно соответствующей линии поглощения. Остальные линии являются фононными повторениями *A*-линии с испусканием характеристических фононов: *TA* — поперечного акустического, *LA* — продольного акустического, *TO* — поперечного оптического и *LO* — продольного оптического. Символы

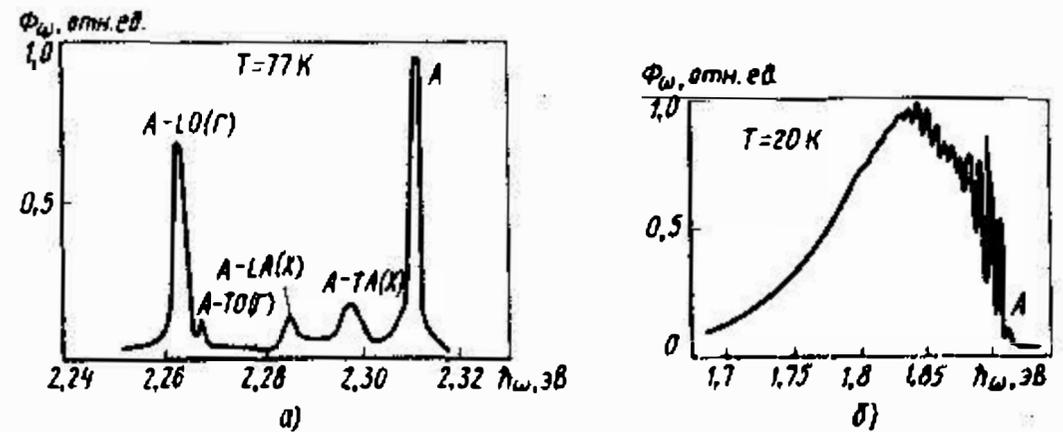


Рис. 4.39. Спектры излучательной рекомбинации связанных экситонов:

а) экситоны, связанные на изоэлектронной ловушке N в GaP; *A* — бесфононная линия, остальные — её фононные повторения с указанием обозначения на рисунке фононов; б) экситоны, связанные на «молекулярном» изоэлектронном комплексе (Cd — O) в GaP; *A* — бесфононная линия

в скобках у обозначений типов фононов указывают на их принадлежность точкам зоны Бриллюэна.

Излучательная рекомбинация связанных экситонов может быть эффективной не только при низких, но и при высоких температурах (300 К и выше). Например, в фосфиде галлия она определяет люминесценцию в зеленой области спектра при комнатной температуре.

Другим примером изоэлектронной ловушки может служить комплекс Zn-O в GaP. Атомы цинка и кислорода замещают атомы галлия и фосфора, располагаясь в ближайших узлах кристаллической решетки фосфида галлия и образуя в целом нейтральный комплекс, действующий как изоэлектронная ловушка. Энергия связи экситона на комплексе значительно больше, чем на азоте в GaP ($\approx 0,4$ эВ). Поэтому вероятность термической диссоциации экситонов, связанных на комплексах Zn-O или Cd-O, значительно меньше и соответственно эффективность излучательной рекомбинации даже при комнатной температуре очень высока (до 10%). Спектр излучения экситонов, связанных на комплексах типа Zn-O или Cd-O, при низких температурах также состоит из узкой «бесфононной» *A*-линии и серии её фононных повторений. Однако поскольку комплексы Zn-O и Cd-O дают более глубокие ловушки, их взаимодействие с решеткой, в том числе с локальными колебаниями самого комплекса, будет более сильным, и в спектре люминесценции проявится большое число близко расположенных линий, как показано на рис. 4.39, б. С повышением температуры или в несовершенных кристаллах эти линии уширяются и, накладываясь друг на друга, образуют широкую, бесструктурную полосу люминесценции.

4. Излучательная рекомбинация свободных носителей заряда на связанных состояниях. Излучательная рекомбинация свободных носителей заряда на связанных состояниях, называемая так-

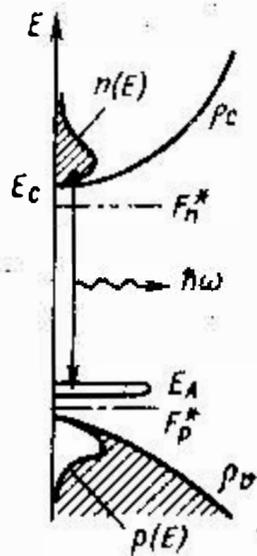


Рис. 4.40. Излучательная рекомбинация свободного электрона с дыркой на акцепторе

же излучательной рекомбинацией зона — примесь (или примесь — зона), обусловлена переходами носителей из одной энергетической зоны на примесные уровни вблизи противоположной зоны, как показано на рис. 4.40 для случая рекомбинации свободного электрона с дыркой на акцепторе. Такие переходы наиболее эффективны в полупроводниках с прямой структурой зон, содержащих примеси. В кристаллах арсенида галлия они совместно с межзонами переходами обеспечивают внутренний квантовый выход люминесценции при комнатной температуре, близкий единице.

Применяя принцип детального равновесия для переходов с участием примесных состояний, аналогично тому, как это было сделано ранее, нетрудно показать, что спектры поглощения и люминесценции с участием примесных состояний, так же как и для переходов зона — зона, связаны между собой соотношением (4.85).

Спектр излучения, обусловленного переходами электронов из зоны проводимости на мелкие акцепторные уровни, расположенные вблизи валентной зоны, в простейшем случае прямых разрешенных переходов для параболических зон и небольшой концентрации примесных центров (в отсутствие взаимодействия между ними) очень похож на спектр межзонаной излучательной рекомбинации и определяется выражением

$$\Phi_{\omega} = C\omega^2 (\hbar\omega - E_g + E_A)^{1/2} \frac{1}{e^{(\hbar\omega - E_g + E_A)/kT} - 1}. \quad (4.88)$$

Как и следовало ожидать, спектр излучения для переходов зона — примесь сдвинут от ширины запрещенной зоны на энергию ионизации. Максимум спектра расположен при энергии фотонов $\hbar\omega = E_g - E_A + kT/2$, а его ширина, как и для переходов зона — зона, в значительной степени определяется температурой.

На форму спектров люминесценции с участием глубоких примесей (как и на форму спектров поглощения) большое влияние оказывает электрон-фононное взаимодействие. При этом спектры поглощения и люминесценции сдвинуты относительно друг друга.

Поясним сказанное с помощью конфигурационной диаграммы рис. 4.41, где энергия основного E_0 и возбужденного E_1 состояний центра представлена как функция его геометрического положения в решетке (ср. с рис. 1.5). Равновесным положениям

центра в основном и возбужденном состояниях соответствуют различные конфигурационные координаты x_0 и x_1 . При оптическом возбуждении (поглощении фотона с энергией $\hbar\omega_{\text{погл}}$) система переходит из основного состояния a в состояние b , поскольку за время оптического перехода не происходит смещения атомов. Но точка b не отвечает минимуму энергии в возбужденном состоянии, поэтому вся система релаксирует в состояние c , например, путем испускания N_1 фононов с энергией $\hbar\Omega_1$. Переход из возбужденного состояния c в основное состояние a сопровождается испусканием фотона с энергией $\hbar\omega_{\text{лом}}$, а затем снова происходит релаксация в наинизшее состояние a путем испускания N_2 фононов с энергией $\hbar\Omega_2$. При этом $\hbar\omega_{\text{лом}} < \hbar\omega_{\text{погл}}$. Такое смещение спектра называется *стоксовым смещением*. Оно аналогично стоксову смещению при комбинационном рассеянии света (см. § 1.7).

Атомы решетки и примесного центра колеблются вблизи равновесных положений, и энергия термической ионизации $E_{\text{терм}}$ будет определяться минимальным энергетическим расстоянием между кривыми E_1 и E_0 на конфигурационной диаграмме. Как видно из рис. 4.41, оптическая энергия ионизации глубокого центра $E_{\text{опт}}$ может быть значительно больше термической энергии $E_{\text{терм}}$. Разницу этих энергий часто называют *сдвигом Франка — Кондона*.

Наблюдаемые в люминесценции полосы излучения являются огибающими групп линий, каждая из которых возникает в результате оптического перехода между колебательными уровнями n и m возбужденного и основного состояний. Расчет показывает, что спектр люминесценции для переходов зона — глубокий уровень (или глубокий уровень — зона) имеет гауссову форму с шириной полосы, определяемой в основном константой электрон-фононного взаимодействия.

В качестве примера на рис. 4.42 изображены теоретические и экспериментальные спектры поглощения и люминесценции с участием глубокого донора — кислорода ($E_d = 0,9$ эВ) в фосфиде галлия. Отчетливо видно стоксово смещение между спектрами поглощения и люминесценции. В данном случае разница между оптической и термической энергиями ионизации $\Delta = E_{\text{опт}} - E_{\text{терм}} = 0,09$ эВ.

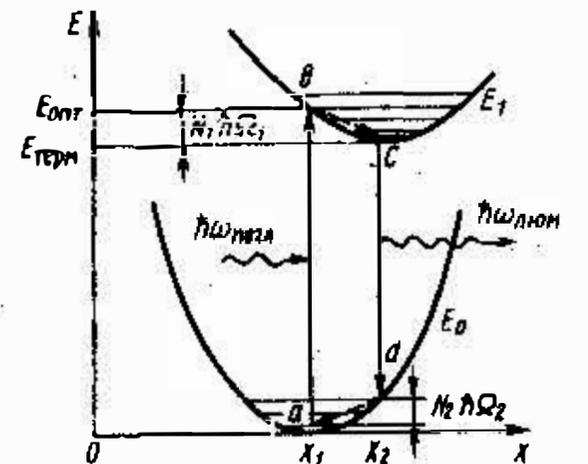


Рис. 4.41. Конфигурационная диаграмма:

x_0 и x_1 — конфигурационные координаты, отвечающие равносильному состоянию глубокого центра в основном E_0 и возбужденном E_1 состояниях; $E_{\text{терм}}$ и $E_{\text{опт}}$ — термическая и оптическая энергии соответственно

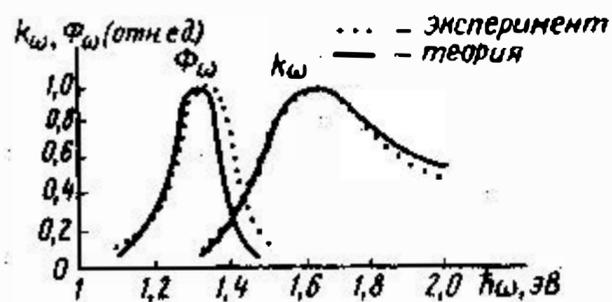


Рис. 4.42. Спектры поглощения k_ω и люминесценции Φ_ω с участием глубокого донора — кислорода в фосфиде галлия

ния сравнительно малы и излучательная рекомбинация может идти за счет межпримесных переходов. Частным и довольно распространенным случаем межпримесной излучательной рекомбинации является излучательная рекомбинация через *донорно-акцепторные пары*, т. е. рекомбинация электрона, захваченного донором, с дыркой, захваченной пространственно разделенным и находящимся на расстоянии r акцептором.

Отметим, что при очень малых расстояниях r (меньше боровских радиусов a_d или a_a) пара донор + акцептор ведет себя как единый молекулярный комплекс. Например, в GaP цинк замещает галлий и является акцептором, а кислород замещает фосфор и является донором. Если расстояние между Zn и O мало (атомы Zn и O располагаются в ближайших друг к другу узлах решетки), то пара Zn-O ведет себя как нейтральный молекулярный комплекс, на котором может связываться экситон. Его спектр люминесценции будет аналогичен представленному на рис. 4.39, б. Если расстояние между донором (O) и акцептором (Zn) $r > a_d + a_a$, то на такие «далекие» центры могут быть по отдельности захвачены электрон и дырка и для них будет проявляться механизм парной люминесценции.

6. Внутрицентровые излучательные переходы. Внутрицентровые излучательные переходы обусловлены переходом примесного центра из возбужденного состояния в основное или более низкое возбужденное состояние с испусканием квантов света. Как отмечалось в § 4.4, типичным примером таких центров могут служить примеси элементов с незастроенными внутренними оболочками в различных кристаллических и аморфных веществах.

Кристаллическая или аморфная основа здесь играет роль матрицы, и электронные переходы, определяющие поглощение и люминесценцию, происходят между «внутренними» состояниями одного и того же атома или иона, как в свободных изолированных атомах и ионах. Спектры излучения при этом определяются в основном свойствами иона-активатора и могут быть уширены как вследствие уширения энергетических уровней центра за счет их расщепления и сдвига под воздействием внутрикри-

5. Межпримесная излучательная рекомбинация. Межпримесная излучательная рекомбинация наблюдается в полупроводниках, содержащих как минимум две группы центров, способных захватывать носители заряда противоположного знака. Если концентрация этих центров велика, то межпримесные расстоя-

сталлических полей, так и вследствие возможного сильного взаимодействия центра с колебаниями решетки. Уширение линии люминесценции в отдельных случаях может изменяться от долей до сотен миллизюльт в зависимости от свойств конкретного центра и матрицы. При наличии сильного электрон-фонового взаимодействия для внутрицентровых переходов, как и для переходов с участием глубоких центров, характерны стоковое смещение и сдвиг Франка — Коидона.

7. Внутризонные излучательные переходы. Внутризонные переходы между подзонами одной и той же зоны, т. е. переходы, противоположные внутризонному междолинному поглощению, в ряде случаев также могут быть излучательными. Однако для их наблюдения необходимо создать «горячие» носители заряда в более высоких подзонах. Скорость релаксации этих носителей внутри зоны очень велика, что уменьшает интенсивность излучения. Излучение, связанное с внутризонными переходами, наиболее часто проявляется при пропускании тока через *p-n*-переход в обратном направлении. Его эффективность обычно очень мала.

Схематическое изображение различных излучательных переходов в полупроводниках, за исключением внутризонных, систематизировано в табл. 4.3. Указанные переходы наблюдались в рекомбинационном излучении различных полупроводников.

Таблица 4.3

Схемы излучательных переходов в полупроводниках

Излучательная рекомбинация	Возбужденное состояние	Основное состояние	Пороговая энергия перехода
Свободного электрона со свободной дыркой	+ -		E_g
Свободного экситона	\pm		$E_{gx} = E_g - E_{ex}$
Свободного электрона с дыркой на акцепторе	$\ominus + -$	\ominus	$E_g - E_A$
Свободной дырки с электроном на доноре	$\oplus - +$	\oplus	$E_g - E_D$
Экситона, связанного на нейтральном изоэлектронном центре	$\circ \pm$	\circ	$E_{gx} - E_{связ}$
Экситона, связанного на заряженном доноре	$\oplus \mp$	\oplus	$E_{gx} - E_{связ}$
Экситона, связанного на заряженном акцепторе	$\ominus \pm$	\ominus	$E_{gx} - E_{связ}$
Внутрицентральной переход на доноре	$\oplus - *$	$\oplus -$	$E_2 - E_1$
Внутрицентральной переход на акцепторе	$\ominus + *$	$\ominus +$	$E_2 - E_1$

Продолжение табл. 4

Излучательная рекомбинация	Возбужденное состояние	Основное состояние	Пороговая энергия перехода
Экситона, связанного на нейтральном доноре	$\oplus - \pm$	$\ominus +$	$E_{gk} - E_{связи}$
Экситона, связанного на нейтральном акцепторе	$\ominus + \mp$	$\ominus +$	$E_{gk} - E_{связи}$
Базиситона на нейтральном центре	$\circ \pm$	\circ	$E_{gk} - E_{связи}$
Электрона на доноре с дыркой на акцепторе	$\oplus - + \ominus$	$\oplus \ominus$	$E_g - (E_D + E_A) + E_{связи}$
Электрона на нейтральном центре с дыркой на акцепторе	$\ominus - + \ominus$	$\circ \ominus$	$E_g - E_A - E_{связи}$
Дырку на нейтральном центре с электроном на доноре	$\oplus - + \circ$	$\oplus \circ$	$E_g - E_D - E_{связи}$

Обозначения: \circ — нейтральный изоэлектронный центр; \oplus — заряженный донор; \ominus — заряженный акцептор; $-$ — электрон; $+$ — дырка; \pm — возбужденное состояние.

Эффективность люминесценции и безызлучательная рекомбинация. Эффективность люминесценции показывает долю излученной энергии относительно затраченной на её возбуждение. Для количественной характеристики излучательных переходов определенного типа в общем числе рекомбинационных переходов вводят понятие *квантового выхода люминесценции*, под которым понимают отношение числа испущенных в заданной спектральной области фотонов к числу созданных в результате возбуждения (генерируемых) носителей заряда или электронно-дырочных пар. Различают внутренний и внешний квантовые выходы люминесценции. При определении *внутреннего квантового выхода* люминесценции учитывают все испущенные фотоны. Его находят обычно расчетным путем из экспериментально измеряемого внешнего квантового выхода, где учитываются лишь фотоны, выведенные наружу, за пределы образца или прибора. За счет явлений внутреннего отражения, поглощения в материале, на контактах и в держателе внешний квантовый выход обычно составляет десятые и даже сотые доли от внутреннего. Он в значительной степени определяется геометрией образца и конструктивными особенностями прибора.

Под *энергетическим выходом люминесценции* понимают отношение энергии электромагнитного излучения, испущенного при люминесценции, к энергии, затраченной на её возбуждение. Как и квантовый выход, энергетический выход может быть внут-

ренним и внешним. Последний по существу определяет коэффициент полезного действия прибора.

Ограничение квантового выхода и эффективности люминесценции происходит из-за наличия процессов *безызлучательной рекомбинации*, которые существенно уменьшают времена жизни неравновесных носителей заряда. Если через $\tau_{изл}$ и $\tau_{бн}$ обозначить времена жизни, определяемые излучательными и конкурирующими безызлучательными переходами в данном материале, то *внутренний квантовый выход люминесценции*

$$\eta_i = \frac{1/\tau_{изл}}{1/\tau} = \frac{\tau}{\tau_{изл}} = \frac{\tau_{бн}}{\tau_{бн} + \tau_{изл}}, \quad (4.89)$$

где τ — экспериментально измеряемое суммарное время жизни избыточных носителей заряда ($1/\tau = 1/\tau_{изл} + 1/\tau_{бн}$).

Очевидно, для получения максимальной эффективности люминесценции необходимо, чтобы значение $\tau_{бн}$ было велико и время жизни неравновесных носителей заряда определялось излучательными переходами, что возможно лишь в кристаллах высокого качества.

Из табл. 4.2 видно, что переходы зона — зона могут давать заметный вклад в люминесценцию лишь в материалах с прямой структурой зон. В полупроводниках с непрямой структурой зон вследствие больших $\tau_{изл}$ конкурирующие процессы безызлучательной рекомбинации при прочих равных условиях будут оказывать существенно большее влияние. Даже в германии, технология производства которого достигла наиболее высокого уровня, экспериментально измеренные времена жизни не превышают 10^{-3} с, в то время как $\tau_{изл макс} \approx 0,4$ с. Таким образом, даже в наиболее чистом материале — германии — остаточные дефекты оказывают сильное влияние на свойства материала, значительно увеличивая скорость рекомбинации. Особенно эффективными центрами рекомбинации могут быть сложные центры, создающие глубокие уровни.

Процессы безызлучательной рекомбинации в полупроводниках (как и в других веществах) изучены слабо. Это связано в первую очередь с трудностью экспериментального исследования безызлучательных процессов, поскольку их непосредственная регистрация затруднена. В элементарном акте безызлучательной рекомбинации должны выполняться законы сохранения энергии и волнового вектора. Заметим, что в материалах с $E_g \geq 0,3$ эВ электрон не может рекомбинировать с дыркой непосредственно безызлучательным путем, т. е. передав свою избыточную энергию решетке. Действительно, колебания решетки квантованы, поэтому энергия рекомбинирующих частиц может выделиться лишь вследствие испускания фононов. Обычно энергия фонона порядка 0,05 эВ. Поэтому для рассеяния энергии, например 2 эВ

($\hbar\omega=2$ эВ соответствует излучению в желтой области спектра с $\lambda=0,62$ мкм), необходимо одновременно испустить 40 фононов. Другими словами, требуется одновременное столкновение 40 частиц, что маловероятно.

В настоящее время определенно можно назвать четыре процесса безызлучательной рекомбинации: 1) Оже-рекомбинацию; 2) поверхностную рекомбинацию; 3) рекомбинацию через макроскопические дефекты, включения и прочие неоднородности; 4) рекомбинацию на точечных микроскопических дефектах. Рассмотрим сущность этих процессов.

1. **Оже-рекомбинация.** Оже-рекомбинация представляет собой ударный процесс, при котором избыточная энергия рекомбинирующих частиц передается третьему носителю заряда, переходящему

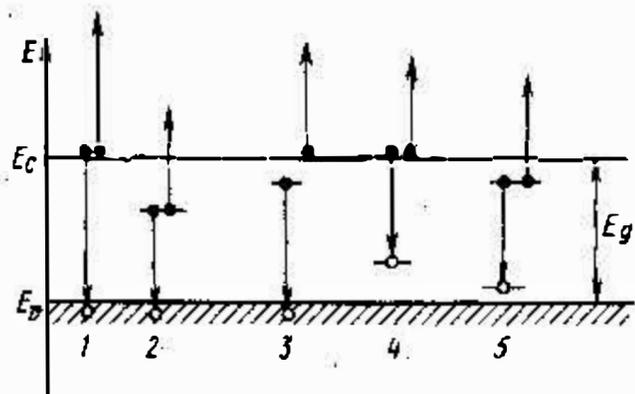


Рис. 4.43. Схематическое изображение процессов Оже для безызлучательной рекомбинации в полупроводнике n-типа электропроводности:

1) межзонная Оже-рекомбинация; 2) Оже-рекомбинация электрона, захваченного на двухзарядном доноре со свободной дыркой; 3) Оже-рекомбинация электрона на доноре со свободной дыркой; 4) Оже-рекомбинация свободного электрона с дыркой на акцепторе; 5) Оже-рекомбинация экситона, связанного на нейтральном доноре

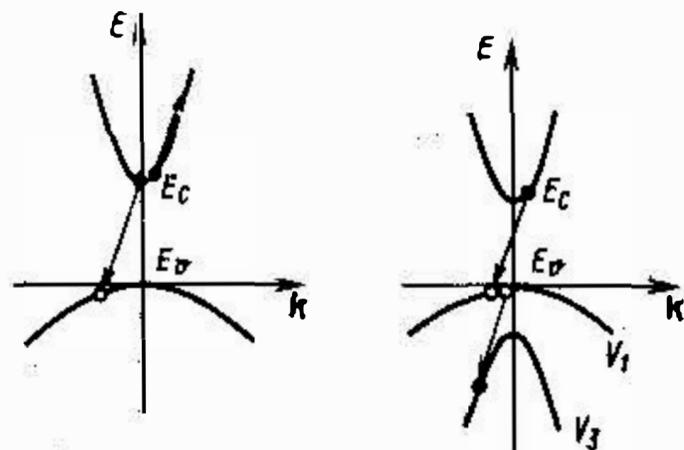


Рис. 4.44. Оже-процессы в полупроводниках с прямой структурой энергетических зон n-типа (а) и p-типа (б) электропроводности

му в состояние с большей кинетической энергией в зоне проводимости (если им является электрон) или в валентной зоне (если им является дырка). Затем эта третья «горячая» частица, находясь глубоко в зоне разрешенных значений энергий, легко релаксирует, рассеивая избыточную энергию путем каскадного испускания фононов. В конечном итоге при таком трехчастичном столкновении, в котором участвуют, например, два электрона и дырка, энергия рекомбинирующих частиц передается решетке, приводя к ее разогреву. В зависимости от начального состояния носителей заряда возможны различные типы Оже-процессов, что для полупроводника n-типа проиллюстрировано на рис. 4.43.

На рис. 4.44, а, б показаны Оже-процессы на диаграммах $E(k)$ в полупроводниках с прямой структурой зон n-типа и p-типа соответственно. Видно, что особен

ности зонной структуры материала, в том числе соотношения эффективных масс m_n, m_p и энергетических зазоров E_g, Δ_{co} , должны существенно влиять на вероятности Оже-рекомбинации. В частности, из рис. 4.44, б нетрудно заключить, что в материале р-типа, в котором $E_g^r \approx \Delta_{co}$, процессы Оже-рекомбинации будут облегчены. Таким примером является антимоид галлия (GaSb).

2. **Поверхностная рекомбинация.** В подавляющем большинстве случаев она происходит безызлучательным путем. Это связано с тем, что поверхность действует как сильное возмущение решетки, приводящее к появлению дискретных и квазинепрерывных энергетических состояний в запрещенной зоне, через которые возможны многоступенчатые (каскадные) безызлучательные переходы. В качестве аналогии можно указать, что при столкновении атома, находящегося в возбужденном состоянии, со стенками сосуда большую вероятность приобретает безызлучательный переход в основное состояние. Этот процесс, в частности, используется для опустошения долгоживущих 1s-уровней неона (см. рис. 1.6) в гелий-неоновом лазере.

3. **Рекомбинация на макроскопических дефектах.** Рекомбинация на макроскопических дефектах (инородных включениях, дислокациях, границах зерен и прочих неоднородностях) аналогична поверхностной рекомбинации, но поверхность в данном случае будет внутренней, создаваемой макроскопическим дефектом. Большая скорость безызлучательной рекомбинации через такие дефекты определяет довольно жесткие требования к однородности и качеству полупроводниковых материалов, применяемых в оптоэлектронике.

4. **Рекомбинация на точечных микроскопических дефектах.** Она может быть пояснена с помощью конфигурационной диаграммы рис. 4.26. Если сдвиг Франка — Кондона большой и соответственно велико значение $\Delta x = x_1 - x_0$, а энергия $E_{\text{терм}}$ сравнительно небольшая, то кривые $E_1(x)$ и $E_0(x)$ могут пересечься вблизи точки с. Тогда из возбужденного состояния b центр может перейти в основное состояние путем каскадных переходов по колебательным подуровням, т. е. безызлучательным путем.

При возрастании концентрации свободных носителей заряда повышается вероятность не только излучательной, но и безызлучательной рекомбинации, в частности Оже-процессов, так как увеличивается вероятность столкновения с третьей частицей. Уменьшение $\tau_{\text{ож}}$ при этом может происходить быстрее, чем $\tau_{\text{изл}}$. Например, для переходов зона — зона вероятность безызлучательной Оже-рекомбинации будет пропорциональна $n^2 p$, в то время как вероятность излучательной рекомбинации пропорциональна np .

При сильном легировании в полупроводниках возможно образование различных макроскопических комплексов, преципитатов и т. д., которые служат центрами безызлучательной реком-

бинации. В арсениде галлия, например, легирование серой или теллуридом до концентрации выше $3 \cdot 10^{18} \text{ см}^{-3}$ приводит к образованию преципитатов Ga_2S_3 и Ga_2Te_3 , на которых происходит интенсивная безызлучательная рекомбинация. Образование «хвостов» плотности состояний (см. рис. 4.5) также облегчает условия безызлучательной рекомбинации. Всё это требует в каждом конкретном случае определять оптимальную концентрацию легирующих центров, при которой эффективность люминесценции будет максимальной.

§ 4.7. ФОТОЭЛЕКТРИЧЕСКИЕ ЭФФЕКТЫ В ОДНОРОДНЫХ КРИСТАЛЛАХ

Фотоэлектрические эффекты (*фотоэффекты*) связаны с изменением электрических свойств вещества под воздействием электромагнитного излучения. В твердых телах может наблюдаться как внешний, так и внутренний фотоэффект. Внешний фотоэффект определяется эмиссией электронов из вещества в вакуум под действием квантов света. Эмитированные электроны затем могут быть собраны анодом. На этом явлении основано действие таких приборов, как вакуумные фотоэлементы, фотоэлектронные умножители, электронно-оптические преобразователи и т. д.

В случае внутренних фотоэффектов носитель заряда (электрон или дырка) под действием кванта света не выходит из вещества, а лишь переходит в более высокое энергетическое состояние (например, из валентной зоны в зону проводимости). При этом первичным процессом является поглощение фотона. Однако не все из рассмотренных в § 4.4 механизмов поглощения приводят к изменению электрических свойств кристалла. Действительно, удельная электрическая проводимость

$$\sigma = e(n\mu_n + p\mu_p)$$

зависит от концентрации и подвижности свободных носителей заряда. Из рассмотренных процессов поглощения только собственное и примесное поглощения приводят непосредственно к изменению концентрации свободных носителей заряда, так как они обусловлены переходами электронов (дырок) под действием квантов света из связанного состояния в свободное. Именно эти процессы поглощения дают наибольший вклад в фотоэлектрический эффект.

Запишем важнейшие уравнения, описывающие внутренние фотоэффекты. К ним относятся прежде всего уравнения непрерывности, согласно которым при равновесии полная скорость генерации дырок G_p и электронов G_n равна полной скорости их рекомбинации:

$$\frac{dp}{dt} = G_p - \frac{\delta p}{\tau_p} - \frac{1}{e} \text{div} j_p, \quad (4.90)$$

$$\frac{dn}{dt} = G_n - \frac{\delta n}{\tau_n} + \frac{1}{e} \text{div} j_n.$$

Здесь $\delta p = p - p_0$; $\delta n = n - n_0$ — отклонения концентраций носителей заряда от равновесных значений n_0 и p_0 ; τ_p и τ_n — времена жизни неравновесных носителей.

Плотности дырочного j_p и электронного j_n токов определяются как сумма дрейфовой и диффузионной составляющих:

$$\begin{aligned} j_p &= e\mu_p p E - eD_p \text{grad} p, \\ j_n &= e\mu_n n E + eD_n \text{grad} n, \end{aligned} \quad (4.91)$$

а плотность полного тока

$$j = j_p + j_n.$$

Здесь μ_p , μ_n , D_p и D_n — соответственно подвижности и коэффициенты диффузии электронов и дырок.

Наиболее сильным и важным фотоэлектрическим эффектом, проявляющимся в однородных полупроводниках, является фоторезистивный эффект.

Фоторезистивный эффект. Этот эффект заключается в изменении электрического сопротивления полупроводника под воздействием оптического излучения. Добавочную проводимость, возникающую в результате фотоактивного поглощения фотонов, называют *фотопроводимостью* и обозначают $\delta\sigma$. Схема опыта для измерения фотопроводимости показана на рис. 4.45.

Плоскопараллельный образец длиной a , шириной b и толщиной d с омическими контактами на торцах подключен к источнику электрического напряжения U . Поверхность образца однородно освещается. Для исключения побочных эффектов при проведении исследований часто затемняют.

В результате фотоактивного поглощения возникает избыточная концентрация электронов δn и дырок δp по сравнению с тепловой. Фотопроводимость можно описать величиной $\delta\sigma$, равной разности проводимости вещества на свету и в темноте:

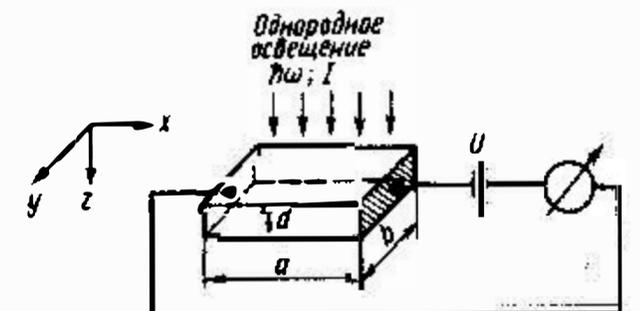


Рис. 4.45. Схема опыта для измерения фотопроводимости

$$\delta\sigma = \sigma_{\text{св}} - \sigma_{\text{т}} = [e\mu_n(n_0 + \delta n) + e\mu_p(p_0 + \delta p)] - [e\mu_n n_0 + e\mu_p p_0] = e\mu_n \delta n + e\mu_p \delta p. \quad (4.92)$$

Как правило, подвижность носителей заряда можно считать неизменной.

Собственное поглощение света приводит к генерации электронов и дырок одновременно. При примесном поглощении генерируются носители заряда только одного знака, причем неравновесные носители являются основными или неосновными. Экситонное поглощение непосредственно не вызывает увеличения концентрации носителей заряда. Но в большинстве случаев энергия связи свободного экситона достаточно мала, так что при комнатной температуре за счет теплового возбуждения экситон диссоциирует на электрон и дырку, давая тем самым вклад в фотопроводимость. Если же экситон до диссоциации рекомбинирует, то экситонное поглощение не приводит к фотопроводимости.

Рассмотрим уравнения непрерывности (4.90) при условии $j_n = j_p \approx 0$. В стационарном состоянии $dn/dt = dp/dt = 0$ и уравнения (4.90) имеют простое и наглядное решение:

$$\delta p = p - p_0 = G_p \tau_p, \quad \delta n = n - n_0 = G_n \tau_n. \quad (4.93)$$

Действительно, стационарная концентрация избыточных носителей заряда равна числу носителей, освобождаемых светом в единицу времени в единице объема, умноженному на среднее время их существования в зоне до рекомбинации, называемое временем жизни. Следовательно, фотопроводимость в стационарном состоянии описывается соотношением

$$\delta\sigma = e\mu_p(G_p \tau_p + bG_n \tau_n) \quad (4.94)$$

и определяется скоростью генерации и временем жизни фотоносителей. Величина $\delta\sigma$ зависит от длины волны падающего света и его интенсивности через $G(\lambda, I)$ и $\tau(L)$.

Скорость генерации G определяется интенсивностью света и показателем поглощения. Если I — интенсивность света, падающего на единицу поверхности в единицу времени, то из соотношения (1.169) следует, что в единице объема поглощается энергия $k_\omega I = -dI/dz$.

Поток фотонов q , измеряемый в $\text{м}^{-2} \cdot \text{с}^{-1}$, легко выразить через интенсивность света: $q = I/\hbar\omega$. Тогда число поглощенных фотонов в единицу времени в единице объема равно

$$-\frac{1}{\hbar\omega} \frac{dI}{dz} = \frac{k_\omega I}{\hbar\omega} = k_\omega q. \quad (4.95)$$

Пусть каждый поглощенный фотон с вероятностью β порождает свободный носитель заряда — электрон или дырку. Тогда скорость генерации носителей заряда, имеющая размерность $\text{м}^{-3} \cdot \text{с}^{-1}$, равна

$$G_n = \beta_n k_\omega q, \quad G_p = \beta_p k_\omega q. \quad (4.96)$$

Величины β_n и β_p называются квантовым выходом фотопроводимости. При поглощении в примесной области спектра, т. е. для примесной фотопроводимости, одна из величин β_n или β_p равна нулю. Это свидетельствует о том, что генерируются носители заряда только одного знака. Так как в одной и той же спектральной области поглощение может определяться различными механизмами, в том числе нефотоактивным (см. § 4.4), то из физических соображений ясно, что квантовый выход фотоионизации, т. е. отношение числа поглощенных фотонов к числу ионизованных центров, может изменяться от нуля до единицы. Под квантовым выходом фотопроводимости подразумевают отношение числа перенесенных зарядов к числу поглощенных фотонов. Его измеряют экспериментально и он может быть как меньше, так и больше единицы. Однако последнее не означает, что один фотон непосредственно может генерировать несколько свободных носителей заряда (или пар носителей), а говорит только о том, что в полупроводнике происходят какие-то вторичные эффекты, в результате которых на каждый фотон приходится эффективное число свободных носителей заряда, превышающих единицу. Например, если энергия фотона $\hbar\omega$ превышает двойную ширину запрещенной зоны, то возникающие «горячие» носители заряда могут обладать достаточной кинетической энергией, чтобы в результате ударной ионизации (вторичный процесс) образовалось еще некоторое количество неравновесных пар. В стационарном состоянии имеем из (4.94) и (4.96)

$$\delta\sigma = ek_\omega \frac{I}{\hbar\omega} (\beta_n \mu_n \tau_n + \beta_p \mu_p \tau_p). \quad (4.97)$$

Если одно из слагаемых в скобках значительно больше другого (за счет примесного фотоэффекта или большой разницы подвижностей), то имеет место неравновесная проводимость, осуществляемая носителями заряда одного знака, которая называется монополярной фотопроводимостью и описывается соотношением

$$\delta\sigma = e\mu\tau\beta k_\omega I/\hbar\omega. \quad (4.98)$$

Из выражений (4.97), (4.98) видно, что фотопроводимость в значительной мере определяется временем жизни фотоносителей τ , т. е. средним временем их существования в зоне. В свою очередь время жизни ограничивается рекомбинацией. Рекомбина-

цией определяются также законы нарастания и спада фотопроводимости при импульсном возбуждении, зависимость фотопроводимости $\Delta\sigma$ от интенсивности возбуждения I и т. д.

Различают *объемную рекомбинацию*, обусловленную объемными свойствами материала: его зонной структурой, наличием в объеме различных центров рекомбинации, захвата и др., и *поверхностную рекомбинацию*, определяемую состоянием поверхности образца. Вклад объемной или поверхностной рекомбинации в ограничение времени жизни фотоносителя зависит от ряда факторов, прежде всего соотношения между толщиной образца d и показателем поглощения k_ω . Рассмотрим два предельных случая: $k_\omega d \ll 1$ и $k_\omega d \gg 1$. При $k_\omega d \ll 1$, как видно из выражения (1.170) и рис. 1.16, интенсивность возбуждающего излучения, а следовательно, скорость генерации остаются приблизительно постоянными во всем образце. Тогда, если диффузионная длина носителей заряда L_p (или L_n) много меньше толщины образца d , то поверхностной рекомбинацией можно пренебречь и время жизни τ_p (или τ_n) будет устанавливаться рекомбинацией в объеме материала. Спектральная зависимость фотопроводимости при этом определяется спектральной зависимостью показателя поглощения k_ω , как показано на рис. 4.46 (кривые 2 и 3). Обычно условие $k_\omega d \ll 1$ выполняется для примесного поглощения, когда значение k_ω мало (меньше 100 см^{-1}). В области собственного поглощения $k_\omega \approx 10^3 \dots 10^5 \text{ см}^{-1}$ и это условие можно выполнить только используя тонкие пленки. Однако диффузионная длина носителей заряда L_p (или L_n) много больше толщины пленки и поверхностной рекомбинацией пренебрегать нельзя. В этом случае влияние поверхностной рекомбинации сводится к уменьшению времени жизни до τ' , определяемого соотношением

$$1/\tau' = 1/\tau + 2S/d, \quad (4.99)$$

где S — скорость поверхностной рекомбинации, м/с.

Таким образом, в тонких пленках концентрация избыточных носителей заряда, а следовательно, фотопроводимость нередко может быть ограничена не объемной, а поверхностной рекомбинацией, скорость которой в значительной мере определяется состоянием поверхности. При $k_\omega d \gg 1$ интенсивность возбуждения, а следовательно, скорость генерации, как видно из выражения (1.170), быстро уменьшается по экспоненциальному закону:

$$G = G_0 e^{-k_\omega z}$$

При этом важную роль начинают играть процессы диффузии неравновесных носителей заряда от поверхности образца, где генерация максимальна, в объем. Спектр фотопроводимости, т. е. зависимость фототока от энергии фотонов (или длины волны) при постоянной интенсивности возбуждения, определяется

теперь не просто спектром поглощения k_ω , как в случае $k_\omega d \ll 1$. Для вычисления фототока нужно решить уравнения непрерывности (4.90) с учетом диффузии, и поскольку $G = G(z)$, провести интегрирование по толщине образца. Граничные условия сводятся к учету поверхностной рекомбинации. Если скорость поверхностной рекомбинации мала (в пределе $S=0$), то с повышением показателя поглощения при увеличении k_ω фототок будет сначала возрастать вслед за увеличением k_ω , а затем достигать насыщения (кривая 4 на рис. 4.46). Насыщение наступает при таких значениях k_ω , когда практически все падающее излучение поглощается материалом. Если скорость поверхностной рекомбинации велика, так что $S\tau_{p,n} > L_{p,n}$, то в спектре фотопроводимости наблюдается максимум (кривая 5 на рис. 4.46). Причина уменьшения $\tau_{p,n}$ с ростом коэффициента поглощения состоит в том, что с увеличением k_ω всё большее число носителей заряда генерируется вблизи поверхности, где время их жизни мало за счет поверхностной рекомбинации, тогда как при меньших значениях k_ω заметное число носителей генерируется в объеме. Этот эффект может сказываться не только в области собственного поглощения, но и в области примесного поглощения, если последнее достаточно велико. При этом в центре полосы интенсивного примесного поглощения может наблюдаться провал фоточувствительности (кривая 6 на рис. 4.46).

Законы нарастания и спада фотопроводимости и зависимость фотопроводимости от интенсивности возбуждения (люкс-амперная характеристика) в значительной мере определяются характером рекомбинации. Рассмотрим два частных случая:

1. *Случай линейной рекомбинации*. Он реализуется, например, при наличии одного типа дырок в полупроводнике p -типа, с которыми непосредственно или через примесные состояния рекомбинируют неравновесные электроны, причем концентрация равновесных дырок p_0 существенно превышает неравновесные концентрации Δn и Δp . В этом случае интенсивность рекомбинации, имеющая размерность $\text{м}^{-3} \text{с}^{-1}$, пропорциональна первой степени

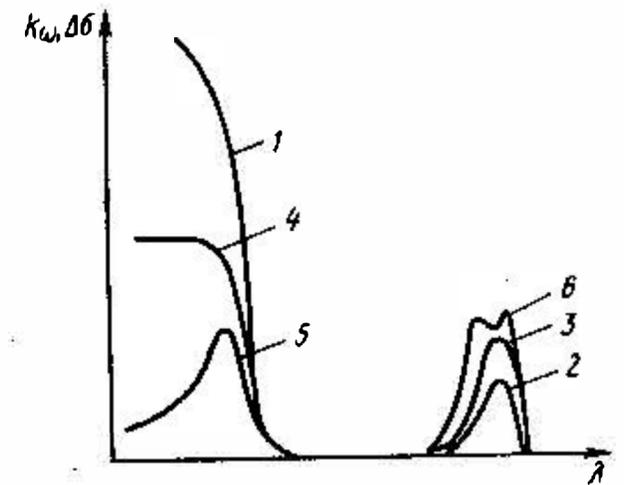


Рис. 4.46. Сравнение спектров фотопроводимости $\Delta\sigma(\lambda)$ при постоянном числе падающих фотонов и фотоактивного поглощения $k_\omega(\lambda)$:

1, 2 — спектры собственного и примесного поглощения; 3 — спектр примесной фотопроводимости при $(k_\omega)_{\text{max}} d < 1$; 4, 5 — спектры собственной фотопроводимости при малой и большой скорости поверхностной рекомбинации; 6 — спектр примесной фотопроводимости при $(k_\omega)_{\text{max}} d \gg 1$ и большой скорости поверхностной рекомбинации

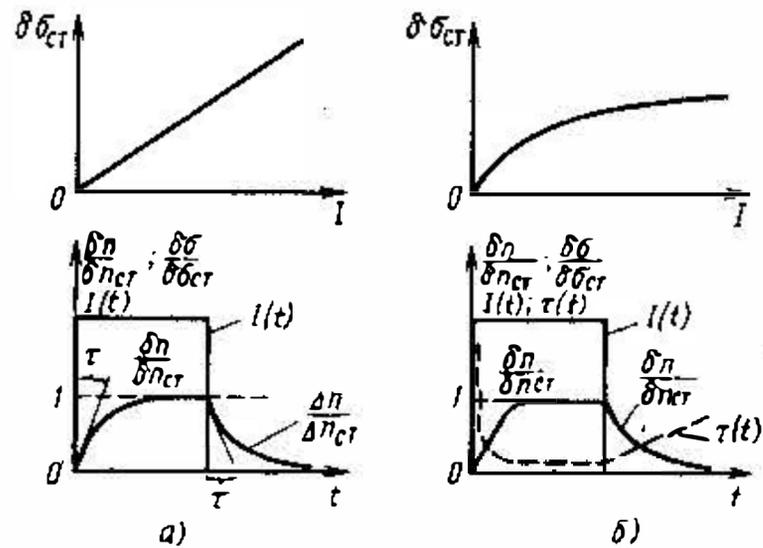


Рис. 4.47. Зависимости стационарной фотопроводимости $\delta\sigma_{ст}$ от интенсивности света I (локс-амперные характеристики) и кривые релаксации фотопроводимости при воздействии прямоугольных импульсов света:

а) в случае линейной рекомбинации; б) в случае квадратичной рекомбинации.

концентрации неравновесных носителей заряда, а их времена жизни не зависят от концентрации δn и δp , т. е. не зависят от интенсивности возбуждения.

При линейной рекомбинации уравнение непрерывности (4.90) для j_n имеет вид

$$\frac{d(\delta n)}{dt} = \beta k_{\omega} \frac{I}{h\omega} - \frac{\delta n}{\tau_n} \quad (4.100)$$

В стационарном режиме $\frac{d(\delta n)}{dt} = 0$, откуда получаем

$$\delta n_{ст} = \beta k_{\omega} \frac{I}{h\omega} \tau = G\tau. \quad (4.101)$$

Интегрирование (4.100) с начальными условиями $\delta n = 0$ при $t = 0$ дает закон нарастания фотопроводимости при быстром включении света:

$$\delta n = G\tau(1 - e^{-t/\tau}). \quad (4.102)$$

Закон спада фотопроводимости после прекращения возбуждения так же легко определяется из уравнения (4.100) с начальными условиями $G = 0$, $\delta n = \delta n_{ст}$ при $t = 0$ и имеет вид

$$\delta n = G\tau e^{-t/\tau}. \quad (4.103)$$

Таким образом, в случае линейной рекомбинации стационарная фотопроводимость (4.101) пропорциональна интенсивности света

в первой степени, а релаксация неравновесной концентрации и фотопроводимости происходит по экспоненциальному закону (4.102) или (4.103), причем «постоянная времени» экспонент совпадает с временем жизни τ (рис. 4.47, а).

Легируя полупроводник примесями, являющимися эффективными центрами захвата для неосновных носителей заряда, можно существенно увеличить эффективное время жизни неравновесных основных носителей и тем самым согласно (4.94) и (4.101) увеличить его fotocувствительность, однако за счет одновременного увеличения инерционности.

2. *Случай квадратичной рекомбинации.* Он реализуется в чистом виде в собственном полупроводнике, когда концентрация равновесных носителей заряда равна нулю и рекомбинация определяется только межзонными переходами. В этом случае интенсивность рекомбинации пропорциональна концентрации как неравновесных электронов, так и неравновесных дырок, т. е. квадрату концентрации неравновесных носителей заряда. Время жизни обратно пропорционально их концентрации, т. е.

$$\tau_n = \tau_p = \tau = \frac{1}{\gamma \delta n} = \frac{1}{\gamma \delta p}. \quad (4.104)$$

Тогда из (4.100) имеем

$$\frac{d(\delta n)}{dt} = \beta k_{\omega} \frac{I}{h\omega} - \gamma (\delta n)^2, \quad (4.105)$$

откуда для стационарного режима $\frac{d(\delta n)}{dt} = 0$ получаем

$$\delta n_{ст} = \sqrt{\frac{G}{\gamma}} = \sqrt{\frac{\beta k_{\omega} I}{\gamma h\omega}}. \quad (4.106)$$

Используя начальные условия, как и для линейной рекомбинации, из интегрирования уравнения (4.105) получаем для случая квадратичной рекомбинации закон нарастания

$$\delta n = \sqrt{\frac{G}{\gamma}} \operatorname{th}[t\sqrt{G\gamma}] \quad (4.107)$$

и закон спада

$$\delta n = \sqrt{\frac{G}{\gamma}} \frac{1}{t\sqrt{G\gamma} + 1} \quad (4.108)$$

неравновесной концентрации носителей заряда, а значит, и фотопроводимости.

Таким образом, в случае квадратичной рекомбинации стационарная фотопроводимость (4.106) пропорциональна корню ква-

дратному из интенсивности света, а законы нарастания и спада различны. Нарастание определяется гиперболической тангенсой (4.107), а спад — значительно более медленной при больших t гиперболой (4.108), как показано на рис. 4.47, б. В противоположность линейной рекомбинации здесь не может быть введено понятие времени жизни как постоянной величины, не зависящей от интенсивности и сохраняющей свое значение в течение всего релаксационного процесса.

Прочие фотоэлектрические эффекты в однородных кристаллах. Кроме фоторезистивного эффекта в однородных полупроводниках могут проявляться значительно более слабые фотоэффекты: фотодиффузионный, фотомагнитный и увлечение носителей заряда фотонами. Все эти эффекты в отличие от фотопроводимости приводят к возникновению фото-ЭДС в однородном до освещения материале, т. е. наблюдаются в отсутствие внешнего источника питания.

1. **Фотодиффузионный эффект (эффект Дембера)** связан с возникновением ЭДС в однородном полупроводнике при освещении его сильно поглощаемым светом. В области собственного поглощения при $k_{\omega}d \gg 1$ поглощение будет происходить в основном вблизи поверхности образца. Возникающий градиент концентрации вызовет появление диффузионных токов электронов и дырок в направлении, перпендикулярном освещаемой поверхности. В случае неравенства их подвижностей в образце возникает электрическое поле, направленное перпендикулярно освещаемой поверхности. Действительно, воспользовавшись соотношением Эйнштейна ($\mu = \frac{e}{kT} D$) и условием $\delta n = \delta p$, для x -компоненты тока из (4.91) получим

$$j_x = j_{nx} + j_{px} = e(bn + p) \mu_p E_x + (b-1)kT \mu_p \partial(\delta p)/\partial x, \quad (4.109)$$

где $b = \mu_n/\mu_p$ — отношение подвижностей носителей.

Если учесть граничное условие для «разомкнутой» цепи $j_x = 0$, то из выражения (4.109) получим, что в образце возникает электрическое поле, напряженность которого

$$E_x = -\frac{kT}{e} \left(\frac{b-1}{bn+p} \right) \frac{\partial(\delta p)}{\partial x}$$

отлична от нуля при $b = \mu_n/\mu_p \neq 1$.

2. **Фотогальваномагнитный эффект (эффект Кикоина — Носкова)** связан с возникновением ЭДС при освещении сильно поглощаемым светом полупроводника, находящегося в магнитном поле. Если параллельно освещаемой поверхности образца, например в направлении y на рис. 4.45, приложить магнитное поле, то на диффундирующие в направлении z от освещаемой поверхности электроны и дырки будет действовать сила Лоренца. Следовательно, наряду с полем вдоль оси z возникнет электрическое поле, направленное перпендикулярно магнитному полю и освещаемой поверхности образца, т. е. вдоль оси x . Этот эффект довольно часто используют для определения малых времен жизни носителей заряда.

3. **Эффект увлечения носителей заряда фотонами** обусловлен передачей импульса фотонов свободным электронам или дыркам в полупроводнике. Ранее в большинстве случаев импульсом фотона было пренебрежено. Хотя его значение $p = \hbar k$ действительно мало, но всё же оно принципиально отлично от нуля. Поэтому, взаимодействуя со свободными носителями заряда, фотон передает им свой импульс, приводя к переносу носителей в направлении распространения

света. В результате между контактами на стороне образца, через которую производится освещение, и на противоположной стороне возникает разность потенциалов, которая называется ЭДС фотонного увлечения. При $\omega \tau_p \gg 1$, где τ_p — время релаксации импульса, для образца толщиной d с показателем поглощения k_{ω} подвижностью свободных носителей заряда μ разность потенциалов U может быть оценена с помощью соотношения

$$U = \frac{\mu \mu_0}{(\omega \tau_p)^2 k_{\omega}} \frac{1 - R}{1 + R} \frac{[1 - \exp(-k_{\omega} d)]}{\exp(-k_{\omega} d)} \quad (4.110)$$

Для интенсивности падающей световой волны $I \approx 10^5$ Вт/см², получаемой от СО₂-лазера, на длине волны 10,6 мкм при $\mu \approx 10^3$ см²/(В·с), $k_{\omega} = 1$ см⁻¹, $R = 0,3$ ЭДС фотонного увлечения $U \approx 1$ мВ. Хотя это значение довольно мало для такой большой интенсивности света, эффект фотонного увлечения обладает двумя существенными положительными факторами. Во-первых, он малоинерционен, поскольку время релаксации импульса порядка 10^{-12} с. Во-вторых, взаимодействие со свободными носителями заряда осуществляется в очень широком спектральном диапазоне, включая дальнюю инфракрасную область. Поэтому этот эффект может быть использован в инфракрасной и дж-субмиллиметровой областях спектра. Важно, что такие приемники могут работать при комнатной температуре.

§ 4.8. КВАНТОВОРАЗМЕРНЫЕ СТРУКТУРЫ: КВАНТОВЫЕ ЯМЫ, НИТИ И ТОЧКИ

При рассмотрении оптических явлений в полупроводниках предполагалось, что размеры объекта (кристалла, эпитаксиальной пленки и т. п.) по всем трем координатам x , y , z настолько велики, что они не влияют на свойства этого объекта. В большинстве случаев рассматривались изотропные материалы, свойства которых одинаковы по всем направлениям. Возможная анизотропия свойств, проявляемая, например, в анизотропных кристаллах (§ 3.5) или при приложении внешних воздействий к изотропным средам учитывалась введением координатно-зависимых макроскопических параметров среды — тензора диэлектрической проницаемости, тензора эффективной массы и т. п.

Объекты, фундаментальные физические свойства которых не зависят от их размеров в трех измерениях, называются трехмерными или 3D*.

В современной электронике, особенно в опто- и наноэлектронике особая роль принадлежит так называемым *системам низкой размерности*, где движение ограничено по одной, двум или трем координатам. Эти объекты, соответственно, имеют размерность 2, 1 и 0 и обозначаются 2D-, 1D- и 0D.

Напомним, что под *размерными эффектами* понимают зависимость физических свойств объекта или вещества от его геометрических размеров. Сами эффекты определяются характерной физической длиной, с которой сравниваются размеры образца.

*От англ. Dimension — измерение, размерность

Если роль этой характерной физической длины выполняют классические величины, например, длина свободного пробега или диффузионная длина, то такие размерные эффекты называются *классическими*. Один из таких эффектов рассмотрев в § 3.4 свойства планарного световода зависят от соотношения его толщины и длины волны оптического излучения.

Если роль характерной физической длины выполняют квантовые величины, то такие эффекты называются *квантоворазмерными*. Для электронов в твердом теле эту роль, как правило, выполняет де-Бройлевская длина волны электрона

$$\lambda_e = \hbar/p_e \quad (4.111)$$

где $p_e = \hbar k$ — квазимпульс электрона, а $k = 2\pi/\lambda_e$ — его волновой вектор. Если характерный геометрический размер соизмерим или меньше λ_e , то проявляются *эффекты размерного квантования* — электронные спектры квантуются и положения каждого из уровней квантования зависят не только от свойств материала, но и от его геометрических размеров. Это означает, что положением уровней энергии и волновыми функциями электрона можно управлять, меняя геометрические размеры объекта, в котором движется электрон. Структуры, в которых движение электрона ограничено по одной, двум и трем координатам, имеют размерности 2, 1 и 0 (2D-, 1D- и 0D) и называются *квантовыми ямами* (КЯ) или квантовыми колодцами, *квантовыми нитями* (КН) или квантовыми проволоками и *квантовыми точками* (КТ) соответственно.

Электрон в одномерной потенциальной яме. Удобным модельным объектом, позволяющим понять свойства квантоворазмерных систем, является прямоугольная потенциальная яма. Задачи о движении электрона в потенциальных ямах различной конфигурации решены в стандартных курсах квантовой механики. В простейшей модели — прямоугольной потенциальной ямы бесконечной глубины или, что то же самое, потенциального ящика со стенками бесконечной высоты, задача имеет простые аналитические решения. Энергетический спектр электрона и одноэлектронные волновые функции для потенциала

$$\begin{aligned} V(z) &= 0 && \text{при} && 0 < z < d \\ V(z) &\rightarrow \infty && \text{при} && 0 > z > d \end{aligned} \quad (4.112)$$

имеют вид

$$E_n = \frac{\pi^2 \hbar^2}{2md^2} n^2, \quad (4.113)$$

$$\psi_n = \sqrt{2/d} \sin(\pi n z/d), \quad (4.114)$$

где квантовое число $n = 1, 2, 3, \dots$ определяет собственные значения (уровни размерного квантования E_1, E_2, E_3, \dots) в прямоугольной яме шириной d . Множитель $\sqrt{2/d}$ является нормировочным.

Часто бывает удобно начало отсчета выбрать не на границе ямы, а в ее центре. Тогда вместо (4.112) имеем

$$\begin{aligned} V(z) &= 0 && \text{при} && -d/2 < z < d/2 \\ V(z) &\rightarrow \infty && \text{при} && -d/2 > z > d/2 \end{aligned} \quad (4.115)$$

Решения для E_n будут иметь тот же вид (4.113), а волновые функции переписутся:

$$\begin{aligned} \psi_n &= \sqrt{2/d} \cos(\pi n z/d) && \text{для } n = 1, 3, 5, \dots \\ \psi_n &= \sqrt{2/d} \sin(\pi n z/d) && \text{для } n = 2, 4, 6, \dots \end{aligned} \quad (4.116)$$

При уменьшении потенциального барьера V (глубины ямы) появляется возможность просачивания электрона за барьер — хорошо известный эффект туннелирования. Для ямы с конечной высотой барьера V_0 и шириной d собственные значения E_n находятся из решения трансцендентного уравнения

$$E_n = \frac{\pi^2 \hbar^2}{2md^2} \left(n - \operatorname{arctg} \sqrt{\frac{V_0}{E_n} - 1} \right)^2, \quad (4.117)$$

где опять-таки $n = 1, 2, 3, \dots$ является хорошим квантовым числом, нумеруя уровни энергии. Собственные волновые функции, как и (4.116) распадаются на четные и нечетные относительно центра ямы, но в отличие от (4.114), (4.116) их амплитуды на границах ямы отличны от «0» и они экспоненциально затухают по мере удаления от барьера. Амплитуда на границе как и постоянная затухания зависят от параметров ямы V_0 и d и от номера уровня n , определяя вероятность туннелирования через барьер. Характер изменения собственных значений E_n и собственных волновых функций ψ_n показаны на рис. 4.48.

Заметный интерес представляют не только одиночные структуры типа квантовых ям, но и их набор. Если ямы отделены друг от друга достаточно широким потенциальным барьером, так что электронные волновые функции не проникают из одной ямы в другую, то их взаимодействием можно пренебречь и свойства структуры, состоящей, например, из набора N одинаковых ям, будут аддитивно складываться. Такие структуры удобно использовать, например, для увеличения поглощательной способности или числа активных центров в N раз.

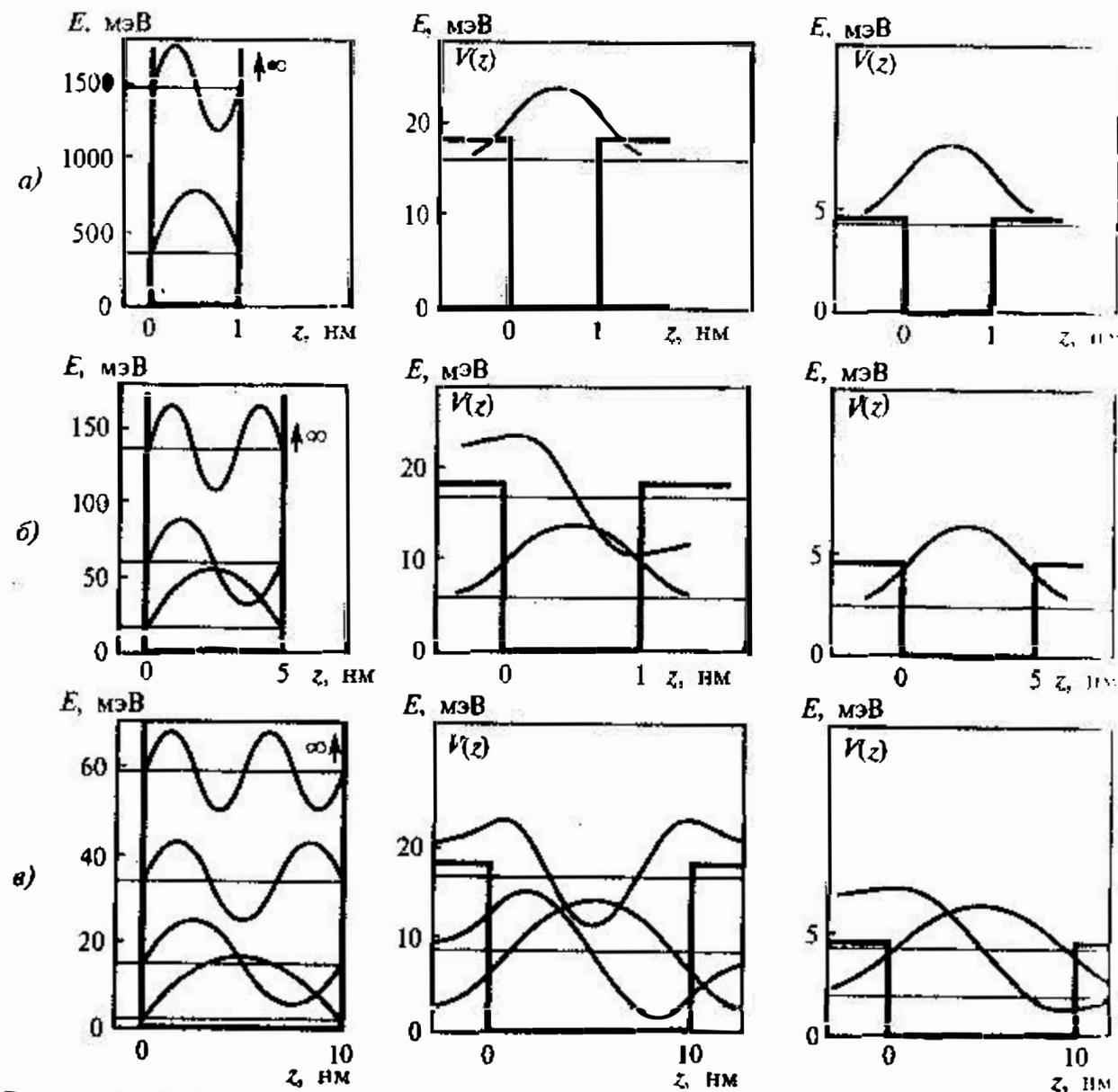


Рис. 4.48. Собственные значения энергии E_n и собственные волновые функции Ψ_n для свободного электрона в прямоугольном потенциальном ящике шириной $d = 1$ нм (а), 5 нм (б), 10 нм (в) бесконечной и конечной глубины

Если ширина потенциального барьера и его высота позволяют электронам проникать (туннелировать) из одной ямы в другую с достаточной эффективностью, то такая структура будет вести себя как единое целое и ее необходимо рассматривать как систему взаимодействующих ям. Простейшая структура состоит из двух одинаковых прямоугольных ям конечной глубины $V_1 = V_2 = V$, разделенных барьером шириной b и высотой U_0 , как показано на рис. 4.49, а.

Такая структура может быть наиболее просто рассмотрена при применении обычной модели теории возмущений.

Пусть ширина барьера b настолько велика, что взаимодействием между ямами можно пренебречь. При сближении ям и уменьшении толщины барьера b экспоненциальные хвосты волновых функций проникают из одной ямы в другую и про-

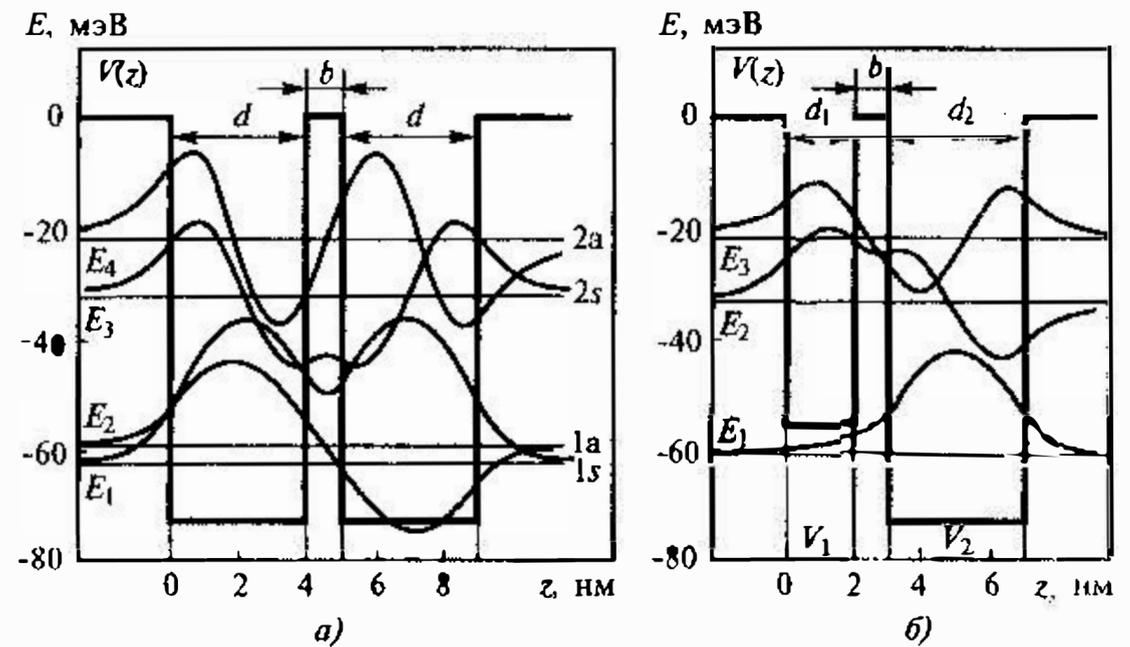


Рис. 4.49. Электронные состояния и волновые функции двух связанных одинаковых (а) и произвольных (б) прямоугольных квантовых ям

исходит их перемешивание. Это взаимодействие рассматривается как малое возмущение. Оно определяет матричный элемент возмущения, который в конфигурации двух ям записывается аналогично (1.77) как

$$V_{12} = (\psi_1 | H | \psi_2) = (\psi_1 | V | \psi_2) = \int \psi_1^* V \psi_2 dz, \quad (4.118)$$

где ψ_1 и ψ_2 — невырожденные невзаимодействующие состояния отдельных ям 1 и 2, а V — связывающий потенциал. В случае одинаковых ям ($\psi_1 = \psi_2$) это взаимодействие приводит к расщеплению энергетического уровня, например E_1 , на два состояния:

$$E_1^{(a,a)} = E_1^{(0)} \pm V_{12}. \quad (4.119)$$

Нижнему состоянию (знак «-») соответствует симметричная (относительно центра структуры) волновая функция (см. рис. 4.49) и это состояние называется *симметричным*, а верхнему (знак «+») — соответствует антисимметричная волновая функция и это состояние называется *антисимметричным*.

Таким образом, в двойной квантовой яме, состоящей из двух одинаковых ям, разделенных барьером b , происходит снятие вырождения по энергии путем расщепления уровней $n = 1, 2, 3 \dots$ на симметричные $1s, 2s, 3s \dots$ и антисимметричные $1a, 2a, 3a \dots$ состояния. Величина расщепления тем сильнее, чем уже барьер и чем ближе уровень расположен к континууму. Это связано с увеличением перекрытия волновых функций как при уменьшении b , так и при увеличении n .

Для двух связанных ям с разными параметрами также будет наблюдаться взаимодействие между разными состояниями и перетекание из одной ямы в другую (рис. 4.49, б). Наиболее сильно это взаимодействие будет проявляться для резонансных состояний, близких по энергиям. Состояния, разделенные большим энергетическим интервалом, можно условно приписать одной или другой яме, хотя при малой ширине барьера b взаимодействие может быть настолько сильным, что применение теории возмущений становится недопустимым и всю систему необходимо рассматривать как единое целое.

При увеличении числа взаимодействующих ям, например, для N одинаковых ям, каждый из уровней будет расщепляться на N состояний. При небольшом значении N каждое из этих состояний возможно различить, но при $N > 7$ они будут перекрываться, образуя полосы разрешенных значений энергии. При больших N это перекрытие и перемешивание состояний будет настолько сильным, что отдельные дискретные уровни неразличимы и можно говорить об образовании зон разрешенных значений энергии, называемых *минизонами*, разделяемых областями запрещенных значений (запрещенными зонами), как это показано на рис. 4.50. Такая структура ведет себя как единое целое и называется *сверхрешеткой* (СР). Понятно, что свойствами такой одномерной сверхрешетки, образованной набором квантовых ям с шириной a , разделенных потенциальными барьерами высотой V и шириной b , можно управлять, меняя как параметры самой ямы, так и период сверхрешетки $A = a + b$. Электрон в такой структуре будет вести себя как в одномерном кристалле с периодом A .

Внешнее электрическое поле F^z , направленное перпендикулярно стенкам ямы, изменяет ее форму и модифицирует как электронный спектр, так и волновые функции. Для одиночной прямоугольной ямы это показано на рис. 4.51.

Для нахождения собственных значений и собственных функций необходимо, как обычно, решить уравнение Шредингера в котором наряду с потенциальной энергией $V(z)$, определяемой формой ямы, добавляется слагаемое $(-qFz)$, учитывающее воздействие поля $F = F_z$. Для одномерной задачи это уравнение запишется в виде

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + V(z) - qFz \right) \psi(E, z) = E\psi(E, z). \quad (4.120)$$

*В настоящем параграфе напряженность внешнего электрического поля, направленного по оси z , будем обозначать буквой $F = E_z$, чтобы не путать ее с энергией E .

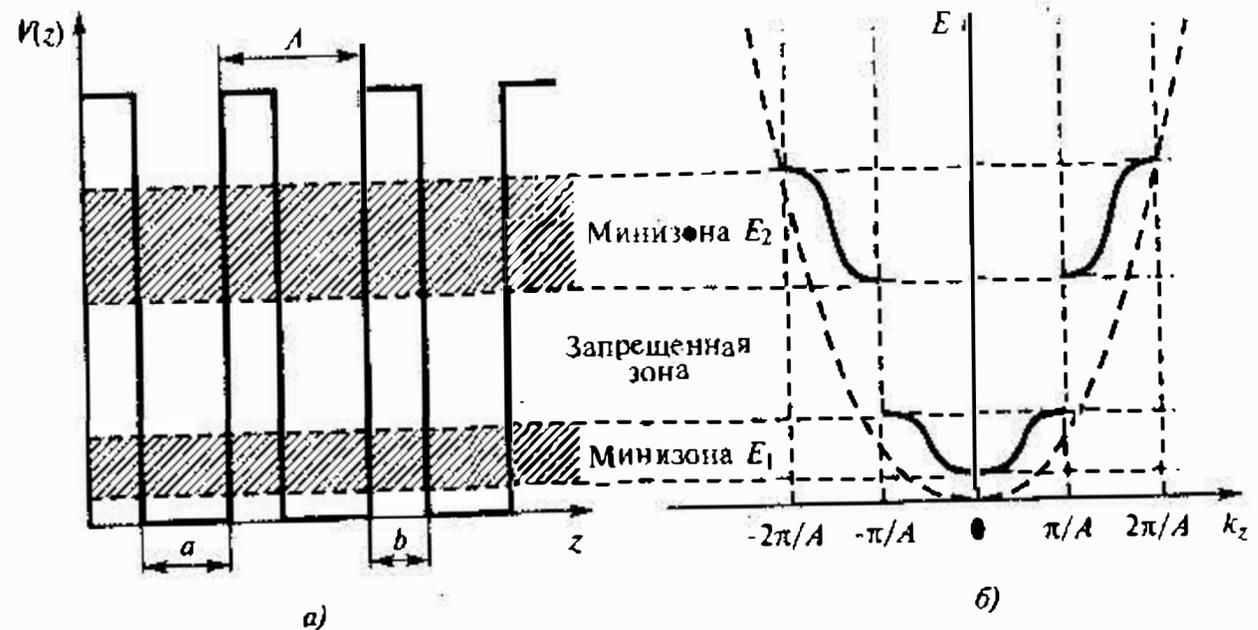


Рис. 4.50. Образование минизон разрешенных значений энергии в одномерной сверхрешетке

Это уравнение решено для ям различной формы $V(z)$, в том числе и для прямоугольной ямы с соответствующими граничными условиями. Результаты решения дают следующее.

1. Энергетический спектр квантовой ямы становится непрерывным во всей области определения. Это означает, что вместо дискретных состояний, нумеруемых квантовым числом n и имеющих вид δ -функций с бесконечно узкой шириной, ста-

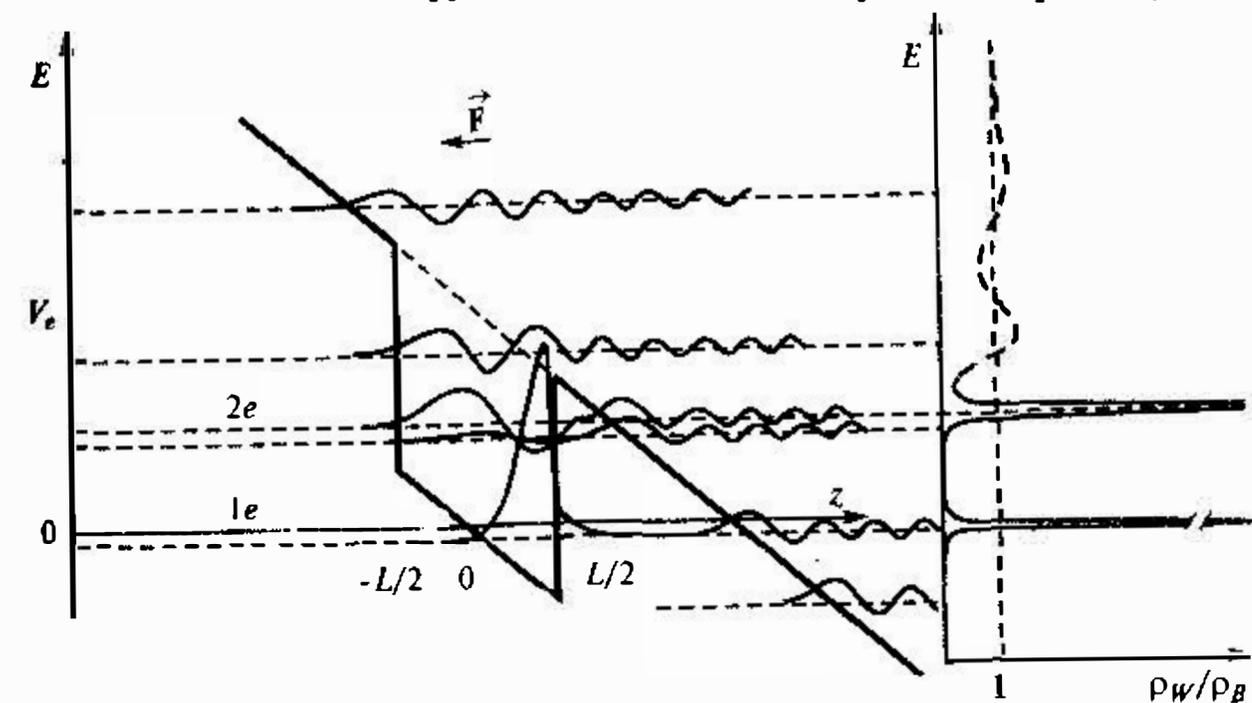


Рис. 4.51. Прямоугольная яма в электрическом поле. Слева показаны распределение потенциала и волновые функции резонансных состояний; справа — распределение по энергии отношения вероятности нахождения электрона в единичном интервале энергий на единицу длины в слое квантовой ямы ρ_w и вне ее ρ_B .

новятся формально разрешенными состояниями с любой энергией. В электронном спектре наблюдаются резонансные пики, называемые резонансами Брейта — Вигнера (см. рис. 4.51). По существу это и есть уровни энергии. При $F \rightarrow 0$ они переходят в дискретные уровни E_n с $\delta E \rightarrow 0$. Контур этих резонансов совпадает с Лоренцевым контуром (1.135), (1.148) и может быть переписан в виде

$$L_n(E) = |I_n(E)|^2 = \frac{C_n}{1 + \left(\frac{E - E_n}{\Gamma_n}\right)^2} \quad (4.121)$$

Здесь C_n — нормировочный коэффициент, n — номер квазисвязанного состояния (номер резонанса E_n), которому можно сопоставить квантовое число n , нумерующее уровни E_n . Ширина резонанса Γ_n определяется мнимой, а энергетическое положение резонанса (резонансная частота $\omega_n = \Gamma_n/\hbar$) — действительной частью соответствующего собственного значения.

Обратим внимание, что размытие электронного спектра и уширение уровней квантования под действием электрического поля вызваны возможностью туннелирования электрона через треугольный потенциальный барьер (см. рис. 4.51). Конечность времени жизни δt приводит к неопределенности в энергии $\delta E = \hbar/\delta t$ и размытию спектральной линии аналогично тому, что это обсуждалось в § 1.6. Чем глубже расположен уровень в яме и чем меньше электрическое поле F , тем шире барьер и тем меньше уширение δE . Примечательно, что контур резонансов (4.121) совпадает с контуром «естественного» уширения спектральной линии (1.148).

2. Положения резонансов E_n зависят от приложенного электрического поля F и при $F \rightarrow 0$ они совпадают с уровнями размерного квантования в яме соответствующей формы $V(z)$. Для оценки положения резонансов (уровней размерного квантования) в прямоугольной яме, находящейся в электрическом поле, можно использовать простую формулу, полученную по теории возмущений для ямы бесконечной глубины:

$$E_n(F) = E(0) - \frac{1}{48} \frac{(qFd)^2}{E_1^{\infty} n^2} \left(1 - \frac{15}{\pi^2 n^2}\right) = E_n(0) + CF^2. \quad (4.122)$$

Здесь d — ширина ямы, E_1^{∞} — энергия основного ($n=1$) уровня в яме бесконечной глубины, а $E_n(0)$ — положения уровней размерного квантования в яме конечной глубины в отсутствие поля

($F=0$). Значения E_1^{∞} и $E_n(0)$ находятся из (4.113) и (4.117). Квантовое число n , как и ранее, может принимать значения 1, 2, 3... Из (4.122) видно, что при $n > \sqrt{15/\pi}$, т. е. для $n \geq 2$ происходит квадратичный сдвиг в область больших энергий ($C > 0$), а при $n=1$ — в область меньших энергий ($C < 0$). Напомним, что формула (4.122) получена с помощью теории возмущений, а потому она справедлива лишь в области слабых электрических полей.

3. Решения уравнения (4.120) для волновых функций записываются через линейную комбинацию функций Эйри первого $A_i(\eta)$ и второго $B_i(\eta)$ рода:

$$\psi(E, z) = a_1 A_i(\eta) + a_2 B_i(\eta), \quad (4.123)$$

где

$$\eta = \left(z + \frac{E - V(z)}{E_z}\right) \left(\frac{2mqF}{\hbar^2}\right)^{1/2}$$

принимает различное значение на участках с различным значением кусочно-постоянного потенциала $V(z)$. Коэффициенты a_1 и a_2 на этих участках определяются из граничных условий и условия нормировки для непрерывного спектра. Тот факт, что ψ зависит не только от z , но и от энергии, как раз отражает непрерывность энергетического спектра. Примеры волновых функций (4.123) для некоторых состояний в прямоугольной яме приведены на рис. 4.51, а.

4. Под действием электрического поля происходит смещение волновых функций и смещение центра тяжести электронного облака — своего рода электронная поляризация. Это отчетливо видно, например, для состояния $n=1$ на рис. 4.51, а: под действием электрического поля центр тяжести электронного облака смещается как по энергии (сдвиг уровня — Штарковский сдвиг), так и по координате. В симметричной яме электрическое поле снимает вырождение относительно направления движения.

Возможность в определенных пределах управлять электронным облаком в квантовой яме с помощью внешнего электрического поля открывает пути создания уникальных приборов опто- и наноэлектроники.

5. В континууме состояний над квантовой ямой [$E > (V - qFd/2)$] появляется дополнительная серия резонансов (см. рис. 4.51, а), положение которых близко к положению уровней в треугольной яме с бесконечной стенкой:

$$E_n = -qF \frac{d}{2} + \left[\frac{3\pi}{4} \left(2n - \frac{1}{2}\right) \sqrt{\frac{qF\hbar}{2m}} \right]^{2/3}. \quad (4.124)$$

6. В двойной квантовой яме (рис. 4.52) поле нарушает симметрию структуры. Волновые функции перекрываются между ямами так, что центры тяжести симметричных и антисимметричных состояний оказываются смещенными в разные стороны.

Сдвиг уровней складывается из сравнительно слабого изменения их положения относительно дна той ямы, в которой преимущественно сконцентрирована соответствующая волновая функция и изменения энергии дна ямы. Последний механизм в большинстве случаев преобладает. При некоторых величинах электрического поля создается ситуация, когда энергетические уровни соседних ям вновь будут иметь одинаковую энергию. Их значения соответствуют процессу «резонансного туннелирования». Этот процесс аналогичен резонансной передаче возбуждения при столкновении атомов и используется для селективного заселения рабочих лазерных уровней в газовых лазерах и в каскадных полупроводниковых лазерах (см. §§ 7.2, 7.4 и 9.10). В качестве примера на рис. 4.53 представлены зависимости от электрического поля уровней размерного квантования (резонансных энергий) в системе двух квантовых ям, разделенных барьером шириной b , составляющей 0,2 от ширины ямы. Эти зависимости построены в безразмерных единицах: энергия — в единицах первого уровня ($n=1$) квантования для ямы бесконечной глубины (E_1^∞) в отсутствие электрического поля ($F=0$), а само поле — в единицах (E_1/qL). Видно, что при некоторых значениях электрического поля, показанных стрелками на рис. 4.53, происходит перемешивание разных состояний из разных ям и перераспределение этих состояний. Возникающее взаимодействие уровней снимает вырождение. Происходит перераспределение волновых функций взаимодействующих уровней так, что они как бы «меняются ямами». С ростом поля увеличивается ширина резонансных пиков и они начинают взаимодействовать на более дальних расстояниях. Процесс перемешивания по-

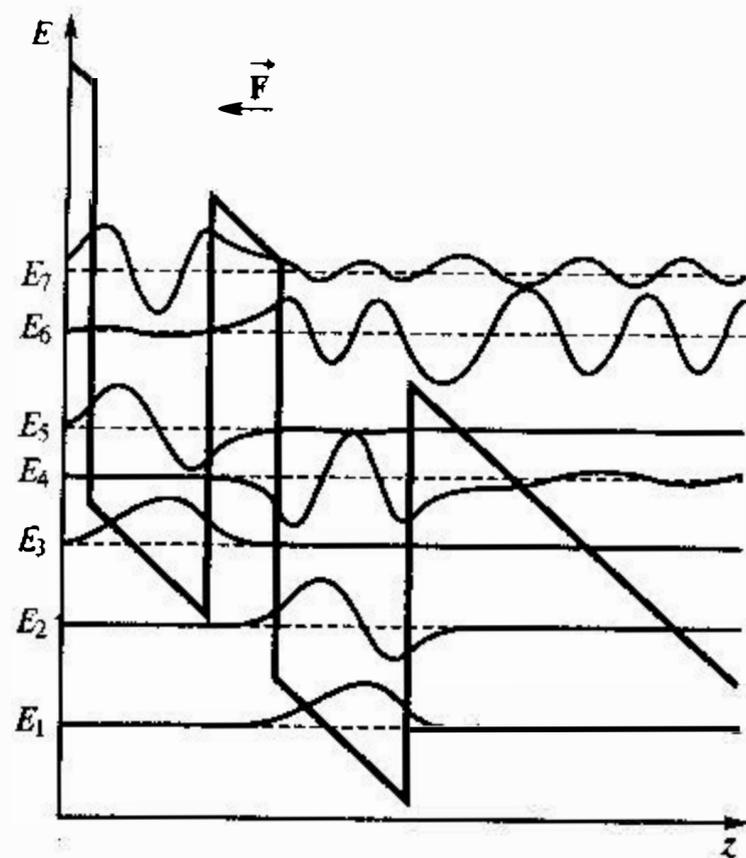


Рис. 4.52. Двойная квантовая яма в электрическом поле

казан на рис. 4.54, где для той же структуры при значении поля $F=4,5 (E_1^\infty/qL)$, соответствующей процессу резонансного туннелирования, приведены профиль потенциальной энергии, волновые функции и энергетический спектр. Вблизи каждого из резонансных уровней приведены величины их уширения, измеренные в единицах E_1^∞ . Напомним, что время жизни τ связано с величиной однородного уширения δE простым соотношением (1.136).

Структуры с двумерным электронным газом. Выше было изложено поведение электрона в одномерной потенциальной яме в системе ям. Эти вопросы представляли чисто теоретический интерес до того момента, пока не были созданы реальные структуры, в которых наблюдались эффекты размерного квантования. Казалось бы, простейшей такой структурой могла бы быть тонкая металлическая пленка. Сам факт, что электроны не выходят из металла, говорит, что они находятся в потенциальном ящике, высота барьеров которого соответствует работе выхода электрона ϕ и имеет порядок величины 4...5 эВ, т. е. весьма велика. Однако в металлах электронный газ сильно вырожден и де-Бройлевская длина волны (4.111) определяется волновым вектором k_F на поверхности Ферми: $\lambda_e = \lambda_F = 2\pi/k_F$. В вырожденном электронном газе она зависит от концентрации свободных носителей заряда:

$$\lambda_F = (8\pi/3n)^{1/3} \text{ для 3D-систем,}$$

$$\lambda_F = (2\pi/n)^{1/2} \text{ для 2D-систем,}$$

$$\lambda_F = (4/n) \text{ для 1D-систем.}$$

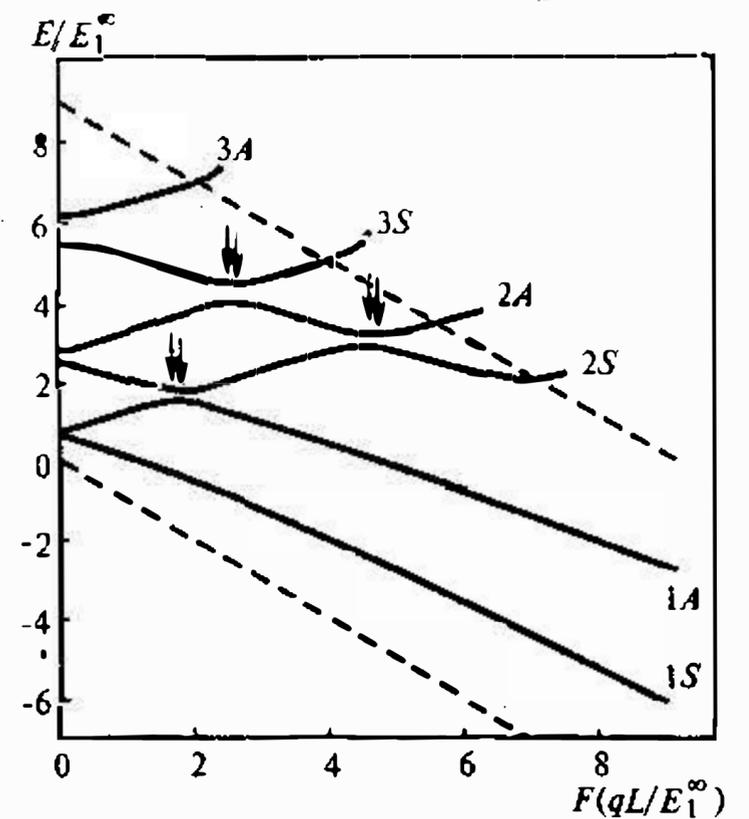


Рис. 4.53. Зависимость от электрического поля уровней размерного квантования в двойной квантовой яме $Al_{0,3}Ga_{0,7}As/GaAs$ ($V/E_1^\infty=9$) с шириной барьера $b=L_B=2$ нм и шириной ямы $a=L=10$ нм. Пунктирными линиями отмечено положение правого треугольного барьера. Стрелки показывают поля, при которых происходит резонансное туннелирование

казалось бы, простейшей такой структурой могла бы быть тонкая металлическая пленка. Сам факт, что электроны не выходят из металла, говорит, что они находятся в потенциальном ящике, высота барьеров которого соответствует работе выхода электрона ϕ и имеет порядок величины 4...5 эВ, т. е. весьма велика. Однако в металлах электронный газ сильно вырожден и де-Бройлевская длина волны (4.111) определяется волновым вектором k_F на поверхности Ферми: $\lambda_e = \lambda_F = 2\pi/k_F$. В вырожденном электронном газе она зависит от концентрации свободных носителей заряда:

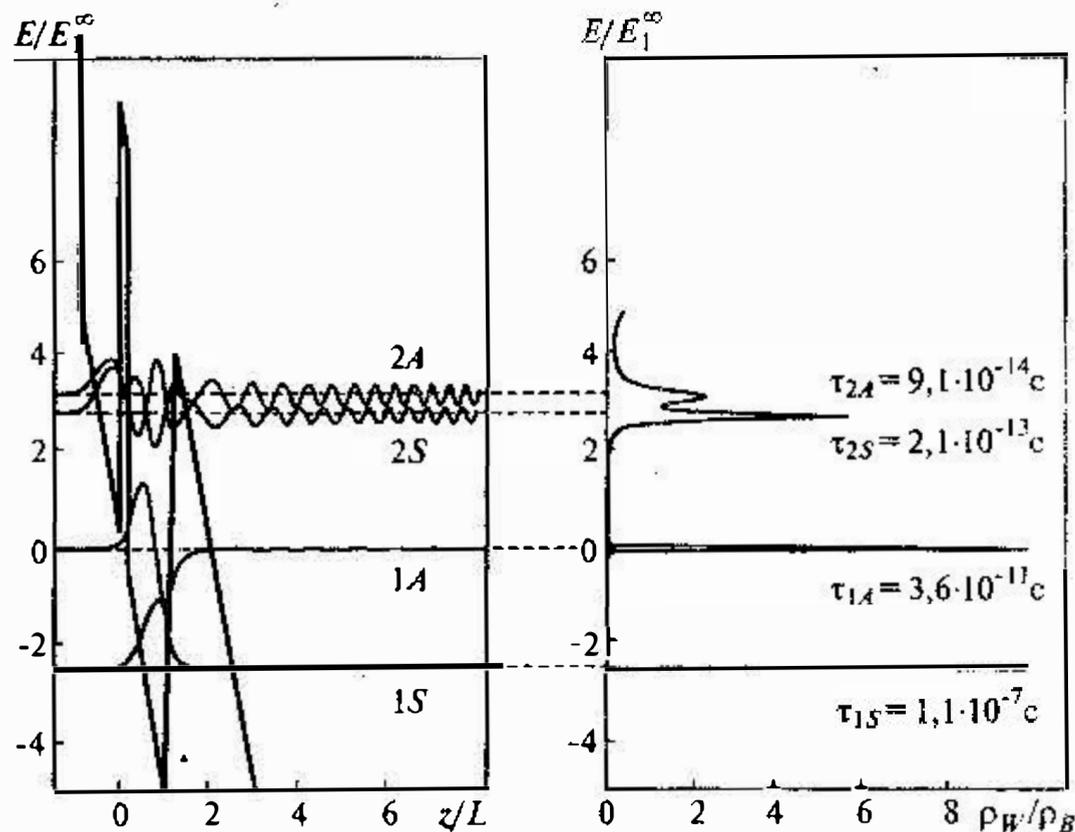


Рис. 4.54. Профиль потенциальной энергии, энергетический спектр и электронные волновые функции, соответствующие процессу резонансного туннелирования в двойной квантовой яме с параметрами:

$$V/E_1^\infty = 9; b = L_B = 0,2L = 0,2d \text{ в электрическом поле } F = 4,5 \left(\frac{E_1^\infty}{qL} \right)$$

Подставив типичные значения n , нетрудно убедиться, что в металлах де-Бройлевская длина волны свободных электронов имеет порядок величины межатомных расстояний, энергетические уровни размыты и эффекты размерного квантования в реальных структурах на основе металлических пленок не проявляются.

Обратим внимание на сходство процессов образования собственных типов колебаний в оптических резонаторах (§ 2.3), оптических волноводах (§ 3.4) и в квантовых ямах (сравните рис. 2.20 и рис. 3.9 с рис. 4.48). Это сходство носит фундаментальный характер, поскольку и в том и в другом случае квантование энергии (собственных частот) определяется процессами интерференции и образованием устойчивых состояний — стоячих электромагнитных или электронных волн. В квантовых ямах роль «зеркал резонатора» для электронных волн выполняют потенциальные барьеры, высота которых задает коэффициент отражения так, что при $V \rightarrow \infty$ величина $R \rightarrow 1$. По тем же причинам, что были рассмотрены в § 2.3, требования к «отражающим поверхностям» в квантоворазмерных структурах весьма высоки. Учитывая, что электронная длина волны λ_e имеет порядок единиц и десятков

нанометров, качество границы раздела, формирующей потенциальный барьер, должно быть исключительно высоким, а ее однородность по площади выдерживаться с точностью $0,1\lambda_e$, т. е. с точностью межатомных расстояний.

Полупроводники представляются наилучшими объектами для наблюдения квантоворазмерных эффектов по трем причинам. Во-первых, де-Бройлевская длина волны электрона λ_e велика для невырожденного электронного (или дырочного) газа определяется его средней энергией, соответствующей максимуму функции распределения $n(E)$ или $p(E)$ и равной $kT/2$ (см. рис. 4.35).

Этой энергии отвечает $\lambda_e = 2\hbar v / \sqrt{mkT}$, имеющая порядок величины десятки нанометров при комнатной температуре, что вполне достаточно для наблюдения интерференционных эффектов. Во-вторых, в полупроводниках имеется уникальная возможность управлять типом и концентрацией носителей заряда путем легирования донорными или (и) акцепторными примесями, а также путем инжекции или экстракции носителей через гомо- или гетеропереходы и барьеры, путем возбуждения светом, электронным пучком и т. п. При этом диапазон изменения концентрации фантастически велик и может достигать 15 порядков и больше.

И, в-третьих, современная технология полупроводниковых материалов и структур развита настолько хорошо, что позволяет воспроизводимо выращивать структуры с точностью до одного молекулярного слоя, повторяя эту процедуру желаемое количество раз. Как правило для этих целей используются эпитаксиальные методы выращивания: молекулярно-пучковая эпитаксия (МПЭ), газофазная эпитаксия из металлоорганических соединений (МОГФЭ) и их комбинации. Важно также, что наряду с развитием высокотехнологических методов, прекрасно развиты методы диагностики полупроводниковых наноструктур, позволяющие контролировать их параметры с точностью опять-таки в один молекулярный (атомный) слой. К ним относятся прежде всего методы электронной, атомно-силовой, рентгеновской и оптической спектроскопии. Все это вместе взятое делает полупроводниковые структуры основными объектами для наблюдения квантоворазмерных эффектов и их использования в современной опто- и наноэлектронике.

В качестве таких структур, где возможно формирование двумерного электронного газа, могут выступать структуры металл — диэлектрик — полупроводник (МДП) или гетероструктуры. В первом случае одной стенкой, формирующей потенциальную яму, служит граница с диэлектриком, а роль второй стенки выполняет электростатический потенциал eFz , возникающий в инверсном слое МДП-структуры при приложении обратного напряжения U и прижимающий электроны к границе с диэлектри-

ком, как показано на энергетической диаграмме (рис. 4.55).
Напряженность поля в инверсном слое

$$F \approx \frac{4\pi e n_s}{\epsilon \epsilon_0}, \quad (4.124)$$

где ϵ — диэлектрическая проницаемость полупроводника, n_s — двумерная концентрация [см⁻²] электронов в слое. В такой треугольной потенциальной яме уровни энергии задаются приближенным соотношением, аналогичным (4.124):

$$E_n = \left[\frac{\hbar^2}{2m_n} \right]^{1/3} \left[\frac{2\pi e F}{2} (n - 1/4) \right]^{2/3} \quad (4.125)$$

и справедливым для треугольной ямы бесконечной глубины. Квантовое число n , как и для прямоугольной ямы, может принимать значения $n = 1, 2, 3, \dots$

Отметим, что в соотношения (4.125) и (4.126) входят диэлектрическая проницаемость и эффективная масса электронов, учитывающие влияние среды на движение свободных носителей заряда.

Из рис. 4.55, а видно, что в инверсном слое МДП-структуры концентрации электронов n_s и параметрами потенциальной ямы, а следовательно — параметрами уровней размерного квантования, можно управлять, изменяя приложенное напряжение U на затворе. Характерная ширина инверсного канала Δz может быть оценена как ширина классической разрешенной области для электронов с энергией E_n

$$\Delta z = \frac{E_n}{eF}. \quad (4.126)$$

Типичные значения n_s для МДП-структур, сформированных на поверхности (100) кремния и SiO₂-диэлектрика, составляют 10¹²...10¹³ см⁻², а расстояния между нижними уровнями размерного квантования — десятки мэВ, т. е. больше кТ при комнатной температуре. Из соотношений (4.125), (4.126) и (4.127) видно, что эффективная ширина ямы $\Delta z \sim F^{-1/3}$ т. е. убывает с ростом поля.

Потенциальная яма, близкая по форме к треугольной, может быть реализована вблизи гетерограницы (интерфейса), образованной широкозонным и узкозонным полупроводником за счет разрывов зоны проводимости ΔE_c для электронов и ΔE_v для дырок (подробнее см. § 9.1). Для резкого изотипного N-p-гетероперехода это показано на рис. 4.55, б. Один потенциальный барьер

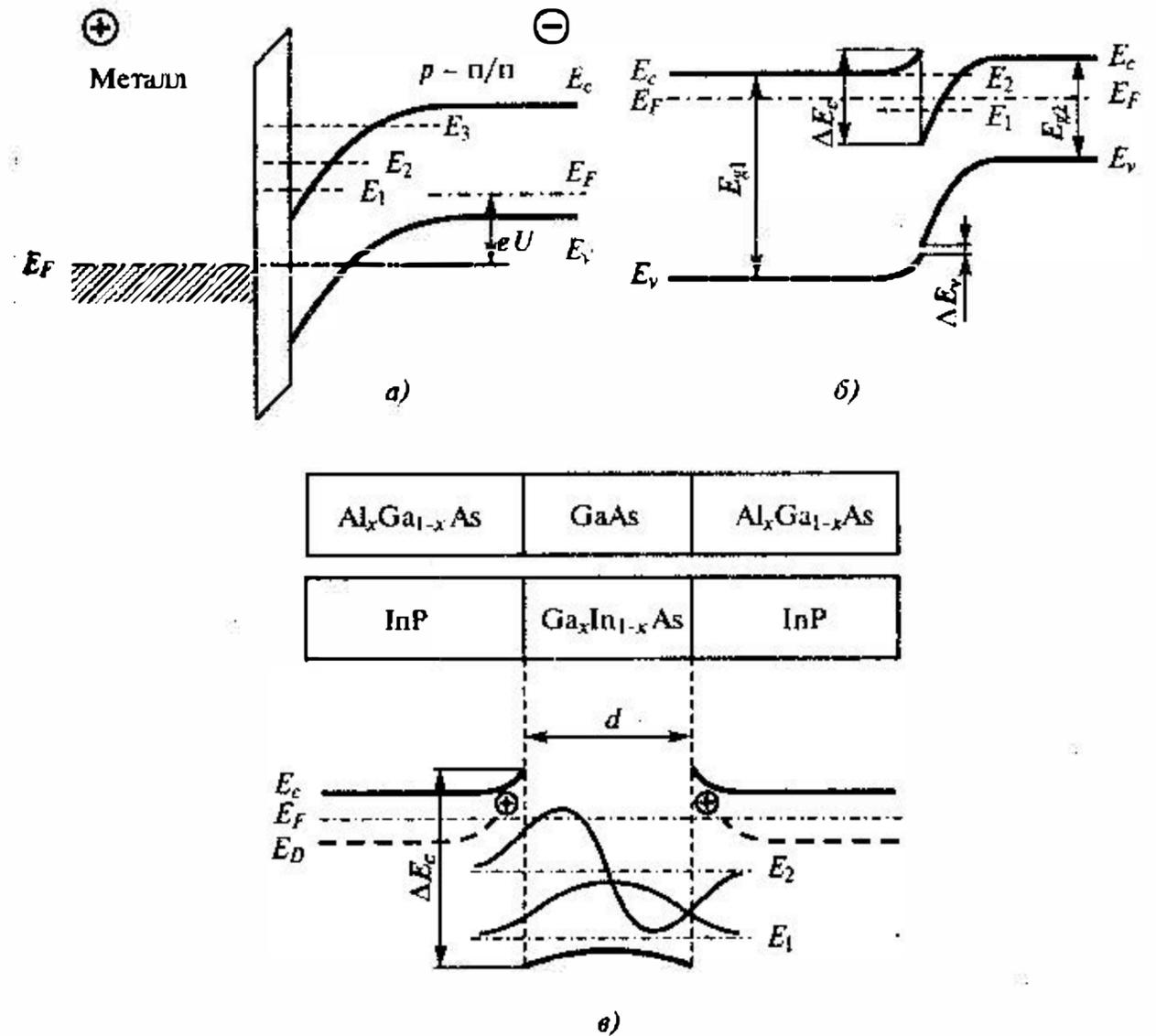


Рис. 4.55. Структуры с двумерным электронным газом:
а — металл — диэлектрик — полупроводник (МДП); б — изотипный N-p-гетеропереход; в — двойная гетероструктура (ДГС)

образован скачком потенциала ΔE_c , вызванным разницей в величинах электронного сродства χ_1 и χ_2 широкозонного и узкозонного материалов, а второй потенциальный барьер, как и в случае МДП-структуры, создается электростатическим потенциалом. В широкозонном эмиттере вблизи гетерограницы мелкие доноры ионизованы, образуя тонкий положительно заряженный слой, который компенсируется электронами, инжектированными в узкозонную область. Образуется двойной электрический слой, как показано на рис. 4.55, б.

Тот факт, что один из барьеров ямы образован электростатическим потенциалом, часто является недостатком: этот барьер довольно пологий и его крутизна изменяется при изменении концентрации носителей, т. е. барьер может «плавать» (впрочем, это может иногда рассматриваться как достоинство).

Реализовать квантовую яму, близкую к прямоугольной, можно более просто с помощью двойной гетероструктуры (ДГС). Действительно, достаточно к одиночному гетеропереходу, изображенному на рис. 4.55, б, добавить со стороны узкозонной области еще один барьер, образованный на контакте со вторым широкозонным материалом, как носители заряда окажутся запертыми в потенциальном ящике, стенки которого образованы двумя гетерограницами. В случае резких гетеропереходов эти стенки близки к вертикальным, а сам тонкий узкозонный слой полупроводника, помещенный между двумя широкозонными слоями, представляет собой прямоугольную квантовую яму, в которой движение носителей заряда в плоскости слоя (x, y) происходит без ограничений, а в направлении, перпендикулярном гетерограницам, это движение ограничено и при соблюдении рассматриваемых выше условий происходит квантование энергетического спектра. Тот факт, что электроны здесь двигаются не в свободном пространстве (вакууме), а в материальной среде (например, в эпитаксиальном монокристаллическом слое), учитывается введением диэлектрической проницаемости среды ϵ и заменой массы свободного электрона m_0 в соотношениях (4.113) — (4.124) на эффективную массу m_n или m_p аналогично тому, как это делается в обычных кристаллах (см. §§ 4.1 и 4.2).

Подчеркнем, что диэлектрическая проницаемость ϵ и эффективная масса свободных носителей заряда m_n и m_p являются макроскопическими параметрами, которые учитывают взаимодействие электронных и электромагнитных волн со сложным полем среды, в которой они распространяются. Дискретность, т. е. атомарным строением этой среды пренебрегается и считается, что эти волны распространяются в континууме. Если для объектов, размеры которых существенно больше соответствующей длины волны и межатомных расстояний, это вполне обосновано, то для объектов, образованных несколькими (или несколькими десятками) атомных слоев, использовать те же значения ϵ , n и m^* , что и в объемных материалах, надо с большой осторожностью. Более того, за счет эффектов туннелирования электронная волна проникает за барьеры ямы конечной глубины (см. рис. 4.48), т. е. в широкозонные области (рис. 4.55, в), параметры которых ϵ и m_n отличаются от аналогичных параметров узкозонного материала. В простейшем случае это обстоятельство учитывается введением «эффективных» ϵ , n и m^* , зависящих как от параметров материалов, так и от параметров квантовой ямы и даже номера уровня размерного квантования.

Итак, в структурах с двумерным электронным газом движение носителей в плоскости (x, y) ямы не ограничено и подчиняется

тем же законам, рассмотренным для 3D-кристаллов (см. §§ 4.1 и 4.2). В направлении z поперек потенциальной ямы поведение электронов подчиняется законам, рассмотренным в настоящем параграфе для одномерной квантовой ямы. Таким образом, полная энергия электрона в квантовой яме (квантовом желобе) носит смешанный дискретно-непрерывный характер, представляя собой сумму дискретных энергий, связанных с движением в направлении квантования, и непрерывной компоненты, описывающей движение в плоскости слоя как свободных носителей с эффективной массой m^* :

$$E = E_n + \frac{p_x^2}{2m^*} + \frac{p_y^2}{2m^*} = E_n + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m^*}, \quad (4.128)$$

где p_x , p_y и k_x , k_y — соответствующие компоненты квазиимпульса и волнового вектора. За счет непрерывной компоненты электроны, принадлежащие одному и тому же уровню E_n , могут иметь любую энергию в интервале от E_n до бесконечности (рис. 4.56). Такую совокупность состояний для данного квантового числа n называют *подзоной размерного квантования*. Поютно, что в силу сложного строения зоны проводимости и валентной зоны конкретных полупроводниковых материалов, закон дисперсии $E(k)$ в плоскости (x, y) может отклоняться от простейшего квадратичного закона (4.128) со скалярной эффективной массой $m_x = m_y = m_n$ и определяется факторами, рассмотренными в §§ 4.1 и 4.2.

Плотность состояний. Важной характеристикой любой электронной системы наряду с ее энергетическим спектром является плотность состояний $\rho(E)$. Для двумерной системы плотность состояний $\rho^{2D}(E)$ представляет собой число состояний в единичном интервале энергий, приходящихся на единицу площади. Она находится аналогично тому, как это делается в обычных 3D-кри-

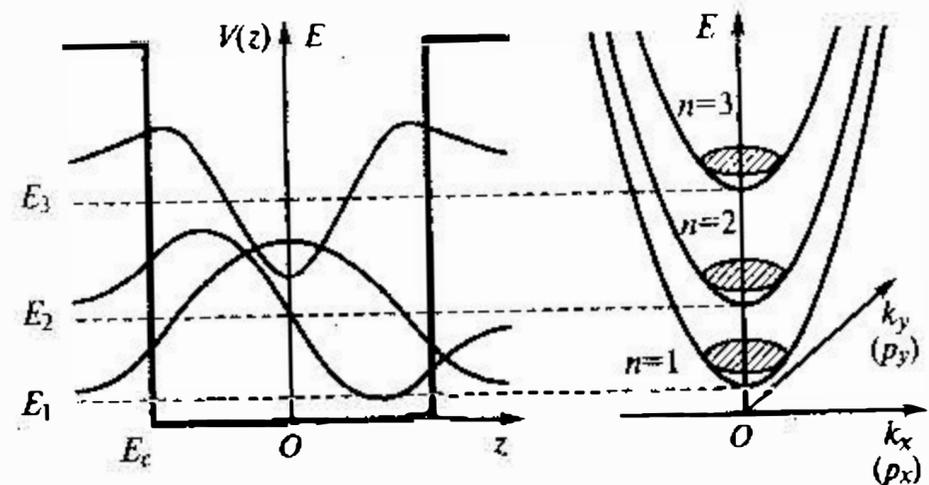


Рис. 4.56. Подзоны размерного квантования в двумерном электронном газе

сталлах. Определим, например, плотность электронных состояний в зоне проводимости ρ_c^D для простейшего параболического закона дисперсии (4.128) для диапазона $E_c < E < E_2$, т. е. для первой $n=1$ подзоны размерного квантования.

Найдем сначала число разрешенных состояний как функцию энергии электрона, а затем возьмем производную по энергии и разделим эту величину на площадь — это и будет плотность состояний. Из рис. 4.56 видно, что в диапазоне $E_c < E < E_2$ число разрешенных состояний, а потому в этом диапазоне $\rho_c^{2D} = 0$. В области $E_1 \leq E < E_2$ энергию E могут иметь электроны, чей полный квазиимпульс равен

$$p = (p_x^2 + p_y^2)^{1/2} = [2m_n^* (E - E_1)]^{1/2}.$$

В четырехмерном фазовом пространстве (x, y, p_x, p_y) эти электроны занимают объем

$$L_x L_y \pi p^2 = 2\pi L_x L_y m_n^* (E - E_1),$$

где L_x и L_y — размеры образца. На каждое состояние двумерного электронного газа согласно принципам статической физики приходится элементарный объем $(2\pi\hbar)^2$ в этом четырехмерном пространстве. С учетом двукратного вырождения по спину получаем полное число состояний с энергией E

$$G(E) = \frac{m_n^* L_x L_y}{\pi \hbar^2} (E - E_1) \quad (4.129)$$

и плотность состояний

$$\rho_c^{2D} = \frac{1}{L_x L_y} \frac{\partial G(E)}{\partial E} = \frac{m_n^*}{\pi \hbar^2} \quad (4.130)$$

для первой подзоны размерного квантования. Аналогично получаем для остальных подзон с $n \geq 2$. Каждая подзона будет давать дополнительный вклад в ρ_c^{2D} , равный (4.130) при неизменной эффективной массе электрона m_n^* :

$$\rho_c^{2D} = \frac{m_n^*}{\pi \hbar^2} \sum_n \Theta(E - E_n), \quad (4.131)$$

где $\Theta(x)$ — единичная ступенчатая функция Хэвисайда, равная 1 при $x > 0$ и 0 при $x < 0$. Плотность состояний будет испытывать скачки, равные $\frac{m_n^*}{\pi \hbar^2}$ всегда, когда энергия E сравнивается с очередным

уровнем размерного квантования a_n , т. е. с дном очередной подзоны, как это показано на рис. 4.57, б. Сопоставляя полученные данные для ρ_c^{2D} с плотностью состояний для объемного кристалла (4.9) нетрудно заметить, что при энергиях, совпадающих с положением уровней размерного квантования, т. е. при $E = E_n$ имеем $\rho_c^{2D}(E_n) = d \cdot \rho_c^{3D}(E_n)$, где d — толщина слоя, т. е. ширина квантовой ямы.

Таким образом, плотность состояний двумерной электронной системы имеет ступенчатый вид (4.131). Размерное квантование понижает плотность состояний и только при определенных энергиях $E = E_n$ в идеальных двумерных структурах наблюдается корреляция ρ_c^{2D} и ρ_c^{3D} , как показано на рис. 4.57, а, б. В реальных структурах каждая ступенька в плотности состояний ρ_c^{2D} сглаживается за счет наличия различных механизмов уширения. Появляются «хвосты плотности состояний», обсуждавшиеся в §§ 4.1 и 4.2. Влияние уширения Γ_n можно учесть заменой ступенчатой функции $\Theta(E - E_n)$, входящей в (4.131), на функцию

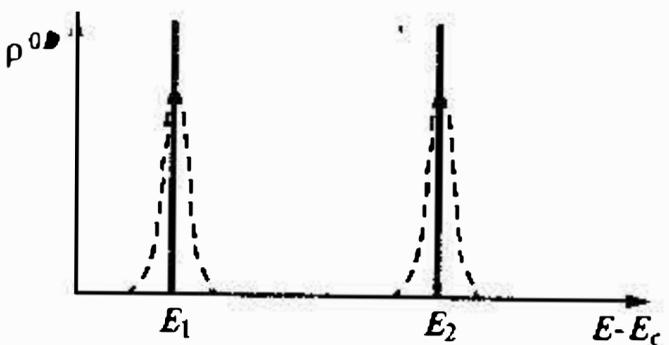
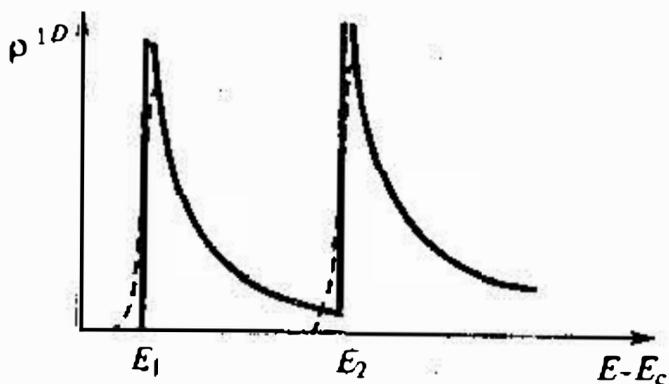
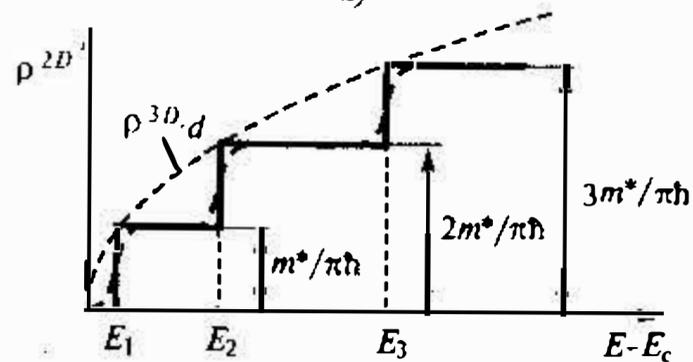
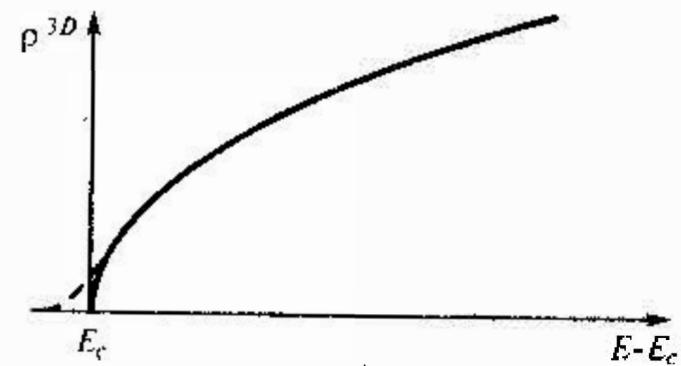
$$\Theta_\Gamma(E - E_n) = \frac{1}{\pi} \left[\frac{\pi}{2} + \arctg \left(\frac{E - E_n}{\Gamma_n} \right) \right]. \quad (4.132)$$

Это проиллюстрировано на рис. 4.57 пунктирными линиями. Процессы уширения сглаживают проявления квантоворазмерных эффектов.

Квантовые шты и квантовые точки. Если ограничить движение электрона не по одной, а по двум координатам, то получим двумерную потенциальную яму, квантование энергии электрона в которой будет подчиняться закономерностям, аналогичным рассмотренным выше. Собственные значения энергии электрона в двумерной квантовой яме будут зависеть от ее параметров, в том числе — от ширины, не только в направлении z , но и во втором направлении, например y , и будут определяться не одним, а двумя квантовыми числами n и m : $E = E_{n,m}$. По третьей координате x в такой структуре, называемой *квантовой проволокой* или *квантовой нитью*, возможно свободное движение так, что энергия электрона в такой 1D-структуре будет равна

$$E = E_{n,m} + \frac{\hbar^2 k_x^2}{2m^*}. \quad (4.133)$$

Нетрудно показать аналогично тому, как мы только что сделали для двумерного электронного газа, что в квантовой нити плотность электронных состояний как число разрешенных состояний, приходящихся на единичный интервал энергий и на единицу длины, определяется как



$$\rho_c^{1D} = \frac{\sqrt{2m^*}}{\pi\hbar} \sum_{n_1, n_2} \frac{\Theta(E - E_{n_1, n_2})}{\sqrt{E - E_{n_1, n_2}}} \quad (4.134)$$

$$\rho_c^{0D} = 2 \sum_{n_1, n_2, n_3} \delta(E - E_{n_1, n_2, n_3}), \quad (4.136)$$

Эта зависимость представлена на рис. 4.57, в. В пределах каждой подзоны фиксированной энергии размерного квантования E_{n_1, n_2} плотность состояний уменьшается с ростом энергии пропорционально

$1/\sqrt{E - E_{n_1, n_2}}$. При E_{n_1, n_2} плотность состояний в идеальной квантовой нити расходится, но в реальных ситуациях благодаря эффектам уширения это значение оказывается конечным.

Если ограничить движение электрона по трем координатам, то свободное движение будет невозможно и такая структура называется *квантовой точкой* и обладает размерностью 0: электроны заперты в трехмерной потенциальной яме и по всем трем координатам происходит квантование их энергии:

$$E = E_{n_1, n_2, n_3}. \quad (4.135)$$

Плотность состояний нульмерной системы (0D), т. е. квантовой точки, состоит из набора δ -функций, положение которых совпадает с энергиями уровней размерного квантования

как это показано на рис. 4.57, г. Это дискретный спектр разрешенных значений энергий, подобный спектру атомов и молекул. Квантовые точки как нульмерные системы по существу представляют собой как бы искусственные атомы. Но если в атомах собственный энергетический спектр в конечном итоге задается Природой, то в квантовой точке мы можем в определенной степени по своему усмотрению формировать ее энергетический спектр, изменяя параметры трехмерной потенциальной ямы.

Напомним, что при рассмотрении энергетических состояний «обычных» атомов и молекул в § 1.2, собственные значения энергии представлялись в виде δ -функций аналогично (4.136) с бесконечно узкой шириной уровня. Аналогичная ситуация наблюдалась при рассмотрении объемного резонатора, в котором отсутствуют потери и для которого добротность $Q \rightarrow \infty$ (§ 2.3). Учет затухания, связанного с конечностью времени атома в возбужденном состоянии (§ 1.6) или с конечностью времени жизни фотона в резонаторе (§ 2.3), приводит к уширению энергетического спектра, «размазыванию» его по энергии и к уширению спектральной линии, характеризуемому форм-фактором $g(\omega)$ (1.147). То же самое наблюдается в квантовых точках. Даже в идеальной ситуации, когда мы будем иметь набор идеальных квантовых точек с абсолютно одинаковыми параметрами, их уровни энергии будут иметь конечную ширину, определяемую конечностью времени жизни в возбужденном состоянии. Форма спектральной линии в этом случае в силу причин, изложенных в § 1.6, будет определяться функцией Лоренца (1.148).

Квантовые точки (КТ), как правило, формируются за счет эффектов самоорганизации при эпитаксиальном выращивании структур на основе материалов с существенно различными параметрами решетки. Например, при монослойном эпитаксиальном наращивании арсенида индия (InAs) и твердого раствора $(\text{In}_{1-x}\text{Ga}_x\text{As})$ на подложке арсенида галлия (GaAs) при определенных технологических условиях возможно образование как отдельных КТ, в виде нанокристаллов, разбросанных по поверхности слоя, так и их определенное упорядочивание, включая наложение друг на друга, так называемое «складирование» квантовых точек. Такие КТ, имеющие поперечные размеры порядка единиц и десятков нанометров и образованные из полупроводникового материала одного состава (как правило, с меньшей шириной запрещенной зоны), оказываются внедренными в матрицу, образованную полупроводником другого состава (как

Рис. 4.57. Плотность состояний в объемном 3D-кристалле (а), двумерном 2D (б), одномерном 1D (в) и нульмерном 0D (г) электронном газе. Пунктиром показано влияние эффектов уширения на функции плотности состояний

ций, положение которых совпадает с энергиями уровней размерного квантования

правило, с большей шириной запрещенной зоны). По причинам, указанным выше, гетерограница между этими двумя материалами должна быть высокого качества, иначе интерференционные эффекты, определяющие квантование, будут выражены слабо.

Основная причина уширения энергетического спектра КТ, как правило, связана с трудностью получения их одинаковых размеров. Этот разброс характеризуется определенной величиной дисперсии ΔR , приводящей к неоднородному уширению спектральных линий и энергетического спектра. По причинам, изложенным в § 1.6, форма спектральной линии и форма функции плотности состояний КТ в этом случае будет определяться функцией Гаусса.

В одномерных квантовых ямах, рассмотренных в начале настоящего параграфа, теоретически всегда должно существовать хотя бы одно связанное состояние. В отличие от этого, трехмерная квантовая яма, т. е. квантовая точка, характеризуется некоторым критическим размером, определяющим возможность существования по крайней мере одного связанного состояния. В качестве такого характерного размера выступает критический радиус

$$R_c = \frac{\pi \hbar}{\sqrt{8m^* V_b}} \quad (4.138)$$

где $V_b = V_b(r)$ — высота потенциального барьера (глубина прямоугольной ямы, обладающей сферической симметрией). Если размер нульмерной системы меньше R_c , то в ней не существует связанных состояний. Значение R_c зависит от эффективной массы частицы m^* и различно для электронов и дырок.

Примесные состояния. Эпитаксиальные слои, формирующие квантовую яму, могут быть легированы донорными или (или акцепторными) примесями. Потенциальная яма, ограничивая движение свободных электронов, изменяет свойства не только свободных, но и связанных с примесью носителей заряда. Очевидно, что для состояний малого радиуса (глубокие центры) это изменение будет минимальным, в то время как для водородоподобных состояний, боровский радиус которых (4.14) может быть существенно больше ширины ямы, изменение может быть кардинальным.

Рассмотрим для определенности водородоподобный донорный эффективный боровский радиус которого a_D в массивном образце, определяемый соотношением (4.14), существенно меньше ширины квантовой ямы d ,

$$d \ll a_D = \frac{4\pi\epsilon_0 \hbar^2}{m_n e^2} \quad (4.138)$$

Тогда в кулоновском потенциале (4.11) можно опустить зависимость r от координаты z и уравнение (4.12) для слабосвязанного примесного электрона в 2D-системе переписывается в виде

$$-\frac{\hbar^2}{2m_n} \nabla^2 \Psi(x, y, z) - \frac{e^2}{4\pi\epsilon_0 \sqrt{x^2 + y^2}} \Psi(x, y, z) = E \Psi(x, y, z) \quad (4.139)$$

Здесь переменные z и $\rho = \sqrt{x^2 + y^2}$ разделяются, что позволяет искать решение (4.139) в виде $\Psi = \varphi(\rho) \chi(z)$. Для $\chi(z)$ имеем обычное уравнение для частицы в яме с собственными значениями (4.113), (4.117) в случае прямоугольной ямы или (4.126) для треугольной ямы. В плоскости (x, y) волновая функция $\Psi(\rho)$ определяется из уравнения Шредингера для «плоского», «двумерного атома водорода» с потенциалом

$$V(\rho) = -\frac{e^2}{4\pi\epsilon_0 \rho} = -\frac{e^2}{4\pi\epsilon_0 \sqrt{x^2 + y^2}}$$

Решение этого уравнения известно. Собственные значения энергии E для него отличаются от результата (4.13), полученного для водородоподобной примеси в объемном материале, лишь заменой квантового числа n на $(i-1/2)$, причем $i = 1, 2, 3, \dots$ *, т. е.

$$E_i^{2D} = -\frac{m_n e^4}{80(\epsilon_0 \hbar)^2 (i-1/2)^2} = -\frac{E_D^{3D}}{(i-1/2)^2} \quad (4.140)$$

Здесь E_D^{3D} — энергия связи водородоподобной примеси (энергия ионизации) в объемном 3D-кристалле. За начало отсчета энергии здесь, как и в (4.13), принято состояние свободного электрона, т. е. уровни E_n размерного квантования для 2D-электронного газа.

Таким образом, собственные значения энергии для водородоподобной примеси в узкой квантовой яме бесконечной глубины с учетом (4.113) равны

*Еще раз обращаем внимание, что входящие в (4.140) макроскопические параметры m_n и ϵ_r относятся к объемному материалу, из которого изготовлена квантовая яма, и не учитывают как атомарное строение, так и возможность туннелирования электрона за стенки ямы конечной глубины.

$$E_{i,n}^{2D} = \frac{\pi^2 \hbar^2}{2m_n d^2} n^2 - \frac{E_D^{3D}}{(i-1/2)^2} \quad (4.140)$$

Это означает, что под каждой подзоной размерного квантования n расположена серия водородоподобных примесных состояний, характеризуемых квантовым числом $i=1, 2, 3, \dots$. В запрещенную зону попадут только состояния, связанные с первой подзоной $n=1$. Именно они будут заметны, поскольку остальные уровни попадут в континуум разрешенных состояний, время жизни электронов на них чрезвычайно мало и они будут сильно уширены. Напомним, что такие примесные состояния называются *резонансными*. Наинизшему примесному уровню, называемому основным состоянием, соответствует $n=1$ и $i=1$. Из (4.140) и (4.141) легко определить, что основное состояние примесного центра в двумерном случае E_D^{2D} имеет энергию связи в 4 раза большую, чем в трехмерном E_D^{3D} .

Причина возрастания энергии связи понятна: ограничение стенками ямы по координате z приводит к локализации волновой функции вблизи заряженного ядра примесного центра и к заглавлению примесного состояния.

При увеличении ширины ямы d соотношение (4.138) не будет выполняться и энергия связи E^{2D} будет уменьшаться от $4E_D^{3D}$ до E_D^{3D} при $d \rightarrow \infty$, как это показано пунктирной кривой на рис. 4.58 для основного состояния донора. Для возбужденных состояний $i \geq 2$ боровский радиус больше и условие (4.138) выполняется для больших значений d .

Соотношение (4.140) получено в предположении, что волновая функция локализована по координате z , обеспечивая выполнение граничных условий $\Psi(0) = \Psi(d) = 0$. Это справедливо лишь для ямы бесконечной глубины (с бесконечными стенками). При уменьшении глубины ямы происходит туннелирование электронов в прилегающие к яме широкозонные области. Чем мельче яма, тем глубже проникновение через барьер (см. рис. 4.48). Если характерная длина, на которой спадают волновые функции в широкозонном материале $z_c \sim \hbar / \sqrt{m(V_0 - E_i)}$ начинает превосходить боровский радиус, то локализирующее действие потенциальных барьеров ослабевает и энергия связи E_i^{2D} уменьшается, стремясь к E_D^{3D} при $d \rightarrow 0$, как это показано сплошной кривой на рис. 4.58.

Поскольку эффективная масса электронов в широкозонном полупроводнике m_2 как правило больше, чем в узкозонном m_1 [см. соотношение (4.6)], то при $d \rightarrow 0$ и малой высоте барьера величина E^{2D} будет стремиться к $\frac{m_2}{m_1} E_D^{3D}$.

В яме конечной ширины энергия примесного состояния зависит от местоположения примесного центра z_0 относительно стенок ямы. В центре энергия связи E_i обычно больше, чем у краев. Зависимость $E_i(z_0)$ приводит к неоднородному уширению примесных состояний и появлению примесной зоны. Сильное легирование, как и для 3D-кристалла, приводит к взаимодействию примесей, образованию примесных зон и «хвостов» плотности состояний аналогично тому, как это было рассмотрено в § 4.1.

Экситонные состояния. При генерации электронно-дырочных пар, например светом, в низкоразмерных системах — квантовых ямах, проволоках и точках — противоположно заряженные электрон и дырка за счет сил кулоновского взаимодействия образуют единую квазичастицу-экситон аналогично тому, как это рассматривалось в § 4.4 для объемных кристаллов.

Рассмотрим экситонные состояния в двумерной системе — квантовой яме. Аналогично тому, как это было только что сделано для водородоподобной примеси, в кулоновском слагаемом уравнения Шредингера (4.53) можно пренебречь зависимостью от координаты z . Это будет справедливо для достаточно узкой квантовой ямы, когда ее ширина d значительно меньше экситонного радиуса a_{ex} , определяемого (4.56). Тогда уравнение (4.53) для волновой функции $\Psi(\rho)$, описывающий внутреннее движение электрона относительно центра масс в плоскости ямы (x, y) , запишется в виде

$$-\frac{\hbar^2}{2m_r} \nabla^2 \Psi(x, y, z) - \frac{e^2}{4\pi\epsilon\epsilon_0 \rho} \Psi(x, y, z) = E \Psi(x, y, z), \quad (4.142)$$

где $m_r = (m_e m_p / m_n + m_p)$ — приведенная масса электрона и дырки.

Уравнение (4.142) отличается от уравнения Шредингера для водородоподобной примеси в двумерной системе (4.139) лишь заменой m_n (или m_p для акцепторов) на приведенную массу m_r . Поэтому собственные значения энергии для двумерных экситонов записываются по аналогии с (4.140):

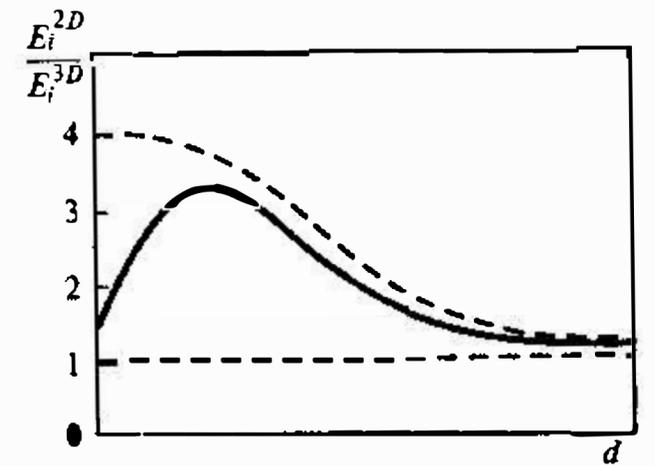


Рис. 4.58. Зависимость энергии ионизации примесного состояния в квантовой яме от ее ширины для ямы бесконечной (пунктир) и конечной (сплошная линия) глубины

$$E = E_{ex}^{2D} = \frac{m_e e^4}{8(\epsilon\epsilon_0 \hbar)^2} \cdot \frac{1}{(i-1/2)^2} = \frac{R_{ex}^{3D}}{(i-1/2)^2}, \quad (4.14)$$

где R_{ex}^{3D} — экситонный ридберг в объемном полупроводнике, определяемый соотношением (4.55), а квантовое число i может принимать значения $i=1, 2, 3, \dots$. Для наивысшего состояния $i=1$ получим, что энергия связи экситона, так же как и водородоподобной примеси, в двумерном электронно-дырочном газе возрастает в 4 раза по сравнению с обычным 3D-кристаллом.

В отличие от связанного на мелком доноре или акцепторе носителя заряда, экситон как единое целое может свободно перемещаться в плоскости ямы (x, y), что описывается движением центра масс (4.52), а на диаграмме $E(k)$ изображается аналогично рис. 4.21. Для каждой подзоны размерного квантования имеется своя экситонная серия аналогично (4.141), а каждый экситонный уровень уширяется как вследствие конечности времени жизни экситона (однородное Лоренцево уширение), так и вследствие возможных флуктуаций параметров ямы (неоднородное Гауссово уширение).

При увеличении ширины ямы d энергия связи 2D-экситона будет уменьшаться от $4R_{ex}^{3D}$ для достаточно глубокой ямы до R_{ex}^{3D} по тем же причинам, что обсуждалось выше для водородоподобной примеси. Аналогично будет сказываться и изменение глубины ямы. Закономерности, отраженные на рис. 4.58, справедливы и для экситонных состояний.

Таким образом, за счет увеличения энергии связи (для двумерной системы в пределе до 4 раз) в квантоворазмерных системах экситонные эффекты оказываются выражены более сильно, чем в объемном кристалле.

Оптические и фотоэлектрические явления в квантоворазмерных структурах определяются теми же фундаментальными процессами взаимодействия света (фотонов) с веществом (атомными системами), которые были рассмотрены в гл. 1 и 4. Основные особенности оптического поглощения, люминесценции и фотоэлектрических явлений будут связаны с рассмотренными выше особенностями энергетического спектра, функции плотности состояний и заполнения носителями разрешенных состояний, а также изменениями в величинах матричных элементов (вероятностей) соответствующих квантовых переходов в системах низкой размерности. Понятно, что в квантовых ямах и квантовых проволоках должна проявляться сильная анизотропия их свойств в зависимости от направления распространения электромагнитной волны и ее поляризации. Поскольку характерные размеры, определяющие эффекты квантования (4.11) значительно меньше длины световой волны (за исключением коротковолновой области

1...10 нм, примыкающей к рентгеновскому диапазону), то использование макроскопических параметров, применяемых для описания распространения света в массивных образцах (§ 1.8) в квантоворазмерных системах, строго говоря, недопустимо. Но эти параметры — комплексный показатель преломления \hat{n} (1.163), включающий вещественную n и мнимую части, комплексная диэлектрическая проницаемость $\epsilon = \epsilon_1 - i\epsilon_2$ весьма удобны и привычны для макроскопического описания процессов распространения света в веществе, в том числе в объектах, имеющих сложную неоднородную структуру. Поэтому часто этими же макроскопическими параметрами пользуются для описания свойств квантоворазмерных структур. Делать это надо с большой осторожностью, понимая, что за понятиями «коэффициент (показатель) преломления», «диэлектрическая проницаемость», «коэффициент отражения», «коэффициент (показатель) поглощения» и т. п. в квантоворазмерных системах стоят некоторые эффективные значения, зависящие как от объемных свойств материалов, формирующих структуру, так и от параметров самой структуры.

Проиллюстрируем сказанное на примере распространения света в плоскости квантовой ямы, т. е. в плоскости (x, y). Пусть показатель преломления (объемный) материала, из которого изготовлена квантовая яма n_2 , больше показателей преломления n_1 и n_3 , примыкающих к ней широкозонных областей (рис. 4.59). Ситуация аналогична рассмотренной в § 3.4 для планарного оптического волновода, с той принципиальной разницей, что в случае, изображенном на рис. 3.8, толщина волновода больше, а в случае, рассматриваемом нами, существенно меньше длины волны оптического излучения в веществе λ/n_2 . По этой причине световая волна не может быть локализована в квантовой яме, как это наблюдается в оптическом волноводе, а захватывает примыкающие к ней области, просачиваясь (туннелируя) аналогично тому, как это происходит с электронной волной в узкой квантовой яме с невысоким потенциальным барьером. Из рис. 4.59 видно, что для световой волны $\lambda/2n_{эф} > d$ и эффективный показатель преломления $n_{эф}$ формируется не только (а при малых d не столько) показателем преломления материала ямы n_2 , но и прилегающими к ней областями. Вклад каждой из областей зависит как от разницы показателей преломления Δn , так и от толщины слоя ямы d . Более того, размерное квантование приводит к изменению функции плотности состояний во всей области энергий (см. рис. 4.57). Это изменяет спектральную зависимость мнимой части диэлектрической проницаемости ϵ_2 на рис. 4.32 и, как следствие, коэффициенты G_1, G_2, A и E_0 в соотношении (4.69) для спектральной зависимости вещественной части показателя преломления n .

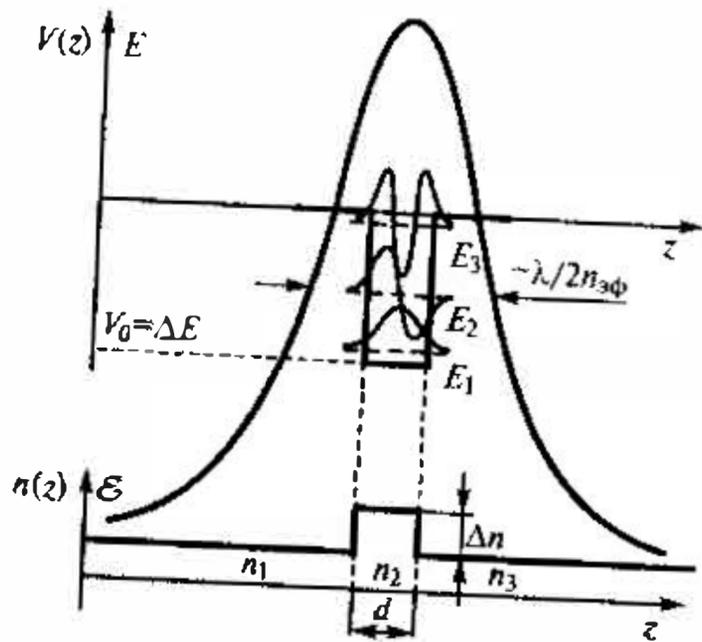


Рис. 4.59. Распределение электронной и световой волны в квантоворазмерной гетероструктуре

Для электронной волны имеем $\lambda_e < d$, и она распространяется в квантовой яме как в волноводе, незначительно просачиваясь за его стенки.

Если световая волна распространяется под углом к плоскости ямы, то происходит ее частичное отражение от границ раздела. Величина коэффициента отражения не будет определяться простейшими формулами Френеля (3.19), (3.20), но сложным образом зависит от параметров материалов, геометрии структуры, поляризации и угла падения световой волны.

При распространении световой волны в структуре, образованной многими квантовыми ямами, или в сверхрешетке, в ее поле действия могут попасть от нескольких единиц до десятков и даже сотен слоев или нитей. С теми же оговорками, что были сделаны выше, им можно приписать эффективные значения $n, k, \epsilon_1, \epsilon_2$.

Сказанное относится не только к квантовым ямам (2D), но и к квантовым нитям (1D) и квантовым точкам (0D), внедренным в виде нанокристаллов в соответствующие матрицы.

Оптическое поглощение в квантоворазмерных структурах в принципе определяется теми же процессами, что были рассмотрены в § 4.4. Перестройки в электронном спектре и функциях плотности состояний естественным образом скажутся, например, на спектрах собственного поглощения. Эти изменения могут быть радикальными. Например, спектр собственного поглощения квантовыми точками будет представлять собой не плавную функцию вида (4.42) с резкой границей при энергии фотона, соответствующей ширине запрещенной зоны 3D-кристалла, а серию узких дискретных линий аналогичных атомным спектрам. Их энергетическое положение, интенсивность и ширина будут зависеть от комбинированной плотности состояний электронов и дырок (которая в идеальных 0D-системах представляет собой набор δ -функций), от вероятности соответствующих переходов и от разброса параметров квантовых точек (нанокристаллов).

Для 2D-электронно-дырочного газа в квантовых ямах будут действовать те же правила отбора по волновому вектору k ,

k , в плоскости ямы, как и в объемном материале. Это означает, что в первом порядке теории возмущений возможны только «прямые», вертикальные переходы на диаграмме $E(k)$ аналогично (4.33) и рис. 4.14.

Ступенчатый вид функции плотности состояний (см. рис. 4.57, б) для электронов и дырок в квантовой яме приводит к ступенчатому виду комбинированной функции плотности состояний (4.41) и к ступенчатому виду спектра собственного поглощения. Правила отбора по квантовому числу n разрешают только переходы между подзонами размерного квантования электронов и дырок с одинаковыми значениями n . Поскольку структуру валентной зоны в большинстве полупроводников, как говорилось в § 4.1, можно представить состоящей из подзоны легкой (lh) и тяжелой (hh) дырок, то разрешены будут переходы между ($1e - 1lh$), ($1e - 1hh$), ($2e - 2lh$), ($2e - 2hh$) и т. п. состояниями. В этих обозначениях цифры указывают квантовое число для уровней размерного квантования электронов (e) в зоне проводимости, легких (lh) и тяжелых (hh) дырок в валентной зоне.

Как говорилось, экситонные эффекты в квантоворазмерных структурах выражены сильнее, чем в 3D-кристаллах. Их учет в квантовых ямах приводит к появлению характерных интенсивных пиков вблизи ступеней в спектрах собственного поглощения, как это показано на рис. 4.60. Резкие границы в спектре поглощения могут быть размыты в силу разных причин, рассмотренных выше, в том числе вследствие температурного уширения экситонных пиков. Но в низкоразмерных структурах температурное уширение будет проявляться менее заметно, поскольку энергия связи экситонов здесь существенно больше. Для квантовых точек, обладающих дискретным атомно-подобным спектром, температурное уширение в идеале вообще не должно проявляться.

Кроме переходов между состояниями из заполненной валентной зоны в свободную зону проводимости, определяющими собственное поглощение, в квантоворазмерных структурах, как и в 3D-кристаллах, возможны внутризонные переходы, например, между подзонами размерного квантования для электронов в зоне проводимости, а также переходы из связанного состояния, например, $n = 1$ в яме в континуум электронных состояний, расположенных выше потолка ямы. Последние процессы отсутствуют в 3D-кристаллах и могут наблюдаться лишь на границе раздела, например, с вакуумом, соответствуя выбросу электрона из полупроводника в вакуум, определяя внешний фотоэффект.

Отметим, что различного рода отклонения от идеальности, связанные с флуктуацией состава, флуктуациями геометрических размеров, наличием внутренних полей, механических напряжений

фективными центрами безызлучательной рекомбинации (§ 4.6), то к качеству квантоворазмерных структур и прежде всего к качеству гетерограниц предъявляются довольно жесткие требования. Контролировать их можно как по спектрам люминесценции, так и по спектрам поглощения, применяя, в частности методы модуляционной оптической спектроскопии.

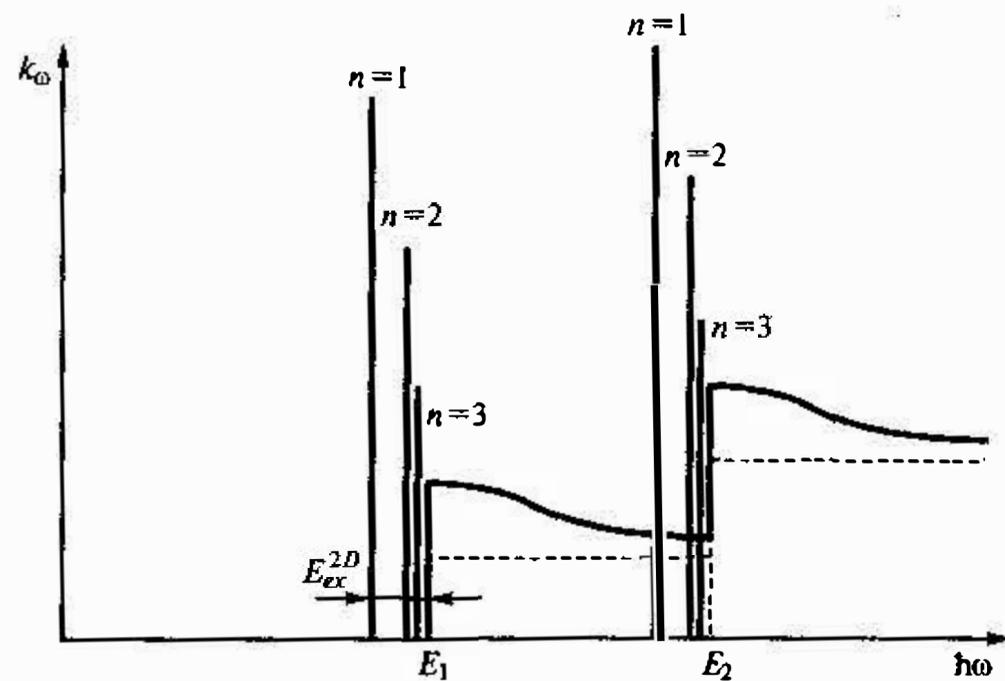


Рис. 4.60. Спектр собственного поглощения квантоворазмерной ДГС с учетом (сплошная линия) и без учета (пунктир) 2D-экситонов

и т. п. в квантоворазмерных структурах приводят к «смягчению» правил отбора для оптических переходов аналогично тому, как это обсуждалось в § 4.2 для твердых растворов.

Люминесценция квантоворазмерных структур, как и 3D-кристаллов, определяется процессами, противоположными оптическому поглощению, с учетом заполнения разрешенных энергетических состояний неравновесными носителями заряда. Как и в 3D-кристаллах, время внутрizonной релаксации в квантоворазмерных структурах, как правило, существенно меньше времени рекомбинации, а потому распределение по энергии внутри зоны для неравновесных носителей, созданных в результате возбуждения (накачки), можно рассматривать как квазиравновесное, пользуясь равновесной функцией Ферми и квазиуровнями Ферми F_n^* и F_p^* отдельно для электронов и для дырок, как это было сделано в § 4.6. Связь между поглощением и люминесценцией устанавливается соотношением Ван-Русбрека-Шокли (4.85), остающаяся справедливой и для квантоворазмерных структур. Поскольку в квантовых ямах, нитях и особенно в квантовых точках волновые функции как электрона, так и дырки локализованы, то вероятность их рекомбинации, в том числе с испусканием фотона, возрастает. Поэтому внутренний квантовый выход люминесценции может быть близок к 1 при условии, что на гетерограницах и в самих структурах концентрация центров безызлучательной рекомбинации мала. Так как различного рода структурные дефекты, дислокации, глубокие центры являются исключительно эф

ГЛАВА 5 НЕЛИНЕЙНЫЕ ОПТИЧЕСКИЕ ЭФФЕКТЫ

Под нелинейными процессами или эффектами в широком смысле понимают процессы или эффекты, нелинейно зависящие от интенсивности воздействия. Соответственно в оптике под нелинейными оптическими эффектами понимают эффекты, определяемые процессами, нелинейно зависящими от интенсивности света. Область оптики, которая исследует и применяет эти эффекты, называется *нелинейной оптикой*. Первые работы по нелинейной оптике были выполнены С. И. Вавиловым еще в 30-х годах. В частности, в 1925 г. он наблюдал уменьшение поглощения света при увеличении его интенсивности (эффект просветления) в урановом стекле. В дальнейшем значительный вклад в развитие нелинейной оптики внесли русские (Р. В. Хохлов, С. А. Ахманов) и американские (П. Франкен, Н. Бломберген) ученые.

Нелинейные оптические эффекты обычно проявляются лишь в сильных световых полях, так что их детальное изучение и применение стало возможным только после создания лазеров. В современной квантовой и оптической электронике нелинейные эффекты широко используют для дискретной и плавной перестройки частоты излучения (сложение, умножение и параметрическое преобразование частот), детектирования оптических сигналов, управления световыми потоками и т. д.

§ 5.1. ОБЩИЕ ПОНЯТИЯ

До сих пор в большинстве случаев считалось, что интенсивность взаимодействия оптического излучения с веществом пропорциональна интенсивности этого излучения, а характер взаимодействия не зависит от интенсивности. Например, число поглощенных фотонов пропорционально числу падающих фотонов, а поляризованность среды \vec{P} , вызванная электрическим полем световой волны, пропорциональна напряженности этого поля E . Это означает, что предполагалась линейная связь между \vec{D} и \vec{E} , \vec{H} , \vec{j} и \vec{E} в уравнениях Максвелла (1.3). Величины ϵ , μ и σ в (1.4) и (1.159), отражающие реакцию среды, считались не зависящими от \vec{E} и \vec{H} . Только при этом решения уравнений Максвелла получаются в виде (1.15), (1.23) и (1.161). Этот случай соответствует приближению линейной оптики.

Исходя из общих соображений, можно утверждать, что линейная зависимость одной физической величины от другой почти всегда является приближением, справедливым в более или менее ограниченной области. Хорошо известна, в частности, нелинейная связь между \vec{D} и \vec{E} в постоянном или низкочастотном поле. Наиболее отчетливо она проявляется у сегнетоэлектриков. На оптических частотах вклад в ϵ дает лишь электронная и иногда ионная поляризация. Поскольку напряженность электрического поля в «обычной» (нелазерной) световой волне существенно меньше напряженностей внутриатомных или внутрикристаллических полей, приближение линейной оптики в этом случае вполне оправдано. Появление лазера дало возможность получать световые пучки, напряженность электрического поля в которых достигает 10^8 В/см и более, т. е. сравнима с внутриатомными полями. Естественно, что в таких полях линейная связь между \vec{D} и \vec{E} должна нарушаться.

Так как в оптическом диапазоне $\mu_r \approx 1$, то нелинейные оптические эффекты могут определяться нелинейной связью между \vec{D} и \vec{E} (нелинейной поляризацией) или между \vec{j} и \vec{E} (нелинейным поглощением).

Диэлектрическая проницаемость среды ϵ определяется вектором поляризации \vec{P} и связана с относительной диэлектрической восприимчивостью χ соотношениями

$$\vec{D} = \epsilon_0 \epsilon_r \vec{E} = \epsilon_0 \vec{E} + \vec{P}; \quad \vec{P} = \epsilon_0 \chi \vec{E}; \quad \epsilon_r = 1 + \chi. \quad (5.1)$$

В общем случае $\chi = \chi(E)$. Ограничиваясь для простоты скалярными соотношениями, можно записать

$$P = \epsilon_0 (\chi_1 E + \chi_2 E^2 + \chi_3 E^3 + \dots) \quad (5.2)$$

или

$$P = P_{л} + P_{нл},$$

где первое слагаемое $\epsilon_0 \chi_1 E$ определяет линейную поляризацию $P_{л}$, а остальные слагаемые — нелинейную поляризацию $P_{нл}$. Все коэффициенты χ_n не зависят от E и характеризуют свойства вещества.

Отношение каждого последующего члена ряда к предыдущему имеет порядок E/E_0 , где E_0 для электронной поляризации

определяется напряженностью внутриатомного электрического поля: $E_a \approx 10^8$ В/см. Именно поэтому в «обычном» световом поле создаваемом «классическими» источниками, где E не превышает 10 В/см, нелинейными эффектами можно пренебречь. В среде обладающей центром инверсии, в разложении (5.2) исчезнут все слагаемые с четными степенями по E . Действительно, для такой среды при любом E должно выполняться равенство $P(-E) = -P(E)$, что возможно лишь при $\alpha_2 = \alpha_4 = \dots = 0$. По этой причине в нелинейной оптике обычно используют кристаллы, не обладающие центром инверсии.

Рассмотрим для определенности среду, в которой существует квадратичная нелинейность, т. е. ограничимся первым нелинейным членом в разложении (5.2):

$$P^{(2)} = \epsilon_0 \alpha_2 E^2. \quad (5.3)$$

Пусть в среде распространяются две электромагнитные волны, начальные фазы которых для простоты считаем одинаковыми:

$$\begin{aligned} E_1(r, t) &= E_0^{(1)} \cos(\omega_1 t - \mathbf{k}_1 r), \\ E_2(r, t) &= E_0^{(2)} \cos(\omega_2 t - \mathbf{k}_2 r). \end{aligned} \quad (5.4)$$

Подставляя (5.4) в (5.3), получаем

$$\begin{aligned} P^{(2)} &= \epsilon_0 \alpha_2 [E_0^{(1)2} \cos^2(\omega_1 t - \mathbf{k}_1 r) + E_0^{(2)2} \cos^2(\omega_2 t - \mathbf{k}_2 r) + \\ &+ 2E_0^{(1)} E_0^{(2)} \cos(\omega_1 t - \mathbf{k}_1 r) \cos(\omega_2 t - \mathbf{k}_2 r)]. \end{aligned}$$

Используя известные тригонометрические соотношения $\cos^2 \alpha = 1/2(1 + \cos 2\alpha)$ и $\cos \alpha \cos \beta = 1/2[\cos(\alpha + \beta) + \cos(\alpha - \beta)]$, находим, что поляризация состоит из нескольких спектральных компонент с различными частотами:

$$P(2\omega_1) = 0,5 \epsilon_0 \alpha_2 E_0^{(1)2} \cos(2\omega_1 t - 2\mathbf{k}_1 r), \quad (5.5a)$$

$$P(2\omega_2) = 0,5 \epsilon_0 \alpha_2 E_0^{(2)2} \cos(2\omega_2 t - 2\mathbf{k}_2 r), \quad (5.5a)$$

$$P(\omega_1 + \omega_2) = \epsilon_0 \alpha_2 E_0^{(1)} E_0^{(2)} \cos[(\omega_1 + \omega_2)t - (\mathbf{k}_1 + \mathbf{k}_2)r], \quad (5.5b)$$

$$P(\omega_1 - \omega_2) = \epsilon_0 \alpha_2 E_0^{(1)} E_0^{(2)} \cos[(\omega_1 - \omega_2)t - (\mathbf{k}_1 - \mathbf{k}_2)r] \quad (5.5b)$$

и постоянного слагаемого, описывающего статическую поляризацию:

$$P(0) = 0,5 \epsilon_0 \alpha_2 (E_0^{(1)2} + E_0^{(2)2}). \quad (5.5g)$$

В § 1.8 отмечалось, что уменьшение скорости света в веществе и появление показателя преломления вызвано поляризацией среды под действием электрического поля световой волны. В линей-

ном приближении волна поляризации вызывает излучение электромагнитной волны той же частоты, что и у падающей волны. В отсутствие поглощения волна переизлучается без потерь. При нелинейной поляризации, как видно из соотношений (5.5), волна поляризации имеет сложный вид и содержит члены с нулевой частотой (статическая поляризация), суммарной и разностной частотами, а также вторые гармоники основных частот. Эти компоненты нелинейной поляризации ответственны за излучение электромагнитных волн на частотах $(\omega_1 + \omega_2)$, $(\omega_1 - \omega_2)$, $2\omega_1$ и $2\omega_2$. Это означает, что часть энергии падающей волны, которая идет на возбуждение нелинейной поляризации, будет переизлучаться на одной или нескольких частотах, отличных от основной.

Очевидно, если рассматривать следующие нелинейные члены в (5.2), например $\alpha_3 E^3$, то в поляризации появятся более сложные комбинации частот, в том числе третьи гармоники: $3\omega_1$, $3\omega_2$ и т. д.

Таким образом, в основе нелинейной оптики лежат нелинейные соотношения (1.159), приводящие к нелинейности всей системы уравнений, описывающих распространение световых волн в веществе. Каждый нелинейный оптический процесс условно можно представить себе состоящим из двух этапов: сначала свет большой интенсивности вызывает нелинейный отклик среды, а затем эта реакция среды нелинейным образом изменяет оптические поля. Первый этап описывается материальными уравнениями, а второй — уравнениями Максвелла. Следствием этой нелинейности является невыполнение принципа суперпозиции волн. Распространяясь в такой среде, световые волны взаимодействуют друг с другом, что проявляется в обмене энергией между разными волнами, появлении одних волн за счет ослабления других и т. д. С квантовой точки зрения это означает, что фотоны, распространяясь в нелинейной среде, могут взаимодействовать друг с другом, причем взаимодействие происходит через нелинейную поляризацию среды.

§ 5.2. ГЕНЕРАЦИЯ ГАРМОНИК И УСЛОВИЕ ФАЗОВОГО СИНХРОНИЗМА

Рассмотрим для определенности генерацию второй гармоники. Схема опыта для наблюдения генерации второй гармоники, т. е. удвоения частоты оптического излучения, показана на рис. 5.1. Мощное лазерное излучение на частоте ω , проходя через фильтр, поступает на нелинейный оптический элемент, в качестве которого могут быть использованы различные диэлектрические кристаллы. За счет нелинейной квадратичной поляризации, возбуждаемой лазерной волной на частоте ω , в кристалле возникает волна поляризации с удвоенной частотой 2ω , которая в свою очередь генерирует электромагнитную волну

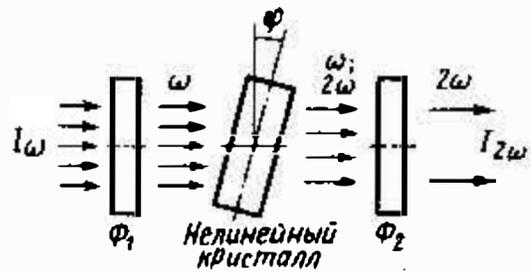


Рис. 5.1. Схема опыта для наблюдения удвоения частоты световой волны

кой форме, как в (5.5), а в показательной. Будем считать, что волна распространяется в направлении z . Тогда

$$P(2\omega) = P_0 \exp[i(2\omega t - 2k(\omega)z)], \quad (5.6)$$

где

$$k(\omega) = 2\pi/\lambda_\omega = \omega n(\omega)/c \quad (5.7)$$

— волновой вектор на частоте ω ; λ_ω — длина световой волны с частотой ω в веществе; $n(\omega)$ — показатель преломления на частоте ω ; P_0 — амплитуда поляризации при $z=0$.

Из соотношений (5.5) имеем $P_0 = \epsilon_0 \alpha_2 E_0^2$.

Генерируемую волной поляризации (5.6) электромагнитную волну можно записать в виде

$$E(2\omega) = E_0(z) \exp[i(2\omega t - k(2\omega)z)], \quad (5.8)$$

где волновой вектор для этой волны (на частоте 2ω)

$$k(2\omega) = 2\pi/\lambda_{2\omega} = 2\omega n(2\omega)/c. \quad (5.9)$$

Здесь $E_0(z)$ — медленно изменяющаяся функция z , не зависящая от t ; $n(2\omega)$ — показатель преломления на частоте 2ω [а не ω , как в (5.7)]. Считая, что поглощение на частотах ω и 2ω в диэлектрике отсутствует, положим $\sigma=0$. Включая поляризационный член в уравнения Максвелла (1.3), имеем

$$\frac{\partial^2 E}{\partial z^2} - \epsilon_0 \epsilon_r \mu_0 \frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P}{\partial t^2}. \quad (5.10)$$

В этом соотношении, по сравнению с (1.160), учитывается движение связанных зарядов, т. е. поляризация. Здесь $\epsilon_r = n^2(2\omega)$ — относительная диэлектрическая проницаемость на частоте 2ω . Вводя величину $\Delta n = n(2\omega) - n(\omega)$ и подставляя (5.6) и (5.8) в (5.10), получаем

$$\frac{\partial^2 E_0(z)}{\partial z^2} - i \frac{2\omega n(2\omega)}{c} \frac{\partial E_0(z)}{\partial z} = \mu_0 \omega^2 P_0 \exp\left(i2\omega z \frac{\Delta n}{c}\right). \quad (5.11)$$

на частоте 2ω , отфильтровываемую от основной волны с помощью фильтра Φ_2 .

Очевидно, возникновение волны поляризации с удвоенной частотой можно рассматривать с помощью соотношений (5.5), положив $\omega_1 = \omega$, $\omega_2 = 0$ или $\omega_1 = \omega_2 = \omega$. Выразим волну поляризации на частоте 2ω не в тригонометрической

Поскольку амплитуда $E_0(z)$ слабо меняется на расстоянии длины волны, то в уравнении (5.11) можно пренебречь второй производной $\partial^2 E_0(z)/\partial z^2$. Тогда решение этого уравнения, удовлетворяющее граничному условию $E_0(0)=0$, приобретает вид

$$E_0(z) = \frac{P_0}{2E_0 n \Delta n} \left[\exp\left(i \frac{2\omega \Delta n}{c} z\right) - 1 \right]. \quad (5.12)$$

Амплитуда вектора напряженности электрического поля определяется модулем $E_0(z)$, и из (5.12) получаем

$$|E_0(z)| = \frac{P_0}{\epsilon_0 n \Delta n} \sin\left(\frac{\omega \Delta n}{c} z\right). \quad (5.13)$$

Интенсивность волны на удвоенной частоте

$$I(2\omega) = \rho(2\omega)c = \epsilon_0 c |E_0|^2 = \frac{P_0^2 c}{\epsilon_0 n^2 (\Delta n)^2} \sin^2\left(\frac{\omega \Delta n}{c} z\right) \sim \\ \sim I(\omega)^2 \alpha_2^2 \frac{1}{(\Delta n)^2} \sin^2\left(\frac{\omega \Delta n}{c} z\right). \quad (5.14)$$

Из этих соотношений видно, что амплитуда генерируемой волны изменяется с расстоянием по синусоидальному закону вследствие интерференции с поляризационной волной. Действительно, волна поляризации перемещается в среде со скоростью электромагнитной волны, возбуждающей эту поляризацию:

$$v_{\text{н}} = v(\omega) = c/n(\omega), \quad (5.15)$$

а фазовая скорость световой волны на второй гармонике определяется показателем преломления на частоте 2ω :

$$v_{\text{ф}} = v(2\omega) = c/n(2\omega). \quad (5.16)$$

Волна второй гармоники получает свою энергию от основного излучения через волну поляризации (5.6). Обмен энергией между основной волной и ее второй гармоникой будет максимален, если разность фаз $\Delta\phi$ между волной поляризации и генерируемой на частоте 2ω электромагнитной волной сохранится на достаточно больших расстояниях. Для этого необходимо, чтобы $v_{\text{н}} = v_{\text{ф}}$, т. е.

$$\Delta n = n(2\omega) - n(\omega) = 0. \quad (5.17)$$

При выполнении этого условия, как видно из (5.13), амплитуда напряженности электрического поля линейно возрастает с расстоянием

$$E_0(z) = \frac{\mu_0 c \omega P_0}{n} z. \quad (5.14)$$

Тогда с учетом $P_0 \sim \varepsilon_2 E_0^2(\omega)$ получаем, что интенсивность излучения на второй гармонике $I(2\omega)$ пропорциональна квадрату интенсивности возбуждающей волны $I(\omega)$, квадрату оптической длины пути взаимодействия l и квадрату нелинейной восприимчивости ε_2 :

$$I(2\omega) \sim \varepsilon_2^2 k^2(\omega) l^2 I^2(\omega). \quad (5.15)$$

В обычных случаях условие (5.17) не удается выполнить вследствие дисперсии показателя преломления света в веществе. В области поглощения $n(2\omega) > n(\omega)$, что соответствует нормальной дисперсии (см. рис. 1.17). Поэтому между волной поляризации и световой волной на частоте 2ω будет наблюдаться фазовый сдвиг, который на длине l равен

$$\Delta\varphi = l [k(2\omega) - 2k(\omega)]. \quad (5.16)$$

Фазовые соотношения считают нарушенными, когда сдвиг фаз $\Delta\varphi \geq \pi/2$. Длину l , при которой $\Delta\varphi = \pi/2$, т. е. длину, при которой фазовые соотношения можно условно считать сохраняющимися, в нелинейной оптике называют *длиной когерентности**:

$$l_{\text{кор}} = \frac{\pi}{2 [k(2\omega) - 2k(\omega)]} \approx \frac{\lambda_{2\omega}}{2\Delta n}. \quad (5.17)$$

Из выражений (5.13), (5.14) видно, что при $z = l_{\text{кор}}$ интенсивность второй гармоники достигает максимума. При дальнейшем увеличении $l > l_{\text{кор}}$ происходит уменьшение E_0 и $I(2\omega)$ и последующие их осцилляции в соответствии с (5.13) и (5.14). Это иллюстрируется экспериментальными данными, представленными на рис. 5.2.

Из сказанного следует, что накопление нелинейных эффектов невозможно на расстояниях, больших $l_{\text{кор}}$. Даже если $\Delta n = 0,01$, то из (5.17) получаем для $\lambda = 1$ мкм $l_{\text{кор}} \approx 50$ мкм. Естественно, что в такой малой области взаимодействия преобразовываться во вторую гармонику будет лишь очень незначительная часть основной волны. Чтобы осуществить эффективное преобразование мощности исходной волны во вторую гармонику, необходимо

*Это понятие не следует путать с аналогичным понятием, введенным в § 2.6 при определении когерентности электромагнитного излучения.

выполнить условия (5.16), (5.17), которые можно записать в виде

$$k(2\omega) = 2k(\omega). \quad (5.22)$$

При выполнении равенства (5.22) $\Delta\varphi = 0$ и $l_{\text{кор}} \rightarrow \infty$. Поэтому условие (5.22), так же как и условия (5.16) и (5.17), называется *условием фазового (волнового) синхронизма*. При его выполнении наблюдается согласование фаз возбуждающей волны и ее второй гармоники и происходит наиболее эффективное преобразование. С квантовой точки зрения условие волнового синхронизма (5.22) представляет собой условие сохранения импульса для фотонов, участвующих в данном процессе:

$$\hbar k(2\omega) = \hbar k(\omega) + \hbar k(\omega). \quad (5.23)$$

Условие волнового синхронизма можно выполнить в анизотропном кристалле, используя взаимодействие волн с разной поляризацией. Рассмотрим одноосный кристалл, в котором наблюдается двойное лучепреломление. В таком кристалле показатель преломления, а следовательно, фазовая скорость волны зависят не только от частоты ω , но и от поляризации (см. § 3.5). В одноосном кристалле для обыкновенной волны показатель преломления n_o не зависит от направления распространения. Сечения поверхностей показателя преломления для этой волны (сферы) дают окружность, как показано на рис. 5.3 для двух частот ω и 2ω . При этом вследствие дисперсии $n_o(2\omega) > n_o(\omega)$. Для необыкновенной волны показатель преломления n_e зависит от направления распространения, так что поверхность показателя преломления для этой волны представляет собой эллипсоид.

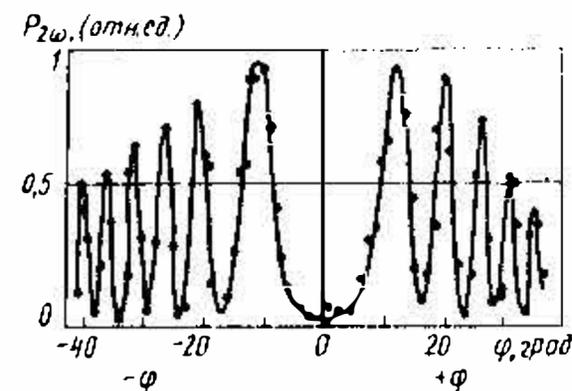


Рис. 5.2. Зависимость мощности второй гармоники от длины оптического пути l , пройденного световой волной в кристалле кварца (величина l изменялась за счет изменения угла наклона φ кварцевой пластинки относительно оси пучка лазера)

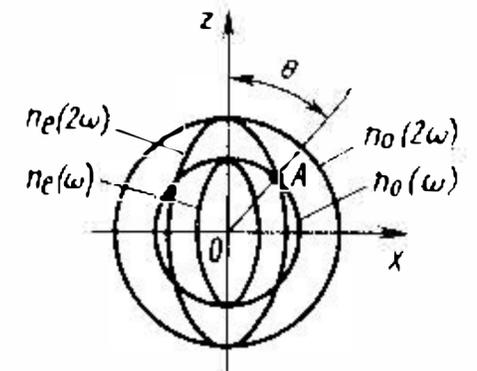


Рис. 5.3. Сечения поверхностей показателя преломления: сферы для обыкновенной волны n_o и эллипсоида для необыкновенной волны n_e в кристалле КДР для частоты рубинового лазера ω и его второй гармоники 2ω

Соответствующие сечения дают эллипсы, которые изображены на рис. 5.3 для двух частот ω и 2ω . В направлении оптической оси (вдоль оси Oz на рис. 5.3) значения n_o и n_e совпадают. В соответствии с определением, введенным в § 3.5, случай, изображенный на рис. 5.3, соответствует оптически отрицательному кристаллу ($n_e \leq n_o$).

Важным является то, что сечение поверхности показателя преломления для необыкновенной волны на частоте 2ω [эллипс $n_e(2\omega)$] пересекает сечение поверхности показателя преломления для обыкновенной волны на частоте ω [окружность $n_o(\omega)$]. В точке пересечения $n_e(2\omega) = n_o(\omega)$. Это означает, что для излучения распространяющегося вдоль направления OA , составляющего угол θ с оптической осью кристалла, фазовая скорость для обыкновенной волны на частоте ω равна фазовой скорости для необыкновенной волны на частоте 2ω . Это направление называется *направлением фазового (волнового) синхронизма*. Очевидно, в одноосном кристалле совокупность таких направлений образует конус с вершиной в точке O и углом раствора θ .

Таким образом, условие фазового синхронизма может быть реализовано в одноосном кристалле для волн различной поляризации, распространяющихся вдоль определенных направлений по отношению к оптической оси кристалла. В частности для оптически отрицательного кристалла исходная возбуждающая волна на частоте ω должна быть поляризована перпендикулярно оптической оси кристалла, а генерируемая волна на удвоенной частоте 2ω будет иметь поляризацию параллельно оптической оси.

Для эффективного удвоения частоты активная среда (кристалл) должна удовлетворять ряду жестких условий:

- 1) кристалл должен быть оптически прозрачен как на частоте ω , так и на частоте 2ω ;
- 2) двойное лучепреломление ($n_e - n_o$) должно быть достаточно большим, чтобы сечения $n_o(\omega)$ и $n_e(2\omega)$ для отрицательного кристалла или $n_e(\omega)$ и $n_o(2\omega)$ для положительного пересекались;
- 3) нелинейная восприимчивость χ_2 должна быть по возможности максимальной.

При соблюдении этих условий интенсивность второй гармоники для направления, совпадающего с направлением волнового синхронизма, будет определяться (5.19).

Перечисленным условиям наиболее полно удовлетворяют кристаллы дигидрофосфата калия KH_2PO_4 (KDP), дигидрофосфата аммония $NH_4H_2PO_4$ (ADP), ниобата лития $LiNbO_3$, танталата лития $LiTaO_3$ и др. Укажем также, что показатели преломления для обыкновенной и необыкновенной волн могут иметь различные температурные зависимости, поэтому условие волно-

вого синхронизма в ряде кристаллов может быть выполнено лишь в ограниченном температурном интервале.

Благодаря выполнению условия волнового синхронизма уже в настоящее время в кристаллах достигнут коэффициент преобразования лазерного излучения во вторую гармонику более 50%. Используя нелинейности более высокого порядка, можно получить излучение на третьей, четвертой и т. д. гармониках, т. е. осуществить эффективное преобразование (умножение) частоты лазерного излучения.

§ 5.3. ПАРАМЕТРИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ И ПАРАМЕТРИЧЕСКАЯ ГЕНЕРАЦИЯ СВЕТА

Из соотношений (5.5) видно, что в нелинейной среде возможны сложения и вычитания двух волн, приводящие к преобразованию частот. Пусть на нелинейную среду в направлении z падает мощная световая волна на частоте ω_n , которую назовем волной накачки:

$$E_n = E_0^{(n)} \cos(\omega_n t - k_n z) \quad (5.24)$$

и одновременно две слабые волны

$$E_1 = E_0^{(1)} \cos(\omega_1 t - k_1 z), \quad (5.24a)$$

$$E_2 = E_0^{(2)} \cos(\omega_2 t - k_2 z) \quad (5.24b)$$

с частотами ω_1 и ω_2 , удовлетворяющими соотношению

$$\omega_n = \omega_1 + \omega_2. \quad (5.25)$$

Пусть, как и ранее, поляризация среды обладает квадратичной нелинейностью (5.3). Тогда, подставляя (5.24) в (5.3), получим компоненты нелинейной поляризации на частотах ω_1 и ω_2 :

$$P(\omega_1) = 0,5\epsilon_0\chi_2 E_0^{(n)} E_0^{(2)} \cos[\omega_1 t - (k_n - k_2)z], \quad (5.26)$$

$$P(\omega_2) = 0,5\epsilon_0\chi_2 E_0^{(n)} E_0^{(1)} \cos[\omega_2 t - (k_n - k_1)z]. \quad (5.26a)$$

Эти соотношения указывают, что за счет взаимодействия волн на частотах ω_n и ω_2 возникает переизлучение на частоте ω_1 , а за счет взаимодействия волн на частотах ω_n и ω_1 — переизлучение на частоте ω_2 . Аналогично генерации второй гармоники для эффективного взаимодействия волны нелинейной поляризации $P(\omega_1)$ со световой волной этой частоты ω_1 необходимо, чтобы фазовый сдвиг между ними сохранялся с расстоянием z . Это приводит к необходимости выполнения условия волнового синхронизма

$$k_n = k_1 + k_2. \quad (5.27)$$

Такое же условие получается для частоты ω_2 .

Условие (5.27), как и (5.22), представляет собой условие сохранения импульса для фотонов, участвующих во взаимодействии, и в общем виде можно записать

$$\hbar k_n = \hbar k_1 + \hbar k_2. \quad (5.28)$$

Отметим, что полученные ранее соотношения (5.22), (5.23) можно рассматривать как частный случай более общих соотношений (5.27), (5.28).

Условие синхронизма (5.27) может быть выполнено в анизотропном кристалле подобно тому, как это рассматривалось в § 5.2. При этом направление волнового синхронизма OA на рис. 5.3 будет зависеть от частот взаимодействующих волн.

Если условие синхронизма выполнено, то энергия волны накачки эффективно передается волнам с частотами ω_1 и ω_2 , которые усиливаются в нелинейной среде. Этот процесс аналогичен процессу параметрического преобразования частоты в радиотехнике и в оптике называется *параметрическим преобразованием оптического излучения*. Но в отличие от радиодиапазона параметрическое преобразование в оптике носит волновой характер, поэтому кроме обычной «частотной» настройки (5.25) требует соответствующей «волновой» настройки (5.27). Существенным является то, что в оптике все нелинейные эффекты развиваются не только во времени, но и в пространстве.

Поскольку нелинейная среда за счет параметрического преобразования способна усиливать излучение на частотах ω_1 и ω_2 , вводя положительную обратную связь путем помещения кристалла в оптический резонатор, можно получить генерацию света на этих частотах. Принцип действия параметрического генератора состоит в следующем (рис. 5.4). Нелинейный кристалл 1 помещают в оптический резонатор, создаваемый зеркалами 2. На кристалл действует мощная электромагнитная волна накачки на частоте ω_n , распространяющаяся вдоль оптической оси резонатора OO . В результате параметрического преобразования с превышением усиления над потерями (при выполнении условий самовозбуждения) в резонаторе возникает генерация на частотах ω_1 и ω_2 . Начальные сигналы обусловлены собственными флуктуациями (шумами) на частотах ω_1 и ω_2 , всегда присутствующими в системе.

Важно то, что частоту генерации в параметрическом генераторе света можно плавно изменять в широких пределах. Перестройку частоты генерации можно осуще-

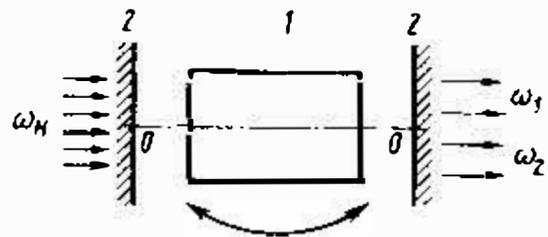


Рис. 5.4. Принцип действия параметрического генератора света: 1 — нелинейный кристалл; 2 — зеркала резонатора

ствлять, например, поворотом кристалла, как показано на рис. 5.4. При повороте кристалла, как видно на рис. 5.3, условие волнового синхронизма будет выполняться для другой совокупности длин волн (частот) излучения. Для перестройки частоты генерации параметрического генератора в принципе пригодны любые эффекты, приводящие к изменению оптической индекса и направлений синхронизма. Кроме поворота кристалла возможна температурная перестройка. Можно также использовать электрооптические и фотоупругие эффекты.

Поскольку эффективное преобразование происходит лишь в мощном монохроматическом световом потоке, накачка параметрических генераторов обычно осуществляется лазерным излучением. КПД преобразования достигает 10%. В качестве нелинейных сред могут служить те же кристаллы, что и для генерации гармоник.

§ 5.4. ПРОЧИЕ НЕЛИНЕЙНЫЕ ЭФФЕКТЫ

1. Из выражения (5.5г) видно, что нелинейная поляризация и переизлученное поле содержат постоянную составляющую. Это означает, что в нелинейной среде происходит *оптическое детектирование* (оптическое выпрямление).

2. В нелинейной среде возможно осуществить так называемое *параметрическое преобразование частоты вверх*. Для этого мощное излучение накачки (обычно на частоте ω_n , соответствующей видимому диапазону) в нелинейном кристалле смешивается с преобразуемым инфракрасным излучением частотой ω_1 . В результате параметрического преобразования на выходе получается сигнал с частотой $\omega_n + \omega_1$ в видимой области спектра, который легко отфильтровывается от сигнала накачки. Таким способом возможно преобразовывать инфракрасное изображение (например, с $\lambda = 10,6$ мкм) в видимое.

3. В сильном световом поле происходит изменение показателя преломления среды. Причины этого явления могут быть довольно разнообразными: высокочастотный эффект Керра, нелинейная добавка к показателю преломления за счет поляризуемости третьего порядка $\{\alpha_3 E^3$ (5.2)], нагрев среды лазерным лучом, электрострикция и т. д. В результате этих эффектов показатель преломления среды будет зависеть от распределения интенсивности распространяющейся волны по сечению луча. Обычно области максимальной интенсивности соответствует большая оптическая плотность (большее значение n). Поэтому световой луч по мере распространения начинает втягиваться в эту область, где поле максимально. Это явление называется *самофокусировкой лазерного излучения*.

4. Поскольку напряженность электрического поля в мощной световой волне при ее фокусировке может достигать 10^8 В/см и более, т. е. превышать пробивную напряженность, при фокусировке лазерного излучения в газах, жидкостях или твердых телах возможен электрический пробой с образованием «световой искры». Этот эффект называется *оптическим пробоем*.

5. Рассмотренный в § 1.7 эффект комбинационного рассеяния света по существу является нелинейным оптическим эффектом. В нелинейной среде световые волны могут взаимодействовать не только друг с другом, но и с молекулярными колебаниями. Последнее и приводит к комбинационному рассеянию света. Кроме обычного (спонтанного) комбинационного рассеяния в сильно нелинейной среде при воздействии на нее мощного светового поля накачки возможно *вынужденное комбинационное рассеяние*, т. е. усиление и генерация на комбинационных частотах. С классической точки зрения вынужденное комбинационное рассеяние можно рассматривать как параметрический процесс, при котором за счет энергии волны накачки возбуждается «волна» молекулярных колебаний и когерентная световая волна стоксовой или антистоксовой частоты. Этот эффект может быть использован для дискретной перестройки частоты лазерного излучения.

6. Электромагнитная волна в кристалле может взаимодействовать также с колебаниями кристаллической решетки (фононами). Рассеяние световой волны на акустических колебаниях кристаллической решетки называется *рассеянием Мандельштама — Бриллюэна*. При большой интенсивности возбуждающей электромагнитной волны в нелинейной среде, кроме спонтанного рассеяния, возможно *вынужденное комбинационное рассеяние Мандельштама — Бриллюэна*. При этом одновременно с возбуждением когерентной световой волны происходит возбуждение когерентных акустических колебаний с частотой порядка 10^{10} Гц (гиперзвук). Избыточное звуковое давление в поле лазерного луча может достигать нескольких тысяч мегапаскалей.

7. Кроме рассмотренных эффектов, связанных с нелинейностью поляризуемости среды, существуют нелинейные оптические эффекты, обусловленные нелинейным поглощением, т. е. эффектами, определяемые зависимостью показателя оптического поглощения k_{ω} от интенсивности света. В частности, уменьшение показателя поглощения света при увеличении его интенсивности называется *оптическим просветлением среды*. Этот эффект вызван изменением населенностей уровней под действием внешнего излучения. Для двухуровневой системы это видно из рис. 2.2. На основе эффекта самопросветления работают так называемые *пассивные модуляторы добротности* лазеров.

8. При большой интенсивности света возможны процессы двухфотонного (в общем случае — многофотонного) поглощения. Сущность этих процессов сводится к следующему. Поглощая фотон $\hbar\omega_1$, атом из начального состояния E_n переходит в виртуальное состояние E_l . Взаимодействуя со вторым фотоном $\hbar\omega_2$, он переходит в состояние E_m . Если выполняется условие $(\hbar\omega_1 + \hbar\omega_2) = E_m - E_n$, то атом перейдет из одного стационарного состояния E_n в другое E_m . В частном случае $\hbar\omega_1 = \hbar\omega_2$ имеем $\hbar\omega_1 = (E_m - E_n)/2$. Вероятность двухфотонных процессов пропорциональна квадрату интенсивности света и при обычных интенсивностях мала.

ГЛАВА 6 МАЗЕРЫ

Мазеры — это квантовые усилители и генераторы СВЧ-радиодиапазона, т. е. приборы, работающие в области длин волн больше 1 мм. Расстояние между рабочими энергетическими уровнями активной среды должно быть мало (≤ 1 мэВ). Мазеры были первыми приборами, работающими на квантовых принципах. Хотя в современной технике мазеры используются реже, чем приборы оптического диапазона, но они сыграли принципиально важную роль в развитии квантовой электроники и до сих пор с успехом применяются в радиофизике, радиоастрономии, метрологии. Квантовые приборы радиодиапазона обладают двумя существенными достоинствами по сравнению со своими классическими аналогами: 1) низким уровнем собственных шумов, что позволяет их использовать в качестве первых, наиболее важных каскадов СВЧ-приемников (квантовые парамагнитные усилители) и 2) высокой стабильностью частоты, что определяет их применение в качестве стандартов частоты (пучковые мазеры).

§ 6.1. ПУЧКОВЫЕ МАЗЕРЫ

Молекулярный мазер на пучке молекул аммиака. Аммиачный мазер — это первый прибор, принцип действия которого основан на использовании вынужденного излучения для усиления и генерации когерентных электромагнитных волн. Принцип его действия сегодня представляется весьма простым. Но напомним, что за создание именно этого прибора А. М. Прохорову, Н. Г. Басову и Ч. Таунсу была присуждена Нобелевская премия.

Молекула аммиака NH_3 представляет собой правильную пирамиду, основание которой образовано ядрами водорода, а в вершине находится ядро атома азота. Такой тип молекулы называется *симметричным волчком*. Атом азота может занимать два устойчивых положения: над плоскостью атомов водорода ($+z_0$) и под этой плоскостью ($-z_0$), как показано на рис. 6.1. Этими состояниям отвечают два минимума на кривой потенциальной

энергии (рис. 6.2). Очевидно, что точка $z=0$ (атом азота находится в плоскости атомов водорода) есть точка неустойчивого равновесия и ей соответствует максимум потенциальной энергии. При $z \rightarrow \infty$ взаимодействие отсутствует и $U \rightarrow 0$. Переход молекулы из состояния ($+z_0$) в состояние ($-z_0$) называется инверсией: молекула как бы «выворачивается наизнанку». Обе конфигурации молекулы эквивалентны и имеют одинаковые вращательные и колебательные уровни энергии. Наличие двух близко расположенных минимумов на кривой $U(z)$ приводит к расщеплению каждого из этих уровней на два. В качестве примера на рис. 6.2 показано инверсионное расщепление колебательных уровней с разными квантовыми числами q . Аналогично происходит расщепление вращательных уровней энергии. Это расщепление называется инверсионным. Симметрия волновых функций верхнего и нижнего инверсионных уровней различна. Верхние уровни образуют антисимметричные $\Psi_a(z) = -\Psi_a(-z)$, а нижние — симметричные $\Psi_s(z) = \Psi_s(-z)$ состояния. Переходы между этими состояниями разрешены в дипольном приближении как переходы между состояниями с различной симметрией. Расстояние между инверсионными подуровнями мало. Для наинизшего состояния $\Delta \tilde{\nu} \approx 0,8 \text{ см}^{-1}$, что соответствует $\lambda \approx 1,25 \text{ см}$. С увеличением колебательного квантового числа q инверсионное расщепление возрастает.

При комнатной температуре вследствие малости $\Delta \tilde{\nu}$ приблизительно половина молекул NH_3 находится в симметричном, а половина — антисимметричном состоянии. Если каким-либо путем удастся пространственно разделить молекулы, находящиеся в состоянии 0_a от молекул в состоянии 0_s , то в объеме с молекулами 0_a мы получим состояние с инверсионной заселенностью уровней. Оказы-

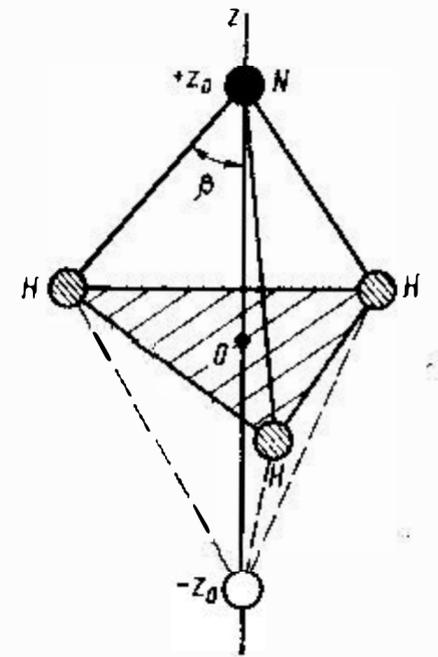


Рис. 6.1. Геометрия молекулы аммиака NH_3 . Состояния ($+z_0$) и ($-z_0$) соответствуют двум возможным устойчивым конфигурациям молекулы. Расстояние между ядрами атомов водорода $r_{\text{HН}} = 10,14 \text{ нм}$, а угол $\beta = 67^\circ 58'$ для молекулы $^{14}\text{N} \text{H}_3$

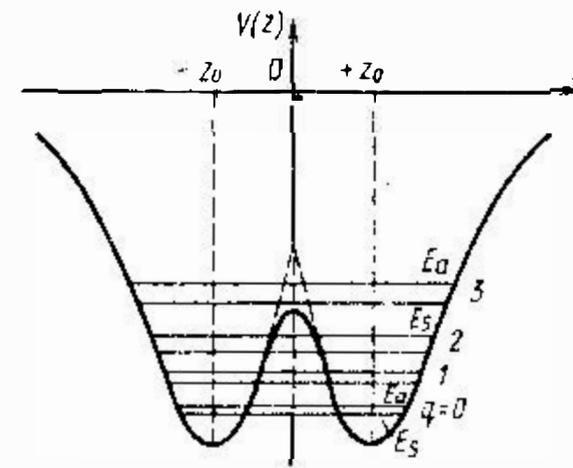


Рис. 6.2. Потенциальная кривая и некоторые собственные значения энергии молекулы аммиака

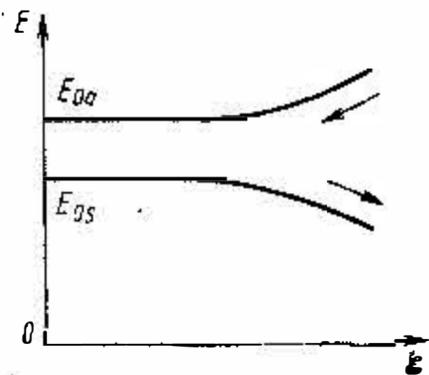


Рис. 6.3. Эффект Штарка для нижних энергетических уровней молекулы аммиака

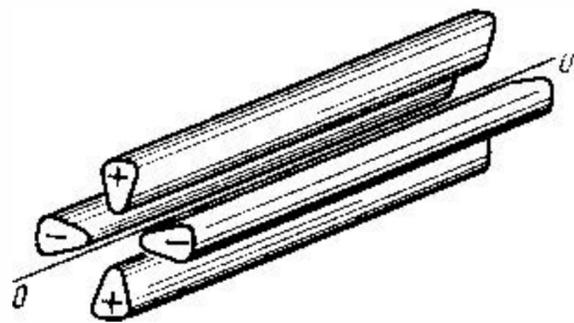


Рис. 6.4. Квадрупольный конденсатор

валяется, что разделить молекулы NH_3 в симметричном и антисимметричном состояниях можно в неоднородном электрическом поле. Действительно, под действием внешнего электрического поля происходит смещение уровней (эффект Штарка), как показано на рис. 6.3. В неоднородном поле, стремясь к минимуму потенциальной энергии, молекулы в антисимметричном состоянии будут двигаться в область минимальной напряженности поля, а молекулы в симметричном состоянии — в область максимальной E . Тогда функцию сортирующей системы может выполнять квадрупольный конденсатор, изображенный на рис. 6.4. В его центре — на оси конденсатора — напряженность поля равна нулю. При удалении от оси она возрастает.

Структурная схема мазера на пучке молекул аммиака показана на рис. 6.5. От источника 1 с помощью системы диафрагм 2 формируется молекулярный пучок 3 с малой угловой расходимостью, который пропускается вдоль оси квадрупольного конденсатора 4, выполняющего роль сортирующей системы. Молекулы NH_3 , находящиеся в нижнем состоянии E_{00} , выталкиваются полем, а молекулы в верхнем рабочем состоянии E_{01} втягиваются к оси конденсатора и через отверстие попадают в резонатор 5, настроенный на рабочую частоту, $\omega = (E_{01} - E_{00})/\hbar$. Если время пролета молекул меньше времени их релаксации, то в резонаторе реализуется состояние с инверсной населенностью. Резонатор, как обычно, выполняет функции обратной связи. Взаимодействие с электромагнитным полем, отсортированный пучок молекул NH_3 за счет вынужденных переходов будет поддерживать электромагнитные колебания, когерентные с полем в резонаторе. Такая настройка (юстировка) резонатора осуществляется настроенным винтом, а отбор мощности — петлей связи 6. Отработанные молекулы выводятся из системы.

Частота генерации определяется энергетическим расстоянием между используемыми инверсионными подуровнями и равна 23870 МГц. Поскольку добротность спектральной линии здесь существенно выше добротности резонатора, то рабочая частота будет определяться в основном свойствами молекулы NH_3 и при-

бор может служить эталоном (стандартом) частоты.

Выходная мощность генератора очень мала и не превышает 10^{-9} Вт. Она ограничена тем, что для работы можно использовать пучок NH_3 очень малой плотности, чтобы избежать столкновения частиц между собой и устранить влияние эффекта Доплера (такой пучок называется молекулярным пучком). Минимальный путь, который должен пройти пучок, не разрушаясь, определяется размером сортирующей системы и резонатором и составляет 20...30 см.

Это вызывает необходимость поддерживать в приборе вакуум на уровне 10^{-4} Па. Так как собственная частота колебательных и вращательных движений молекулы зависит от ее приведенной массы, то в качестве рабочих отбираются молекулы NH_3 , содержащие изотоп N^{14} . Относительная нестабильность частоты аммиачного мазера составляет $10^{-8} \dots 10^{-10}$.

Мазер на пучке атомов водорода. Принцип действия квантового генератора на пучке атомов водорода аналогичен принципу действия аммиачного мазера, но как стандарт частоты он обладает рядом преимуществ. В водородном мазере используется магнитный дипольный переход между уровнями сверхтонкой структуры атома водорода. Эта структура обязана своим появлением взаимодействию магнитного момента ядра атома с внутриатомным магнитным моментом, вызванным орбитальным движением электрона. В атоме водорода в результате такого взаимодействия основное состояние расщепляется на два подуровня с расстоянием между ними по частоте 1420,4 МГц ($\lambda \approx 21$ см). Переходы между этими подуровнями определяются изменением магнитного дипольного момента, а не электрического, как в аммиачном мазере. Квадрат модуля матричного элемента магнитного дипольного момента примерно в 10^4 раз меньше, чем квадрат модуля матричного элемента электрического дипольного момента. Примерно во столько же раз меньше вероятность перехода и уже естественная ширина спектральной линии (см. § 2.6), а следовательно, при прочих равных условиях выше стабильность частоты. Повышению стабильности способствует также то обстоятельство, что линия рабочего перехода в водородном мазере «чистая», т. е. не имеет дополнительной структуры.

В общих чертах устройство генератора на пучке атомов водорода аналогично устройству аммиачного мазера, за исключением следующих особенностей:

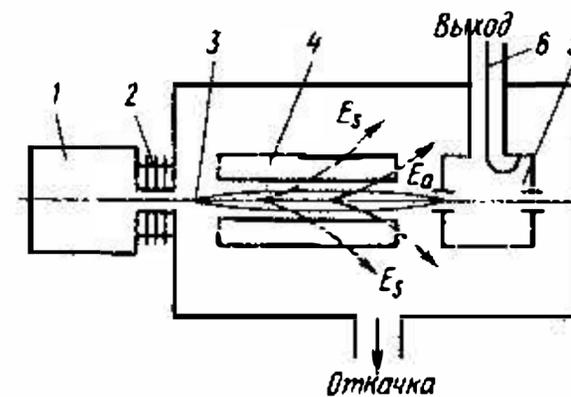


Рис. 6.5. Блок-схема квантового генератора на пучке молекул аммиака: 1 — источник NH_3 ; 2 — система диафрагм; 3 — молекулярный пучок; 4 — сортирующая система; 5 — объемный резонатор; 6 — петля связи

1. Источник пучка атомов водорода Н содержит дополнительное устройство для диссоциации молекул водорода H_2 на атомы. Обычно это осуществляется в электрическом разряде.

2. Сортировка атомов производится не в электрическом, а в неоднородном магнитном поле магнитной сортирующей системы.

3. Поскольку вероятность перехода между рабочими уровнями мала ($\tau \approx 1$ с) то необходимо существенно (примерно в 10^4 раз) увеличить число активных частиц в резонаторе. Добиться этого простым увеличением плотности потока атомов невозможно. Увеличение среднего времени пребывания частиц в резонаторе водородного лазера достигается путем применения специальной накопительной колбы с небольшим отверстием для влета активных частиц. Эта колба размещается внутри резонатора. Чтобы соударения со стенками колбы не приводили к изменению состояний активных атомов (т. е. к безызлучательной релаксации), на внутреннюю поверхность стенок наносится специальное тефлоновое покрытие. Диаметр колбы делается меньше длины волны излучения, в результате чего уширение спектральной линии, вызванное эффектом Доплера первого порядка, не возникает.

Квантовые генераторы на пучке атомов водорода работают на длине волны $\lambda \approx 21$ см ($\nu_{\text{раб}} = 1420,405$ МГц) с выходной мощностью около 10^{-11} Вт при относительной стабильности частоты порядка 10^{-12} .

§ 6.2. КВАНТОВЫЕ ПАРАМАГНИТНЫЕ УСИЛИТЕЛИ

Работа квантовых парамагнитных усилителей (КПУ) основана на усилении слабых СВЧ-сигналов вынужденным излучением, вызванным электронными переходами между энергетическими уровнями парамагнитных ионов в диэлектриках, расщепленными внешними или внутренними полями. Как правило, используются квантовые переходы между зеемановскими подуровнями парамагнитных ионов в кристаллах во внешнем магнитном поле.

Напомним, что под действием внешнего магнитного поля происходит квантование магнитного момента атома, в результате чего каждый из уровней с определенным квантовым числом J расщепляется на $(2J+1)$ подуровней с разными значениями проекции M_J на направление поля: магнитное поле расщепляет вырожденные по квантовому числу M_J уровни (§ 1.2). Расщепление $\Delta E = -\mu_J B = g_J \mu_B M_J B$ пропорционально внешнему магнитному полю B и магнитному моменту атома μ_J . Между этими уровнями в дипольном приближении разрешены переходы, удовлетворяющие следующим правилам отбора:

$$\Delta J = 0, \pm 1; \Delta M_J = 0, \pm 1.$$

Переходы с $\Delta J = \pm 1$ и $\Delta M_J = 0, \pm 1$ попадают в оптическую область спектра и определяют сущность эффекта Зеемана. Переходы между подуровнями одного уровня с $\Delta J = 0$ и $\Delta M_J = \pm 1$ лежат в СВЧ-радиодиапазоне и определяют сущность электронного парамагнитного резонанса.

В отличие от изолированных атомов в кристаллах типа $NaCl$, Al_2O_3 и т. п. орбитальные и спиновые моменты отдельных атомов связаны и взаимно компенсируют друг друга, вследствие чего такие кристаллы являются диамагнетиками. Если в эти кристаллы ввести ионы с незаполненными внутренними оболочками (элементы группы железа: V^{3+} ; Cr^{3+} , Mn^{2+} , Co^{2+} , Fe^{3+} ... или редкоземельные элементы — лантаноиды), то появятся нескомпенсированные магнитные моменты (нескомпенсированные спины) и кристаллы становятся парамагнетиками. Именно такие кристаллы используются в качестве активных элементов КПУ. Рабочие переходы осуществляются между зеемановскими подуровнями парамагнитного иона-активатора. Кристаллическая основа, выполняя роль матрицы, со своей стороны влияет как на энергетический спектр иона-активатора, так и на процессы обмена энергией (релаксационные процессы). Это естественно, поскольку активный ион находится под воздействием сильного внутрикристаллического поля. Это воздействие тем больше, чем слабее экранированы внешними электронами внутренние незаполненные $3d$ - или $4f$ -оболочки, внутри которых собственно и происходят рабочие переходы (подробнее см. в § 8.1).

Накачка, обеспечивающая инверсную населенность между рабочими уровнями КПУ, осуществляется вспомогательным СВЧ-излучением. Поэтому по двухуровневой схеме, как это было в случае пучковых мазеров, прибор работать не может. Как правило, работа происходит по трехуровневой схеме первого или второго типов (рис. 2.7).

В качестве примера активного диэлектрика КПУ рассмотрим рубин, представляющий собой кристалл окиси алюминия Al_2O_3 , в котором часть атомов алюминия изоморфно замещена атомами хрома ($Al_2O_3:Cr^{3+}$). Чистые бесцветные кристаллы Al_2O_3 , называемые *а-корундом* и иногда не совсем точно *сапфиром*^{*}, обладают кубической симметрией. Каждый ион алюминия Al^{3+} находится в окружении шести ионов кислорода O^{2-} , образующих октаэдр. Окружение иона кислорода составляют четыре иона алюминия, которые образуют тетраэдр. Корунд — бесцветный кристалл, прозрачный в оптической области от 0,17 до 6,5 мкм и в СВЧ-диапазоне. Без легирующих примесей он обладает диамагнитными свойствами. Радиус иона хрома (0,065 нм) больше радиуса иона алюминия (0,057 нм). Поэтому при легировании хромом параметры решетки изменяются. Ион хрома, замещающий алюминий, оказывается не в центре октаэдра ионов O^{2-} решетки корунда, а несколько смещенным вдоль тригональной оси. Следствием этого является возникновение напряжений

^{*}Название «сапфир» от греч. sappheiros обозначает синий камень. Этот цвет корунду придают ионы титана, содержащиеся в виде примесей. Минералоги прошлого использовали название «сапфир» в более широком смысле для обозначения всех разновидностей драгоценных корундов, включая и рубин.

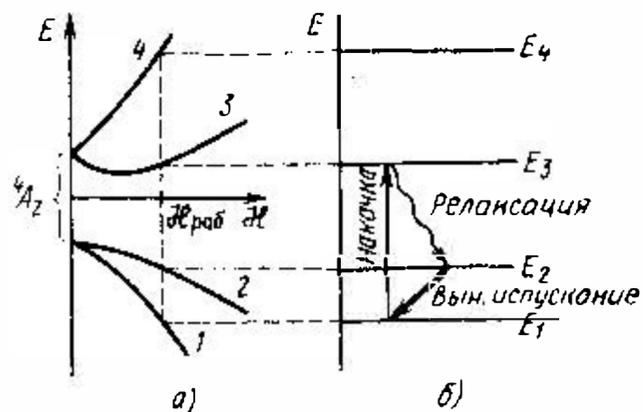


Рис. 6.6. Расщепление основного энергетического состояния 4A_2 иона хрома Cr^{3+} в рубине при перпендикулярной ориентации магнитного поля (а) и схема работы трехуровневого КПУ (б)

В отличие от α -корунда кристалл рубина анизотропен и обладает ромбоэдрической симметрией (пространственная группа D_{3d}^6). Оптическая ось кристалла (ось C) совпадает с осью симметрии третьего порядка.

Основное состояние иона Cr^{3+} в рубине — 4A_2 . Под действием магнитного поля оно расщепляется на четыре зеемановских подуровня, как показано на рис. 6.6, а. Это расщепление заметно отличается от простых линейных зависимостей ($\Delta E \sim H$), что обусловлено влиянием сильного кристаллического поля. При $H=0$ это кристаллическое поле расщепляет четырехкратно вырожденное состояние 4A_2 на два двукратно вырожденных. Вследствие анизотропии кристаллического поля расстояние между магнитными подуровнями иона Cr^{3+} в рубине существенно зависит не только от величины магнитного поля, но и от его ориентации относительно оптической оси кристалла. В КПУ наиболее часто используют ориентацию магнитного поля под углами $\theta=90^\circ$ (перпендикулярная ориентация) и $\theta=54^\circ 44'$ к тригональной оси кристалла. В последнем случае энергетические уровни смещаются симметрично под действием магнитного поля (рис. 6.7, а).

Инверсия населенности может быть достигнута между различными подуровнями состояния 4A_2 иона Cr^{3+} в рубине. В слабых полях (200...700 Гс) для усиления используется переход между уровнями 4 — 3, а вспомогательное излучение накачки осуществляет переход 2 — 4. В этом режиме реализуется усиление в дециметровом диапазоне длин волн от $\lambda=10$ до 30 см. При этом длина волны излучения накачки меняется слабо — от 2,2 до 2,5 см. Перестройка рабочего диапазона производится магнитным полем.

Для работы в сантиметровом диапазоне ($\lambda=3...10$ см) реализуют другой режим работы КПУ на рубине, создавая инвер-

и структурных искажений, возрастающих с увеличением концентрации хрома и приводящих к понижению симметрии кристалла.

В мазерной и лазерной технике, как правило, используется бледно-розовый рубин с содержанием хрома около 0,05%, что соответствует $1,6 \cdot 10^{19}$ ионов Cr^{3+} в 1 см^3 . Увеличение содержания хрома изменяет цвет кристалла до красного и вишнево-красного*. В отличие

от рубина, а накачку осуществляя за счет переходов 1 — 3, как это показано на рис. 6.6, б. В этом режиме требуются большие магнитные поля.

Инверсию населенностей можно получить также между уровнями 3 и 2, используя так называемый «пушпульный» режим накачки (рис. 6.7, б). Он применяется в приборах, работающих в диапазоне $\lambda < 5$ см. Для его реализации магнитное поле должно быть направлено под углом $\theta=54^\circ 44'$ к тригональной оси кристалла рубина. При такой ориентации расстояния между уровнями 4 — 2 и 3 — 1 одинаковы и излучение накачки осуществляет переходы как между уровнями 1 — 3, так и между уровнями 2 — 4 (рис. 6.7, б), что приводит к увеличению населенности уровня 3 и уменьшению населенности уровня 2.

Для работы КПУ требуется охлаждение до температур жидкого гелия (4,2 К при атмосферном давлении и 1,5...1,8 К при откачке паров гелия до $P=5 \cdot 10^2...3 \cdot 10^3$ Па). Необходимость такого глубокого охлаждения вызвана тем, что уменьшение температуры препятствует термическому заселению рабочих уровней энергии; увеличивает времена спин-решеточной релаксации и тем самым уменьшает вероятность безызлучательных переходов между рабочими уровнями; уменьшает уровень собственных шумов усилителя.

Как прибор КПУ может быть выполнен либо в виде усилителя бегущей волны, либо в виде резонаторного усилителя. Квантовый парамагнитный усилитель бегущей волны в сущности представляет собой отрезок волновода, заполненный активной средой. Естественно, что в активной среде должна быть обеспечена такая степень инверсии населенности между рабочими мазерными уровнями, при которой усиление за счет вынужденного излучения превышает все потери. Для увеличения времени взаимодействия поля сигнала с активной средой и соответственно увеличения коэффициента усиления в КПУ бегущей волны используют специальные замедляющие системы, например, штыревого типа. Эти усилители работают в режиме усиления малых сигналов и способны регистрировать отдельные фотоны. Эффективность усиления, равная произведению полосы пропускания на корень квадратный из коэффициента усиления, составляет для них несколько сотен мегагерц, достигая ГГц.

В резонаторных усилителях для повышения эффективности взаимодействия поля сигнала с активной средой ее помещают в резонатор. Резонатор должен быть настроен как на частоту сигнала, так и на частоту накачки. Резонаторные КПУ в свою очередь могут быть двух типов: проходного и циркуляторного (отражательного). В проходном резонаторном усилителе вход и выход усиленного сигнала разнесены. В циркуляторном (отражательном) усилителе входной

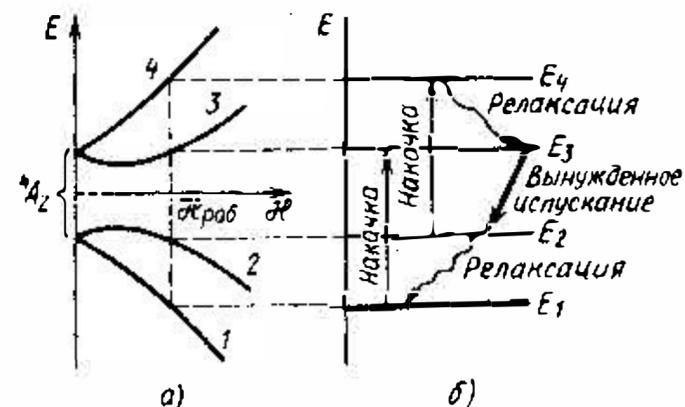


Рис. 6.7. Расщепление основного энергетического состояния 4A_2 иона хрома Cr^{3+} в рубине при ориентации магнитного поля под углом $\theta=54^\circ 44'$ к тригональной оси кристалла (а) и схема работы КПУ при пушпульной накачке (б)

*Название «рубин» происходит от лат. ruber — красный.

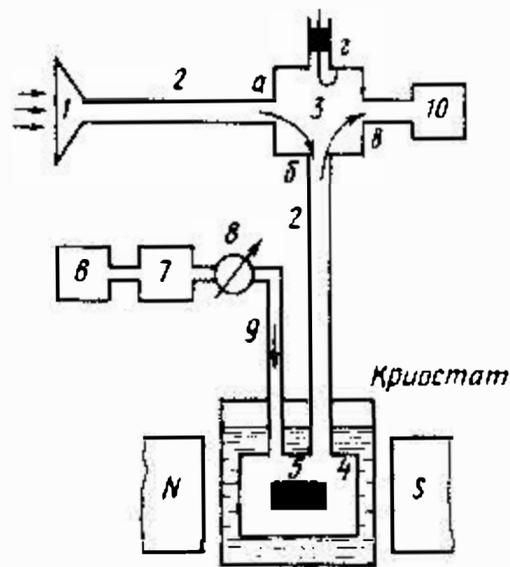


Рис. 6.8. Структурная схема резонаторного КПУ циркуляторного типа:

1 — приемная антенна; 2 — волноводный тракт сигнала; 3 — ферритовый циркулятор; 4 — резонатор; 5 — парамагнитный кристалл; 6 — источник накачки; 7 — ферритовый клапан; 8 — аттенуатор; 9 — волноводный тракт накачки; 10 — выходное устройство

пропускающим СВЧ-излучение только в одном направлении. Распространение сигнала в нем возможно из плеча «а» только в плечо «б», а из плеча «б» только в плечо «в». Для гашения паразитных колебаний, которые могут возникать из-за неполного согласования цепей и из-за отражения от неоднородностей, в плечо «б» циркулятора включается согласованная нагрузка.

Активный кристалл помещают в однородное магнитное поле и вместе с резонатором охлаждают в специальном криостате. Внешнее магнитное поле напряженностью (0,1...5) кГс обеспечивает нужное расщепление между энергетическими уровнями активного вещества. Оно создается магнитами со сверхпроводящими обмотками, которые размещаются в криостате с жидким гелием. Открытие явления высокотемпературной сверхпроводимости (при $T \geq 90$ К) в ряде керамических материалов, вероятно, позволит в ближайшее время разработать сверхпроводящие магниты, охлаждаемые жидким азотом.

Для глубокого охлаждения КПУ применяют специальные криостаты. Наиболее удобны гелиевые ожижители непрерывного действия с замкнутым циклом. Ранее охлаждение проводилось путем погружения КПУ в сосуд Дьюара с жидким гелием, который в свою очередь помещался внутри второго сосуда Дьюара, залитого жидким азотом.

Основное преимущество квантовых парамагнитных усилителей — низкий уровень их собственных шумов. Шумы усилителя принято оценивать по той мощности излучения $P = kT_{ш}\Delta\nu$, которую надо подать на вход идеального шумящего усилителя

и выходной сигналы распространяются по одному волноводу в противоположных направлениях и разделяются с помощью специального циркулятора. Усилители отражательного типа обладают большей эффективностью по сравнению с проходными усилителями, так как в проходном усилителе мощность усиленного сигнала распределяется между входным и выходным волноводами и часть ее теряется. Блок-схема резонаторного КПУ циркуляторного типа представлена на рис. 6.8. Входной сигнал, улавливаемый приемной антенной 1 по волноводному тракту сигнала 2 через ферритовый циркулятор 3 направляется в резонатор 4, в который помещен активный парамагнитный кристалл 5. Вспомогательное СВЧ-излучение заданной частоты от источника накачки 6, в качестве которого обычно применяют клистронный генератор, через ферритовый клапан 7, пропускающий излучение лишь в одном направлении, и аттенуатор 8, регулирующий мощность накачки, по волноводному тракту накачки 9 поступает в резонатор 4. Сигнал, усиленный активным элементом 5, возвращается по волноводному тракту сигнала 2 и через ферритовый циркулятор 3 поступает в выходное устройство 10 (например, на вход обычного СВЧ-приемника). Ферритовый циркулятор 3 является невзаимным элементом,

с полосой пропускания $\Delta\nu$, чтобы получить на выходе сигнал, равный шуму реального усилителя. Величина $T_{ш}$ называется эффективной шумовой температурой усилителя.

Кроме эффективной шумовой температуры для характеристики шумовых свойств усилителя вводится понятие шум-фактора $F_{ш}$, который равен отношению полной мощности шумов к мощности чисто тепловых шумов на выходе усилителя. При этом считается, что тепловые шумы создаются входным генератором, находящимся при $T = 290$ К, т. е. при комнатной температуре. Для идеального шумящего усилителя $F_{ш} = 1$. Шум-фактор связан с эффективной шумовой температурой соотношением

$$T_{ш} = (F_{ш} - 1) 290 \text{ К.}$$

В обычных СВЧ-электронных приборах, работающих на классических принципах, процесс усиления происходит при взаимодействии электромагнитного поля с потоком электронов. В электронном потоке всегда существуют так называемые дробовые шумы, определяющие внутренние шумы классических усилителей и ограничивающие их чувствительность. Для приемников сантиметрового и дециметрового диапазонов они эквивалентны шумовой температуре $T_{ш} \approx 1000 \dots 3000$ К. В приборах, работающих на квантовых принципах, эти шумы отсутствуют.

Основными источниками шумов КПУ являются спонтанное излучение атомов активного вещества и тепловое излучение от элементов усилителя (стенок волновода, резонатора, циркулятора, шумов антенны и т. п.). Тепловое излучение описывается функцией Планка (1.107), откуда находим мощность тепловых шумов в единичном интервале частот, равную средней энергии, приходящейся на один тип колебаний при температуре T :

$$P_{\Delta\nu} = \frac{h\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1}$$

Тепловой шум можно уменьшить путем охлаждения резонатора и стенок волновода. Шум, обусловленный спонтанными переходами, в СВЧ-диапазоне также мал, поскольку коэффициент Эйнштейна A_{mn} пропорционален кубу частоты [см. (1.106) и (1.143)]. Если в видимой области спектра на частоте $\approx 10^{15}$ Гц для разрешенных в дипольном приближении переходов $A_{mn} \approx 10^8 \text{ с}^{-1}$ (§ 1.5), то в СВЧ-диапазоне на частоте ≈ 10 ГГц ($\lambda = 3$ см) эта величина на 15 порядков меньше. Обусловленная спонтанным излучением $T_{ш}$ в СВЧ-диапазоне не превышает (1...5) К.

При рациональной конструкции элементов шумовая температура для усилителей резонаторного типа составляет 20...100 К,

а для усилителей бегущей волны — 10 К и меньше. Это определяет основную область применения КПУ как низкошумящих усилителей в приемных системах сверхдальних линий связи, в радиотелескопах и радиолокаторах в радиоастрономии.

Кроме рубина в качестве активного вещества КПУ применяют рутил (TiO_2), легированный хромом ($\text{TiO}_2 : \text{Cr}^{3+}$) или железом ($\text{TiO}_2 : \text{Fe}^{3+}$), а также $\text{Al}_2\text{O}_3 : \text{Fe}^{3+}$. В этих веществах расщепление энергетических уровней, вызванное кристаллическим полем, больше, чем в рубине, и потому они используются для работы в более коротковолновом диапазоне. Современные КПУ перекрывают диапазон длин волн от 3 мм до 1 м.

ГЛАВА 7 ГАЗОВЫЕ ЛАЗЕРЫ

Лазеры (квантовые усилители и генераторы оптического диапазона) — наиболее важные и широко применяемые приборы квантовой электроники. Это первые и пока что единственные источники интенсивного оптического излучения, обладающего высокой степенью когерентности, монохроматичности и направленности (см. § 2.6). Первый лазер был создан в 1960 г. Т. Мейманом на основе рубина. В конце этого же года был изготовлен первый газовый лазер на смеси газов гелия и неона, в 1962 г. — первый полупроводниковый лазер на арсениде галлия и в 1966 г. — первый жидкостный лазер на органическом красителе. За прошедший сравнительно небольшой период времени появилось много новых типов лазеров, использующих различные активные вещества и перекрывающих по спектру почти весь оптический диапазон. Конструкции лазеров были существенно улучшены и, что самое главное, были найдены оптимальные активные материалы, благодаря чему эффективность приборов (КПД) возросла от сотых долей до десятков процентов (30...50% для полупроводниковых лазеров на гетероструктурах, молекулярных CO_2 - и CO -газовых лазеров и некоторых типов химических лазеров).

В зависимости от типа используемого активного вещества различают: газовые, твердотельные (на твердых диэлектриках), жидкостные и полупроводниковые лазеры.

Хотя по агрегатному состоянию активного вещества полупроводниковые лазеры формально можно отнести к твердотельным, особенности происходящих в них процессов, а также особенности их конструирования, изготовления и эксплуатации требуют вынесения этих приборов в отдельную группу.

§ 7.1. ОБЩАЯ ХАРАКТЕРИСТИКА И ОСОБЕННОСТИ ГАЗОВЫХ ЛАЗЕРОВ

Активное вещество газовых лазеров находится в газообразном состоянии. Это могут быть как обычные газы (вещества, находящиеся в газообразном состоянии при нормальных услови-

ях), так и пары различных веществ (например, металлов) и смеси. В газовых лазерах в качестве рабочих могут быть использованы весьма разнообразные переходы между электронными колебательными и вращательными уровнями энергии нейтральных или ионизованных атомов и молекул. Это определяет весьма широкий спектральный диапазон их работы, включая ультрафиолетовую, видимую, инфракрасную и субмиллиметровую области. На рис. 7.1 показаны длины волн, на которых генерируются наиболее распространенные типы газовых лазеров. В общей сложности в настоящее время получена генерация более чем в 1000 различных переходах в газах и газовых смесях в интервале длин волн от $\sim 0,1$ мкм до ~ 1 мм и более. Только газовые среды могут быть прозрачны в таком широком диапазоне.

Газ, как активная среда, обладает высокой оптической однородностью. Вследствие слабого взаимодействия между активными частицами уширение уровней энергии в газах мало, а спектральные линии излучения — узкие. Поэтому излучение, генерируемое газовыми лазерами, обладает наивысшим «качеством» по сравнению с другими типами лазеров: диаграмма направленности узкая и определяется свойствами резонатора; излучение по спектру сосредоточено в одной или нескольких модах и сравнительно просто реализуется одночастотный режим работы. При принятии специальных мер в газовых лазерах может быть достигнута относительная стабильность частоты излучения на уровне $10^{-13} \dots 10^{-14}$, т. е. лучше, чем в пучковых лазерах, работающих в качестве стандартов частоты. С другой стороны, узость спектральных линий не дает возможность получать в газовых лазерах сверхкороткие импульсы света (§ 2.5) и перестраивать диапазон их работы. (Это не относится к газовым лазерам высокого давления.)

В газовых средах существует большое число процессов, приводящих к обмену энергией между частицами. Это приводит к возможности осуществлять различные методы накачки для создания инверсной населенности между рабочими лазерными уровнями активных частиц. По методам накачки газовые лазеры разделяют на газоразрядные, газодинамические и химические. В принципе в газовых лазерах может быть использована и оптическая накачка, но для ее реализации необходимы источники излучения в узком спектральном интервале, совпадающем со спектром поглощения активного газа. Поэтому в газовых лазерах она неэффективна.

Наиболее широкое распространение получили газоразрядные лазеры, которые в свою очередь подразделяют на три группы: атомарные, ионные и молекулярные. Газовый разряд, как правило, создается непосредственно в самой активной среде, хотя возможны схемы работы, при которых разряд осуществляется во вспомогательном газе, а затем атомы вспомогательного газа

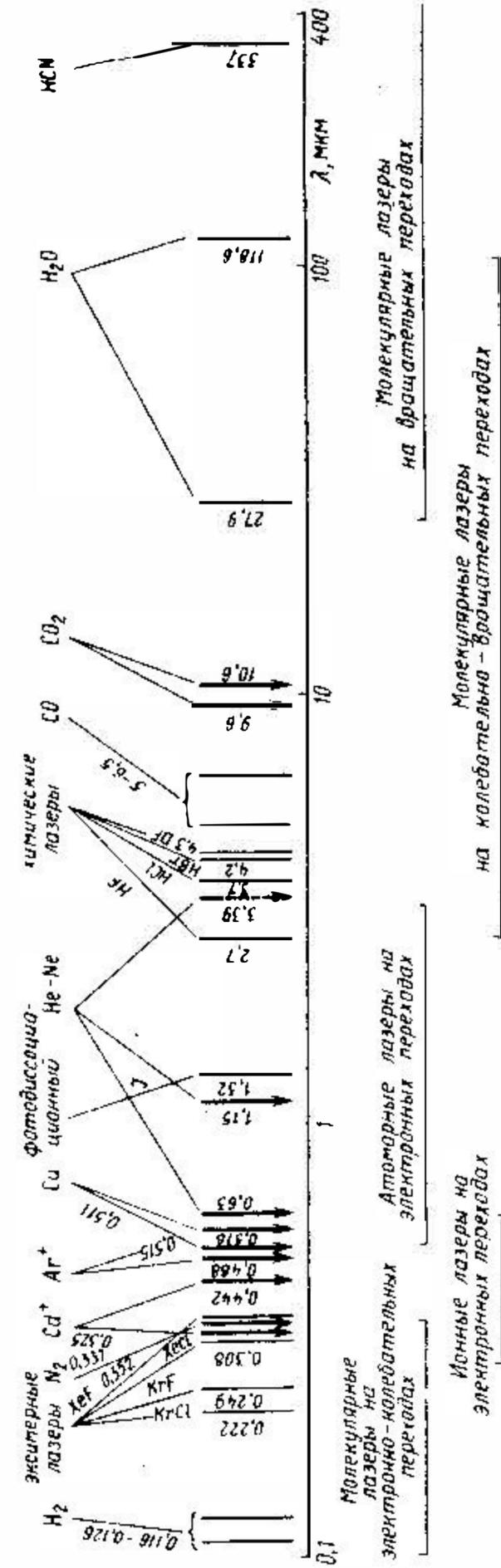


Рис. 7.1. Длины волн генерации газовых лазеров. Жирными стрелками обозначены наиболее распространенные лазеры, выпускаемые промышленностью

в рабочем объеме передают свою энергию атомам рабочего газа. Используют различные виды газовых разрядов: самостоятельный и несамостоятельный, импульсный и стационарный, дуговой и тлеющий, высокочастотный разряд и разряд на постоянном токе.

При *несамостоятельном разряде* ионизация газа производится внешним воздействием, а процесс возбуждения осуществляется независимо от условий пробоя при оптимальном значении напряженности электрического поля в разрядном промежутке. Такой разряд необходимо применять при больших давлениях газа в рабочем объеме, когда зажечь самостоятельный разряд бывает трудно.

Для *импульсного разряда* существенны как крутизна фронтов импульсов тока, так и интервал времени между ними. Разряд считается импульсным, если время нарастания (спада) импульса тока меньше характерного времени заселения (опустошения) рабочих состояний, а период времени между токовыми импульсами существенно больше времени опустошения рабочих состояний в плазме газового разряда. В этом случае инверсия населенности возникает либо на фронте, либо на спаде импульса тока, а сами процессы создания инверсии связаны со свойствами нестационарной плазмы. В противном случае, при плавном нарастании и спаде токового импульса, в каждый интервал времени в плазме успевает установиться стационарное (точнее — квазистационарное) состояние. Такой разряд называется квазистационарным.

Для импульсного разряда, используемого в газоразрядных лазерах, характерны высокие значения плотности тока порядка 1 кА/см^2 в импульсе, времена нарастания и спада токового импульса менее 1 мкс при скважности $T_{\text{н}}/\tau_{\text{н}} \geq 1000$.

Стационарный разряд, применяемый для возбуждения газовых лазеров непрерывного или квазинепрерывного действия, разделяют на *дуговой* и *тлеющий*. Для *дугового разряда* характерны большие плотности тока до 1 кА/см^2 , высокая эффективная температура газа, яркий пшур разряда и высокая степень ионизации, достигающая 10% и более. Для *тлеющего разряда*, напротив, характерны небольшие плотности тока порядка $10^{-5} \dots 10^{-4} \text{ А/см}^2$, низкая температура газа, примерно равная температуре окружающей среды, и слабая степень ионизации, составляющая всего $10^{-4} \dots 10^{-2}\%$, так что концентрация заряженных частиц (электронов и ионов) в тлеющем разряде мала.

Используют две разновидности тлеющего разряда: разряд на постоянном токе и *высокочастотный разряд*. При высокочастотном разряде, осуществляемом обычно на частотах $10 \dots 50 \text{ МГц}$, электроды располагаются на внешней стороне разрядной трубки в виде металлических поясков. Это упрощает конструкцию разрядной трубки и создает ряд удобств при ее изготовлении и эксп

луатации. Однако при высокочастотном разряде стенки трубки разогреваются и происходит дополнительное газовыделение в рабочий объем. Поэтому срок службы трубок с высокочастотным разрядом меньше, чем трубок с разрядом на постоянном токе. Кроме того, высокочастотный разряд создает помехи в приемных и других электронных устройствах. Все это привело к тому, что в последнее время в промышленных газоразрядных лазерах используется тлеющий разряд на постоянном токе. В этом случае электроды (катод и анод) располагаются внутри трубки. К ним через балластное сопротивление подводится постоянное напряжение порядка $(1 \dots 2) \text{ кВ}$ на метр разрядного промежутка.

Характерной особенностью газов является многообразие процессов обмена энергией между частицами и возможность управления некоторыми из них путем изменения состава газовой смеси, ее плотности и условий электрического разряда. Все процессы обмена энергией в плазме газового разряда могут быть вызваны взаимодействием частиц (электронов, ионов, атомов и молекул) между собой или их взаимодействием с электромагнитным полем. Последнее, как мы знаем, может быть представлено тремя элементарными актами: спонтанным испусканием, вынужденным испусканием и поглощением кванта электромагнитного поля — фотона.

Элементарные акты обмена энергией при взаимодействии частиц газа между собой разделяют на упругие и неупругие. При упругом взаимодействии (*упругом столкновении*) двух частиц их суммарная кинетическая энергия не меняется. Процессы осуществляются по законам механики взаимодействия абсолютно упругих тел, откуда и возник термин «упругое столкновение». При неупругом взаимодействии (*неупругом столкновении*) суммарная кинетическая энергия взаимодействующих частиц изменяется за счет изменения внутренней (потенциальной) энергии одной из них. Если кинетическая энергия системы убывает, приводя к увеличению внутренней энергии одной из частиц, то такой процесс называется неупругим взаимодействием первого рода. Напротив, если кинетическая энергия возрастает за счет уменьшения внутренней энергии атома, иона или молекулы, то такой процесс называется неупругим взаимодействием второго рода.

Каждый из процессов взаимодействия характеризуется своей вероятностью или сечения взаимодействия. Понятие поперечного сечения взаимодействия с фотоном мы ввели в § 1.8 на примере поглощения фотона. Аналогичным образом вводят *эффективные сечения взаимодействия* для других процессов, в том числе для упругих и неупругих столкновений в плазме газового разряда. Смысл этого термина легко понять из его названия и размерности. Если известно сечение взаимодействия σ_{AB} двух частиц *A* и *B*, движущихся со скоростью v_{AB} относительно друг друга, то

число актов взаимодействия в единице объема в единицу времени равно

$$(dN/dt)_{AB} = N_{AB} = N_A N_B v_{AB} \sigma_{AB}, \quad (7.1)$$

где N_A и N_B — концентрации частиц A и B соответственно.

Поскольку величина сечения взаимодействия зависит и иногда довольно сложным образом от относительной скорости взаимодействующих частиц, а частицы распределены по скоростям, то полное число актов взаимодействия

$$\begin{aligned} N_{AB} &= N_A N_B \langle v_{AB} \sigma_{AB} \rangle = N_A N_B \int_0^{\infty} \sigma_{AB}(v) v_{AB} f_{v_{AB}} dv_{AB} = \\ &= N_A N_B \int_0^{\infty} \sigma_{AB}(E) v_{AB} f_E dE, \end{aligned} \quad (7.2)$$

где $f_{v_{AB}}$ — функция распределения частиц по скоростям v_{AB} , а f_E — функция распределения по энергиям.

При упругом взаимодействии существенным является масса частицы и ее скорость. С этой точки зрения в плазме газовой разряда можно выделить две группы частиц: легкие (электроны) и тяжелые (атомы, молекулы, ионы). Соответственно между ними возможны три типа упругих взаимодействий: 1) взаимодействие между электронами; 2) взаимодействие между тяжелыми частицами и 3) взаимодействие вне между электронами и тяжелыми частицами.

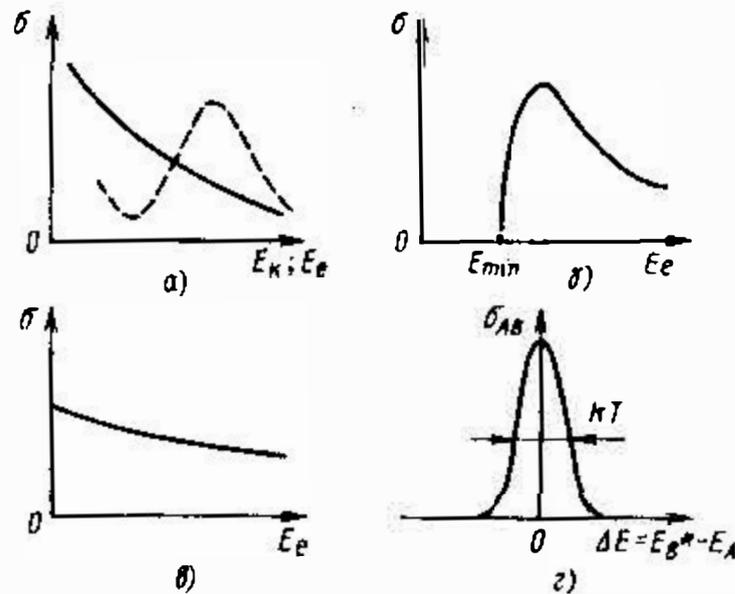


Рис. 7.2. Характерные зависимости сечения взаимодействия от кинетической энергии частиц в газе для процессов:

а) упругого взаимодействия (пунктирной кривой показана зависимость $\sigma(E_e)$ при рассеянии электрона на некоторых атомах инертных газов); б) неупругого взаимодействия первого рода; в) девозбуждения и деполюаризации; г) передачи возбуждения при столкновении

между электронами; 2) взаимодействие между тяжелыми частицами и 3) взаимодействие вне между электронами и тяжелыми частицами. Сечение взаимодействия для всех упругих процессов, как правило, уменьшается при увеличении кинетической энергии E_e (скорости) частицы, как показано сплошной линией на рис. 7.2, а. Для третьего процесса сечение упругого рассеяния электрона на некоторых атомах (например, на атомах инертных газов Хе, Кг, Аг и Ne) и молекулах (типа N_2 ; O_2 ; CO_2) может зависеть от его кинетической энергии более

сложным образом, как показано пунктирной линией на рис. 7.2, а.

В результате упругих взаимодействий за счет обмена энергией и импульсом в стационарном режиме между одинаковыми частицами устанавливается определенное изотропное распределение их по скоростям (энергиям). О едином законе распределения частиц по скоростям имеет смысл говорить в том случае, если характерные размеры сосуда, в данном случае — диаметр газоразрядной трубки, существенно превышают длину свободного пробега частиц. Такой режим называется *диффузионным*. Если при диффузионном режиме преобладают упругие взаимодействия, то там устанавливается максвелловское распределение частиц по скоростям (энергиям).

Напомним, что функция распределения частиц A по скоростям определяет долю общего числа частиц f_v , приходящуюся на интервал скоростей от v_A до $v_A + dv_A$:

$$dN_A/N_A = f_v dv_A. \quad (7.3)$$

При максвелловском распределении функция f_v имеет вид

$$f_v = \frac{4}{\sqrt{\pi}} \frac{v^2}{v_0^3} \exp \left[-\left(\frac{v}{v_0} \right)^2 \right]. \quad (7.4)$$

При $v = v_0$ функция f_v имеет максимум, так что v_0 имеет смысл наиболее вероятной скорости частиц. Она зависит только от температуры среды T и массы частиц M_A :

$$v_0 = \sqrt{2kT/M_A}. \quad (7.5)$$

Напомним, что при максвелловском распределении средняя арифметическая скорость ансамбля частиц равна

$$\bar{v} = \sqrt{(8kT)/(\pi M)}, \quad (7.6)$$

а средняя квадратичная скорость

$$(\bar{v}^2)^{1/2} = \sqrt{3kT/M}. \quad (7.7)$$

Аналогичным образом максвелловская функция распределения частиц по энергиям

$$f_{E_x} = \frac{2}{\sqrt{\pi}} \frac{E_x}{E_0^2} \exp \left(-\frac{E_x}{E_0} \right), \quad (7.8)$$

где $E_0 = kT$, а максимум функции наблюдается при кинетической энергии $E_x = E_0/2 = kT/2$. Функция распределения (7.8) зависит только от температуры. На рис. 7.3 функции распределения (7.4) и (7.8) представлены в обобщенных безразмерных координатах.

Максвелловское распределение реализуется при преобладании процессов упругого взаимодействия. Оно, как правило, выполняется для тяжелых частиц (естественно, при соблюдении диффузионного режима). Распределение электронов может отличаться от максвелловского в силу заметной роли неупругих процессов в их обмене энергией с тяжелыми частицами.

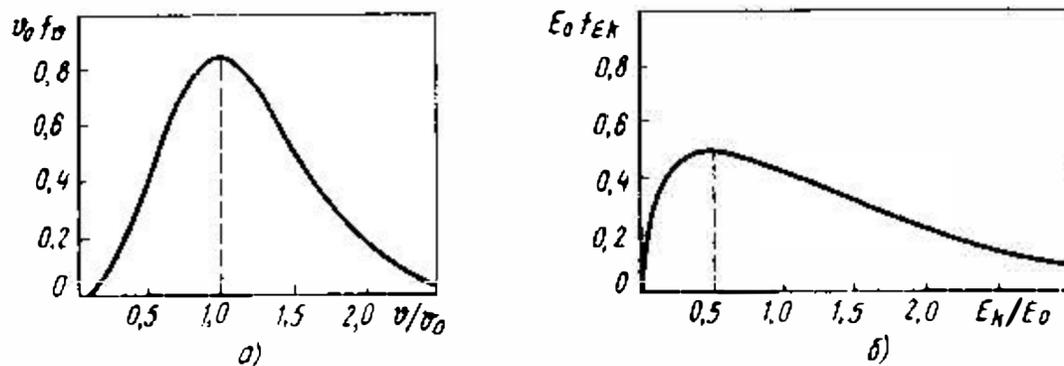


Рис. 7.3. Обобщенные кривые максвелловского распределения по скоростям (а) и энергиям (б)

В отсутствие электрического разряда в газе кинетические энергии всех частиц равны между собой и определяются только температурой окружающей среды. В газовом разряде заряженные частицы ускоряются в электрическом поле, увеличивая свою кинетическую энергию. За счет упругих взаимодействий эта дополнительная энергия передается нейтральным частицам. Поэтому средняя кинетическая энергия частиц в газовом разряде возрастает, что эквивалентно увеличению температуры. (Хотя по тем же причинам, что и в § 4.6, говорить о температуре здесь можно лишь с оговорками.) Среди каждой из групп частиц (электронов, ионов, нейтральных атомов и молекул) устанавливается свое квазиравновесие со своей средней кинетической энергией и, следовательно, со своей эффективной температурой.

Электроны, обладая малой массой и большой подвижностью, ускоряются в электрическом поле. При упругом столкновении с атомами они мало изменяют свою энергию вследствие большой разницы масс. Ионы, напротив, обладают малой подвижностью и активно обмениваются энергией при упругих столкновениях с нейтральными атомами и молекулами. Поэтому эффективная температура электронов в газовом разряде (электронная температура T_e) существенно превышает температуру тяжелых частиц, а температура ионов T_n , хотя и превышает температуру нейтральных частиц T , но близка к ней:

$$T_e \gg T_n \approx T.$$

Электронная температура T_e как параметр, характеризующий среднюю кинетическую энергию электронов, зависит от следующих параметров газового разряда: 1) от давления газа: при увеличении давления уменьшается длина свободного пробега и уменьшается T_e ; 2) от энергии ионизации газа: чем выше потенциал атома, тем, как правило, выше расположены его возбужденные состояния и тем меньше электронов теряет свою энергию при неупругих столкновениях первого рода, т. е. тем выше T_e ; 3) от градиента потенциала в активной области разряда

ного промежутка — кинетическая энергия электрона пропорциональна напряженности поля и длине свободного пробега, поэтому T_e растет при увеличении градиента потенциала.

Кроме того, электронная температура зависит от диаметра разрядной трубки d , увеличиваясь при его уменьшении. Это связано с тем, что при уменьшении d уменьшается степень ионизации газа за счет интенсивного взаимодействия ионов со стенками разрядной трубки. При неизменном значении тока этот процесс должен компенсироваться увеличением интенсивности процесса ионизации атомов, что возможно при больших T_e .

Типичные значения электронной температуры в газоразрядной трубке диаметром 7 мм, заполненной неонам при давлении 100 Па, имеют порядок 50000 К, а в той же трубке, заполненной гелием при том же давлении, — 70000 К.

Рассмотрим теперь основные процессы неупругого взаимодействия, благодаря которым осуществляется возбуждение атомов рабочего газа и достигается инверсная населенность между рабочими лазерными уровнями. Из неупругих столкновений первого рода важны следующие три:

1. *Прямое электронное возбуждение.* Этот процесс символически может быть записан в виде:



Электрон, ускоренный электрическим полем (стрелка над символом e), взаимодействует с атомом, молекулой или ионом A , переводя его в возбужденное состояние (звездочка у символа A) и теряя свою кинетическую энергию.

2. *Ступенчатое электронное возбуждение:*



В результате этого процесса центр A , который до взаимодействия уже находился в возбужденном состоянии, переходит в более высокое возбужденное состояние.

3. *Ионизация электронным ударом:*



В результате этого процесса в столбе разряда появляются дополнительные заряженные частицы — электрон и ион (знак «+» у символа A).

Для осуществления этих трех процессов необходимо, чтобы энергия электрона превышала некоторое минимальное пороговое значение $E_{\text{мин}}$, которое определяется состоянием атома (молекулы или иона) для первого процесса, энергетическим расстоянием между возбужденными состояниями для второго процесса и энергией ионизации для третьего процесса. Поэтому на зависимо-

стях сечений взаимодействия для этих процессов от энергии электрона наблюдается порог, как показано на рис. 7.2, б. Такие зависимости иногда называют функциями возбуждения.

Сечение возбуждения электронным ударом, как правило, определяют экспериментально, поскольку методы его расчета для многоэлектронных атомов развиты слабо. Известно, однако, что для процесса (7.9) сечение взаимодействия будет велико, если оптический переход между основным и возбужденным состоянием, в которое происходит переход за счет столкновения с электроном, разрешен в дипольном приближении. Т. е. сечение взаимодействия будет велико для переходов между такими состояниями, для которых дипольный матричный элемент (1.139) отличен от нуля. В этом случае σ_{\max} имеет порядок $\sim 10^{-16}$ см².

Кроме этого, важны следующие процессы взаимодействия второго рода:

4. *Девозбуждение и деионизация при столкновении с электронами:*



Очевидно, эти процессы противоположны первым трем. Сечения взаимодействия для них убывают при увеличении кинетической энергии электронов, как показано на рис. 7.2, в. Процессы (7.12) и (7.13) наряду со спонтанными переходами приводят к разрушению возбужденных состояний. Отметим, что процесс (7.14) ион-электронной рекомбинации в объеме разряда маловероятен, поскольку для его осуществления необходим отвод энергии рекомбинации ΔE (например, излучательным путем) за короткое время столкновения. Поэтому процесс деионизации (7.14) происходит преимущественно на стенках сосуда.

5. *Резонансная передача возбуждения при столкновении частиц одного сорта A с частицами другого сорта B :*



Этот процесс исключительно важен для избирательного заселения возбужденных состояний, поскольку он носит резонансный характер, как показано на рис. 7.2, г. Передача возбуждения будет эффективной лишь тогда, когда разница между энергиями возбужденных состояний E_{A^*} и E_{B^*} меньше kT :

$$\Delta E = |E_{A^*} - E_{B^*}| < kT, \quad (7.16)$$

разница энергий не должна превышать среднюю энергию теплового движения одной частицы.

6. *Ионизация с перезарядкой:*

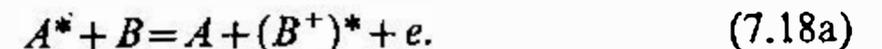


Энергия ионизации атома A переходит в энергию ионизации и возбуждения атома B . Взаимодействие (7.17), как и (7.15), носит резонансный характер: энергия ионизации атома A должна быть равна сумме энергий ионизации и возбуждения атома B с точностью $\Delta E \leq kT$. Этот процесс, как и (7.15), особенно эффективен, если состояние иона A^+ является метастабильным.

7. *Ионизация Пеннинга*:*



или



Конечное состояние иона B^+ может быть как основным (7.18), так и возбужденным (7.18a). Разумеется, что эти процессы возможны лишь в случае, когда энергия возбуждения атома A превышает энергию ионизации атома B . Избыточная энергия передается электрону. В отличие от резонансных процессов (7.15) и (7.17), при которых не происходит испускание электрона, ионизация Пеннинга представляет собой нерезонансный процесс. Избыток энергии возбуждения переходит в кинетическую энергию испущенного электрона. Процесс будет эффективным, если возбужденное состояние A^* обладает большим временем жизни, т. е. является метастабильным.

8. Кроме рассмотренных выше процессов обмена энергией в газовых смесях, возможно получение атомов и молекул в возбужденных состояниях при некоторых видах химических реакций. Простейшим примером может служить элементарная реакция замещения



при которой продукт реакции — молекула AB получается в возбужденном состоянии. Эти специфические процессы используются для накачки химических лазеров.

Итак, используя перечисленные основные процессы обмена энергией в плазме газового разряда, нетрудно представить общую схему работы газового лазера. Рассмотрим двухкомпонентную газовую смесь из частиц A и B , энергетические уровни которых изображены на диаграмме рис. 7.4. Газ B является рабочим. Инверсия населенности создается между его рабочими лазерными уровнями 2 и 1 по четырехуровневой схеме (см. рис. 2.9). Газ A — вспомогательный. Он служит для селективной пе-

*Процесс назван по имени голландского физика Ф. М. Пеннинга.



Рис. 7.4. Общая схема работы газового лазера: газ А — вспомогательный, газ В — рабочий

денным состоянием атома А и между ним и основным состоянием E_0^A отсутствовали бы какие-либо разрешенные состояния, особенно — близкие по энергии к состоянию E_2^B рабочего атома В.

3. Время жизни атома А в возбужденном состоянии было бы достаточно большим (уровень E_1^A — метастабильный), во всяком случае, больше среднего времени между столкновениями частиц рабочего и вспомогательного газа.

4. Система уровней рабочего газа В должна удовлетворять требованиям, предъявляемым к четырехуровневой (в крайнем случае — трехуровневой) схеме работы лазера (§ 2.2).

Из сказанного выше вытекает общая конструктивная схема газоразрядного лазера, изображенная на рис. 7.5. Активный элемент представляет собой трубку 1, заполненную смесью газов, в которой с помощью впаянных в нее электродов — анода 2 и катода 3 и источника питания ИП зажигается электрический разряд. Трубка помещается в оптический резонатор, образованный двумя зеркалами 4. Зеркала резонатора могут располагаться или внутри трубки, или вне ее. В первом случае необходимо специальное устройство (сильфон), которое позволяло бы перемещать зеркала внутри разрядной трубки для их точной юстировки. Во втором случае при расположении зеркал резонатора вне рабочей трубки необходимо исключить потери при выводе лазерного излучения через ее торцы, так как эти потери происходят внутри

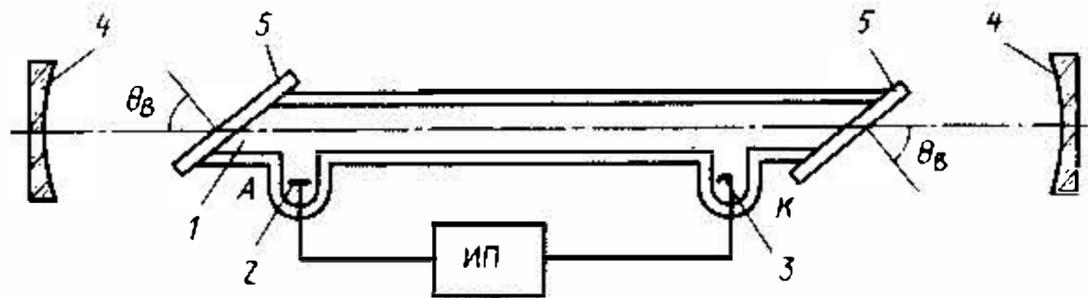


Рис. 7.5. Общая конструктивная схема газоразрядного лазера: 1 — разрядная трубка; 2 — анод; 3 — катод; 4 — зеркала резонатора; 5 — окна Брюстера

редачи возбуждения на уровни E_3 рабочего газа. Для этого необходимо чтобы:

1. Энергии возбужденных состояний E_1^A и E_3^B совпадали (отличались не более чем на kT).

2. Состояние E_1^A было бы нижним возбужденным состоянием атома А и между ним и основным состоянием E_0^A отсутствовали бы какие-либо разрешенные состояния, особенно — близкие по энергии к состоянию E_2^B рабочего атома В.

3. Время жизни атома А в возбужденном состоянии было бы достаточно большим (уровень E_1^A — метастабильный), во всяком случае, больше среднего времени между столкновениями частиц рабочего и вспомогательного газа.

4. Система уровней рабочего газа В должна удовлетворять требованиям, предъявляемым к четырехуровневой (в крайнем случае — трехуровневой) схеме работы лазера (§ 2.2).

Из сказанного выше вытекает общая конструктивная схема газоразрядного лазера, изображенная на рис. 7.5. Активный элемент представляет собой трубку 1, заполненную смесью газов, в которой с помощью впаянных в нее электродов — анода 2 и катода 3 и источника питания ИП зажигается электрический разряд. Трубка помещается в оптический резонатор, образованный двумя зеркалами 4. Зеркала резонатора могут располагаться или внутри трубки, или вне ее. В первом случае необходимо специальное устройство (сильфон), которое позволяло бы перемещать зеркала внутри разрядной трубки для их точной юстировки. Во втором случае при расположении зеркал резонатора вне рабочей трубки необходимо исключить потери при выводе лазерного излучения через ее торцы, так как эти потери происходят внутри

резонатора. Исключить эти потери, обусловленные в основном френелевским отражением, можно, расположив выходные окна 5 на торцах трубки не перпендикулярно ее оси, а под углом, равным углу Брюстера θ_B . В этом случае, как следует из формулы (3.22) и рис. 3.5, коэффициент отражения для волны, поляризованной параллельно плоскости падения, равен 0 (точнее, достаточно мал) и френелевские потери будут отсутствовать. Такие окна изображены на рис. 7.5. Они называются окнами Брюстера. Очевидно, что излучение лазера, внутри резонатора которого расположены окна Брюстера, будет плоскополяризованным, поскольку условие самовозбуждения будет в первую очередь выполняться для волны, вектор Е которой параллелен плоскости падения и перпендикулярен оптической оси резонатора.

Одно из основных требований, которое предъявляется к энергетическим состояниям рабочего газа в схеме рис. 7.4, — большая скорость опустошения нижнего лазерного уровня E_1^B . Инверсия населенности в стационарном режиме может быть получена, если время жизни уровня E_2^B определяется излучательными переходами на уровень E_1^B , вероятность которых меньше вероятности опустошения нижнего уровня E_2^B , т. е. $\tau_2 > \tau_1$. В противном случае при $\tau_2 \ll \tau_1$ инверсию населенностей между уровнями E_2 и E_1 в стационарном режиме получить невозможно, но ее можно получить в нестационарном, импульсном режиме с помощью так называемых самоограниченных переходов.

Общая схема работы лазера на самоограниченных переходах представлена на рис. 7.6. Временные диаграммы рис. 7.7 поясняют механизм получения инверсии в нестационарном режиме между рабочими уровнями 2 и 1. Накачка осуществляется импульсным дуговым разрядом. Переход $0 \rightarrow 2$ из основного состояния в верхнее рабочее разрешен, а в нижнее рабочее состояние $0 \rightarrow 1$ — запрещен в дипольном приближении. Уровень 1 — метастабильный. Поэтому сечение возбуждения электронным ударом для перехода

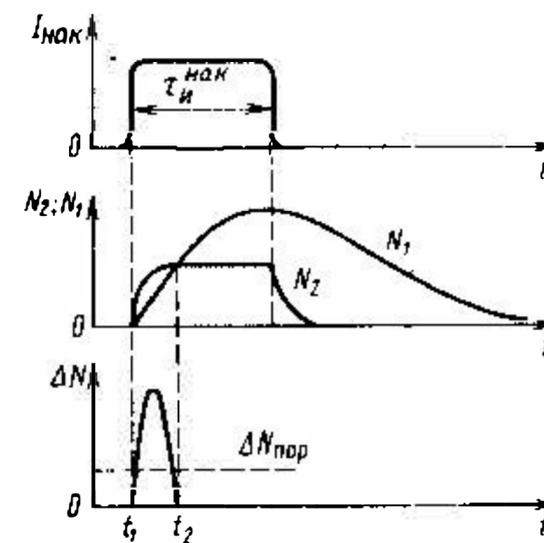


Рис. 7.7. Диаграммы, поясняющие механизм образования инверсии населенностей на самоограниченных переходах 2—1

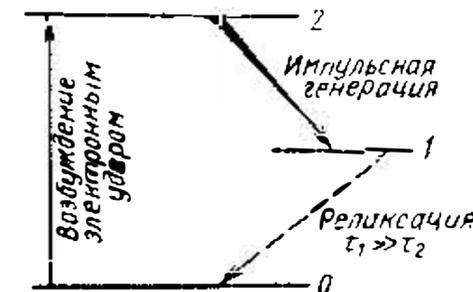


Рис. 7.6. Схема работы лазера на самоограниченных переходах

Общая схема работы лазера на самоограниченных переходах представлена на рис. 7.6. Временные диаграммы рис. 7.7 поясняют механизм получения инверсии в нестационарном режиме между рабочими уровнями 2 и 1. Накачка осуществляется импульсным дуговым разрядом. Переход $0 \rightarrow 2$ из основного состояния в верхнее рабочее разрешен, а в нижнее рабочее состояние $0 \rightarrow 1$ — запрещен в дипольном приближении. Уровень 1 — метастабильный. Поэтому сечение возбуждения электронным ударом для перехода

$0 \rightarrow 2$ существенно больше, чем для перехода $0 \rightarrow 1$. Оптический переход $2 \rightarrow 1$ разрешен. При этих условиях на резком переднем фронте токового импульса происходит преимущественно заселение верхнего уровня 2. Уровень 1 заселяется медленнее и накопление на нем частиц ограничивает инверсию населенностей. На диаграммах рис. 7.7 показаны изменения во времени: *a* — импульса тока накачки; *b* — населенностей верхнего N_2 и нижнего N_1 рабочих уровней; *в* — инверсии населенностей $\Delta N = N_2 - N_1$. Понятно, что лазер может работать только в импульсном режиме с длительностью лазерного импульса меньше времени жизни уровня 2. Верхний предел частоты повторения импульсов ограничен скоростью релаксации уровня 1, которая может определяться столкновением со стенками, атом-атомной дезактивацией и др.

Перейдем к рассмотрению конкретных примеров газовых лазеров.

§ 7.2. АТОМАРНЫЕ ГАЗОВЫЕ ЛАЗЕРЫ

В качестве примеров газоразрядных лазеров на нейтральных атомах рассмотрим два прибора — гелий-неоновый лазер и лазер на парах меди.

(He-Ne)-лазер является типичным и наиболее распространенным газовым лазером. Это был первый газоразрядный лазер, в котором в конце 1960 г. была получена генерация в инфракрасной области на длине волны 1,15 мкм. Особую популярность он получил, когда была осуществлена генерация в видимой (красной) области спектра ($\lambda = 0,63$ мкм). За прошедшее время он превратился из экзотического лабораторного прибора в обычный инструмент, используемый практически во всех областях науки, техники и даже искусства.

Лазер на парах меди представляет интерес прежде всего потому, что он излучает в важной для практического применения зеленой области спектра ($\lambda = 0,51$ мкм), где имеются самые чувствительные фотоприемники и где наблюдается максимум прозрачности морской воды. Он является примером импульсных лазеров, в которых инверсия населенности может быть получена только в импульсном режиме на самоограниченных переходах.

Гелий-неоновый лазер. Схема работы (He-Ne)-лазера, представленная на рис. 7.8, в общих чертах соответствует схеме рис. 7.4. Первые возбужденные состояния атома гелия (см. § 1.2 и рис. 1.4) 2^3S_1 и 2^1S_0 образуются электронной конфигурацией $1s^22p^2$ и расположены выше основного состояния 1^1S_0 на расстояниях 19,82 и 20,61 эВ соответственно. Оптические переходы в основное состояние запрещены, и потому они являются метастабильными состояниями, хорошо накапливая энергию, получаемую при возбуждении электронным ударом. Времена жизни 2^3S_1 и 2^1S_0 со-

ставляются 10^{-4} с и $5 \cdot 10^{-6}$ с соответственно.

Система энергетических уровней атома неона несколько сложнее, чем гелия. Основное состояние 1^1S_0 отвечает замкнутой оболочке $1s^22s^22p^6$. Нижним возбужденным состояниям соответствует переход одного электрона $2p$ в состояние $3s$, т. е. электронная конфигурация $1s^22s^22p^53s^1$. Этим конфигурациям отвечают четыре разрешенных уровня энергии (см. § 1.2). Используя полу-

эмпирические обозначения Пашена, каждому из этих уровней присвоим свой номер от 2 до 5 по мере убывания энергии, а всей совокупности уровней — индекс $1S$, так что нижние возбужденные состояния Ne будут обозначаться $1S_5, 1S_4, 1S_3$ и $1S_2$. Аналогичные системы уровней будут для электронных конфигураций $2p^54s^1$ и $2p^55s^1$, которые обозначаются как $2S$ и $3S$. Оптические переходы как между этими состояниями, так и в основное состояние запрещены в дипольном приближении. Важно, что верхние уровни $2S_2$ и $3S_2$ по энергии расположены близко к уровням 2^3S_1 и 2^1S_0 атома гелия. Дефицит энергии примерно равен 35 мэВ.

Кроме уровней, образуемых электронными конфигурациями $2p^5ns^1$, в Ne имеются уровни, отвечающие переходу одного электрона в $3p, 4p, \dots$ состояния. Каждой из электронных конфигураций $2p^53p^1$ и $2p^54p^1$ соответствует группа уровней, обозначаемая для краткости $2P$ и $3P$. Каждая группа состоит из 10 уровней, нумеруемых по мере уменьшения энергии от 1 до 10. Оптические переходы между s - и p -состояниями разрешены (точнее, из 40 возможных комбинаций разрешены 30). При этом время жизни $2S$ - и $3S$ -состояний ($\tau, \approx 0,1$ мкс) на порядок превышает время жизни $2P$ - и $3P$ -состояний ($\tau, \approx 0,01$ мкс). Это позволяет получить инверсную населенность между уровнями $2S - 2P$; $3S - 3P$ и $3S - 2P$ при электрическом разряде в чистом неоне. (Это справедливо и для других благородных газов — Ar, Xe и Kr.) Добавление к неону большого количества гелия обеспечивает селективное заселение верхних рабочих уровней $2S$ и $3S$, увеличивает инверсию и существенно облегчает получение генерации.

Таким образом, в He-Ne лазере He выполняет функцию вспомогательного газа, а Ne — рабочего в соответствии с общей

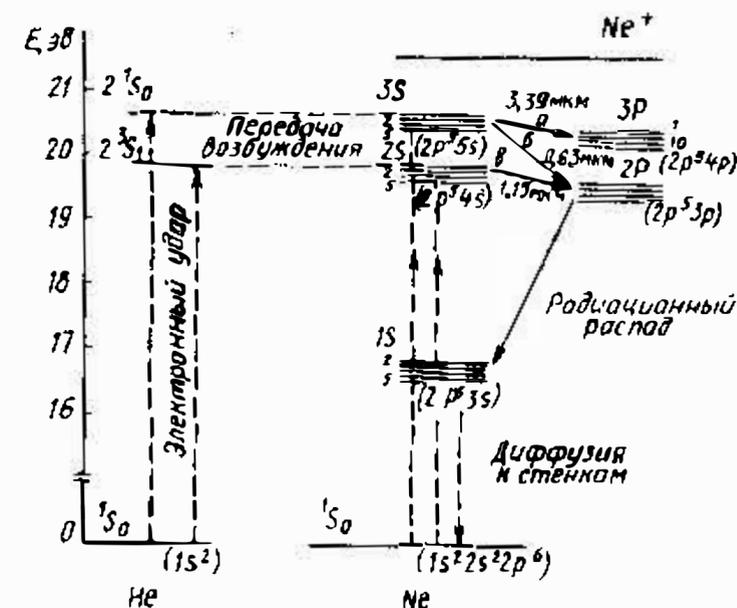


Рис. 7.8. Энергетические уровни атомов гелия и неона и схема работы (He-Ne)-лазера

схемой рис. 7.4. Однако его работа осуществляется не по четырехуровневой, как на рис. 7.4, а по трехуровневой схеме: накачка производится на верхний рабочий лазерный уровень. Усиление и генерация возможны на трех группах переходов, обозначенных a , b и c на рис. 7.8. Им соответствуют длины волн 3,39; 0,63 и 1,15 мкм. В этих группах наиболее сильными являются следующие переходы: $3S_2 — 3P_4$ ($\lambda = 3,3913$ мкм); $3S_2 — 3P_4$ ($\lambda = 0,6328$ мкм) и $2S_2 — 2P_4$ ($\lambda = 1,15228$ мкм). Наибольшее усиление — до 20 дБ/м — может быть получено для переходов в области 3,39 мкм. Переходам в области 1,15 и 0,63 мкм соответствуют значительно меньшие усиления — 10...20%/м и 5...6%/м. Для того чтобы осуществить генерацию на этих переходах, необходимо в резонаторе применять селективные зеркала, обладающие большим коэффициентом отражения в заданной области и большими потерями (малым отражением) в области конкурирующих переходов. Наиболее капризен в этом отношении переход $3S_2 — 2P_4$, поскольку верхнее рабочее состояние для него совпадает с верхним рабочим состоянием самого сильного $3S_2 — 3P_4$ -перехода. Поэтому получить генерацию в области 0,63 мкм в He — Ne-лазере наиболее сложно.

Опустошение нижних лазерных уровней $3P$ и $2P$ в He-Ne-лазере происходит далеко не оптимальным образом. Эти уровни быстро (что хорошо) опустошаются за счет излучательных переходов в $1S$ -состояния, которые являются долгоживущими (что плохо). В состоянии $1S$, расположенном на 16 эВ выше основного состояния, происходит накопление частиц. Это весьма нежелательно, ибо из состояния $1S$ велики вероятности переходов в состояния $2P$ и $3P$ при столкновении с электронами, согласно реакции (7.10). Переход атома неона из состояний $1S$ в основное состояние, как правило, происходит при его столкновении со стенками сосуда, т. е. за счет медленных процессов диффузии к стенкам. По этой причине работа (He-Ne)-лазера критична к диаметру разрядной трубки D и не допускает применения трубок с $D > 10$ мм.

Из всех газовых лазеров гелий-неоновый лазер наиболее прост по своей конструкции, которая полностью соответствует схеме рис. 7.5. Разрядная трубка изготавливается, как правило, из стекла и заполняется смесью газов He + Ne при оптимальном давлении. В трубку впаяны электроды. Часто применяют холодный катод из оксидированного тантала. Иногда в дополнительный отрогек помещается геттер. При внешнем расположении зеркал резонатора с торцов трубки под углом Брюстера к ее продольной оси вакуум плотно приклеиваются плоские окна толщиной 3...5 мм, изготовленные из оптического стекла высокого качества. Зеркала резонатора и газоразрядная трубка зафиксированы в специальной арматуре, основу которой составляют инваровые стержни, имеющие малый температурный коэффициент линейного расширения. Юстировочные устройства по-

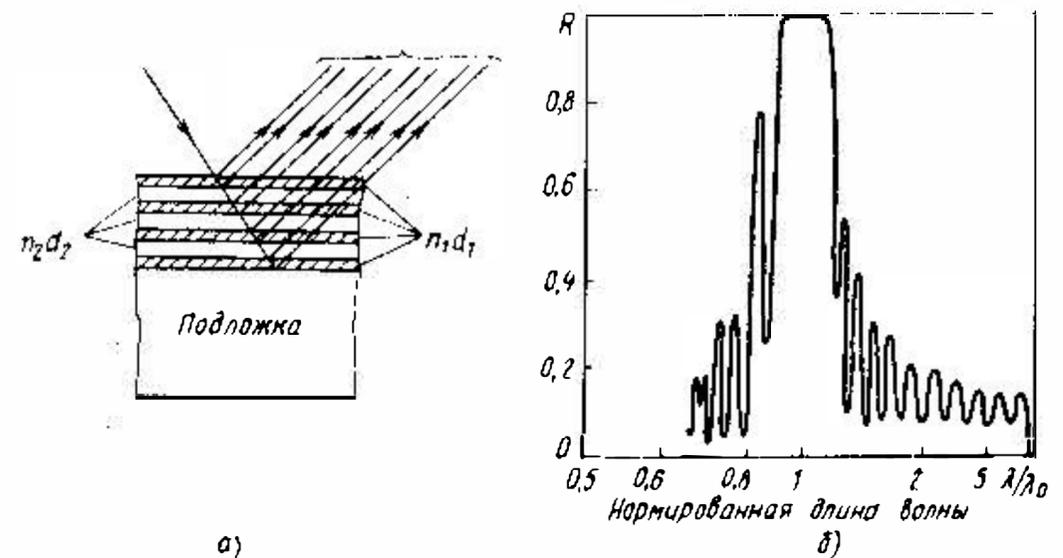


Рис. 7.9. Диэлектрические зеркала:
а) структура и схема работы (интерференция); б) спектральная зависимость коэффициента отражения

зволяют настраивать одно из зеркал резонатора и перемещать трубку. При расположении зеркал внутри рабочей трубки необходимость в специальной арматуре, естественно, отпадает.

Наибольшее распространение получили малогабаритные маломощные He-Ne лазеры с длиной разрядной трубки 8...20 см и внутренним диаметром 2...4 мм, работающие в красной области спектра на длине волны 0,6328 мкм в непрерывном режиме с выходной мощностью 1...5 мВт.

Поскольку коэффициент усиления в (He-Ne)-лазере мал, особенно для $\lambda = 0,63$ мкм, то его резонатор должен обладать высокой добротностью. Это накладывает жесткие требования к качеству зеркал резонатора. Коэффициент отражения на рабочей длине волны для одного из них должен быть близок к 100% (глухое зеркало), а для другого — 99,5...98% при пропускании 0,5...2%. Металлические покрытия этим требованиям не удовлетворяют и в резонаторе (He-Ne)-лазера всегда применяют многослойные интерференционные диэлектрические зеркала. Они представляют собой плоские или сферические пластины, изготовленные из оптического стекла или плавленного кварца, на которые методом напыления или химического осаждения нанесены чередующиеся диэлектрические слои толщиной d_1 и d_2 с разными показателями преломления n_1 и n_2 , как показано на рис. 7.9. Отражаясь от каждой из границ раздела, световые волны интерферируют между собой. Оптическая толщина каждого слоя n_1d_1 и n_2d_2 подбирается такой, чтобы интерференция шла на максимум (или минимум). Например, при соблюдении условия

$$n_1d_1 = n_2d_2 = \lambda/4 \quad (7.20)$$

на длине волны λ будет наблюдаться максимум отражения и минимум пропускания в многослойном покрытии из диэлектрических

ких слоев, как показано на рис. 7.9, б. Обычно применяют 5...13-слойные покрытия. Для видимой и ближней инфракрасной области — слои могут быть изготовлены из $ZnS - MgF_2$, $TiO_2 - SiO_2$ и др. Для других областей применяют другие материалы. Изменение коэффициента отражения в максимуме достигается, например, изменением числа слоев.

Рабочие характеристики (He-Ne)-лазера зависят от общего давления и соотношения компонент газовой смеси, от диаметра и длины газоразрядной трубки, от коэффициентов отражения зеркал и от разрядного тока. Эти зависимости определяются особенностями процессов возбуждения и релаксации активных атомов. Для каждого из перечисленных параметров существует свой оптимум. Некоторые типичные характеристики He-Ne лазера приведены на рис. 7.10. Качественно эти зависимости совпадают для всех трех рабочих диапазонов.

Наличие оптимумов по давлению P как рабочего (Ne), так и вспомогательного (He) газа связано с конкуренцией двух факторов: с одной стороны, при увеличении P растет число активных частиц, а с другой — при больших давлениях начинает сказываться снижение электронной температуры. Величина T_e определяется произведением давления газа P на внутренний диаметр трубки D . Для (He — Ne)-лазера, работающего на длине волны 0,63 или 3,39 мкм, оптимальные значения PD и P_{He}/P_{Ne} совпадают, поскольку верхний лазерный уровень для этих переходов один и тот же. Они составляют $PD \approx 500$ Па·мм и $P_{He}/P_{Ne} \approx 5:1$. Для лазера, работающего на длине волны 1,15 мкм, оптимальные значения $PD \approx 1500$ Па·мм и $P_{He}/P_{Ne} \approx 10:1$.

Внутренний диаметр разрядной трубки ограничен процессами диффузии атомов Ne к стенкам, так что обычно $D \approx 3...10$ мм.

В зависимости выходной мощности от плотности тока разряда (He-Ne)-лазера наблюдается характерный максимум, что связано с накоплением атомов Ne в долгоживущих 1s-состояниях. С увеличением плотности тока мощность излучения скачком

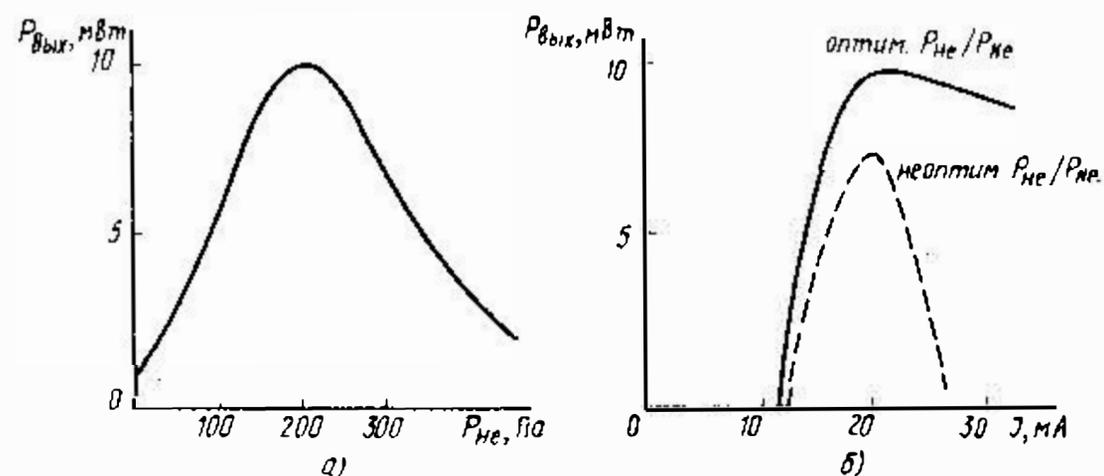
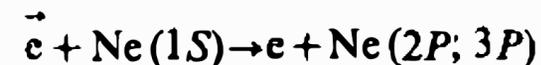


Рис. 7.10. Рабочие характеристики (He-Ne)-лазера

возрастает, начиная от $J = J_{\text{пор}}$. Однако дальнейший рост выходной мощности ограничивают процессы двухступенчатого электронного возбуждения [по реакции (7.10)] атомов Ne из состояний 1S в состояния 2P и 3P, приводя к заселению нижних рабочих лазерных уровней, уменьшению инверсии и в конечном итоге к срыву генерации. Так как эти переходы разрешены в дипольном приближении, то для них характерны большие сечения взаимодействия при столкновении с электронами. Скорость этого процесса



приблизительно пропорциональна концентрации свободных электронов и концентрации атомов неона в возбужденном состоянии 1S, которая в свою очередь примерно пропорциональна концентрации электронов (плотности разрядного тока). Таким образом, процесс паразитного заселения нижних рабочих состояний 2P и 3P примерно пропорционален квадрату плотности тока. Поэтому в He-Ne лазерах увеличение тока разряда выше оптимального значения приводит к падению мощности излучения.

Каких-либо эффективных путей дезактивации возбужденных состояний 1S Ne, кроме взаимодействия со стенками разрядной трубки, пока что не найдено. Естественно, что это ограничивает мощность излучения этого лазера единицами и десятками милливатт в областях 0,63 и 1,15 мкм и сотнями милливатт в области 3,39 мкм. Срок службы (He-Ne)-лазера ограничен процессами в разряде и исчисляется годами. Из-за сорбции атомов в стенках и электродах с течением времени меняются парциальные давления гелия и неона, происходит процесс «жестчения» газа. Для многих промышленных образцов (He-Ne)-лазеров характерно так называемое «складское» старение: приборы стареют в нерабочем состоянии. Поэтому рекомендуют по крайней мере раз в день дать прибору поработать некоторое время.

Существенным недостатком (He-Ne)-лазера является его низкий КПД (0,1...0,01%), что связано с особенностью энергетических диаграмм He и Ne (см. рис. 7.8). Рабочие уровни расположены выше основного состояния примерно на 20 эВ. Это требует очень высокой электронной температуры в разряде. Но даже при $T_e \approx 60000$ К лишь небольшая часть электронов будет в состоянии возбудить атомы гелия (см. рис. 7.3, б). Более того, для получения каждого из фотонов с энергией ~ 1 эВ, испускаемого, например, при оптическом переходе 2S — 2P, необходимо затратить энергию 20 эВ. Остальные 19 эВ в конечном итоге переходят в тепло.

Излучение (He-Ne)-лазера обладает высоким качеством. Его диаграмма направленности практически полностью определяется резонатором (см. §§ 2.3 и 2.6) и вводить какие-либо поправки на

оптическую неоднородность активного вещества нет необходимости.

Остановимся подробнее на спектральном составе излучения определяющем его временную когерентность. Естественная ширина линии излучения неона на рабочих переходах составляет $\Delta\nu_0 \approx 20$ МГц (при $\tau_{2,3} \approx 100$ нс, $\tau_{2p,3p} \approx 10$ нс). Общая ширина линии определяется эффектом Доплера и для атомов неона при 400 К равна $\Delta\nu_D \approx 1500$ МГц (см. § 1.6). Расстояние между собственными аксиальными типами колебаний резонатора определяется соотношениями (2.26) и (2.44). Нетрудно подсчитать, что при длине резонатора (He-Ne)-лазера 10...20 см в пределах контура спектральной линии расположится 1 — 2 собственных осевых типа колебаний. Поэтому в таких лазерах сравнительно просто реализовать одночастотный режим работы.

Поскольку в (He-Ne)-лазере преобладает неоднородное доплеровское уширение линии, то возможно проявление эффекта «частотного выжигания дырок», упоминавшегося в § 2.4. Для лазера, работающего в одночастотном режиме, это приводит к появлению в спектральной зависимости коэффициента усиления характерного провала, называемого «провалом Лэмба». Существование эффекта сводится к тому, что в системах с неоднородно уширенной спектральной линией монохроматическая волна на частоте ω взаимодействует только с частицами, находящимися в резонансе с этой электромагнитной волной, т. е. с частицами попадающими в пределы контура однородно уширенной спектральной линии. Это приводит к насыщению показателя усиления вблизи частоты ω , так что в спектре усиления α_ω образуется характерный провал.

Этот эффект может быть использован для тонких спектроскопических измерений и для разработки эффективных методов стабилизации частоты лазера. В (He-Ne)-лазере стабилизация частоты по минимуму лэмбовского провала позволяет получить относительную нестабильность порядка 10^{-9} , а с использованием специальных мер — до 10^{-13} .

Используя переходы между верхними уровнями состояний $3d$ и нижними уровнями состояний $2P$ (рис. 7.8), в He-Ne-лазере была получена генерация в оранжевой ($\lambda=612$ нм), желтой ($\lambda=594$ нм) и зеленой ($\lambda=543$ нм) областях видимого спектра с мощностью излучения порядка 1 мВт.

Лазерная генерация может быть получена на других нейтральных атомах, в том числе и на большинстве инертных газов (Ne, Ar, Kr, Xe). Эти лазеры работают в диапазоне длин волн 1...10 мкм примерно по той же схеме, что и рассмотренный He-Ne-лазер. В связи с низким КПД и малой выходной мощностью они не получили широкого распространения.

Лазер на парах меди. К числу атомарных лазеров относится целый класс лазеров на парах металлов (Pb, Cu, Au, Ca, Sr, Mn).

Все они генерируют только в импульсном режиме, поскольку их работа основана на самоограниченных переходах по схеме рис. 7.6. Наибольшее распространение среди них получил лазер на парах меди, который в настоящее время является распространенным источником когерентного излучения в зеленой области спектра.

Диаграмма энергетических уровней атома меди, участвующих в процессе создания инверсии, представлена на рис. 7.11. Нижние уровни $2D_{3/2}$ и $2D_{5/2}$, отвечающие электронной конфигурации $3d^9 4s^2$, являются метастабильными с временем жизни порядка 10 мкс. Оптический переход с них в основное состояние $2S_{1/2}$ с электронной конфигурацией $3d^{10} 4s^1$ запрещен. Сечения возбуждения электронным ударом на уровни $2P_{1/2}$ и $2P_{3/2}$ весьма велики и достигают в максимуме $\sigma \approx 10^{-15}$ см². Генерация возникает на переходах с этих уровней на метастабильные состояния $2D_{3/2}$ и $2D_{5/2}$. Наиболее сильной является зеленая линия 510,5 нм (переход $2P_{3/2} \rightarrow 2D_{5/2}$). Желтая линия 578,2 нм (переход $2P_{1/2} \rightarrow 2D_{3/2}$) на порядок слабее. Для увеличения населенностей верхних лазерных уровней в лазере на парах меди используют процесс реабсорбции излучения $0 \rightarrow 2$: спонтанное излучение, вызванное переходами $2P_{3/2, 1/2} \rightarrow 2S_{1/2}$ (эти переходы разрешены), поглощается атомами меди, находящимися в основном состоянии, переводя их в возбужденное. Этот эффект называется эффектом «пленения излучения». Он повышает эффективные времена жизни верхних уровней, что значительно облегчает требования к источнику импульсного питания. Для эффективной реабсорбции излучения показатель поглощения должен быть достаточно большим, для чего давление газа должно быть достаточно высоким.

В силу того, что рабочие переходы разрешены, а число активных частиц велико, лазеры на самоограниченных переходах обладают большим усилением на единицу длины (на один проход). Оно существенно выше, чем у (He-Ne)-лазеров. Кроме того, рабочие уровни по энергии расположены ниже, чем в (He-Ne)-лазере. Поэтому КПД Cu-лазера может быть весьма высок.

Конструкция лазеров на парах металлов следует общей схеме рис. 7.5, хотя технических трудностей в ее реализации здесь существенно больше, чем в He-Ne-лазере. В Cu-лазере для достижения требуемой плотности паров металла активное вещество должно находиться при высокой температуре (~ 1500 °С). Поэтому трубка изготавливается из окиси алюминия, алюминиевой или бериллиевой керамики на основе Al_2O_3 и BeO . Для предотвращения конденсации паров металла на холодных торцевых окнах в трубку добавляют неон или гелий при давлении в несколько мм рт. ст. Трудности, связанные с высокотемпературным режимом разрядной трубки, могут

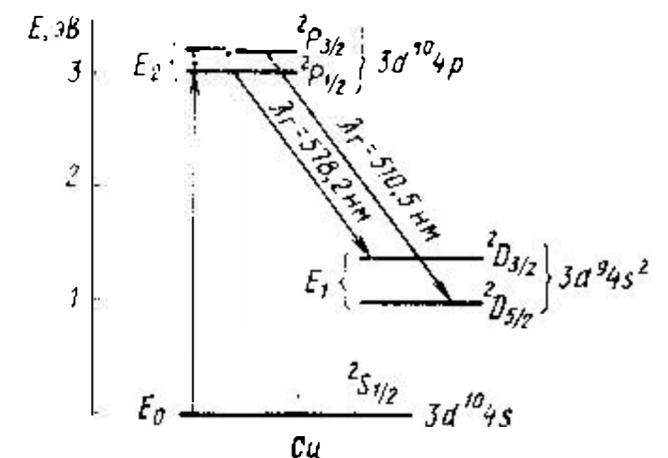


Рис. 7.11. Схема рабочих уровней лазера на парах меди

быть уменьшены, если вместо чистого металла использовать летучие металлсодержащие молекулы, например CuBr . В этом случае рабочая температура существенно понижается ($\sim 550^\circ\text{C}$ для CuBr). Однако для образования атомов меди необходимо диссоциировать молекулы. Для этого может быть использован метод двойного разряда: первый импульс диссоциирует молекулы CuBr , а вторым создается инверсия населенностей.

Естественно, что к источнику питания Cu -лазера предъявляются весьма жесткие требования, в первую очередь по фронтам нарастания импульса.

Cu -лазеры с разрядными трубками длиной 80...100 см и диаметром 1,5...2,5 см генерируют на зеленой линии 510,5 нм световые импульсы длительностью 5...10 нс с импульсной мощностью порядка 200 кВт. При частоте следования импульсов 20 кГц средняя мощность составляет 40...50 Вт, а КПД достигает 1...2%.

§ 7.3. ИОННЫЕ ГАЗОВЫЕ ЛАЗЕРЫ

В ионных газоразрядных лазерах рабочие переходы осуществляются между энергетическими состояниями одно-, а иногда двукратно-ионизированных атомов. Для поддержания их высокой концентрации плотность тока в ионных лазерах должна быть намного больше, чем в атомарных. Как правило, накачка осуществляется в дуговом разряде.

Шкала энергетических уровней ионов значительно шире, чем нейтральных атомов. Это связано с тем, что каждый из электронов в ионизированном атоме испытывает действие со стороны эффективного заряда порядка $(+2e)$, а в нейтральном атоме — $(+e)$. Такое расширение энергетической шкалы приводит к тому, что ионные лазеры эффективно работают в коротковолновой области видимого спектра и в ультрафиолетовом диапазоне (рис. 7.1). Вероятности оптических переходов между энергетическими состояниями ионов также несколько выше, чем нейтральных атомов.

Ионные лазеры, как и атомарные газовые лазеры, можно разделить на две группы: 1) лазеры на благородных и других газах и 2) лазеры на парах металлов. Типичным представителем первой группы является аргоновый (Ar^+), а второй — гелий-кадмиевый (He-Cd^+)-лазер.

Аргоновый лазер. Основное состояние иона аргона соответствует электронной конфигурации $3s^23p^5$. Это два близко расположенных уровня $^2P_{1/2}$ и $^2P_{3/2}$. Возбуждение одного электрона с $3p$ - на $4s$ - и $4p$ -оболочки дает сложный набор разрешенных уровней, между которыми возможно получение инверсной населенности в стационарном режиме. Поэтому аргоновый лазер

может генерировать на нескольких длинах волн. Наиболее интенсивная генерация происходит на голубой ($\lambda = 488$ нм) и зеленой ($\lambda = 514,5$ нм) линиях.

Использование ионов как активных лазерных частиц приводит к определенным техническим особенностям ионных лазеров. Плазма разряда должна быть высокоионизованной. Это достигается в сильноточном дуговом разряде. Для увеличения плотности тока разрядная трубка делается малого диаметра (1...3 мм) в виде капилляра. Пороговый ток ионных лазеров составляет несколько ампер, в рабочий ток в номинальном режиме достигает нескольких десятков ампер. Плотность тока — 1000 А/см² и более. Ионы, ускоряемые в электрическом поле разряда, становятся очень «горячими». Их эффективная температура составляет $T_e \approx 3000$ К. Эксплуатация такого разряда сопряжена с рядом технических трудностей. Разрядная трубка-капилляр выполняется из бериллиевой (на основе BeO) керамики или графита. Отметим, что окись бериллия обладает примерно в 100 раз лучшей теплопроводностью, чем графит, и поэтому ее применение предпочтительнее. В первых конструкциях трубка изготавливалась из кварца, и потому срок службы таких приборов был ограничен несколькими десятками часов. В большинстве конструкций трубка, анод и катод требуют водяного охлаждения. Вследствие высокой плотности тока в газоразрядной трубке происходит интенсивная перекачка ионов Ar^+ по направлению к катоду (*катафорез*). (Напомним, что катафорезом называется направленное движение ионов в газовом разряде постоянного тока под действием электрического поля.) В трубках большого диаметра этот эффект компенсируется обратной диффузией, которая препятствует образованию больших градиентов концентрации. В трубках малого диаметра, характерных для ионных лазеров, в дуговом разряде катафорез преобладает над диффузией, приводя к возникновению больших градиентов концентрации вдоль оси трубки. Для компенсации вызванного катафорезом перепада давлений в конструкции лазера предусмотрена специальная обводная трубка, соединяющая анодную и катодную части рабочего объема.

Газоразрядную трубку Ar^+ -лазера помещают в постоянное магнитное поле, направленное вдоль ее оси. Это поле напряженностью порядка 1000 Э, создаваемое соленоидом, «сжимает» разряд и препятствует движению электронов и ионов к стенкам рабочей трубки, предохраняя ее от разрушения ионным пучком. Это повышает мощность излучения и увеличивает долговечность прибора.

Выходная мощность Ar^+ -лазера быстро увеличивается с ростом тока разряда и в отличие от (He-Ne)-лазера не испытывает насыщения. Это обусловлено эффективным опустошением нижнего рабочего уровня, так что насыщение инверсии возможно при плотностях тока, существенно превышающих достигаемые на практике. Поэтому в Ar^+ -лазере можно получить высокие выходные мощности, вплоть до нескольких сотен ватт в непрерывном режиме при диаметре разрядной трубки около одного сантиметра и ее длине до двух метров. Как правило, Ar^+ -лазеры генерируют на нескольких длинах волн в сине-зеленой области спектра. Для осуществления селекции одной длины волны в резонатор лазера может быть введена диспергирующая призма. Одномодовый или одночастотный режим работы реализуется с большим трудом.

Типичные характеристики аргоновых лазеров: мощность излучения в непрерывном режиме — единицы и десятки ватт на длинах волн 488 и 514,5 нм при сроке службы сотни часов. Эти лазеры наряду с импульсными Cu - и N_2 -лазерами, широко при-

меняются для накачки перестраиваемых жидкостных лазеров на органических красителях, для подводной связи и локации, в лазерной хирургии, для разметки, в научных экспериментах.

Существенным недостатком ионных газовых лазеров, как и (He-Ne)-лазера, является их очень низкий КПД, составляющий 0,1...0,01%.

Из ионных лазеров на благородных газах применяется также Kг⁺-лазер, излучающий в красной области спектра на длине волны 647,1 нм.

Гелий-кадмиевый лазер. В этом лазере He является вспомогательным газом, а пары Cd — рабочим. По механизму возбуждения верхних лазерных уровней он похож на (He-Ne)-лазер, но передача возбуждения от атомов He к атомам Cd осуществляется за счет ионизации Пеннинга по реакции (7.18а) в результате чего появляется ион Cd⁺ в возбужденном состоянии. Интенсивная лазерная генерация возникает в ультрафиолетовой ($\lambda = 325$ нм) и синей ($\lambda = 441,6$ нм) областях спектра. Выходная мощность в непрерывном режиме при длине разрядной трубки 1...1,5 м и ее диаметре 2...2,5 мм достигает 50...200 мВт на длине волны 325 нм.

Конструктивно (He-Cd)-лазер выполняется аналогично He-Ne-лазеру по схеме рис. 7.5. Основное отличие и основные трудности связаны с необходимостью поддерживать однородное распределение паров кадмия в активном объеме разрядной трубки. Эта задача решается следующим образом. В газоразрядной трубке вблизи анода в небольшом отступлении (резервуаре) помещают металлический Cd. Этот резервуар нагревается до температуры 230...250 °С, что обеспечивает нужное давление паров металла. Когда пары достигают области разряда, то за счет катафореза ионы Cd⁺ прогоняются через разрядный промежуток от анода к катоду. За счет энергии разряда температура трубки-капилляра повышается, что препятствует конденсации на ее поверхности паров металла. Вблизи катода делается второй резервуар, на холодных стенках которого происходит осаждение металлического Cd. Диффузия кадмия к окнам Брюстера препятствует гелий, давление которого на два-три порядка превышает давление кадмия. Для обеспечения длительной работоспособности трубки надо иметь запас кадмия около 1...10 г на 1000 часов работы. Современные конструкции позволяют после полного испарения кадмия из резервуара поменять местами катод и анод и продолжить работу.

Выходная мощность (He-Cd)-лазера на длине волны 325 нм достигает 50...200 мВт в непрерывном режиме при длине разрядной трубки 1...1,5 м, ее диаметре 2...2,5 мм и оптимальной прозрачности выходного зеркала резонатора 5...7%. Оптимальное давление гелия составляет несколько сотен паскаль. Вначале зажигается тлеющий разряд в He, затем включается подогрев резервуара с Cd до 230...250 °С. Давление паров Cd составляет около 0,5 Па, а поток ионов кадмия — около 10 мг/ч при разрядном токе 100 мА и напряжении на трубке (4...5) кВ. Энерговыводление на единицу длины трубки порядка 3 Вт/см, что позволяет применять стеклянные трубки и воздушное охлаждение.

Поскольку процесс образования и возбуждения рабочих ионов кадмия проходит по реакции (7.18а) и является одноступен-

чатым, то в отличие от Ag⁺-лазера скорость накачки He-Cd⁺-лазера пропорциональна плотности тока разряда. Это приводит к существенно меньшим плотностям порогового тока накачки, чем в Ag⁺-лазере.

Самый большой недостаток He-Cd⁺-лазера — низкий КПД, составляющий 0,1...0,01%. Причина этому та же, что в (He-Ne)- и Ag⁺-лазерах. He-Cd⁺-лазер представляет интерес как источник непрерывного монохроматического излучения в УФ- и синей областях спектра для оптической спектроскопии, лазерной фотохимии, биологии, фотолитографии, неразрушающего контроля с использованием процессов молекулярного рассеяния света (комбинационного, рэлеевского, рассеяния Мандельштама — Бриллюэна), интенсивность которых пропорциональна $1/\lambda^4$ и др.

В отличие от Ag⁺-лазера в (He-Cd⁺)-лазере сравнительно просто может быть реализован одномодовый режим работы. Это связано с тем, что в нем активная среда нагревается слабо и доплеровская ширина линии на рабочих переходах составляет 1...1,5 ГГц. В спектре генерации хорошо видно изотопическое расщепление линий кадмия. При необходимости получения одночастотного режима работы лазера используют один изотоп ¹¹⁴Cd.

Кроме (He-Cd⁺)-лазера существуют ионные лазеры на парах других металлов: Sn, Pb, Zn, Se. Из них интерес представляет He-Se⁺-лазер, дающий интенсивную (до 200 мВт) непрерывную генерацию по крайней мере на 19 — 20 длинах волн, перекрывающих почти весь видимый диапазон. Для передачи возбуждения от He к Se в нем используется не ионизация Пеннинга, а резонансный процесс (7.17) ионизации с перезарядкой.

§ 7.4. МОЛЕКУЛЯРНЫЕ ЛАЗЕРЫ

Энергетический спектр молекул значительно богаче, чем атомов и ионов, и может быть представлен тремя частями: электронной, колебательной и вращательной (§ 1.2). Многообразие и сложность энергетического спектра приводит к богатым спектральным возможностям молекулярных систем и облегчает выбор подходящих уровней энергии. Молекулярные газовые лазеры перекрывают наиболее широкий диапазон из всех других типов лазеров (см. рис. 7.1).

В зависимости от типа участвующих в генерации переходов молекулярные лазеры разделяют на три класса. 1. *Лазеры на колебательно-вращательных переходах*, использующие переходы между колебательными уровнями одного и того же (как правило, основного) электронного состояния. Поскольку расстояние между колебательными уровнями энергии имеет порядок десятых и сотых долей электронвольта (см. § 1.2), то эти лазеры работают

в среднем ИК-диапазоне (5...100 мкм). 2. Лазеры на электронно-колебательных переходах, использующие переходы между колебательными уровнями различных электронных состояний и работающие в видимой и ближней УФ-областях спектра. 3. Лазеры на чисто вращательных переходах, использующие переходы между вращательными уровнями одного и того же электронно-колебательного состояния и работающие в далеком ИК-диапазоне (25...1000 мкм).

Многообразие внутренних движений в молекулах открывает новые каналы релаксации, которые могут быть использованы в частности, для эффективного опустошения нижнего лазерного уровня. (Вспомним, что именно отсутствие эффективных процессов релаксации с уровнями 1S неона в (He-Ne)-лазере существенно ухудшает параметры этого прибора.)

Рассмотрим два типичных представителя молекулярных лазеров: 1) лазеры на углекислом газе, работающие на колебательно-вращательных переходах, и 2) импульсный азотный лазер, работающий на электронно-колебательных переходах, а также СО- и Н₂-лазеры.

Газоразрядные СО₂-лазеры. Лазеры на углекислом газе являются наиболее важными газовыми лазерами, разработке и совершенствованию которых уделяется исключительное внимание во всех технически развитых странах. Это определяется тем, что СО₂-лазеры: а) обладают очень высоким КПД; б) способны генерировать исключительно большие мощности как в непрерывном, так и в импульсном режиме и в) спектр их излучения совпадает с окном прозрачности атмосферы.

Молекула СО₂ представляет собой линейную симметричную молекулу. В ней возможны три типа колебательных движений (рис. 7.12). Им соответствуют три колебательные моды: а) симметричная валентная ν_1 ; б) деформационная ν_2 и в) асимметричная валентная ν_3 . Деформационная мода

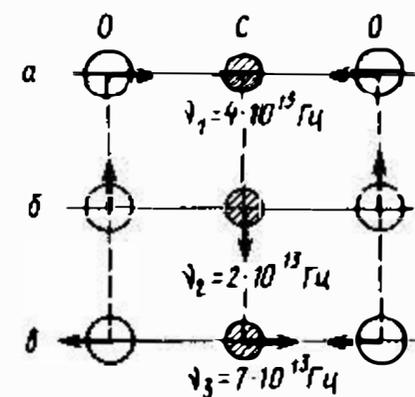


Рис. 7.12. Типы колебаний молекулы СО₂: а) симметричное валентное; б) деформационное; в) асимметричное валентное

дважды вырождена: колебания могут происходить как в плоскости рисунка, так и перпендикулярно этой плоскости. Колебательное состояние молекулы описывается тремя квантовыми числами q_1, q_2 и q_3 , которые определяют число квантов с собственными частотами ν_1, ν_2 и ν_3 в каждой колебательной моде аналогично (1.57). Соответствующий колебательный уровень обозначается этими тремя квантовыми числами в последовательности $q_1 q_2 q_3$. Справа вверху при квантовом числе q_2 ставится дополнительный индекс. Он возникает вследствие вырождения деформационной моды и указывает на

определенную комбинацию этих двух колебаний. Величина $l\hbar$ представляет собой угловой момент колебания относительно оси молекулы. Например, состояние 01^10 соответствует колебанию, деформационная мода которого имеет один колебательный квант, а в состоянии 02^00 два вырожденных колебания комбинируются таким образом, что угловой момент $l\hbar=0$.

Энергетическая диаграмма низших колебательных уровней основного электронного состояния молекулы СО₂ приведена справа на рис. 7.13. Генерация осуществляется на переходах между уровнями 00^01 и 10^00 ($\lambda=10,6$ мкм). Возможно получение генерации с использованием переходов $00^01 \rightarrow 02^00$ ($\lambda=9,6$ мкм). Верхний рабочий лазерный уровень 00^01 молекулы СО₂ почти точно ($\Delta E=2,2$ мэВ) совпадает с энергией первого колебательного состояния $q=1$ молекулы N₂. Это состояние является метастабильным, поскольку переход $1 \rightarrow 0$ для молекулы N — N запрещен в дипольном приближении. В то же время процесс возбуждения молекулы N₂ из основного состояния $q=0$ в состояние $q=1$ очень эффективен. Эффективен также процесс резонансной передачи энергии от молекулы N₂ к молекуле СО₂, как показано на рис. 7.13. Нетрудно убедиться, что схема работы газоразрядного СО₂-лазера соответствует общей схеме рис. 7.4, а молекулы СО₂ и N₂ наиболее полно удовлетворяют требованиям, предъявляемым к рабочему и вспомогательному газу, которые были сформулированы в § 7.1.

Хотя схемы работы СО₂-лазера и (He-Ne)-лазера похожи (ср. рис. 7.13 и 7.8), эффективность первого на три порядка превышает эффективность второго. Это в основном определяется двумя причинами. Во-первых, схема энергетических уровней оптимальна и на генерацию фотона с энергией 0,1 эВ надо затратить энергию всего в два раза больше (а не в 20...40 раз, как в (He-Ne)- и Ar⁺-лазерах). Во-вторых, в СО₂-лазере происходит исключительно эффективное как заселение верхнего рабочего 00^01 , так и опустошение нижнего 10^00 лазерных уровней. Остановимся на этих процессах более подробно. Отметим, что энергетическая диаграмма на рис. 7.13 существенно упрощена. На ней не показаны вращательные подуровни для каждого из колебательных уровней, а также более высоколежащие колебательные уровни с $q \geq 2$.

Эффективное заселение верхнего рабочего уровня

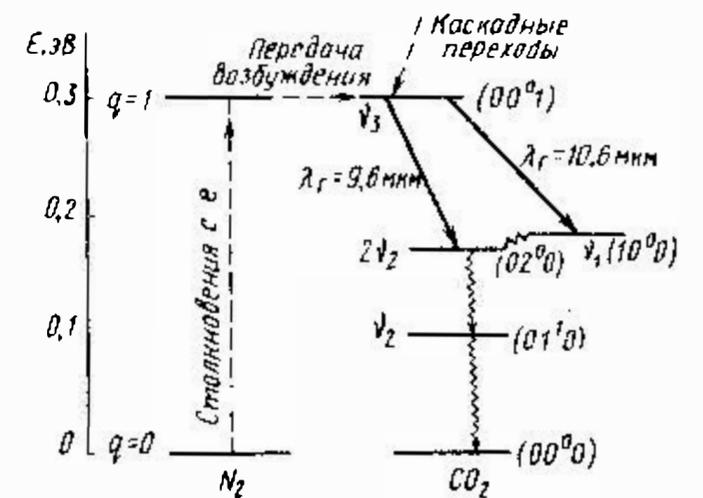
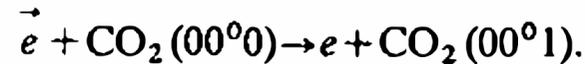


Рис. 7.13. Схема работы СО₂-лазера

00^01 в газоразрядном CO_2 -лазере осуществляется благодаря следующим трем процессам:

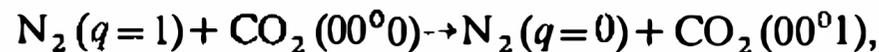
1. Столкновение с электронами



2. Резонансная передача возбуждения от молекулы N_2



Оценим эффективность этого процесса. Типичное значение приведенной напряженности электрического поля в плазме тлеющего разряда составляет $E/p \approx 4$ В/м Па. При энергии электронов около 2 эВ происходит возбуждение колебаний молекул N_2 в диапазоне $q = 1 \dots 8$. Сечение возбуждения молекул азота электронным ударом велико и составляет $3 \cdot 10^{-16}$ см², так что при концентрации электронов $(0,5 \dots 5) \cdot 10^{10}$ см⁻³ будут возбуждены от 40 до 80% молекул азота. Молекулы N_2 , находящиеся в колебательном состоянии $q = 1$, сталкиваясь с невозбужденными молекулами CO_2 , переводят их непосредственно в верхнее рабочее состояние согласно реакции



как это показано на рис. 7.13.

Оказалось, что более высокие колебательные уровни молекулы N_2 ($q = 2, 3, 4, 5, \dots$) находятся почти в резонансе ($\Delta E < kT$) с соответствующими уровнями молекулы CO_2 вплоть до уровня (00^05) . Поэтому эффективная передача возбуждения идет по реакции (7.15) не только между низшими возбужденными состояниями, но и между гармониками 00^0n молекул CO_2 и N_2 вплоть до $n = 5$. Скорость столкновительной передачи энергии возбуждения от N_2 к CO_2 имеет порядок 10^2 с⁻¹ Па⁻¹.

Таким образом, значительная часть молекул CO_2 возбуждается до колебательных состояний (00^0n) с $n \geq 2$. Их переход в состояние (00^01) осуществляется за счет еще одного резонансного процесса:

3. Столкновение молекул CO_2 в состоянии (00^0n) с $n \geq 2$ с молекулами CO_2 , находящимися в основном состоянии:



Этот очень эффективный процесс в конечном итоге переводит все возбужденные молекулы CO_2 в верхнее рабочее состояние (00^01) . Например, молекула CO_2 , находящаяся в состоянии (00^04) , после трехкратного столкновения с невозбужденными молекулами переводит их в состояние (00^01) и сама оказывается в этом же состоянии. Поскольку расстояние между колебательными уровнями энергии в силу (1.57) не зависит [точнее, слабо зависит по

(1.59)] от колебательного квантового числа, то этот процесс носит почти резонансный характер.

Совокупность трех рассмотренных процессов приводит к очень эффективной накачке верхнего лазерного уровня, обеспечивая высокий КПД CO_2 -лазера.

Рассмотрим теперь процессы опустошения нижнего лазерного уровня. При генерации на длине волны 10,6 мкм им будет уровень 10^00 , а при $\lambda = 9,6$ мкм — уровень 02^00 . На самом деле эти два уровня очень сильно взаимодействуют между собой за счет случайного совпадения частот ν_1 и $2\nu_2$. Это взаимодействие, называемое *резонансом Ферми*, обусловлено тем, что деформационное колебание (см. рис. 7.12) стремится изменить расстояние между атомами кислорода вдоль оси молекулы, вызывая тем самым ее симметричное растяжение и сжатие, т. е. вызывая симметричное валентное колебание. Поэтому в кинетических процессах (но не в оптических) эти два состояния смешиваются и выступают как одно целое.

Оказалось, что первый возбужденный уровень деформационной моды (01^10) эффективно релаксирует при столкновениях с атомами гелия. Поэтому добавление гелия существенно увеличивает эффективность опустошения нижних лазерных уровней, уменьшая их время жизни до ~ 20 мкс. Кроме того, гелий обладает высокой теплопроводностью, что способствует охлаждению CO_2 за счет отвода тепла к стенкам газоразрядной трубки.

Определенную роль в опустошении нижних лазерных уровней играют также процессы релаксации по вращательным подуровням молекулы CO_2 (на рис. 7.13 они не показаны), расстояние между которыми существенно меньше kT .

Таким образом, молекулярный CO_2 -лазер — это лазер на смеси углекислого газа, азота и гелия. CO_2 является рабочим газом, N_2 способствует эффективному заселению верхнего лазерного уровня, а He — обеднению нижнего. Оптимальное соотношение давлений $\text{CO}_2 : \text{N}_2 : \text{He}$ колеблется от 1 : 1 : 8 для лазеров низкого давления ($P_{\text{He}} \approx 2000$ Па) до 1 : 1 : 1,5 для лазеров высокого давления ($P > 1$ атм). Лазер может генерировать на переходе $(00^01) \rightarrow (10^00)$ ($\lambda = 10,6$ мкм) или на переходе $00^01 \rightarrow 02^00$ ($\lambda = 9,6$ мкм). Первый из них обладает большим коэффициентом усиления, и поскольку оба перехода имеют один и тот же верхний уровень, то генерация обычно возникает в области 10,6 мкм. Для получения генерации в области 9,6 мкм необходимо внутри резонатора осуществлять частотную селекцию, используя, например, селективный резонатор.

Конструкции CO_2 -лазеров могут быть весьма разнообразны. В зависимости от метода возбуждения различают *газоразрядные* и *газодинамические* лазеры. В свою очередь газоразрядные

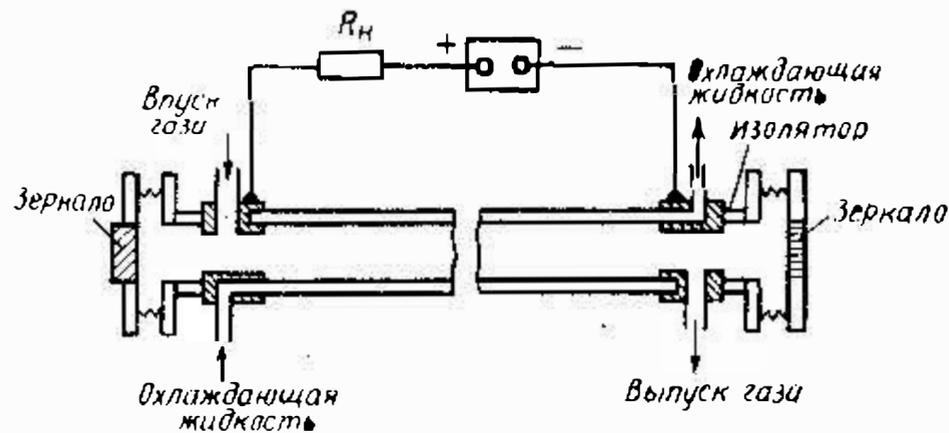


Рис. 7.14. Схема устройства CO_2 -лазера с продольным разрядом и прокачкой газа

CO_2 -лазеры можно разделить на пять типов: 1) лазеры с продольной прокачкой газа, 2) отпаянные лазеры, 3) волноводные лазеры, 4) лазеры с поперечной прокачкой газа и 5) лазеры высокого давления с поперечным возбуждением.

В качестве примера на рис. 7.14 показано устройство CO_2 -лазера с продольной прокачкой газа. Зеркала резонатора могут находиться или внутри разрядной трубки, как показано на рис. 7.14, или вне ее. В последнем случае применяются окна Брюстера, изготовленные из прозрачного в области $10,6 \text{ мкм}$ материала. Основной целью прокачки газовой смеси, состоящей из CO_2 , N_2 и He в соотношении примерно $1:1:8$, является удаление продуктов диссоциации, особенно молекул CO , которые ухудшают работу лазера. Мощность, получаемая с единицы длины разряда в этих лазерах, достигает $50 \dots 100 \text{ Вт/м}$ при диаметре трубки около $1,5 \text{ см}$.

Более удобны в эксплуатации отпаянные лазеры, не требующие непрерывной прокачки газа. Их реализация стала возможной после того, как были найдены катализаторы, регенерирующие молекулы CO_2 из образующихся в газовом разряде молекул CO . Роль катализатора могут выполнять пары воды, добавляемые в количестве около 1% в газовую смесь, или нагретый до $300 \text{ }^\circ\text{C}$ специальный никелевый катод. Использование этих методов привело к созданию в промышленном масштабе CO_2 -лазеров с отпаянными трубками со сроком службы 10000 часов и более. Рабочие характеристики этих приборов близки к характеристикам лазеров с продольной прокачкой газа. Весьма удобны в эксплуатации отпаянные лазеры с коротким резонатором, работающие в одномодовом режиме с выходной мощностью $1 \dots 10 \text{ Вт}$.

В остальных конструкциях газоразрядных CO_2 -лазеров газовая смесь находится при значительно больших давлениях, вплоть до атмосферного. Это позволяет на порядок и более увеличить съем энергии с единицы объема, но требует специальных систем для поддержания однородного газового разряда в активной сме-

си. Такие конструкции, как правило, применяются в мощных лазерах. Существуют системы, генерирующие световые импульсы длительностью менее 1 нс с выходной энергией в одном импульсе до 100 кДж и импульсной мощностью до 10^{14} Вт .

Газодинамические CO_2 -лазеры. Это особый тип лазеров, использующий неэлектрический метод накачки. Инверсия населенностей в нем создается за счет адиабатически быстрого расширения предварительно нагретой до высокой температуры газовой смеси $\text{CO}_2 + \text{N}_2 + \text{H}_2\text{O}$. Такие лазеры генерируют самые высокие мощности излучения.

Принцип работы газодинамического лазера пояснен на рис. 7.15. Газовая смесь, находящаяся в резервуаре при высокой температуре (примерно 1400 К) и высоком давлении (более 10 атм), выпускается через специальные сверхзвуковые сопла с малой высотой критического сечения (менее 1 мм). В высокотемпературной камере газ находится в термодинамическом равновесии. Из-за большой температуры около 10% молекул CO_2 находится

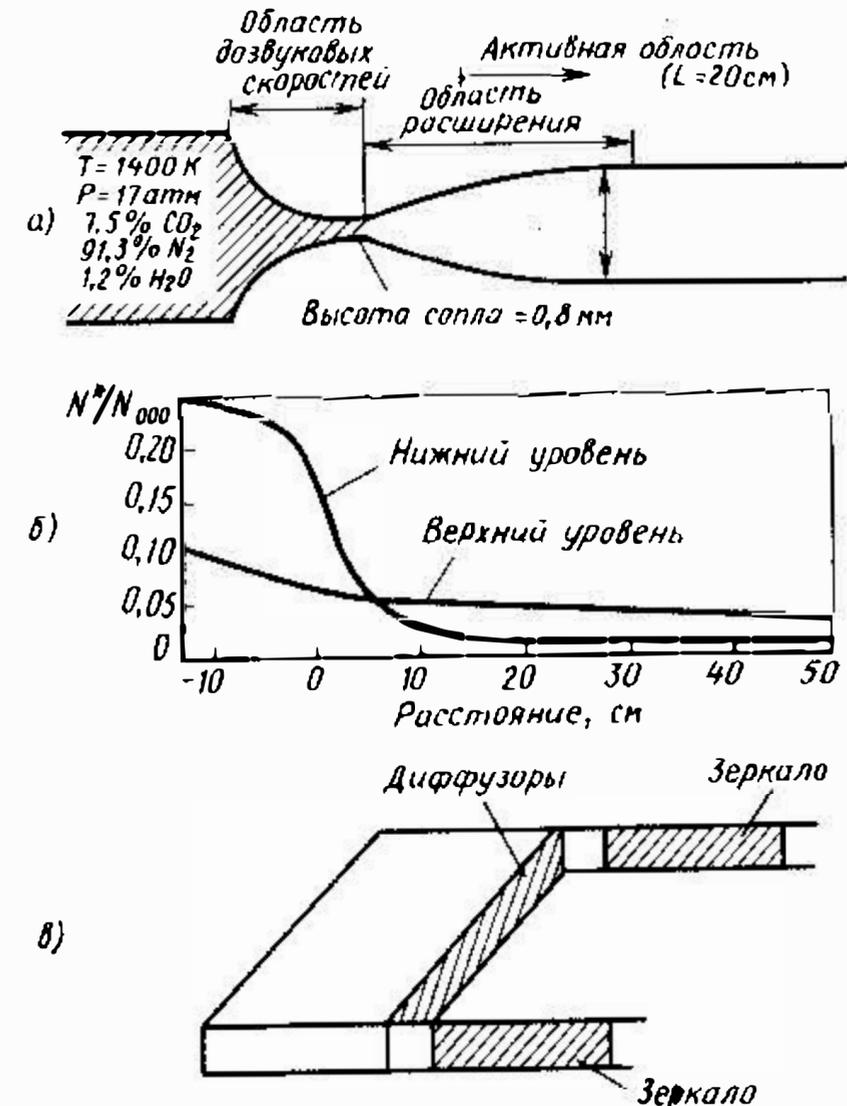


Рис. 7.15. Принцип работы газодинамического CO_2 -лазера

в состоянии 00^0_1 . Разумеется, населенность нижних рабочих уровней больше, чем верхних, и составляет около 25% от общего числа молекул CO_2 . Проходя через сопла со сверхзвуковой скоростью (число Маха $M \approx 4$), газ адиабатически быстро расширяется и его эффективная температура резко понижается. Населенности уровней будут стремиться к новым равновесным значениям. Поскольку время жизни верхнего лазерного уровня 00^0_1 больше времени жизни нижнего 10^0_0 (или 02^0_0), то нижний уровень достигнет состояния равновесия раньше, чем верхний, и в направлении по потоку от зоны расширения будет существовать довольно широкая область с инверсией населенностей, как показано на рис. 7.15, б. Протяженность этой активной области L определяется временем передачи возбуждения от молекул N_2 молекулам CO_2 . Зеркала резонатора располагаются вдоль газового потока так, что ось резонатора направлена перпендикулярно движению газа.

Начальную высокую температуру газовой смеси получают сжигая специально подобранные топлива (например, горение CO и H_2 или бензола C_6H_6 , а также окиси азота N_2O) в камерах типа ракетных двигателей. Опустошению нижнего уровня способствуют пары воды, образующиеся в процессе горения.

Газодинамические лазеры излучают мощности сотни киловатт в непрерывном (точнее — квазинепрерывном) режиме. Их химический КПД, определяемый как отношение выходной энергии лазерного излучения к полной химической энергии, которую можно получить при сгорании топлива, составляет единицы процентов. Лазер работает короткое время (доли или единицы секунд), что ограничивается лучевой прочностью используемых элементов. Такие лазеры представляют собой сложное инженерное сооружение.

Азотный лазер. N_2 -лазер является типичным примером молекулярных лазеров, использующих в качестве рабочих переходы между электронно-колебательными уровнями энергии. Он работает на длине волны $\lambda = 337,1$ нм, представляет интерес как источник интенсивного излучения в ультрафиолетовой области спектра и широко применяется для накачки перестраиваемых по длине волны жидкостных лазеров на красителях, для возбуждения люминесценции органических и неорганических веществ, для инициирования химических реакций, в физических и биологических исследованиях. Азотный лазер относится к типу лазеров на самоограниченных переходах и потому работает только в импульсном режиме.

Классификация электронных состояний двухатомных молекул, к которым принадлежит молекула N_2 , в силу их осевой симметрии проводится по молекулярному орбитальному квантовому числу Λ (см. § 1.2). Переходы разрешены в дипольном приближении, если выполняется правило отбора

$$\Delta\Lambda = 0, \pm 1. \quad (7.21)$$

Для работы молекулярных лазеров на электронно-колебательных переходах существенным является выполнение принципа Франка — Кондона (см. рис. 1.5 и 4.51), согласно которому электронный переход, сопровождающийся перестройкой электронных оболочек, происходит очень быстро по сравнению с движением ядер, и потому за время перехода относительное положение ядер и их скорости не меняется. На кривых потенциальной энергии, как функции межатомных расстояний, такой переход изображается вертикальной линией между так называемыми точками поворота колебательного движения молекулы, как показано на рис. 7.16. Скорость в этих точках равна нулю, а плотность состояний велика (максимальна для больших значений колебательного квантового числа q).

Схема нижних энергетических уровней молекулы N_2 и кривые потенциальной энергии, как функции межатомного расстояния R , приведены на рис. 7.17. Расположение потенциальных кривых основного ($X^1\Sigma_g^+$), верхнего ($C^3\Pi_u$) и нижнего ($B^3\Pi_g$) рабочих состояний таково, что при электронном возбуждении в соответствии с принципом Франка — Кондона эффективно заселяются уровни $q=0$ в $q=1$ состояния C , а состояние B почти не заселяется. Излучательный переход происходит из правых поворотных точек терма $C^3\Pi_u$ в состояние B , как показано на рис. 7.17. Таким образом, каналы возбуждения и генерации оказываются разнесенными.

Излучательное время жизни верхнего рабочего C -состояния равно 40 нс, а нижнего рабочего B -состояния — существенно больше и составляет 10 мкс. Поэтому азотный лазер не может работать в непрерывном режиме. Однако возможна генерация в импульсном режиме на самоограниченных переходах по схеме рис. 7.6 и 7.7. Время существования усвения при этом в несколько раз меньше времени жизни верхнего лазерного уровня и в азотном лазере составляет 3...10 нс. Длительность возбуждающих электрических импульсов не должна превышать это время.

Накачка азотного лазера осуществляется в импульсном разряде, создаваемом специальным импульсным генератором на полосковой линии. Давление рабочего газа внутри трубки составляет $10^3 \dots 10^4$ Па, что требует больших напряженностей

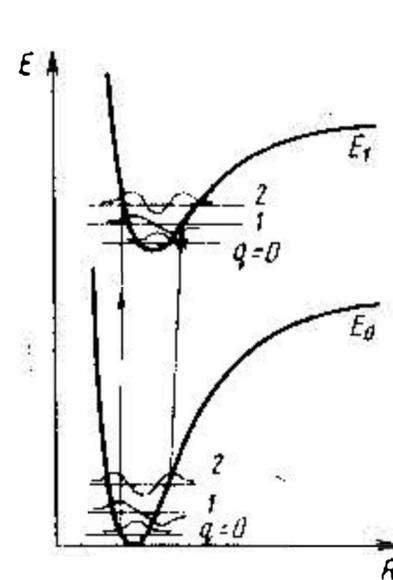


Рис. 7.16. Переходы между колебательными состояниями в соответствии с принципом Франка—Кондона

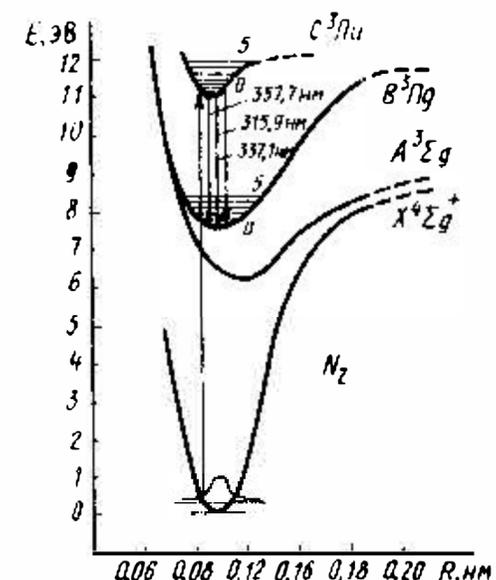


Рис. 7.17. Кривые потенциальной энергии в зависимости от расстояния между ядрами азота и энергетические уровни молекулы N_2 . Стрелками указаны лазерные переходы на соответствующих длинах волн

электрического поля (≥ 10 кВ/см) при плотности тока порядка 10^4 А/см². Поэтому, как правило, используется схема с поперечным электрическим разрядом (ТЕ-лазер). Генерация происходит преимущественно на нескольких вращательных линиях перехода $S^3\Pi_u(q=0) \rightarrow V^3\Pi_g(q=0)$, соответствующего $\lambda = 337,1$ нм. Длина разрядной трубки — около 1 м.

Особенностью азотного лазера является то, что за время существования инверсии в нем не успевает реализоваться обратная связь (напомним, что за 1 нс свет проходит в вакууме расстояние 30 см). Лазерное излучение представляет собой усиленное спонтанное излучение (сверхлюминесценцию), которое возникает благодаря высокому коэффициенту усиления активной среды. Такие приборы называются *суперлюминесцентными лазерами* или *лазерами на сверхсветимости*. Хотя они могут работать без зеркал резонатора, обычно с целью уменьшения пороговой мощности накачки и увеличения направленности излучения на одном из торцов газоразрядной трубки устанавливается «глухое» зеркало. Азотные лазеры генерируют импульсную мощность 50...500 кВт (до 20 МВт) при длительности импульса 5...10 нс и частоте повторения импульсов до 100 Гц. Их КПД составляет 0,01 — 1%, а срок службы — более 10000 часов.

СО- и Н₂-молекулярные лазеры. Эти лазеры обладают рядом особенностей. Как и СО₂-лазер, СО-лазер работает на колебательно-вращательных переходах в основном электронном состоянии. Он во многом подобен СО₂-лазеру, но работает в области длин волн 5...6,5 мкм. Могут быть использованы те же методы накачки и аналогичные конструкции. В СО-лазере получены выходные мощности в сотни киловатт при КПД выше 50% в квазинепрерывном режиме. Однако для получения таких высоких характеристик газовую смесь необходимо охлаждать до температуры около 100 К. В непрерывном режиме, если не принимать специальных мер, лазер работает на нескольких частотах одновременно, а в импульсном режиме наблюдается переключение частот генерации.

Интерес к СО-лазеру определяется его исключительно высокими энергетическими параметрами. Схема уровней энергии такова, что до 90% энергии электронов в разряде может быть преобразовано в колебательную энергию молекул СО. Во всех линиях генерации может быть получен суммарный КПД до 75%.

Благодаря добавлению ксенона в активную среду СО-лазера (СО + Хе + Не) стало возможным работать при комнатной температуре в отпаянных системах с продольным разрядом. Такие конструкции аналогичны маломощным СО₂-лазерам.

Водородный лазер и по принципу работы, и по конструкции аналогичен азотному лазеру. Он представляет интерес прежде всего тем, что работает в очень коротковолновой области спектра — вблизи $\lambda = 116$ нм и $\lambda = 160$ нм. Эти длины волн находятся в области вакуумного ультрафиолета (поглощение в атмосфере

настолько велико, что это излучение может распространяться лишь в вакууме или в Не-атмосфере). Трудность создания водородного лазера обусловлена тем, что излучательное время жизни верхнего лазерного уровня у него примерно в 50 раз меньше, чем у азотного лазера, и составляет 0,8 нс. Поэтому сконструировать систему возбуждения такого лазера весьма не просто.

Отметим, что при переходе в коротковолновый УФ-диапазон возникают трудности не только технического, но и принципиального характера. Уменьшение длины волны и увеличение частоты приводят к уменьшению времени жизни и падению усиления.

§ 7.5. ЭКСИМЕРНЫЕ И ХИМИЧЕСКИЕ ЛАЗЕРЫ

Эксимерные лазеры представляют особый класс молекулярных лазеров на электронно-колебательных переходах между устойчивым возбужденным и химически неустойчивым основным состояниями некоторых молекул. Для пояснения принципа работы эксимерного лазера рассмотрим двухатомную молекулу A_2 , кривые потенциальной энергии которой для основного E_0 и одного из возбужденных E_1 состояний приведены на рис. 7.18. Такая форма кривых означает, что молекула может существовать только в возбужденном состоянии, а в основном состоянии она химически неустойчива: при любых R атомы отталкиваются. Такая молекула A_2^* , которая может существовать только в возбужденном состоянии, называется *эксимером**

Оказалось, что свойствами эксимеров обладают некоторые соединения благородных газов: Хе₂^{*}, Кг₂^{*}, Аг₂^{*}, ХеF^{*}, ХеСl^{*}, КгF^{*}, АгF^{*} и др. Если в рабочем объеме создать большую концентрацию эксимерных молекул, то генерация может быть получена на переходах между верхним связанным E_1 и нижним неустойчивым E_0 состояниями. Такой лазер называется *эксимерным*. Рабочие переходы $E_1 \rightarrow E_0$ в эксимерном лазере обладают двумя важными особенностями: 1) при переходе в основное состояние молекула очень быстро (за время порядка 10^{-14} с) диссоциирует, т. е. нижний лазерный уровень опустошается исключительно быстро; 2) переход $E_1 \rightarrow E_0$ сопровождается излучением в сравнительно широком спектральном диапазоне (рис. 7.18), поскольку в нижнем состоянии отсутствует дискретная вращательно-колебательная структура уровней энергии. Это позволяет перестраивать спектр излучения лазера в пределах данного широкополосного перехода.

По своему устройству эксимерный лазер пока что представляет собой сложное сооружение. Его накачка осуществляется пучком электронов, ускоренных до энергий 0,1...1 МэВ при плот-

* От англ. excited dimer — возбужденный димер.

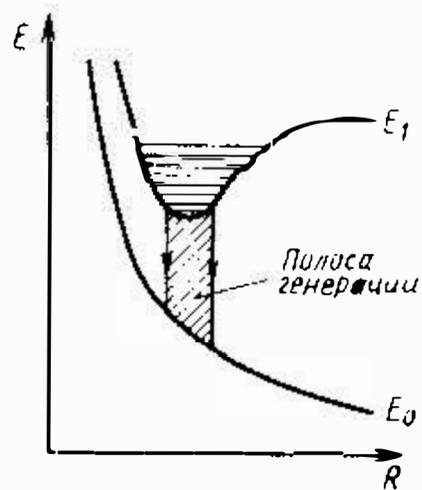


Рис. 7.18. Принцип работы эксимерного лазера

ностях тока 1...10 кА. Поскольку время жизни в возбужденном состоянии эксимеров мало (1...10 нс), то такие лазеры работают, как правило, в импульсном режиме. Они способны генерировать короткие световые импульсы очень большой энергии (до 10^5 Дж) при высоком КПД (до 10%). Основное достоинство эксимерных лазеров — возможность получения больших мощностей излучения в ультрафиолетовой области от 0,125 до 0,35 мкм при высоком КПД и возможность плавной перестройки спектра в диапазоне до $\Delta\lambda = 0,005$ мкм. Это весьма важно для таких применений, как

иницирование фотохимических реакций, разделение изотопов, стимулирование биохимических процессов и др. Лабораторные образцы эксимерных лазеров работают вблизи следующих длин волн: Ar_2^* — 126 нм; He_2^* — 172 нм; ArF^* — 193 нм; KrCl — 222 нм; KrF — 249 нм; HeCl — 308 нм; XeF — 352 нм.

Химическими лазерами называют лазеры, инверсия населенностей в которых создается непосредственно в процессе химической реакции между компонентами активной среды. Излучение химического лазера возникает как результат прямого преобразования химической энергии в энергию электромагнитного поля. Процессы, приводящие к излучению в таком лазере, по своей природе обратны процессам, приводящим к фотохимическим реакциям, при которых реакция происходит как непосредственный (а не косвенный) результат поглощения фотона атомом или молекулой. В то время как фотохимическое стимулирование осуществляется для проведения эндотермических реакций, химические лазеры могут работать только на основе экзотермических реакций.

Большая часть энергии, выделяемой в результате экзотермических реакций (химических реакций с выделением энергии) в газовых смесях, переходит в колебательную энергию молекул. Поэтому в большинстве случаев химические лазеры работают на колебательно-вращательных переходах в спектральном диапазоне 3...10 мкм. Их преимущества — возможность непосредственного преобразования химической энергии в энергию когерентного лазерного излучения с высоким КПД (более 10%), возможность получения высоких мощностей излучения в непрерывном режиме (более 10 кВт) при высоких значениях энергии, снимаемой с единицы объема активного вещества (до 10 Дж/л) и в ряде случаев — отсутствие громоздких источников питания. Недостатки химических лазеров, существенно ограничивающие их применение, связаны с высокой токсичностью большинства ис-

пользуемых веществ, их взрывоопасностью и необходимостью замены активной среды в процессе работы.

Газовые химические лазеры разделяют на два вида: с внешним инициированием химической реакции и «чисто химические» (без инициирования). Типичным примером химического лазера является HF-лазер. Его генерация происходит на нескольких колебательно-вращательных переходах в диапазоне 2,6...3,3 мкм. Выходная мощность в непрерывном режиме порядка 10 кВт, энергия в импульсе — несколько килоджоулей при химическом КПД около 10%.

Кроме химических лазеров, разработаны так называемые *фотодиссоциативные* (или *фотохимические диссоциативные*) лазеры, типичным примером которых является лазер на атомарном иоде, работающий на длине волны 1,315 мкм.

§ 7.6. РЕНТГЕНОВСКИЕ ЛАЗЕРЫ И ЛАЗЕРЫ НА СВОБОДНЫХ ЭЛЕКТРОНАХ

Коротковолновые лазеры, работающие в области далекого (вакуумного) ультрафиолета и на стыке оптического и рентгеновского диапазонов электромагнитного излучения ($\lambda < 100$ нм, рис. В.1), представляют несомненный интерес для решения целого ряда прикладных задач. Уменьшение длины волны приводит к увеличению разрешающей способности оптических приборов. Обладание когерентным излучением в рентгеновской области спектра позволит достичь субмикронного разрешения в фотолитографии, оптической микроскопии, голографии, включая получения трехмерных изображений с субатомным разрешением.

Создание лазеров, работающих в сверхкоротковолновой оптической и рентгеновской областях спектра, сопряжено с рядом трудностей принципиального характера. Во-первых, это связано с отсутствием подходящих рабочих сред с соответствующим набором энергетических уровней, расстояние между которыми составляло бы десятки, и даже сотни электронвольт (напомним, что длине волны $\lambda = 10$ нм соответствует энергия фотона 124 эВ). Во-вторых, при уменьшении длины волны (увеличении частоты) вероятность оптических переходов возрастает пропорционально кубу частоты для дипольных переходов (см. § 1.5). Отношение коэффициентов Эйнштейна для спонтанных и вынужденных переходов (1.106) резко увеличивается с увеличением частоты: $A_{mn}/B_{mn} \sim \omega^3$, что затрудняет генерацию вынужденного излучения.

Рентгеновские лазеры на высокоионизованных атомах. Как следует из §§ 1.2 и 7.3, энергетические состояния ионизованных атомов по своей структуре аналогичны энергетическим состояни-

ям эквивалентных атомов с соответствующей электронной конфигурацией. Основное отличие заключено в масштабе шкалы энергий. Действительно, из соотношений (1.36) и (1.39) следует, что собственные значения внутренней энергии атома водорода пропорциональны квадрату заряда его ядра Z и квадрату заряда электрона e . Поскольку в атоме водорода эти величины равны, то

$$E_{n,l,m} \sim Z^2 \cdot e^2 = e^4.$$

Спектр разрешенных значений энергии однократно ионизованного атома гелия He^+ оказывается аналогичен таковому для атома водорода с той лишь разницей, что шкала энергий будет увеличена в 4 раза. Действительно, как и в атоме водорода, в ионе He^+ один электрон движется в центрально-симметричном поле ядра, но в отличие от атома водорода, заряд $(Z/e) = 2$, а потому

$$R_{\infty}^{\text{He}^+} = 4R_{\infty}^{\text{H}} \simeq 54 \text{ эВ}.$$

Для многоэлектронных атомов и ионов можно аналогично записать

$$E \sim (Z^*/e)^2, \quad (7.22)$$

где Z^* представляет собой эффективный заряд, значение которого определяется не только степенью ионизации атома, но и экранировкой со стороны оставшихся электронов (в случае He^+ она отсутствует). Обычно $Z^* < Z$ и можно представить

$$Z^* = Z - \sigma_{nl}, \quad (7.23)$$

где σ_{nl} — постоянная экранирования, определяемая конфигурацией электронного облака атома (иона).

Таким образом, спектры разрешенных значений энергий, например, восьмикратно ионизованного атома аргона Ar^{8+} или двадцатичетырехкратно ионизованного селена Se^{24+} будут аналогичны спектру атома неона Ne , поскольку всем им соответствует одинаковая электронная конфигурация $1s^2 2s^2 2p^6$ для основного состояния и серии возбужденных состояний $2p^5 3s$ ($1S$), $2p^5 3p$ ($2P$) и т. п. (рис. 7.8). Ожидаемый порядок изменения масштаба энергий в силу (7.22) составляет 50 и 500 раз для Ar^{8+} и Se^{24+} , соответственно.

Обеспечить приемлемую концентрацию активных ионов возможно лишь в высокоионизированной плазме, что требует применения специальной аппаратуры. В частности, такая плазма может быть получена при фокусировке мощного излучения им-

пульсных твердотельных лазеров на соответствующую, например, селеновую мишень в установках, применяемых для термоядерного синтеза. Понятно, что КПД таких лазеров будет чрезвычайно низок, а их реализация возможна лишь в нескольких специальных лабораториях. Несмотря на это, в силу исключительно важных причин, отмеченных в начале настоящего параграфа, работы по созданию рентгеновских лазеров проводятся весьма интенсивно. Впервые лазерная генерация наблюдалась на длинах волн $\lambda_1 = 20,63$ нм и $\lambda_2 = 20,96$ нм при использовании в качестве активных центров Ne — подобных ионов Se^{24+} . В дальнейшем было опробовано много активных сред, включая Ne — подобные Ar^{8+} и Ag^{37+} , H — подобные C^{5+} и Al^{12+} и др. Получено излучение в области от 3,6 до 47 нм с коэффициентом усиления за один проход от 10 до 1000. Все приборы работали в режиме сверхкоротких импульсов. Была продемонстрирована принципиальная возможность получения усиления в плазме Ar^{8+} , полученной при пропускании короткого мощного импульса тока через капилляр, заполненный Ar . Это направление представляется весьма перспективным.

Лазеры на свободных электронах. Как и рентгеновские лазеры, эти приборы реализуются лишь в единичных экземплярах в специальных лабораториях, обладающих ускорителями электронов, способными разогнать их до релятивистских скоростей. Такой пучок электронов, движущийся со скоростью v , близкой к скорости света c , пропускается через специальное устройство, называемое *ондулятором**. В ондуляторе на достаточно большом протяжении создается периодическое (в частности — знакопеременное) поле с периодом Λ . Например, магнитный ондулятор представляет собой набор расположенных друг за другом магнитов с чередующейся полярностью. Проходя через магнитное поле, направленное перпендикулярно к оси пучка, заряженные частицы (электроны) под действием силы Лоренца изменяют траекторию движения. Кроме поступательного движения со скоростью v электроны под действием периодического поля ондулятора совершают периодические колебания, т. е. осциллируют. Следовательно, они испускают электромагнитное излучение аналогично классическому осциллирующему диполю (§ 1.5).

При дорелятивистских скоростях $v \ll c$ частота испускаемого излучения равна частоте периодического движения:

$$\omega_0 = 2\pi v/\Lambda. \quad (7.24)$$

*франц. Ondulateur (om onde — волна) или andulateur — волнообразный, волнообразователь.

При обычных скоростях электронов эта частота лежит в радиодиапазоне, а сам эффект используется в классических СВЧ электронных приборах, например, в лампах бегущей волны (ЛБВ). В процессе взаимодействия электронного пучка с электромагнитной волной, в которую перекачивается его энергия, происходит фазовая группировка, обеспечивающая выполнение условия синхронизма, и тем самым достигается когерентность генерируемой волны.

При скоростях электронов, близких к скорости света $v \approx c$, происходит релятивистское сокращение масштаба, так что эффективная длина Λ уменьшается в γ раз:

$$\Lambda_{эф} = \Lambda (1 - v^2/c^2)^{1/2} = \Lambda/\gamma, \quad (7.25)$$

где $\gamma = (1 - v^2/c^2)^{-1/2}$ называется *релятивистским фактором*. Кроме того, поскольку пучок осциллирующих электронов движется со скоростью v относительно наблюдателя, то для него будет проявляться эффект Доплера и воспринимаемая в лабораторной системе координат частота излучения будет определяться соотношением (1.149). Учитывая, что для релятивистских электронов $(1 + v/c) \approx 2$, а в направлении скорости их поступательного движения $\theta = 0$, получаем

$$\omega = 2\pi \left[\frac{v \sqrt{1 - v^2/c^2}}{\Lambda_{эф} (1 - v/c)} \right] = 2\omega_0 \gamma^2 \quad (7.26)$$

и длина волны электромагнитного излучения

$$\lambda = \Lambda/2\gamma^2. \quad (7.27)$$

Чтобы это излучение попало в видимый диапазон спектра при разумных значениях периода $\Lambda = 1$ см необходимо обеспечить $\gamma \approx 100$. Оценим, до какой энергии eU надо разогнать электроны, чтобы реализовать эту величину. Для этого воспользуемся известным релятивистским соотношением между энергией свободной частицы E и ее массой m : $E = mc^2$, причем $m = \gamma m_0$, где m_0 — масса покоя частицы. Для свободного электрона $m_0 c^2 = 511$ кэВ. Это означает, что для получения $\gamma = 100$ необходимо разогнать электронный пучок до энергий

$$E = eU = \gamma m_0 c^2 \approx 50 \text{ МэВ}.$$

В современных ускорителях возможно достичь энергии электронов до 500 МэВ и более и тем самым получить когерентное излучение на стыке между коротковолновым оптическим и рентгеновским диапазонами.

Как можно было заметить, лазеры на свободных электронах по своему принципу действия ближе к классическим приборам вакуумной электроники, чем к типичным квантовым приборам. Вынужденное (стимулированное) излучение здесь возникает как результат коллективного движения свободных электронов в электромагнитном поле, а не в результате квантовых переходов между уровнями (или полосами) разрешенных значений энергий отдельных атомов. Но поскольку этот прибор является генератором когерентного электромагнитного излучения, использующий индуцированное испускание света электронным потоком, его естественно называть лазером в силу определения этого слова, данного во введении.

Электронные ускорители, генерирующие пучки сверхвысокой энергии, обычно работают в импульсном режиме и, следовательно, лазеры на свободных электронах также работают в импульсном режиме, генерируя мощности сотен кВт. Несмотря на громоздкость и исключительно высокую стоимость этих приборов, они вызывают заметный интерес, поскольку позволяют осуществить перестраиваемую генерацию во всем оптическом диапазоне спектра, включая субмиллиметровую, инфракрасную, видимую и ультрафиолетовые области при высоком его качестве и большой мощности.

ГЛАВА 8 ТВЕРДОТЕЛЬНЫЕ И ЖИДКОСТНЫЕ ЛАЗЕРЫ

Активным веществом твердотельных и жидкостных лазеров является диэлектрик, находящийся в конденсированной фазе. По сравнению с газами в конденсированных средах можно создать большие плотности (концентрации) активных частиц в единице объема и тем самым достичь большей плотности инверсии и большего удельного энергосъема. В то же время разнообразие процессов, с помощью которых можно изменить энергетическое состояние атомов в таких системах без их разрушения, весьма ограничено. Практически единственным способом активного воздействия остается облучение светом. Поэтому в твердотельных и жидкостных лазерах на диэлектриках применяется только оптическая накачка. В полупроводниках возможно дополнительное воздействие путем пропускания через них электрического тока.

§ 8.1. ОБЩАЯ ХАРАКТЕРИСТИКА И ОСОБЕННОСТИ ТВЕРДОТЕЛЬНЫХ ЛАЗЕРОВ

Рабочий элемент твердотельных лазеров выполнен из кристаллического или аморфного диэлектрика. Такие лазеры надежны, удобны и сравнительно просты в эксплуатации. При небольших габаритах они могут генерировать очень высокие импульсные мощности (вплоть до 10^{12} Вт и более), очень короткие световые импульсы (до 10^{-12} с и менее), а также работать в непрерывном режиме с выходной мощностью от единиц мВт до сотен Вт. Их накачка осуществляется оптическим путем. Для этого электрическая энергия с помощью специальных ламп накачки или с помощью полупроводниковых лазерных диодов преобразуется в оптическое излучение, которое поглощается атомами активного вещества, переводя их в возбужденное состояние. Наличие такого промежуточного процесса преобразования электрической энергии накачки снижает общий КПД твердотельных лазеров, который не превышает нескольких процентов при лам-

повой накачке и достигает 30% при накачке инжекционными лазерами.

Спектральный диапазон работы твердотельных лазеров ограничен оптической прозрачностью активной среды. С коротковолновой стороны он ограничен процессами собственного поглощения, а с длинноволновой — взаимодействием с колебаниями решетки. Поэтому твердотельные лазеры работают в ближней ультрафиолетовой, видимой и ближней инфракрасной областях спектра (рис. 8.1).

Материалы, предназначенные для изготовления лазерных активных элементов с оптической накачкой, называемые *активными диэлектриками*, должны обладать определенной системой энергетических уровней, удовлетворяющей тем требованиям, которые вытекают из трех- или четырехуровневой схемы работы лазера (рис. 2.8 и 2.9). Нетрудно заключить, что эти материалы должны обладать: а) интенсивной флуоресценцией с квантовым выходом на рабочем переходе, близким единице; б) широкими полосами активного поглощения в области излучения источника накачки и в) отсутствием потерь на частоте рабочего перехода. Активный элемент, находясь внутри резонатора и работая в условиях интенсивной световой накачки, должен сохранять высокие оптические качества. Это предъявляет к механическим, оптическим, теплофизическим, химическим и т. п. свойствам *активных* материалов весьма жесткие требования.

Наиболее полно предъявляемым требованиям удовлетворяют так называемые активные диэлектрики, представляющие собой твердые растворы элементов с недостроенными внутренними электронными оболочками в различных кристаллических и аморфных матрицах. Обычно используют ионы одной из групп переходных элементов: элементы группы железа с недостроенной $3d$ -оболочкой или редкоземельные элементы с недостроенной $4f$ -оболочкой (лантаноиды). Актиноиды с недостроенной $5f$ -оболочкой применяются редко вследствие их радиоактивности.

Недостроенные внутренние электронные оболочки этих элементов хорошо экранированы валентными электронами от внешних воздействий, в том числе и от воздействия кристаллического поля. Поэтому при введении таких ионов в конденсированную среду не происходит коренной перестройки их энергетического спектра. Эти ионы, называемые *активаторами*, и являются собственно активными, в то время как кристаллическая или аморфная основа в большинстве случаев выполняет роль матрицы. Все рабочие лазерные переходы осуществляются между уровнями энергии активных ионов.

Активный парамагнитный ион в диэлектрике находится под воздействием сильного электростатического поля окружающих его ионов. В зависимости от степени экранировки внутренних оболочек внешними электронами это поле, называемое внутрикристаллическим полем, по-разному влияет на свойства иона-активатора. При этом различают три случая.

активные центры (ноны-активаторы) одного или разного вида. В результате взаимодействия активных ионов с матрицей изменяется не только структура энергетических уровней иона, но и вероятности переходов, в том числе вероятности безызлучательных переходов, а также соотношения между вероятностями излучательных и безызлучательных переходов. Поэтому нахождение оптимальной матрицы является весьма важной задачей.

Негрудно сформулировать основные требования, которым должны удовлетворять матрицы лазерных активных диэлектриков. Во-первых, матрица должна обладать высоким оптическим качеством, быть оптически однородной и прозрачной для лазерного излучения и не должна обладать неактивным поглощением в области излучения источника накачки. Во-вторых, матрица должна допускать введение заданного активатора в регулируемых количествах без нарушения ее механических свойств и оптической однородности; чем более благоприятны геометрические параметры замещаемого и замещающего ионов, тем это условие легче выполняется. В-третьих, матрица должна обладать высокой теплопроводностью для того, чтобы энергия, передаваемая решетке в результате безызлучательных переходов $E_3 - E_2$ и $E_1 - E_0$, и энергия накачки, переходящая в тепло, могли быстро рассеиваться, не приводя к заметным термооптическим искажениям и термическим деформациям. В-четвертых, матрица должна обладать высокой лучевой стойкостью по отношению к генерируемому лазерному излучению, а также быть фотохимически стойкой к мощному излучению накачки. Наконец, в-пятых, матрица, обладая хорошими механическими свойствами, должна быть технологичной в изготовлении и оптической обработке, поскольку к параметрам активного элемента (качество обработки поверхности, параллельность торцов, однородность) предъявляются очень высокие требования.

В настоящее время генерация получена на нескольких сотнях переходов между энергетическими уровнями различных ионов более чем на 300 кристаллах и нескольких десятках типов стекол. Но не существует материала, который в полной мере удовлетворял бы всем перечисленным требованиям. С различной степенью полноты этим требованиям удовлетворяют монокристаллы на основе кислородных соединений — окислы и соединения с комплексными анионами, фториды элементов II и III групп, стекла, пластмассы и другие кристаллические и аморфные диэлектрики. Большой интерес представляет использование смешанных систем типа твердых растворов.

Оптимальной совокупностью параметров обладают: рубин ($Al_2O_3 : Cr^{3+}$), алюмоиттриевый гранат с неодимом ($Y_3Al_5O_{12} : Nd^{3+}$), стекло с неодимом, иттербий-эрбиевое стекло, алюминат иттрия с неодимом ($YAlO_3 : Nd^{3+}$), натрий-лан-

тан-молибдат с неодимом ($NaLaMoO_4 : Nd^{3+}$), флюорит кальция с диспрозием ($CaF_2 : Dy^{2+}$) и гадолиний-скандий-галлиевый гранат с хромом и неодимом $[(Gd_{3-x}Nd_x)(Sc_{2-y}Cr_y) \cdot Ga_3O_{12}]$. Все лазеры на этих веществах, за исключением рубина, работают по четырехуровневой схеме (см. рис. 2.9). Основные лазерные свойства этих веществ приведены в табл. 8.1. Значения КПД, представленные в таблице, относятся к лазерам с ламповой накачкой. Применение для этих целей инжекционных лазеров и светодиодов повышает КПД на порядок и улучшает эксплуатационные свойства твердотельных лазеров.

Первой лампой накачки была импульсная спиральная лампа-вспышка с ксеноновым наполнением, аналогичная применяемому в фотографии. Активный элемент в виде рубинового стержня помещался на оси этой лампы, через которую производился разряд мощного конденсатора. Зеркала резонатора в виде серебряных покрытий наносились непосредственно на тщательно отполированные торцы рубинового стержня, как показано на рис. 8.2, а. Современные твердотельные лазеры конструктивно выполняются в виде двух блоков: лазерного излучателя, называемого также лазерной головкой, и блока питания. При накачке импульсными лампами блок питания включает в себя накопитель энергии обычно в виде батареи высоковольтных конденсаторов

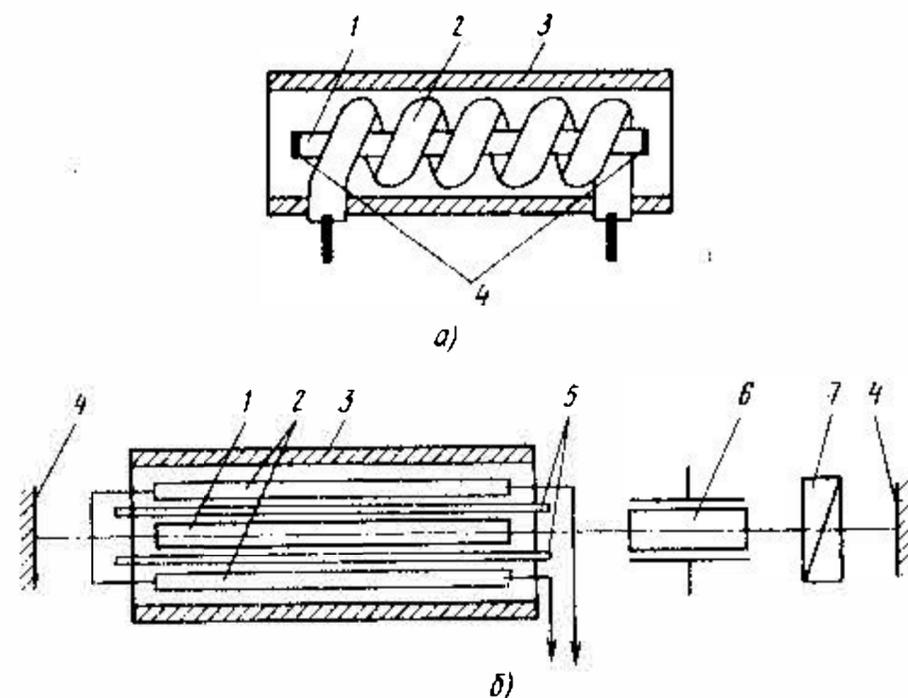


Рис. 8.2. Схемы лазерных излучателей: простейшего рубинового лазера со спиральной лампой накачки (а) и лазера с модулируемой добротностью (б)

1 — активный элемент; 2 — лампы накачки; 3 — отражатель; 4 — зеркала резонатора; 5 — фильтр и система охлаждения; 6 — модулятор добротности; 7 — поляризатор

Таблица 8.1

Основные материалы твердотельных лазеров и их рабочие характеристики

Материалы активной среды	Матрица	Активатор	Концентрация активного центра, мол. %	Длина волны генерации, мкм	Генерационные потери, см ⁻¹	Показатель преломления	КПД, %, режим генерации
Рубин	Al ₂ O ₃	Cr ³⁺	0,03 — 0,05	0,694	0,02	1,76	1%; свободная генерация
Иттрий-алюминиевый гранат (ИАГ) с неодимом	Y ₃ Al ₅ O ₁₂	Nd ³⁺	1 — 3	1,06	0,003 — 0,005	1,83	4%; P _{вых} ≈ 300 Вт; непрерывная генерация
Стекло с неодимом	Стекло	Nd ³⁺	2 — 6	1,06	0,004	1,55	8%; свободная генерация
Стекло с эрбием	Фосфатное стекло с сенсибилизаторами (Yb ³⁺) и (Cr ³⁺)	Er ³⁺	1	1,54	—	—	3%; свободная генерация
Алюминат иттрия с неодимом	YAlO ₃	Nd ³⁺	3	1,06	—	1,95	1%; P _{вых} ≈ 30 Вт; непрерывная генерация
Натрий-лантан-молибдат с неодимом	NaLa (MoO ₄) ₂	Nd ³⁺	2	1,06	0,01	1,82	2,5%; свободная генерация
Флюорит кальция с диспрозием	CaF ₂	Dy ²⁺	0,02	2,36	0,03	1,42	2%; непрерывная генерация
Гадолиний-слюдавый гранат (ГСГГ) с хромом	Gd ₃ Sc ₂ Ga ₃ O ₁₂	Cr ³⁺	—	0,7...0,9	—	2	Лазер с перестраиваемой длиной волны
Сапфир	Al ₂ O ₃	Ti ³⁺	0,1	0,7...0,98	—	1,7	Лазер с перестраиваемой длиной волны. Импульсная генерация, P _{ср} ≈ 0,1 Вт, f ≈ 10 кГц, КПД ≈ 3%. Лазерная накачка.

большой емкости, схему поджига, выполняющую роль ключа, и систему, формирующую заданную длительность и форму импульса тока. В качестве таковой часто применяют индуктивность, включаемую параллельно лампам-вспышкам. Обычно длительность импульса накачки составляет 0,1...3 мс при энергии одного импульса от десятков джоулей до десятков и даже сотен килоджоулей.

Излучатель твердотельного лазера (рис. 8.2, б) состоит из активного элемента 1, источника накачки 2, отражателя 3, зеркал резонатора 4, системы охлаждения 5 и элементов управления излучением 6; 7. Активные элементы на основе монокристаллов обычно выполняются в виде цилиндрических стержней диаметром 3...10 мм и длиной до 10 см. Стекланные активные элементы могут быть как цилиндрической формы диаметром 8...200 мм и длиной более 50 см, так и прямоугольного сечения в виде бруска (в сверхмощных лазерах) длиной до 1 м.

В качестве источников накачки сначала применяли ксеноновые (иногда — криптоновые, аргоновые и ртутные) газоразрядные лампы. Их спектр излучения в первом приближении можно описать излучением черного тела (1.107) с эффективной температурой 6000...10000 К. Естественно, что с увеличением энергии разряда увеличивается температура плазмы и растет доля коротковолнового излучения. Вместо спиральных ламп, применявшихся в первых лазерах, используют прямолинейные трубчатые лампы, работающие в импульсном или непрерывном режимах. Характерные размеры этих ламп: внутренний диаметр кварцевой колбы — от 2 до 30 мм и длина — от 4 см для маломощных ламп непрерывной накачки до 1 м для сверхмощных импульсных ламп. Лампы непрерывного режима имеют мощность 0,5...5 кВт. Они требуют принудительного охлаждения. Импульсные лампы характеризуются предельной энергией одной вспышки $E_{пред}$, которая изменяется в широких пределах от единиц джоулей до десятков килоджоулей. Эта величина определяет предельную энергию вспышки, которую лампа может выдержать без разрушения оболочки. Она связана с внутренним диаметром лампы d , ее длиной l и длительностью импульса τ_n эмпирическим соотношением

$$E_{пред} \approx 1,2 \cdot 10^4 dl \sqrt{\tau_n},$$

где E измеряется в Дж, d — см, l — см, τ_n — с. Отношение рабочей энергии вспышки к предельной определяет срок службы ламп накачки. При $E/E_{пред} = 1$ срок службы не превышает 10 вспышек, при $E/E_{пред} \approx 0,5$ это число увеличивается до 10^3 , а при $E/E_{пред} \approx 0,3$ — до 10^6 .

Хотя КПД импульсных ламп достаточно высок (до 75%), из общего количества излучаемой световой энергии не более

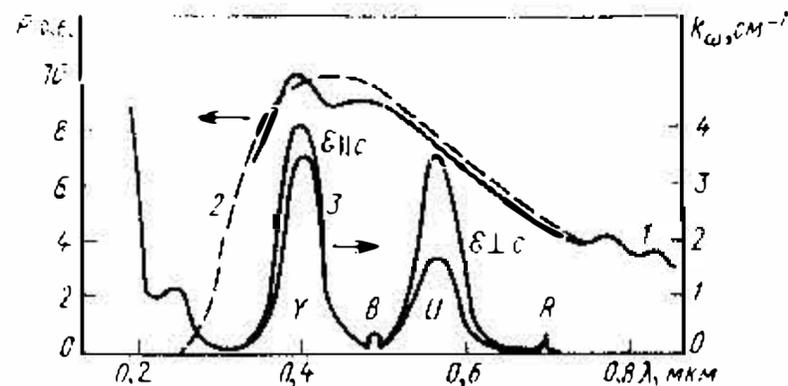


Рис. 8.3. Сравнение спектров излучения импульсной ксеноновой лампы (1), абсолютно черного тела при $T=6000\text{ K}$ (2) со спектром поглощения рубина (3)

25...30% имеет спектральный состав, соответствующий полосам поглощения активной среды (см. рис. 8.3). Остальная энергия не просто теряется, а переходит в тепло, приводя к нагреву активного элемента и элементов конструкции лазерной головки. Более того, коротковолновое ультрафиолетовое излучение накачки с $\lambda < 0,3\text{ мкм}$, соответствующее собственному поглощению матрицы, поглощается в ее тонком приповерхностном слое, приводя к поверхностному разрушению активного элемента. Чтобы избежать этого явления, применяют специальные фильтры, например водяные. Поэтому весьма перспективным является использование полупроводниковых светодиодов и лазеров в качестве источников накачки. Естественно, что их собственная эффективность при этом должна быть достаточно высока.

В маломощных лазерах непрерывного действия в качестве источников накачки могут быть использованы галогенные лампы накаливания с водным циклом, а также солнечное излучение.

Основная функция отражателя — обеспечить максимальную эффективность светопередачи и равномерное освещение активного элемента. Конструкция отражателя зависит от конфигурации активного элемента и от типа и количества ламп накачки. Часто отражатели выполняются в виде круговых или эллиптических цилиндров, в которых активный элемент и лампы накачки устанавливаются параллельно друг другу, как это показано на рис. 8.2, б. В зеркальных отражателях эллиптического типа активный элемент и лампы накачки располагаются на фокальных осях эллиптического цилиндра. В диффузных осветителях обычно используют отражатель в виде кругового цилиндра.

Резонаторы твердотельных лазеров могут иметь одну из конфигураций, рассмотренных в § 2.3. Для более полного использования объема активного вещества и селекции неаксиальных типов колебаний используют неустойчивые резонаторы. Как

и в газовых лазерах, в качестве зеркал резонаторов часто применяют многослойные диэлектрические покрытия.

Твердотельные лазеры могут работать в непрерывном и в импульсном режимах, в том числе в режиме свободной генерации при импульсной накачке и в режиме модулированной добротности. Каждый из них имеет свои особенности. В непрерывном режиме могут работать только некоторые типы лазеров, в частности — лазер на иттрий-алюминиевом гранате с неодимом. При модуляции добротности достигаются наиболее высокие уровни импульсной мощности излучения (до 10^{12} Вт при $\tau \approx 10...30\text{ нс}$), но КПД при этом несколько ниже, чем при свободной генерации.

При выполнении условий синхронизации мод твердотельные лазеры могут генерировать сверхкороткие световые импульсы длительностью 10^{-12} с и менее (см. § 2.5).

Для реализации режима модулированной добротности внутрь резонатора (обычно между непрозрачным зеркалом и активным элементом) вводят элемент управления излучением, выполняющий функции светового затвора. Это могут быть оптико-механические, электрооптические, акустооптические или пассивные (фототропные) модуляторы (см. §§ 10.1 и 10.2).

Спектр излучения твердотельных лазеров и диаграмма направленности определяются активным веществом, резонатором и режимом работы. Оптическая однородность конденсированных сред хуже, чем газообразных. В них возможны заметные вариации показателя преломления, вызванные внутренними напряжениями, неоднородностями состава, дефектами, градиентами температуры и т. д. В процессе работы активный элемент нагревается неравномерно. Появляются большие термоупругие напряжения и вызванные ими неоднородности показателя преломления. Волновой фронт при прохождении через активную среду испытывает большие искажения. Это сказывается как на диаграмме направленности, так и на спектре излучения, которые могут изменяться даже в процессе одного импульса. Все это приводит к тому, что качество излучения, генерируемого твердотельными лазерами, хуже, чем в газовых. Длина когерентности лазера, работающего в импульсном режиме при пиковом характере излучения, составляет всего несколько сантиметров. Основное достоинство твердотельных лазеров — возможность генерации световых импульсов большой мощности, а также непрерывного излучения при сравнительно малых габаритах прибора.

Твердотельные лазеры могут работать не только в режиме генерации излучения, но и как оптические усилители. Наибольший интерес в этом качестве представляют лазеры на стекле, активированном эрбием (табл. 8.1), поскольку их рабочая длина волны (1,54...1,56 мкм) соответствует минимуму потерь в кварцевом волокне (см. § 12.3). Лазерный усилитель выполняется

нии, поскольку уровень \bar{E} заселен больше, чем $2\bar{A}$, и вероятности перехода $\bar{E} \rightarrow {}^4A_2$ также несколько больше, чем $2\bar{A} \rightarrow {}^4A_2$.

Показатель поглощения в U - и Y -полосах накачки составляет $2...4 \text{ см}^{-1}$ при оптимальном содержании ионов Cr^{3+} . Поэтому диаметр активного рубинового элемента не должен превышать 2 см. Обычно применяют стержни диаметром около 1 см и длиной около 10 см. Как правило, выращивают кристаллы, оптическая ось c которых составляет угол 90 или 60° с осью стержня. Излучение такого кристалла и соответственно лазерное излучение линейно поляризованы с вектором E , перпендикулярным плоскости, проходящей через ось c и ось стержня. Пороговое значение объемной плотности энергии накачки в зеленой U -полосе рубинового лазера примерно равно 3 Дж/см^3 . При значительном превышении порога в режиме свободной генерации удельный съем энергии составляет $0,2...0,25 \text{ Дж/см}^3$, а показатель усиления — около $0,2 \text{ см}^{-1}$. Типичные характеристики рубинового лазера приведены в табл. 8.1. Основным его недостатком — большие пороговые энергии накачки, что обусловлено трехуровневой схемой работы и низким КПД.

§ 8.3. ЛАЗЕРЫ НА КРИСТАЛЛАХ И СТЕКЛАХ, АКТИВИРОВАННЫХ НЕОДИМОМ

Основные недостатки рубинового лазера связаны с трехуровневым механизмом его работы, что обусловлено особенностями трехвалентного иона хрома. Более удачными в этом отношении оказались ионы редкоземельных элементов, строение энергетических уровней которых позволяет осуществить работу по четырехуровневой схеме.

Незаполненная $4f$ -оболочка редкоземельных элементов расположена ближе к ядру, чем $3d$ -оболочка элементов группы железа, и хорошо экранирована от внешних полей $5s$ - и $5p$ -электронами. Поэтому здесь наблюдается случай слабого кристаллического поля. Для всех редкоземельных ионов характерно наличие узких линий люминесценции на переходах между состояниями оболочки $4f$ и интенсивных полос поглощения на переходах $4f \rightarrow 5d$, используемых для накачки. Энергетическое положение уровней слабо зависит от типа матрицы.

Из всех 14 редкоземельных элементов наилучшие результаты получены на трехвалентном ионе неодима Nd^{3+} . Этот ион обладает оптимальным набором параметров энергетических уровней, наилучшим образом удовлетворяющим требованиям, предъявляемым четырехуровневой схемой работы лазера. Эффективные процессы электрон-решеточной релаксации осуществляют как быстрое опустошение нижнего лазерного уровня, так и переходы с верхних уровней накачки на верхний рабочий уровень, на

котором происходит накопление частиц.

Генерация с участием иона Nd^{3+} наблюдалась более чем в 100 различных матрицах. Из них оптимальной совокупностью свойств, удовлетворяющей сформулированным в § 8.1 требованиям, обладают кристаллы иттрий-алюминиевого граната $\text{Y}_3\text{Al}_5\text{O}_{12}$ (называемые также ИАГ или YAG), кристаллы гадолиний-скандий-галлиевого граната $\text{Gd}_3\text{ScGa}_3\text{O}_{12}$ (кристаллы ГСГГ) и стекла.

На рис. 8.6 приведена диаграмма энергетических уровней Nd^{3+} в ИАГ. Оптическая накачка переводит ионы Nd^{3+} из основного состояния ${}^4I_{9/2}$ в серию возбужденных состояний, состоящую из большого числа узких, частично перекрывающихся друг с другом уровней. Основные полосы накачки расположены на длинах волн $0,73$ и $0,8 \text{ мкм}$. Последняя полоса очень удобна для накачки полупроводниковыми лазерами и светодиодами на основе арсенида галлия. С этих уровней энергии осуществляется быстрая релаксация на метастабильный уровень ${}^4F_{3/2}$ с временем жизни $\tau_{21} = 0,23 \text{ мс}$. Переходы с этого уровня в состояния ${}^4I_{15/2}$, ${}^4I_{13/2}$, ${}^4I_{11/2}$ и ${}^4I_{9/2}$ сопровождаются эффективной люминесценцией в области длин волн $1,8$; $1,3$; $1,06$ и $0,95 \text{ мкм}$. Наиболее сильным является переход ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ в области $1,06 \text{ мкм}$. Обычно на этом переходе и осуществляется генерация. Уровень ${}^4I_{11/2}$, являющийся нижним рабочим лазерным уровнем при генерации в области $1,06 \text{ мкм}$, расположен примерно на $0,25 \text{ эВ}$ выше основного состояния ${}^4I_{9/2}$. Эти уровни связаны между собой быстрой безызлучательной релаксацией. В то же время расстояние между ними существенно больше kT , что обеспечивает эффективное опустошение нижнего рабочего уровня и определяет четырехуровневый характер генерации неодимового лазера.

Верхний лазерный уровень ${}^4F_{3/2}$ расщеплен на два подуровня, а нижний уровень ${}^4I_{11/2}$ — на шесть подуровней. Поэтому полоса люминесценции $1,06 \text{ мкм}$ имеет сложную структуру. Наиболее сильная линия является однородно уширенной с шириной около $\Delta\nu_{0,5} \approx 6,5 \text{ см}^{-1} \approx 195 \text{ ГГц}$ при комнатной температуре, как показано на рис. 8.7. Аналогичная ситуация наблюдается и для линий излучения в области $1,3$ и $0,9 \text{ мкм}$.

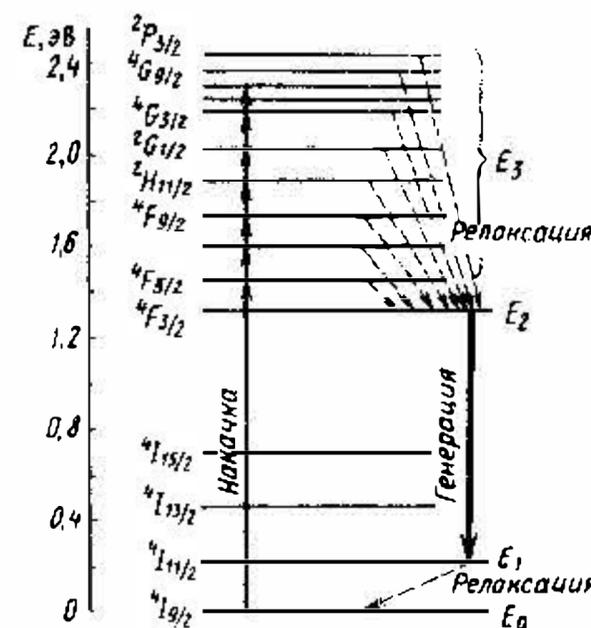


Рис. 8.6. Упрощенная диаграмма энергетических уровней Nd^{3+} в $\text{Y}_3\text{Al}_5\text{O}_{12}$ и схема работы неодимового лазера

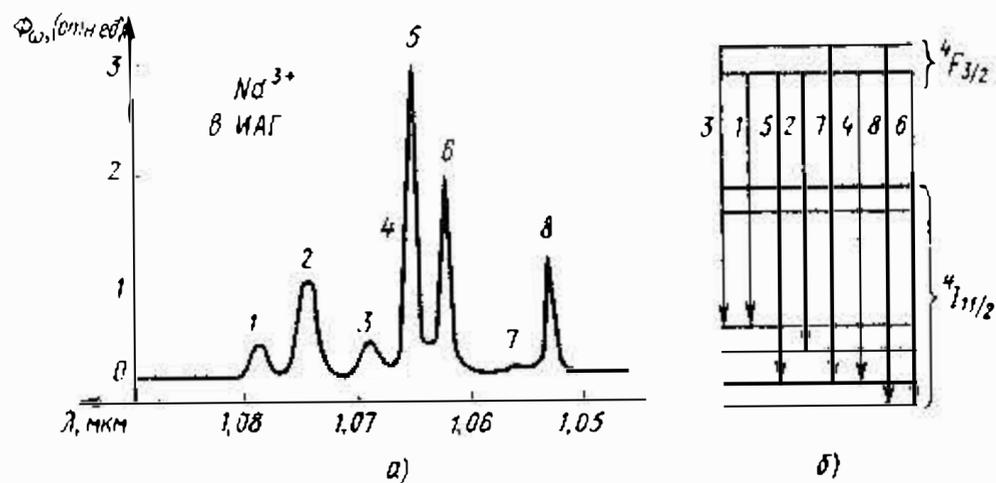


Рис. 8.7. Спектр люминесценции Nd^{3+} в ИАГ при $T=300$ К в полосе ${}^4F_{3/2} - {}^4I_{11/2}$ (а) и соответствующие электронные переходы (б)

Лазеры на гранате (ИАГ : Nd^{3+} -лазеры) работают как в непрерывном, так и в импульсном режимах с большой частотой повторения со средней мощностью до нескольких сотен ватт.

Это оказывается возможным благодаря низким пороговым энергиям накачки, высокой теплопроводности ИАГ и малым оптическим потерям. Лазер с ламповой накачкой конструктивно выполняется в соответствии со схемой (рис. 8.2). Однако, отсутствие широких полос поглощения в кристаллах, активированных элементами с $4f$ — незаполненной оболочкой, делает такую накачку малоэффективной. Поэтому в современных ИАГ : Nd^{3+} лазерах в качестве источников накачки применяют мощные инжекционные лазеры и лазерные линейки (см. § 9.10), спектр излучения которых хорошо согласуется со спектром поглощения ${}^4I_{9/2} \rightarrow {}^4F$. Это повышает КПД лазера почти на порядок по сравнению с данными, приведенными в табл. 8.1. Лазер может быть выполнен в едином блоке с нелинейными элементами — преобразователями частоты, генерируя излучение на второй, третьей и даже четвертой гармониках в области длин волн 0,53; 0,35 и 0,265 мкм при выходной мощности от нескольких мВт до единиц Вт.

Кроме основной линии излучения 1,0648 мкм, обусловленной переходами ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$, возможно получение лазерного эффекта на переходах ${}^4F_{3/2} \rightarrow {}^4I_{13/2}$ в области 1,319 мкм и на переходах ${}^4F_{3/2} \rightarrow {}^4I_{9/2}$ в области 0,946 мкм. Однако пороговые энергии накачки для этих переходов существенно выше.

Заметная разница в величинах ионных радиусов Nd^{3+} и Y^{3+} затрудняет получение кристаллов больших размеров с равномерным распределением активатора. Поэтому проводится интенсивный и весьма успешный поиск новых сред со структурой граната, обладающих такими же удачными сочетаниями физико-химических и спектрально-генерационных свойств, как ИАГ : Nd^{3+} , но лишенных этого недостатка. Одним из таких материалов является

ся гадолиний-скандий-галлиевый гранат с неодимом (ГСГГ : Nd^{3+}).

К недостаткам всех кристаллов, активированных редкими землями, относится отсутствие широких полос поглощения. Для увеличения эффективности накачки при применении источников излучения со спектром, близким к излучению абсолютно черного тела, возможно использовать эффект сенсibilизации (очувствления). Для этого в кристаллическую матрицу основы наряду с активными ионами добавляют ионы другого вида, называемые сенсibilизаторами. Их роль заключается в поглощении энергии возбуждения в широком спектральном диапазоне и передаче ее основным рабочим ионам. Добавление сенсibilизатора приводит к расширению эффективной полосы накачки и к повышению эффективности лазера. Передача энергии возбуждения от сенсibilизатора к основному иону может осуществляться различным образом, в том числе — путем реабсорбции излучения иона-сенсibilизатора ионом активатора, путем резонансного взаимодействия ионов, по каскадной схеме и т. п. Во всех случаях передача энергии от сенсibilизатора к активатору будет наиболее эффективна при совпадении (резонансе) соответствующих энергетических состояний. Сенсibilизаторами могут служить не только один или несколько ионов, специально вводимых в матрицу, но и ионы, входящие в состав основного вещества. В ИАГ : Nd^{3+} в качестве сенсibilизатора могут быть использованы ионы Cr^{3+} .

Наряду с кристаллическими матрицами в лазерной технике широко используются активные среды на стеклянной основе с примесью редкоземельных элементов (наиболее часто — Nd^{3+}). К преимуществам стекол как лазерных активных материалов относятся: высокая оптическая однородность, технологичность и возможность изготовления активных элементов больших размеров, возможность введения активатора в необходимых концентрациях с равномерным его распределением по объему.

В то же время стекла, как активные материалы твердотельных лазеров, обладают двумя существенными недостатками: 1) низкой теплопроводностью и высоким температурным коэффициентом линейного расширения и 2) ограниченной областью оптической прозрачности (0,33...2,5 мкм), которая при введении примесей может еще более сужаться.

Сравнение свойств кристаллических и аморфных матриц показывает, что эти материалы дополняют друг друга и, следовательно, одинаково важны для лазерной техники.

Неоднородность окружения иона-активатора в стеклах вызывает неоднородность локальных электростатических полей, воздействующих на этот ион. Это приводит к сильному неоднородному уширению спектральных линий как в люминесценции, так и в поглощении. Уширение линий поглощения и люминесценции

§ 8.4. ТВЕРДОТЕЛЬНЫЕ ПЕРЕСТРАИВАЕМЫЕ ЛАЗЕРЫ

элементов с $3d$ -незаполненными оболочками в стеклах настолько велико, что они перекрываются друг с другом. Для редкоземельных элементов с $4f$ -оболочкой это уширение значительно слабее и составляет примерно $0,03$ эВ. Стекла по сравнению с кристаллами допускают введение большей концентрации активных центров, что компенсирует некоторое снижение усиления, вызванное неоднородным уширением спектральных линий (см. табл. 8.1).

Возможность изготовления оптически совершенных активных элементов больших размеров (длиной до 1 м с поперечным сечением более 500 см²) при большой концентрации активных центров позволяет получать от стеклянных лазеров очень большие энергии выходного импульса (до нескольких тысяч джоулей). В то же время низкая теплопроводность стекла ограничивает область его применения в основном лазерами с небольшой средней мощностью излучения, т. е. с малой частотой следования импульсов. При необходимости получения сверхвысоких энергий при сохранении удовлетворительного качества излучения используют многокаскадную схему «задающий генератор — усилитель мощности». Задающим генератором может служить гранатовый лазер на ИАГ: Nd^{3+} , а усилителем мощности (один или несколько каскадов) — лазер на неодимовом стекле.

Неоднородное уширение линий люминесценции в стекле позволяет, используя метод синхронизации мод, генерировать сверхкороткие световые импульсы длительностью до $0,5$ пс ($5 \cdot 10^{-13}$ с), т. е. получать сгустки световой энергии в объеме с характерными размерами порядка $0,15$ мм, в которых сосредоточена мощность до 10 ТВт (10^{13} Вт) и более.

При генерации больших световых мощностей (энергий) важное значение имеет *лучевая стойкость* активного элемента, т. е. его способность выдерживать большие потоки лазерного излучения без разрушения. Большое значение здесь имеет однородность активного вещества. Любые, в том числе микроскопические, неоднородности могут явиться теми центрами, на которых начинается лавинообразный процесс разрушения. Для неодимовых стекол, не содержащих металлических включений (примесей технологического происхождения), порог разрушения в режиме свободной генерации ($\tau_n \approx 0,1 \dots 1$ мс) составляет $10^3 \dots 10^4$ Дж/см². В режиме модулированной добротности ($\tau_n \approx 1 \dots 10$ нс) при малом пятне облучения он на порядок меньше, а при поперечном сечении более 1 см² — еще на порядок ниже, т. е. $10 \dots 100$ Дж/см². Уменьшение лучевой стойкости в последнем случае объясняется увеличением вероятности попасть под излучение такому дефекту материала, который инициирует лавинообразный процесс разрушения.

В режиме синхронизации мод пороги разрушения не превышают единиц Дж/см². Рубиновые и гранатовые активные элементы разрушаются в режиме модулированной добротности при энергиях ($10 \dots 30$) Дж/см².

Перестройка частоты (длины волны) излучения лазера может быть осуществлена в пределах контура спектральной линии люминесценции. Для рассмотренных рубинового и неодимовых лазеров этот диапазон очень мал. Даже в лазерах на стекле с неодимом он не превышает нескольких ангстрем. Для создания перестраиваемых лазеров надо использовать активные материалы с широкими интенсивными полосами люминесценции. Такая люминесценция возникает от центров, эффективно взаимодействующих с колебаниями решетки и обладающих энергетическими уровнями, уширенными кристаллическим полем матрицы. Очевидно, редкоземельные элементы с хорошо экранированной $4f$ -оболочкой для этой цели не подходят.

Один из путей создания твердотельных перестраиваемых лазеров — использование в качестве активаторов ионов с незаполненными $3d$ - и $4d$ -оболочками. В качестве примера такого лазера может служить лазер на александрите.

Александрит — это хризоберилл ($BeAl_2O_4$), часть ионов Al^{3+} которого изоморфно замещена ионами Cr^{3+} . Это двухосный кристалл с хорошими оптическими и механическими свойствами. Его теплопроводность всего в два раза хуже теплопроводности рубина (и в два раза лучше теплопроводности ИАГ). Важно, что его оптические качества не ухудшаются при замещении до 40% ионов Al^{3+} ионами Cr^{3+} . Основной недостаток александрита как лазерного материала — сложность технологии выращивания высококачественных образцов приемлемых размеров, что связано с присутствием бериллия и его оксида, являющихся опасными для здоровья.

На рис. 8.8, а приведена конфигурационная диаграмма для ионов хрома в александрите (а) и схема работы александритового лазера (б). Вследствие сравнительно слабой экранировки $3d$ -электронов ионов хрома, энергетические положения, ширина и вероятность переходов между соответствующими уровнями существенно меняются от основы к основе. Если сравнить рис. 8.8 и 8.4, то нетрудно заметить, что энергетическое расстояние между состояниями 2E и 4T_2 в александрите существенно меньше, чем аналогичный зазор ${}^2E - {}^4F_2$ в рубине (энергетические уровни 4F_2 и 4F_1 ионов Cr^{3+} в рубине эквивалентны уровням 4T_2 и 2T_1 хрома в александрите), и равно $0,1$ эВ. Между этими состояниями происходит очень эффективный обмен энергией, так что время термализации ${}^2E - {}^4T_2$ имеет порядок 10^{-13} с. Излучательные времена жизни равны $1,5$ мс для состояния 2E (как и в рубине, оно является метастабильным) и $6,6$ мкс для состояния 4T_2 . Излучательные переходы с уровня 2E

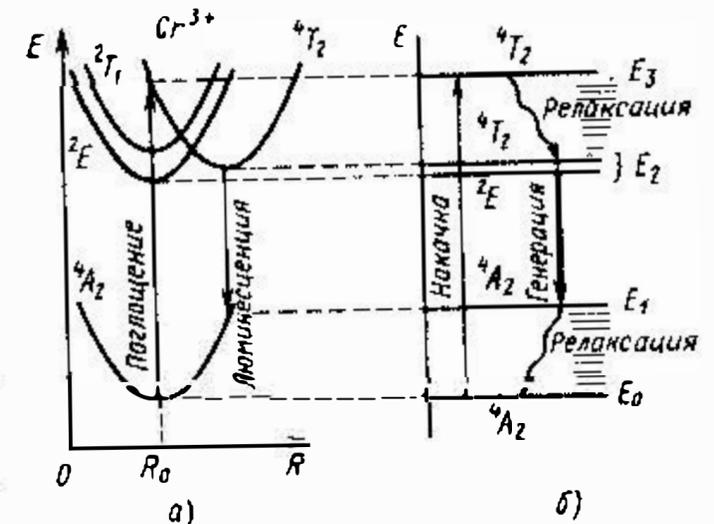


Рис. 8.8. Конфигурационная диаграмма для уровней иона хрома в александрите (а) и схема работы александритового лазера (б)

в основное состояние 4A_2 дают узкую R -линию люминесценции с максимумом при $\lambda=681$ нм, аналогичную R -линиям в рубине. С коротковолновой стороны к этой линии примыкает широкая полоса люминесценции, простирающаяся от 690 до 820 нм. Эта полоса обусловлена излучательными переходами с уровней 4T_2 в состояние 4A_2 . Поскольку минимумы на конфигурационной диаграмме для этих уровней расположены при разных конфигурационных координатах, то в соответствии с принципом Франка — Кондона эти переходы сопровождаются сильным взаимодействием с колебаниями решетки (так называемые вибронные переходы), что приводит к существенному уширению спектра излучения и к возможности перестройки частоты генерации внутри этого спектра.

Работа лазера на александрите пояснена на рис. 8.9, б. Оптическая накачка осуществляется в перекрывающиеся между собой полосы 2T_1 и 4T_2 . За счет интенсивного взаимодействия с решеткой электроны по колебательным подуровням быстро (за время порядка 10^{-13} с) релаксируют в нижние состояния 2E и 4T_2 аналогично тому, как это изображено на рис. 4.41. Верхним лазерным уровнем является уровень 4T_2 . Уровень 2E расположен выше, но излучательное время жизни на нем больше чем на три порядка превышает время жизни 4T_2 -уровня. Поэтому уровень 2E за счет малого зазора $^2E - ^4T_2$ и интенсивного обмена энергией между этими уровнями выполняет функцию накопительного уровня, непрерывно подпитывающего уровень 4T_2 . Нижним рабочим лазерным уровнем служит один из колебательных уровней состояния 4A_2 . Релаксация с этого уровня в основное состояние 4A_2 осуществляется опять-таки за счет взаимодействия с колебаниями решетки. Из рис. 8.8 видно, что при условии $\Delta E = E_1 - E_0 > kT$, что соблюдается в рассматриваемой системе, лазер работает по четырехуровневой схеме. При применении селективного резонатора, настроенного на заданную длину волны, выбор стартового и финишного лазерных уровней в пределах колебательной полосы $^4T_2 - ^4A_2$ осуществляется автоматически.

Накачка лазера на александрите может осуществляться как с помощью другого лазера (например, газового), аналогично тому, как это делается в жидкостном лазере на органических красителях, так и с помощью ксеноновой лампы-вспышки. Область перестройки лазера составляет 700...820 нм. Средняя мощность в импульсно-периодическом режиме с частотой следования 5...100 Гц достигает 50 Вт, а в режиме модулированной добротности получены гигантские импульсы длительностью 30...200 нс перестраиваемые по длине волны в диапазоне 700...820 нм.

По аналогичной схеме работает перестраиваемый лазер на гадолиний-скандий-галлиевом гранате с хромом, а также лазер на сапфире ($Al_2O_3 : Ti^{3+}$). Их параметры приведены в табл. 8.1.

В качестве активаторов люминесценции лазерных кристаллов могут выступать не только элементы с незастроенными внутренними оболочками, но и другие центры, дающие интенсивное излучение с высоким квантовым выходом и обладающие бо-

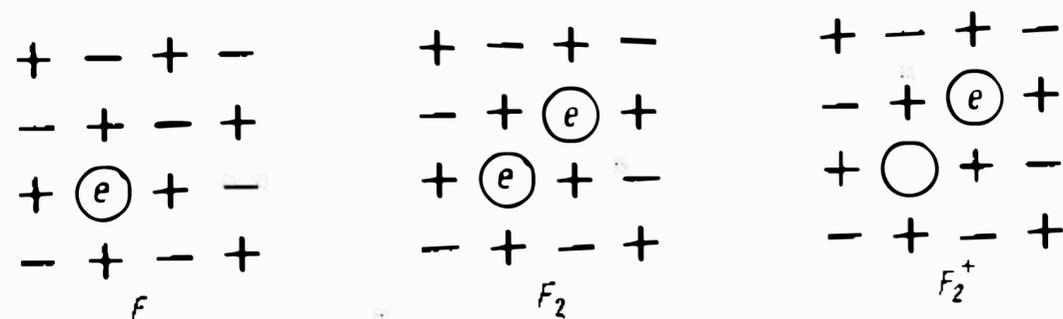


Рис. 8.9. Структура некоторых F -центров окраски в щелочно-галогенидных кристаллах

шими сечениями взаимодействия. Примером таких центров могут служить так называемые центры окраски в ионных кристаллах, например, в щелочно-галогенидных кристаллах типа NaCl, LiF и т. п. Эти центры образуются на базе собственных точечных дефектов кристалла. Разнообразные комплексы, образованные такими дефектами, способны создавать широкие полосы поглощения, придавая различную окраску бесцветным бездефектным кристаллам. Обширный класс таких комплексов принято называть общим термином — F -центры*. Простейший дефект решетки (собственно F -центр) представляет собой анионную вакансию — отсутствие отрицательного иона в соответствующем узле кристаллической решетки. В ионных кристаллах такая вакансия действует как эффективный положительный заряд. Поэтому свободный электрон захватывается анионной вакансией — так образуется простейший центр окраски. Его можно рассматривать как глубокий центр с радиусом локализации порядка межатомного расстояния. Элементарные образования типа «вакансия — электрон» в ионных кристаллах обладают склонностью образовывать комплексы в виде парных центров и центров окраски более высокого порядка. Центры окраски такого типа обозначают символами F_2, F_3, F_4 и т. д. Например, F_2 -центр представляет собой ассоциацию из двух F -центров, находящихся в соседних анионных узлах (рис. 8.9). Он подобен квазимолекуле водорода. F_2^+ -центр представляет собой ионизованный F_2 -центр, т. е. один электрон, движущийся в поле двух соседних анионных вакансий.

Получение центров нужного типа производится путем довольно сложной фотохимической и термической обработки. К сожалению, многие из таких центров характеризуются низкой термической и оптической стабильностью. Поэтому хранить такие кристаллы и работать с ними бывает необходимо при температуре жидкого азота ($77^\circ K$).

Как и прочие глубокие центры, F -центры сильно взаимодействуют с колебаниями решетки. Их спектры поглощения и люминесценции заметно уширены и между ними наблюдается боль-

*Название « F -центр» происходит от нем. die Farbe — цвет, краска.

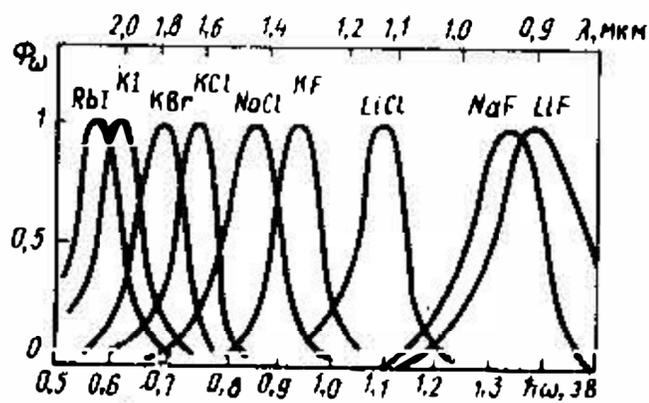


Рис. 8.10. Спектры люминесценции центров окраски в различных щелочно-галогидных кристаллах

состояниями водородоподобного атома. Однако собственно F -центры не являются оптимальными для создания лазеров, поскольку для них наблюдаются существенные потери (малый квантовый выход), а сами центры нестабильны. Лучшие результаты получаются на F_2 , F_2^+ , F_2^- и некоторых других центрах окраски. Сильное взаимодействие с колебаниями решетки обеспечивает быструю релаксацию в состояние E_2 и быстрое опустошение состояния E_1 за время порядка 1 пс при излучательном времени жизни состояния E_2 порядка 10 нс.

Основное достоинство лазеров на центрах окраски — возможность непрерывной перестройки частоты. Их возбуждение (накачки) обычно осуществляется другим лазером с длиной волны, совпадающей со спектром активного поглощения центра окраски. Часто используют Kr^+ -лазер или ИАГ: Nd^{3+} -лазер, применяя схему продольной накачки, как это делается в жидкостных лазерах на органических красителях. Грубая перестройка частоты (длины волны) излучения осуществляется с помощью дисперсионной оптической системы, состоящей из призмы или решетки, аналогично жидкостным лазерам.

Так как оптические сечения взаимодействия для центров окраски велики, то размеры активного элемента могут быть небольшими — порядка нескольких миллиметров. Типичная пороговая мощность накачки составляет несколько десятков милливатт при фокусировке излучения накачки в активном кристалле в пятно диаметром око-

ло 20 мкм. Мощность непрерывной генерации составляет 0,1...1 Вт при КПД 30% и выше. Используя центры окраски различного типа в различных щелочно-галогидных кристаллах, возможно непрерывно перекрыть диапазон от 0,5 до 4,0 мкм (см. рис. 8.1 и 8.11).

Конфигурационная диаграмма на рис. 8.11 поясняет работу лазера. Здесь показаны основное и первое возбужденное состояния центра, в которые осуществляется переход под действием излучения накачки. Для F -центра это переход $1s \rightarrow 2p$ между

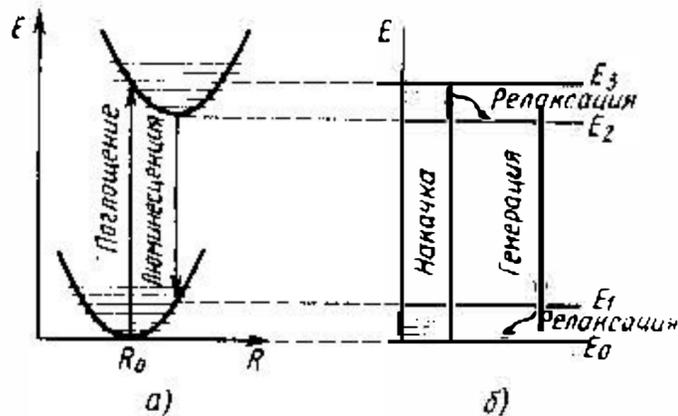


Рис. 8.11. Конфигурационная диаграмма (а) и схема работы перестраиваемого лазера на центрах окраски (б)

ло 20 мкм. Мощность непрерывной генерации составляет 0,1...1 Вт при КПД 30% и выше. Используя центры окраски различного типа в различных щелочно-галогидных кристаллах, возможно непрерывно перекрыть диапазон от 0,5 до 4,0 мкм (см. рис. 8.1 и 8.11).

§ 8.5. ВОЛОКОННЫЕ ЛАЗЕРЫ И ЛАЗЕРНЫЕ УСИЛИТЕЛИ

Развитие техники волоконно-оптической связи (ВОЛС, см. § 12.3) и необходимость передачи огромных объемов информации на расстояния сотни и тысячи километров потребовали создания усилителей оптических сигналов, свободно встраиваемых в волокно, по которому передается этот сигнал. Создание таких оптических усилителей стало возможным благодаря успехам в технологии получения и легирования оптических волокон исключительно высокого качества, созданию мощных инжекционных лазеров с высокими эксплуатационными характеристиками (см. § 9.10), обеспечивающими эффективную накачку; разработке эффективных методов ввода лазерного излучения в волокно, включая системы волоконного суммирования излучения инжекционных лазеров.

Упрощенная схема волоконного лазера приведена на рис. 8.12. Усилительный элемент выполняется в виде кварцевого волокна, легированного ионами-активаторами, в качестве которых используют редкоземельные элементы. В волоконных лазерах, работающие как самостоятельные генераторы (источники) излучения в качестве активаторов могут использоваться, например, ионы Nd^{3+} . Лазерный элемент в этом случае работает по схеме, представленной на рис. 8.6, но система накачки и резонатор выполнены иначе, чем в традиционных твердотельных лазерах с объемными активными элементами. В оптическом волокне, подробнее рассмотренном в § 12.3, свет распространяется как в диэлектрическом волноводе, формируя собственные типы колебаний, обладающие минимальными потерями (см. § 3.4). Диаметр центральной части (ядра) волокна, по которому собственно и распространяется излучение, может изменяться от десятков и сотен микрон для лазерных элементов-источников излучения до единиц микрон для лазерных усилителей, встраиваемых в одномодовые оптические волокна системы ВОЛС. Длина активного волокна при этом может составлять от единиц до нескольких десятков метров. Зеркала резонатора (отражатели) встраиваются непосредственно в волокно и выполняются, например, в виде распределенной обратной связи по схеме (см. рис. 2.26). Собственные типы колебаний в такой системе формируются сложным образом и определяются не только (а часто и не столько) параметрами «обычного» резонатора, рассмотренного в § 2.3, но



Рис. 8.12. Схема волоконного лазера и лазерного усилителя с диодной накачкой

и модовой структурой поля в самом волокне, зависящей от соотношения его диаметра и длины волны.

Обычная оптическая накачка через боковые поверхности активного элемента в волоконном лазере неэффективна. Поэтому ввод излучения накачки в волокно осуществляется методами интегральной оптики, применяя, например, призмное устройство ввода излучения в тонкопленочный волновод (см. рис. 12.21), используя эффекты просачивания (туннелирования) оптического излучения при полном внутреннем отражении света (см. рис. 3.6).

В качестве источников накачки в волоконных лазерах применяют лазерные диоды, спектр излучения которых соответствует спектру поглощения активных ионов (например, переходам ${}^4I_{9/2} \rightarrow {}^4F_{5/2}$ в волоконном Nd^{3+} -лазере). Излучение накачки может быть введено или в торец активного волокна или распределено по его длине. Для увеличения мощности накачки может быть использовано несколько лазерных диодов или лазерная линейка, в которой суммируется излучение отдельных лазеров. Современные мощные инжекционные лазеры способны генерировать излучение мощностью несколько Вт в непрерывном режиме при КПД более 50% и сроке службы 10.000 ч и более (см. § 9.10) и по своим характеристикам идеально подходят в качестве источников накачки твердотельных лазеров.

Волоконные лазеры, выполненные по рассмотренной выше схеме, компактны и удобны в эксплуатации. Активное волокно может быть собрано в моток и залито компаундом. Внутри волокна при необходимости могут быть встроены элементы управления лазерным излучением. Средняя мощность излучения для разных типов волоконных лазеров составляет от нескольких мВт до единиц Вт и более.

В лазерных усилителях, в отличие от генераторов, обратная связь не должна приводить к самовозбуждению, что обеспечивается выбором параметров отражателя. Ввод усиливаемого излучения осуществляется через оптический разъем, изображенный пунктиром на рис. 8.12. Лазерные усилители наиболее важны для

линий волоконно-оптической связи, обеспечивающих передачу оптических сигналов на дальние расстояния. Поэтому они должны работать в области длин волн, соответствующих минимуму потерь в волокне. Как будет показано в § 12.3, наиболее благоприятной для этих целей областью является диапазон спектра вблизи $\lambda = 1,54$ мкм. Поэтому ионы Nd^{3+} как активаторы для этих целей не подходят. Оптимальными являются ионы эрбия (Er^{3+}), обладающие интенсивной люминесценцией как раз в области 1,52...1,57 мкм.

Схема энергетических уровней Er^{3+} в кварцевом волокне приведена на рис. 8.13. Как видно из диаграммы, накачка может осуществляться в одной из трех областей спектра: 1,48 мкм, 0,989 мкм или 0,545 мкм. Из них полосы 1480 нм и 980 нм обеспечивают наилучшую эффективность. Спектры поглощения и люминесценции Er^{3+} в рабочей области, называемой С-областью, и соответствующей минимуму потерь нелегированного волокна, приведены на рис. 8.14.

При накачке в области 1480 нм эффективность оптического преобразования весьма высока, однако шумовые характеристики при этом несколько хуже, чем при накачке на длине волны 980 нм. Поэтому для достижения больших коэффициентов усиления при низком уровне входного сигнала на практике применяют в качестве первой секции, выполняющей функции предусилителя, волокно, легированное Er^{3+} , накачку которого проводят в области

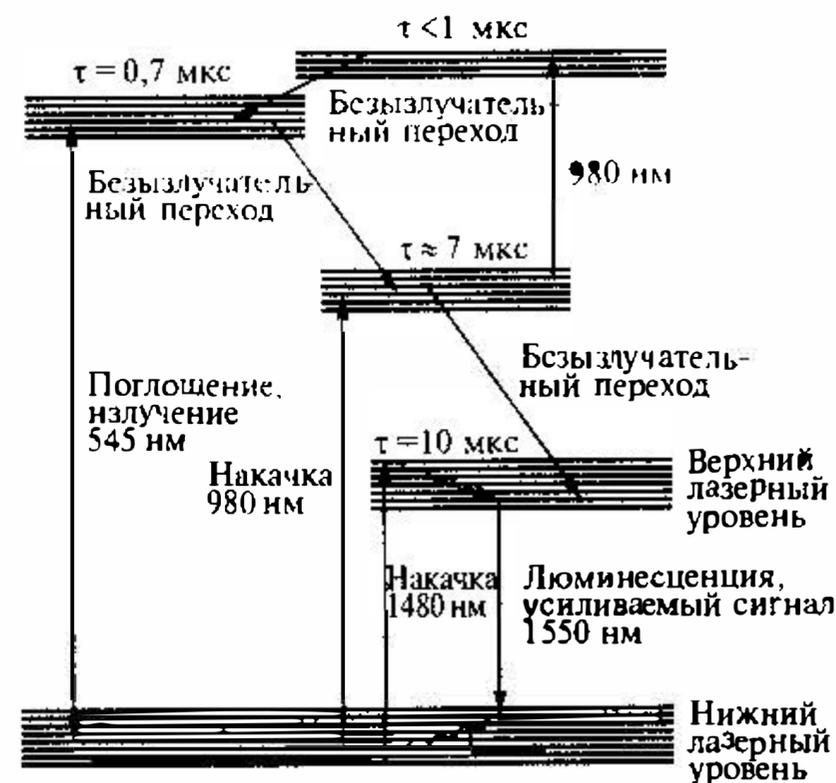


Рис. 8.13. Схема энергетических уровней трехвалентного иона эрбия (Er^{3+}) в кварцевом волокне. Обозначены характерные полосы поглощения и люминесценции, а также времена жизни возбужденных состояний

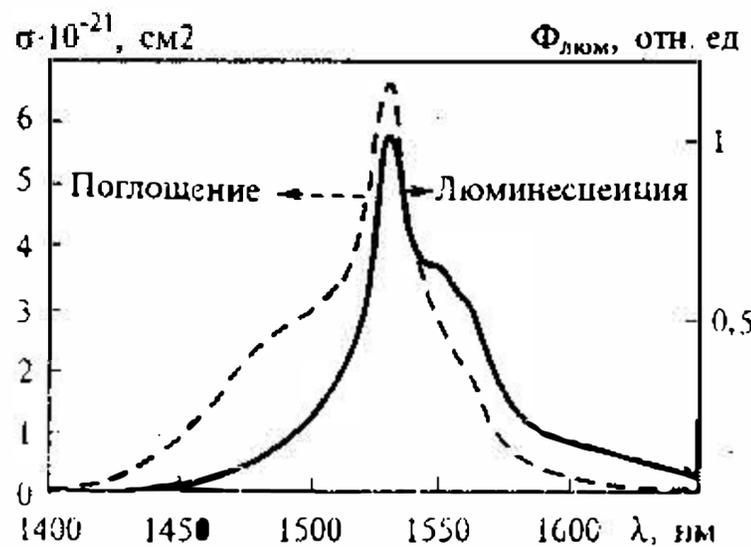


Рис. 8.14. Спектры поглощения (пунктир) и излучения (сплошная кривая) трехвалентного иона эрбия (Er^{+3}) в кварцевом волокне в С-области

в волоконных усилителях ВОЛС, возможно применение трехвалентных ионов иттербия (Yb^{+3}) с дополнительным сдвигом спектра в длинноволновую область при комбинационном рассеянии света (см. рис. 1.14 и § 5.4). Такие усилители удобны в системах многочастотного (многоволнового) мультиплексирования тогда, когда по одному волокну на разных, близко расположенных несущих частотах (рабочих длинах волн) передается несколько сигналов.

Лазерный волоконный усилитель встраивается в сеть ВОЛС с помощью обычных оптических разъемов, как показано на рис. 8.12.

§ 8.6. ЖИДКОСТНЫЕ ЛАЗЕРЫ

Успешное использование редкоземельных элементов (в частности, неодима) как активных центров в твердотельных лазерах наводит на мысль о замене кристаллической или стеклянной матрицы на жидкость, тем более что стекло по существу представляет собой переохлажденную жидкость. Действительно, для редкоземельных элементов характерен случай слабого кристаллического поля, а потому, казалось бы, замена твердотельной матрицы на жидкостную не должна сильно повлиять на свойства активного центра. В то же время использование конденсированной среды — основы в виде жидкости — сулит целый ряд заманчивых преимуществ: легкость получения активной среды и изменения ее состава, возможность прокачки и вытекающая отсюда легкость охлаждения, оптическая однородность и т. п. Однако в действительности оказалось, что подобрать жидкость-растворитель в качестве основы, удовлетворяющей требованиям, пере-

980 нм. Накачку второй секции, выполняющей функцию усилителя мощности, проводят в области 1480 нм. Длина волновая накачка предпочтительней, поскольку энергия, выделяемая в процессе безызлучательной релаксации и приводящая к нагреву активного волокна, будет минимальна.

Кроме ионов Er^{+3} , выступающих в качестве активных центров

численным в § 8.1, очень сложно. Несмотря на то что было опробовано много сотен вариантов различных жидкостей, реализовать лазерную генерацию удалось лишь в немногих из них, в том числе в растворах металлоорганических комплексов — хелатах редкоземельных элементов, в растворах Nd_2O_3 , в хлористом фосфориле POCl_3 , в хлористом селениде SeOCl_2 и некоторых других токсичных и химически активных жидкостях. Эти лазеры, не обладая заметными преимуществами перед твердотельными лазерами на редкоземельных ионах, уступают им по своим рабочим характеристикам, а потому пока не находят широкого применения.

Наибольший интерес и наибольшее значение представляют жидкостные лазеры на органических красителях*. По своим свойствам и механизму работы они наиболее близки к лазерам на центрах окраски. Эти лазеры замечательны прежде всего тем, что допускают плавную перестройку частоты (длины волны) генерации во всей видимой области спектра, захватывая ближний УФ и ИК диапазоны. Каждый из типов красителей допускает перестройку длины волны генерации в диапазоне шириной несколько десятков нанометров при очень высокой монохроматичности, достигающей 1...1,5 МГц. Различные по конструкции, источникам накачки и типу активных веществ лазеры на органических красителях могут работать в непрерывном, импульсном и импульсно-периодическом режимах. Энергия одного импульса может достигать сотен джоулей, а мощность непрерывной генерации — десятков ватт при КПД в несколько процентов при лазерной накачке. В режиме синхронизации мод могут быть получены пикосекундные и субпикосекундные лазерные импульсы.

Весь набор этих свойств делает лазеры на красителях уникальными источниками монохроматического излучения высокого качества с перестраиваемой длиной волны. Рассмотрим их работу более подробно.

Активным веществом лазеров на красителях служат растворы молекул органических красителей в органических растворителях или в воде. Красителями принято называть сложные органические соединения с разветвленной системой сопряженных химических связей. Эти соединения, как правило, обладают интенсивной окраской, что вызвано наличием сильных полос поглощения в видимой области спектра. Эти свойства краситель проявляет лучше в жидких и твердых растворах. В кристаллической форме характерное для него поглощение ослабевает или вообще исчезает. Структура молекулы красителя сложна. Она содержит бензольные (C_6H_6), пиридиновые ($\text{C}_5\text{H}_5\text{N}$), азотные ($\text{C}_4\text{H}_4\text{N}_2$) и дру-

*В зарубежной литературе эти лазеры называют Dye-Lasers.

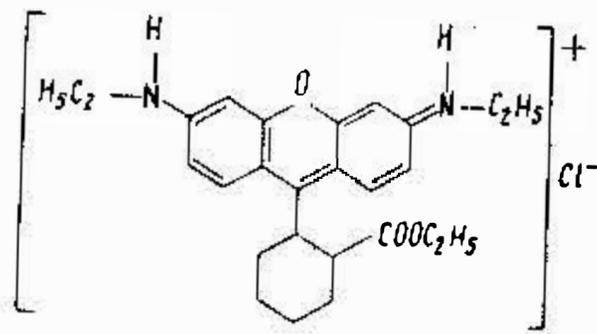


Рис. 8.15. Структурная формула органического красителя родамин 6G

гие кольца. В качестве примера на рис. 8.15 показана структура наиболее распространенного в лазерной технике красителя — родамина 6G. Такая макромолекула обладает богатым набором разрешенных значений энергий электронных, колебательных и вращательных состояний. Напомним (см. § 1.2), что каждому элект-

ронному состоянию соответствует набор колебательных значений энергии, а каждому электронно-колебательному состоянию — набор вращательных уровней. Энергетические расстояния между этими состояниями имеют порядок (1...3) эВ, (0,1...0,01) эВ и 10^{-3} ... 10^{-4} эВ соответственно. Поскольку число атомов в молекуле превышает несколько десятков, то число колебательных степеней свободы может превышать 10^2 . Нетрудно догадаться, что выделить индивидуально каждый из уровней молекулы не представляется возможным. Впрочем, в этом нет необходимости: колебательные и вращательные состояния перекрываются друг с другом, образуя серии разрешенных энергетических полос, соответствующих определенным электронным состояниям. Эти состояния можно разбить на две группы: синглетные (S) и триплетные (T) состояния. К первой группе относятся состояния с антипараллельной ориентацией спинов ($S=0$), а ко второй — с параллельной ($S=1$) ориентацией. Схема уровней органического красителя условно изображена на рис. 8.16. Каждое электронное состояние сопровождается серией колебательных уровней (выделены жирными линиями) и серией вращательных уровней. Согласно правилам отбора по спину оптические переходы разрешены между состояниями с одинаковой мультиплетностью ($\Delta S=0$), т. е. переходы S-S (синглет-синглетные)

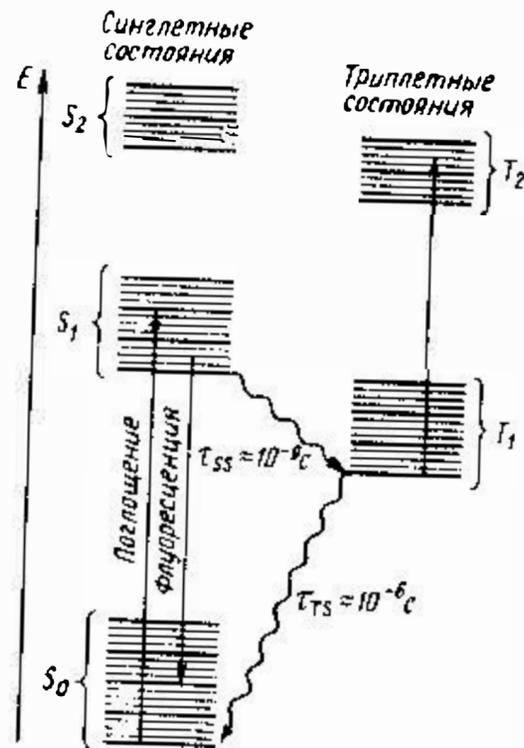


Рис. 8.16 Схема энергетических уровней органического красителя в растворе.

Сплошными линиями изображены оптические (излучательные), а волнистыми — безызлучательные переходы. Колебательные уровни энергии в синглетных S_0 , S_1 и S_2 и триплетных T_1 и T_2 -электронных состояниях выделены жирными линиями

и T-T (триплет-триплетные). Они имеют наибольшую вероятность. Синглет-триплетные переходы S-T и T-S запрещены, и их вероятность примерно на три порядка меньше.

При нормальных условиях молекулы находятся в основном состоянии S_0 , распределяясь по колебательно-вращательным подуровням в пределах kT . В результате поглощения оптического излучения молекула переходит из основного состояния S_0 на один из колебательно-вращательных уровней состояния S_1 . Спектр поглощения, определяемый такими переходами, будет представлять собой широкую бесструктурную полосу (кривая 1 на рис. 8.17). Сечение поглощения в максимуме очень велико и составляет примерно 10^{-16} см². Спектральное положение максимума полосы поглощения определяет цвет красителя и для разных веществ изменяется примерно от 0,3 до 1 мкм. Ширина полосы поглощения также различна для разных красителей и примерно равна 0,2 эВ.

Попадая в результате оптического перехода $S_0 \rightarrow S_1$ (оптическая накачка) в одно из возбужденных состояний полосы S_1 , молекула в результате релаксационных безызлучательных процессов по колебательно-вращательным подуровням внутри состояния S_1 переходит на нижние уровни группы S_1 . Этот процесс термализации происходит очень быстро, за время порядка 1 пс и аналогичен процессам внутрizonной релаксации неравновесных носителей заряда в полупроводниках. Все оптические переходы осуществляются в соответствии с принципом Франка — Кондона по диаграмме рис. 4.41.

Термализованные носители из состояния S_1 излучательно или безызлучательно переходят в основное состояние S_0 . Из нескольких тысяч известных красителей лишь в некоторых из них (около 100) излучательные процессы преобладают над безызлучательными. Именно они и представляют интерес как активные вещества жидкостных лазеров. Излучательное время жизни для переходов $S_1 \rightarrow S_0$ мало (вероятность велика) и составляет примерно 1 нс. Эти переходы определяют люминесценцию (флуоресценцию) красителя. Ее спектр определяется диаграммой (см. рис. 4.41) и представляет собой широкую полосу люминесценции, смещенную в длинноволновую область

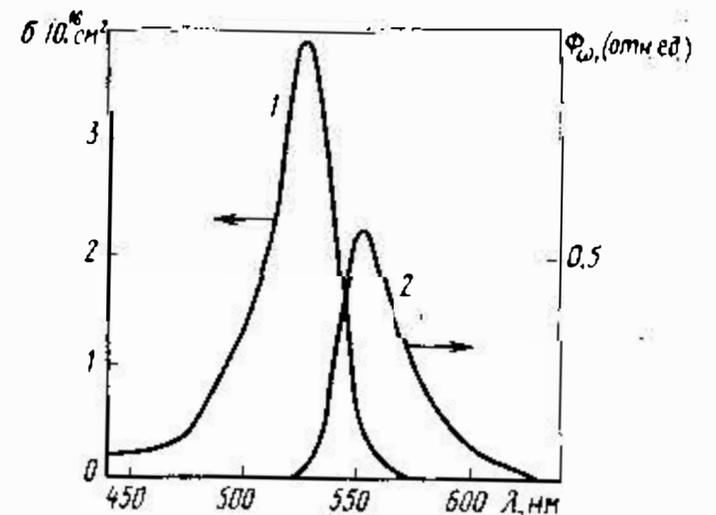


Рис. 8.17. Типичные спектры поглощения (1) и флуоресценции (2) раствора органического красителя родамин 6G в этаноле

относительно спектра поглощения, как показано на рис. 8.17. Это стоксово смещение обсуждалось в § 4.6.

При интенсивной оптической накачке между нижними состояниями полосы S_1 и верхними S_0 может быть достигнута инверсия населенностей. Генерация осуществляется по четырехуровневой схеме между энергетическими состояниями полос S_1 и S_0 аналогично лазеру на центрах окраски. Нижними лазерными уровнями в этой схеме являются высокорасположенные уровни основного состояния S_0 . Эти уровни не заселены термически, а их опустошение осуществляется очень эффективно по колебательно-вращательным подуровням за время порядка 1 пс. Триплетные состояния T_1 и T_2 , изображенные на рис. 8.13, не участвуют в процессе лазерной генерации, а, напротив, препятствуют ей. Действительно, из возбужденных состояний S_1 молекулы могут перейти в состояние T_1 , например, безызлучательным путем. Обычно вероятность этого перехода $S_1 \rightarrow T_1$ всего на порядок меньше, чем вероятность излучательного перехода $S_1 \rightarrow S_0$. Попав в состояние T_1 , молекулы там накапливаются, поскольку переход $T_1 \rightarrow S_0$ является запрещенным (излучательное время жизни для этого перехода имеет порядок 1 мкс). Из состояний T_1 молекулы могут оптически переходить в состояние T_2 , поглощая фотоны соответствующей энергии. К сожалению, в большинстве красителей переходы $T_1 \rightarrow T_2$ по энергии близки к переходам $S_1 \rightarrow S_0$ и, кроме того, они разрешены правилами отбора. Поэтому накопление молекул в состоянии T_1 приводит к появлению дополнительного паразитного поглощения на частотах генерации, что приводит к ее срыву. По этой причине некоторые лазеры на органических красителях могут работать только в импульсном режиме.

Спектр спонтанной люминесценции красителя (флуоресценция), изображенный кривой 1 на рис. 8.17, определяется взаимным расположением термов S_1 и S_0 и требованием выполнения принципа Франка — Кондона для квазиравновеснозаселенных колебательно-вращательных уровней этих термов. При помещении красителя в резонатор и достижении условия самовозбуждения начинается лазерная генерация на частоте, определяемой свойствами как самого красителя, так и резонатора. Если применять обычный оптический резонатор типа плоского интерферометра или конфокального резонатора, то генерация будет осуществляться на многих собственных частотах резонатора, расположенных вблизи максимума спектра флуоресценции красителя. Однако наибольший интерес лазер на органическом красителе представляет как генератор с перестраиваемой длиной волны. Для осуществления этой возможности необходимо применять высокооборотный селективный (называемый также дисперсионным) резонатор, собственную частоту которого можно было бы эффек-

тивно перестраивать. Идеальный вариант представляет одномо-довый одночастотный резонатор. При настройке такого резонатора на заданную частоту в пределах контура линии флуоресценции именно на этой частоте будет происходить преимущественное опустошение вполне определенного уровня из серии уровней S_1 , что опять определяется необходимостью соблюдения принципа Франка — Кондона. Важно, что этот опустошаемый излучательным путем верхний лазерный уровень будет заселяться за счет процессов термализации внутри терма S_1 . В силу высокой скорости этого процесса ($\tau \approx 10^{-12}$ с) в формировании одночастотного излучения будет участвовать практически вся энергия, запасенная термом S_1 .

Аналогичная ситуация будет наблюдаться при опустошении нижнего лазерного уровня в группе уровней S_0 . Энергия, выделяемая при этих релаксационных процессах, переходит в тепло.

Поскольку вероятность оптических переходов $S_0 \rightarrow S_1$ в красителях весьма высока, то показатель поглощения и показатель усиления для этих веществ могут быть очень велики. Они примерно на два порядка превышают показатель усиления малого сигнала для твердотельных лазеров на гранате и рубине. По этой причине лазеры на красителях, обладая высоким коэффициентом усиления, требуют небольшого объема активной среды* с характерными размерами порядка 1 мм. Поглощение интенсивного излучения накачки и последующий нагрев малого объема красителя, вызванный, в частности, релаксационными процессами, приводит к необходимости быстрой непрерывной замены вещества в рабочем объеме. Если этого не делать, происходит во-первых, термическое разложение красителя и, во-вторых, накопление молекул в триплетном состоянии T_1 и срыв генерации. Эти факторы определяют конструктивные особенности лазеров на красителях.

Накачка жидкостных лазеров всегда осуществляется оптическим путем. В принципе может быть применена та же схема, которая применяется для накачки твердотельных лазеров, с использованием специальных стержневых ламп и трубки, по которой прокачивается краситель. Однако эффективность и срок службы такого лазера невелики. В процессе эксплуатации краситель быстро разлагается и «прятается» к окнам кюветы. Чтобы избежать этого эффекта не допустить накопления молекул в состоянии T_1 , необходима прокачка активной жидкости с такой скоростью, чтобы за время нескольких микросекунд произвести ее полную смену в активной области. Наилучшим решением этой задачи оказалось использование струи раствора красителя, которая из специально сконструированного сопла выпускается

*По этому параметру органические красители уступают только полупроводникам, где показатель усиления малого сигнала может быть еще на два порядка выше, достигая 10^4 см⁻¹.

в воздух, где образует ровный плоскопараллельный ламинарный слой.

Так как основной целью лазеров на красителях является перестройка длины волны генерации, то для их накачки наиболее часто применяют лазерное излучение. В зависимости от типа красителя для накачки могут быть использованы газовые лазеры на азоте, аргоновый лазер, эксимерные лазеры, вторые гармоники неодимового и рубинового лазеров и т. п. Важно, чтобы лазерное излучение накачки попадало в область активного поглощения красителя. Пучок излучения лазера накачки фокусируется обычной или цилиндрической линзой в активный объем струи красителя, где почти полностью поглощается. Толщина струи при этом составляет всего 1 мм.

Для перестройки частоты лазерного излучения внутри резонатора вводят специальные частотно-селективные элементы типа призм, специальных клиновых фильтров и т. п. При большом коэффициенте усиления красителя роль диспергирующего элемента и одновременно «глухого» зеркала может выполнять дифракционная решетка (см. рис. 2.24, б).

В лазерах на красителях может быть использована как поперечная, так и продольная схема накачки. В качестве примера на рис. 8.18 показана оптическая схема жидкостного лазера на органическом красителе с продольной накачкой и использованием в качестве диспергирующих элементов призмы и эталона Фабри — Перо. Накачка осуществляется лазерным излучением 1. Резонатор образован зеркалами 2, 4, 6 и 8. Призма 3 выполняет

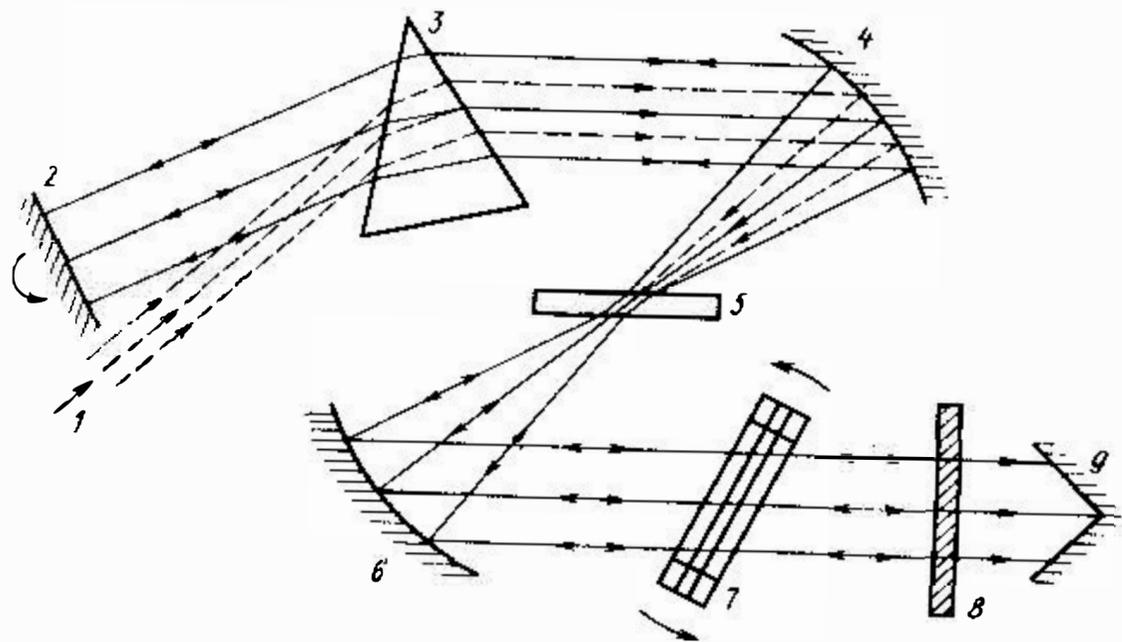


Рис. 8.18. Оптическая схема лазера на органическом красителе с продольной накачкой:
1 — излучение накачки; 2, 4 и 6 — «глухие» зеркала; 3 — призма; 5 — активный элемент; 7 — эталон Фабри — Перо (плоский интерферометр); 8 — полупрозрачное зеркало; 9 — выходное лазерное излучение

две функции: 1) перестройку длины волны лазера и 2) пространственное разделение пучка накачки и пучка лазера на красителе. Излучение накачки фокусируется на активный элемент 5, представляющий собой струю красителя, вытекающего из плоского сопла со скоростью более 10 м/с. На рис. 8.18 показано сечение этой плоской струи, вытекающей перпендикулярно плоскости чертежа. Плоскость струи расположена под углом Брюстера к оси резонатора для уменьшения потерь на отражение. Для тонкой подстройки длины волны выходного излучения и дополнительной селекции типов колебаний внутри резонатора введен дополнительный плоский интерферометр 7 (интерферометр Фабри — Перо) с малой базой. Грубое изменение длины волны генерации осуществляется поворотом зеркала 2, а точное — изменением базы (расстояния между пластинами) интерферометра 7 или его поворотом.

Применяя набор различных красителей, жидкостные лазеры покрывают диапазон длин волн от 0,34 до 1,17 мкм. В качестве растворителей применяют этиловый спирт, дистиллированную воду, этиленгликоль и др. Концентрация красителя обычно не превышает одного молярного процента во избежание образования димеров и концентрационного тушения. На рис. 8.19 приведены спектральные зависимости мощности непрерывного излучения струйного лазера при использовании различных органических красителей при возбуждении аргоновым лазером с мощностью излучения 4 Вт. Диапазон перестройки для каждого из красителей равен 40...80 нм. Ширина спектра лазерного излучения имеет порядок 0,001 нм при применении двух-трех интерферометров Фабри — Перо внутри резонатора по схеме рис. 8.18 и возрастает до 0,1 нм при применении только дифракционной решетки в качестве диспергирующего элемента и до 10 нм при отсутствии диспергирующих элементов внутри резонатора. КПД современных лазеров на органических красителях достигает 25...30% при накачке лазерным излучением и 1% при накачке специальными импульсными лампами. Лазеры могут работать в непрерывном режиме с выходной мощностью несколько Вт, а также в импульсных режимах с импульсной мощностью в несколько МВт при длительности им-

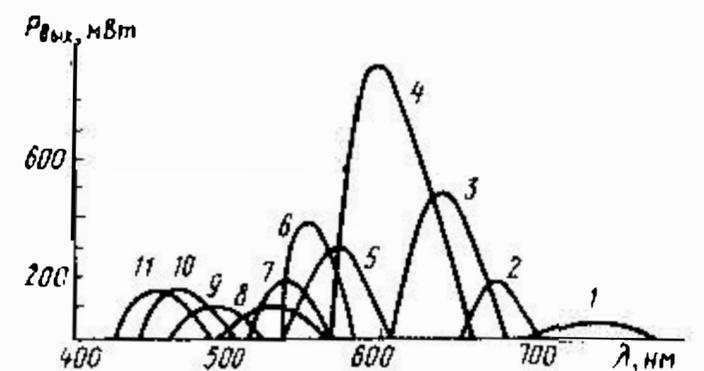


Рис. 8.19. Спектральные зависимости мощности непрерывного излучения жидкостных лазеров на органических красителях при возбуждении Ag^+ -лазером ($P=4$ Вт) Красители: 1 — нальский голубой; 2 — крезил-фиолетперхлорат; 3 — родамин В; 4 — родамин 6G; 5 — родамин 110; 6 — флуоресцина; 7 — кумарин В; 8 — кумарин 7; 9 — кумарин 102; 10 — 7-диэтил-амино-4-метилкумарин; 11 — кумарин 2

пульса порядка 20 нс и частоте повторения до 200 Гц. Расходимость лазерного пучка составляет 2...5 мрад.

К сожалению, большинство красителей токсичны и при работе с ними надо соблюдать осторожность.

Вследствие большой ширины спектра люминесценции органических красителей лазеры на их основе успешно работают в режиме синхронизации мод, генерируют световые импульсы длительностью до 0,03 пс ($3 \cdot 10^{-14}$ с) при непрерывной накачке. Это самые короткие световые импульсы, которые удалось получить с помощью лазера.

Отметим в заключение, что в качестве активной среды возможно применение твердых растворов красителей в полимерах типа полистирола, а также пропитанной раствором красителя мелкопористой губчатой стеклянной матрицы.

ГЛАВА 9 ПОЛУПРОВОДНИКОВЫЕ ЛАЗЕРЫ И СВЕТОДИОДЫ

Отличительной особенностью полупроводников, выделяющей их в отдельный класс материалов, является возможность управляемо изменять (инвертировать) тип их электропроводности. При этом диапазон изменения удельного сопротивления может достигать двадцати и более порядков. Именно эта особенность привела к созданию *p-n*-перехода и развитию полупроводниковой электроники и микроэлектроники. Использование рассмотренных в § 4.6 процессов излучательной рекомбинации в полупроводниках при инжекции неосновных носителей заряда через *p-n*-переход позволило создать новые классы приборов — светодиоды и полупроводниковые инжекционные лазеры. Эти приборы наряду с фотодиодами являются теми элементами, на которых базируется современная оптоэлектроника. Области их применения весьма широки — от простейших световых индикаторов до волоконно-оптических линий связи сверхвысокой емкости и лазерных систем обработки информации. Их тиражи превышают миллионы при номенклатуре в несколько сотен модификаций. Обладая традиционными преимуществами полупроводниковых приборов: малыми габаритами, мгновенной готовностью к работе, низкими рабочими напряжениями, надежностью, совместимостью с интегральной полупроводниковой технологией, экономичностью и низкой стоимостью, — светодиоды и инжекционные лазеры с высокой эффективностью преобразуют электрическую энергию (сигнал) в световую. Светодиоды преобразуют электрический сигнал в некогерентное, а инжекционные лазеры — в когерентное излучение оптического диапазона.

§ 9.1. ПОЛУПРОВОДНИКОВЫЕ ГЕТЕРОПЕРЕХОДЫ И ИНЖЕКЦИОННАЯ ЭЛЕКТРОЛЮМИНЕСЦЕНЦИЯ

В основе действия полупроводниковых светодиодов и инжекционных лазеров лежит электролюминесценция. Как отмечалось в § 4.6, электролюминесценция — это люминесценция, возбужда-

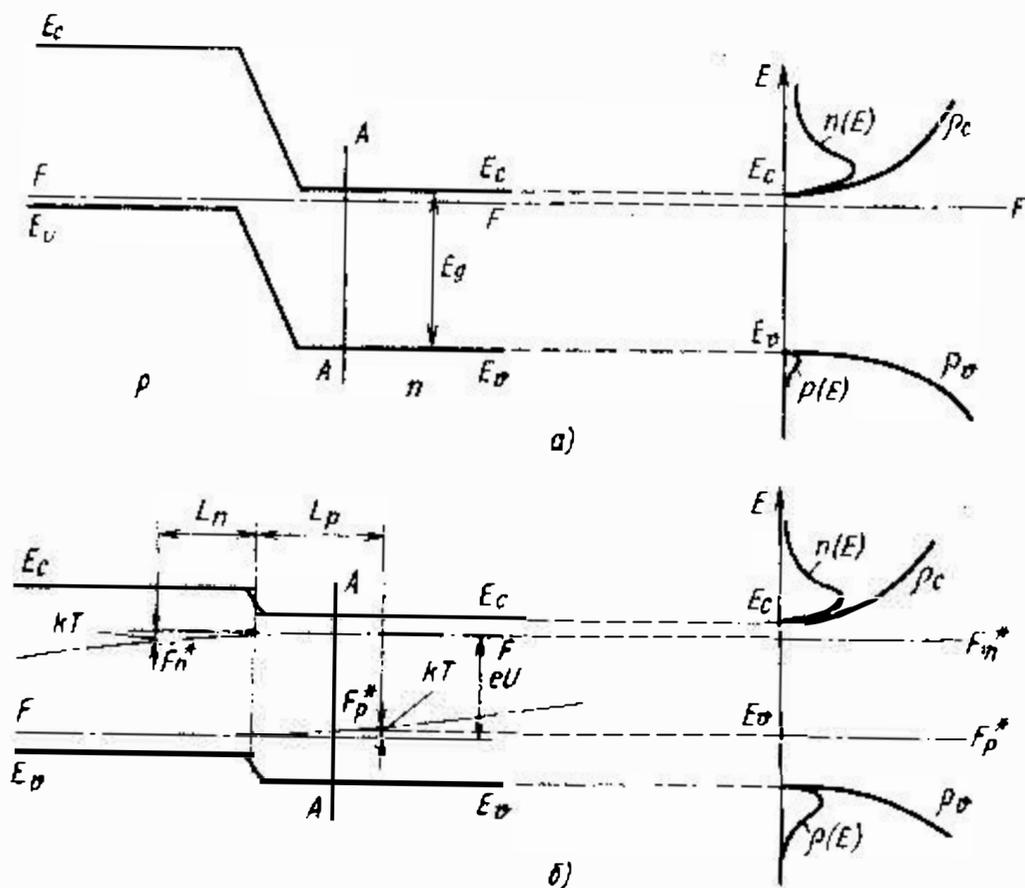


Рис. 9.1. Энергетические диаграммы и распределения носителей заряда по энергиям (по сечению А-А) для p - n -перехода при отсутствии внешнего напряжения (а) и при приложении напряжения U в прямом направлении (б)

емая внешним электрическим полем. Наиболее эффективным методом электрического возбуждения является инжекция неосновных носителей заряда через p - n -переход при приложении к нему напряжения U в прямом направлении. Такая люминесценция называется *инжекционной**. При смещении p - n -перехода в прямом направлении потенциальный барьер снижается на eU , электроны из n -области инжектируются в p -область, а дырки — в n -область. Этот процесс поясняется на диаграммах рис. 9.1.

Инжектированные через p - n -переход неосновные носители заряда диффундируют в глубь материала. За счет процессов рекомбинации их концентрация убывает по мере удаления от области объемного заряда. При постоянной скорости рекомбинации (постоянном времени жизни) концентрация неосновных носителей заряда будет уменьшаться с расстоянием по экспоненциальному закону. Расстояние, на котором их концентрация уменьшится в e раз, равно диффузионной длине. Эту величину можно принять за глубину, на которую проникают инжектированные носители заряда. Таким образом, глубину проникновения электронов в p -область можно считать равной их диффузионной длине

*Напомним, что явление инжекционной электролюминесценции было открыто О. В. Лосевым в кристаллах карбида кремния в 1923 г.

L_n в p -материале, а глубину проникновения дырок в n -область — диффузионной длине L_p в n -материале.

За счет инжекции неосновных носителей заряда через p - n -переход в n - и p -областях будет создано неравновесное распределение носителей. Это распределение при тех же условиях и по тем же причинам, что и ранее (см. § 4.6), может быть охарактеризовано с помощью *квазиуровней Ферми*. Это означает, что уровень Ферми будет расщепляться на два квазиуровня отдельно для электронов F_n^* и дырок F_p^* , как показано на рис. 9.1, б. По мере удаления от области объемного заряда квазиуровни Ферми будут приближаться к равновесному уровню, сливаясь с ним. На расстоянии, равном диффузионной длине L_n или L_p , неравновесный квазиуровень Ферми снижается (для электронов) или повышается (для дырок) приблизительно на kT от максимального значения на границе p - n -перехода.

Распределения носителей заряда по сечению А-А показаны на рис. 9.1. Они по существу не отличаются от соответствующего распределения, приведенного на рис. 4.35.

Энергетическое расстояние между квазиуровнями Ферми ($F_n^* - F_p^*$) вблизи области объемного заряда определяется напряжением, приложенным непосредственно к p - n -переходу (за вычетом омических потерь), и равно eU . Другими словами, расстояние $F_n^* - F_p^*$, как и ранее, определяется уровнем возбуждения. Заметим, что максимальное напряжение, которое может быть приложено к p - n -переходу, ограничено. Для «обычного» p - n -гомоперехода это значение определяется полным спрямлением потенциального барьера. Внешним полем можно почти полностью «убрать» потенциальный барьер, но невозможно сделать его отрицательным. Физически это означает, что в p - n -переходе (точнее, в p - n -гомопереходе) концентрация инжектированных неосновных носителей заряда, как правило, не может превышать концентрацию этих же носителей в эмиттере, где они являются основными. По этой причине для получения инверсии населенностей путем инжекции неосновных носителей заряда через такой p - n -переход [т. е. для удовлетворения условия (4.76)] необходимо, чтобы как минимум одна из областей p - n -перехода была вырожденной. Это проиллюстрировано на рис. 9.2, где приведены диаграммы p - n -перехода в отсутствие смещения (рис. 9.2, а) и при максимально возможном смещении (рис. 9.2, б), соответствующем полному спрямлению энергетического барьера. Распределение носителей заряда по энергиям на рис. 9.2 для простоты показано при температуре $T = 0$ К. «Хвосты» плотности состояний появляются вследствие сильного легирования материала.

По целому ряду причин, которые будут ясны из дальнейшего изложения, характеристики светодиодов и особенно инжекционных лазеров существенно улучшаются при применении не «обыч-

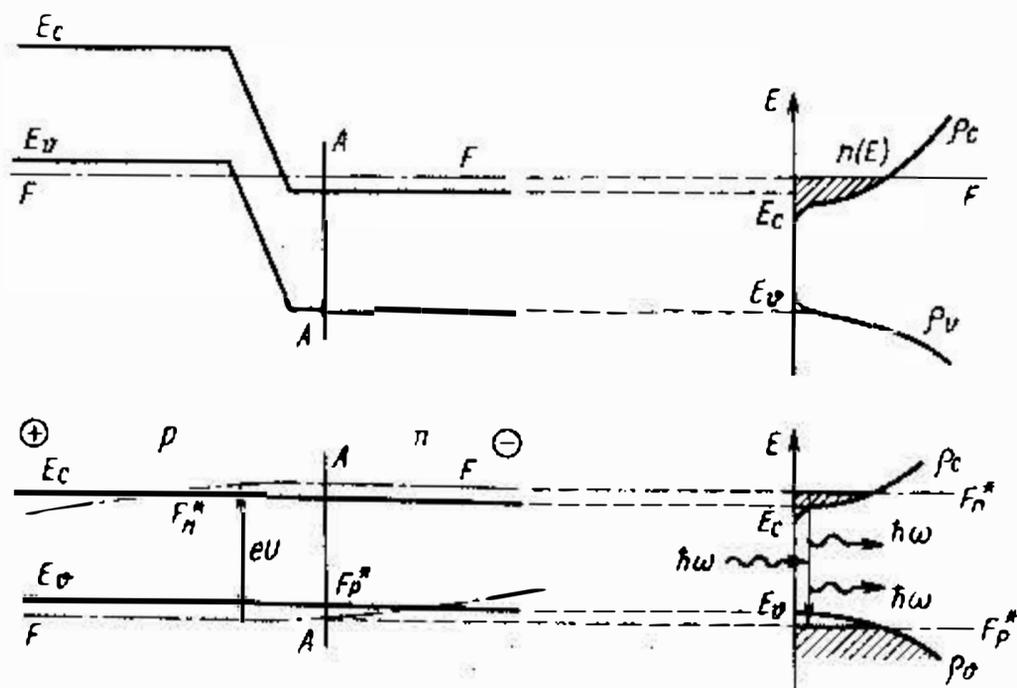


Рис. 9.2. Энергетические диаграммы и распределения носителей заряда по энергиям (для $T=0$ К) в вырожденном p - n -переходе при отсутствии смещения (а) и при максимально возможном смещении в прямом направлении (б)

ных» p - n -переходов, изготовленных на основе гомогенных материалов и называемых иногда гомопереходами, а гетеропереходов. Решающий вклад в создание полупроводниковых гетероструктур и приборов на их основе внесла группа ученых физико-технического института им. А. Ф. Иоффе под руководством Ж. И. Алферова*.

Гетеропереход представляет собой контакт (переход) на атомном уровне двух различных по химическому составу материалов (полупроводников), осуществленный в одном кристалле. Гомопереход, как известно, образуется в полупроводнике за счет искусственно созданного распределения примесей. В гетеропереходах помимо управления концентрацией и типом носителей заряда путем легирования появляется новая возможность управлять такими важными параметрами материала, как ширина запрещенной зоны и показатель преломления. Именно поэтому применение гетеропереходов привело как к созданию принципиально новых полупроводниковых приборов, так и к существенному улучшению параметров ранее созданных устройств.

В зависимости от скорости изменения основных параметров по координате различают *резкие* и *плавные гетеропереходы*. В резком гетеропереходе это изменение происходит на уровне одного или двух-трех молекулярных слоев. В плавном гетеропереходе контактная область простирается на десятки, а иногда и сотни периодов решетки (молекулярных слоев). Если свойства матери-

*Напомним, что за эти работы акад. Ж. И. Алферову была присуждена Нобелевская премия по физике за 2000 г.

ала изменяются плавно в макроскопическом масштабе, то такой полупроводник называют *варизонным полупроводником*, а полупроводниковую структуру — *варизонной структурой*, подчеркивая тем самым плавно изменяющуюся в пространстве ширину запрещенной зоны.

Гетеропереход называется *изотипным*, если он образован полупроводниками с одинаковым типом электропроводности, и *анизотипным*, если образующие его материалы обладают разными типами электропроводности. При обозначении гетероперехода принято указывать тип проводимости полупроводника с узкой запрещенной зоной строчными буквами n или p , а полупроводника с широкой запрещенной зоной — прописными буквами N или P . На первом месте обычно указывают тип электропроводности более узкозонного материала; так, изотипные гетеропереходы обозначают как n - N или p - P , а анизотипные гетеропереходы как n - P или p - N .

В области контакта двух различных по химическому составу материалов происходит изменение не только ширины запрещенной зоны, но и других электрофизических характеристик материала, в том числе диэлектрической проницаемости, электронного сродства, геометрических параметров решетки, эффективной массы носителей заряда и их подвижности. На границе раздела возможно появление различных дефектов, что является нежелательным. Гетеропереход, в котором на границе раздела двух материалов отсутствуют поверхностные состояния, называется *идеальным*. Наиболее сильное влияние на отклонение свойств гетероперехода от идеального оказывает различие в параметрах решеток и в температурных коэффициентах линейного расширения материалов, образующих гетеропереход. По этой причине относительное значение разности параметров решеток двух контактирующих материалов

$$\frac{\Delta a}{a} = \frac{2(a_1 - a_2)}{a_1 + a_2} \quad (9.1)$$

обычно не должно превышать 0,1%. Дополнительным условием является совпадение их кристаллохимической природы.

Наиболее полно перечисленным требованиям удовлетворяют трех- и четырехкомпонентные твердые растворы в группе полупроводниковых соединений $A^{\text{III}}B^{\text{V}}$: GaAs — AlAs ($Al_xGa_{1-x}As$), GaSb — AlSb ($Al_xGa_{1-x}Sb$), GaP — AlP ($Al_xGa_{1-x}P$), Ga — In — As — P ($Ga_xIn_{1-x}P_yAs_{1-y}$), что нетрудно заключить из рис. 4.8. Возможность плавно и в широких пределах изменять основные характеристики твердого раствора путем изменения его состава (см. § 4.2) при минимальном изменении параметров решетки и определила широкое применение этих материалов в квантовой и оптической электронике.

В отличие от идеального в реальном гетеропереходе на границе раздела всегда присутствуют поверхностные состояния. Важно, чтобы их влияние на свойства $p-n$ -перехода было мало. В трехкомпонентных твердых растворах изменение состава приводит одновременно к изменению как периода идентичности решетки, так и ширины запрещенной зоны. В четырехкомпонентных твердых растворах типа $Ga_xIn_{1-x}P_yAs_{1-y}$ изменением x и y можно в определенных пределах независимо изменять параметры решетки и ширину запрещенной зоны. При определенных соотношениях между x и y можно подобрать такой ряд растворов, в котором ширина запрещенной зоны будет изменяться при неизменном периоде идентичности решетки. Такой ряд твердых растворов называют *изопериодным* (§ 4.2). На рис. 4.11 пунктировыми линиями указаны твердые растворы $Ga_xIn_{1-x}P_yAs_{1-y}$, изопериодные фосфиду индия и арсениду галлия. Гетеропереходы, изготовленные на основе изопериодных твердых растворов, по своим характеристикам наиболее близки к идеальным.

Энергетические зонные диаграммы $p-N$ - и $n-P$ -гетеропереходов представлены на рис. 9.3. Эти диаграммы построены в предположении, что объемные свойства полупроводников, образующих гетеропереход, остаются неизменными вплоть до их границы раздела. Вследствие скачка диэлектрической проницаемости и электронного сродства контактирующих материалов в резких гетеропереходах появляются *разрывы* в валентной зоне (ΔE_v) и в зоне проводимости (ΔE_c), причем

$$\Delta E_v + \Delta E_c = E_{g1} - E_{g2}. \quad (9.2)$$

Соотношение между ΔE_v и ΔE_c различно для разных материалов. Для гетеропереходов в системе $GaAs-Al_xGa_{1-x}As$ при $x < 0,4$ установлено, что на долю ΔE_v приходится около 40%, а на долю ΔE_c — около 60% от разности ($E_{g1} - E_{g2}$), т. е.

$$\frac{\Delta E_c}{E_{g2} - E_{g1}} \approx 0,6, \quad (9.3)$$

$$\frac{\Delta E_v}{E_{g2} - E_{g1}} \approx 0,4. \quad (9.3a)$$

Из рис. 9.3 видно, что потенциальные барьеры в гетеропереходах различны для электронов и дырок, что обуславливает особенности электрических, фотоэлектрических и оптических свойств таких структур. К важнейшим эффектам, наблюдаемым в гетероструктурах и определяющим их широкое применение в оптической электронике, относятся: 1) эффект широкозонного окна; 2)

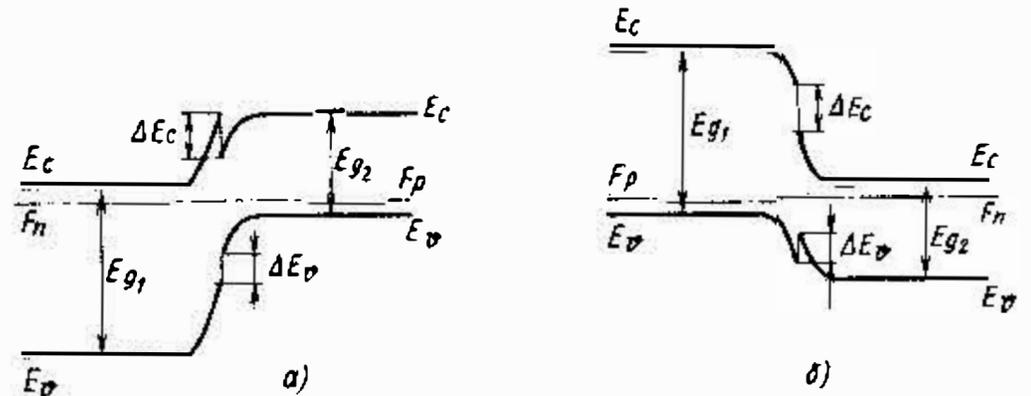


Рис. 9.3. Энергетические диаграммы $p-N$ - (а) и $n-P$ -гетеропереходов (б) в условиях теплового равновесия. ΔE_v и ΔE_c — разрывы в валентной зоне и зоне проводимости

3) эффект сверхинжекции и 4) волноводный эффект.

1. *Эффект широкозонного «окна».* Этот эффект позволяет с минимальными потерями вывести излучение, генерируемое в области $p-n$ -гетероперехода, через его широкозонную часть. Это особенно важно, когда люминесценция определяется межзонными или квазимсжзонными переходами с участием мелких примесей. В «обычном» $p-n$ -переходе излучение с энергией фотонов $\hbar\omega \approx E_g$ поглощается в прилегающих к активному слою p - и n -областях (происходит «самопоглощение»). В гетеропереходе излучение с энергией фотонов $\hbar\omega \approx E_{g2}$ свободно проходит через широкозонную область, где $E_{g1} > E_{g2}$.

2. *Эффект односторонней инжекции* (рис. 9.4). В гетеропереходах в отличие от гомопереходов будет осуществляться преимущественная инжекция носителей зарядов из широкозонной части в узкозонную. Инжекции из узкозонной в широкозонную область препятствует наличие дополнительного энергетического барьера. Так, для $n-P$ -гетероперехода барьер в зоне проводимости препятствует инжекции электронов в широкозонную область, а для $p-N$ -гетероперехода инжекции дырок в n -область препятствует барьер в валентной зоне. Токи инжектированных через гетеропе-

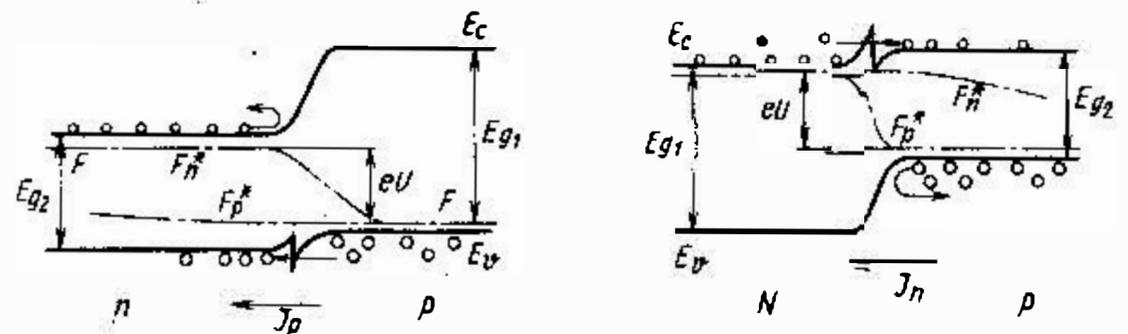


Рис. 9.4. Эффект односторонней инжекции в $n-P$ - и $p-N$ -гетеропереходах

реход электронов и дырок в первом приближении отличаются на фактор $\exp\left(\frac{\Delta E_c + \Delta E_v}{kT}\right) \approx \exp\left(\frac{E_{g1} - E_{g2}}{kT}\right)$. При комнатной температуре $kT = 0,025$ эВ, а $\Delta E_c + \Delta E_v$ может составлять несколько десятых электронвольта, так что отношение электронного и дырочного токов может достигать нескольких тысяч. Эффект односторонней инжекции позволяет осуществлять преимущественную инжекцию неосновных носителей заряда в ту область p - n -перехода, где максимален квантовый выход люминесценции.

3. **Эффект «сверхинжекции»** (рис. 9.5). В гомопереходе концентрация неосновных носителей заряда, инжектированных из эмиттера, не может превышать их концентрации в эмиттере (см. рис. 9.1 и 9.2). В гетеропереходе за счет наличия разрывов в зоне проводимости ΔE_c (для p - N -гетероперехода) или в валентной зоне ΔE_v (для n - P -гетероперехода) при достаточно большом напряжении смещения возможно как бы образование «отрицательного» барьера для электронов (дырок). Это проиллюстрировано на рис. 9.5 для p - N -гетероперехода. Узкий положительный потенциальный барьер на гетерогранице электроны свободно проходят за счет туннельного эффекта, попадая затем в потенциальную яму. Благодаря этому концентрация инжектированных носителей заряда может превышать их концентрацию в эмиттере (на рис. 9.5 уровень Ферми в n -области расположен ниже дна зоны проводимости, а квазиуровень Ферми для электронов в p -области — внутри зоны проводимости). Предельно достижимое отношение концентрации инжектированных электронов n_p в p - N -гетеропереходе к их равновесной концентрации в эмиттере n_N (в N -области) равно

$$\frac{n_p}{n_N} \rightarrow \exp \frac{\Delta E_c}{kT}. \quad (9.4)$$

Аналогично в n - P -гетеропереходе

$$\frac{p_n}{p_P} \rightarrow \exp \frac{\Delta E_v}{kT}. \quad (9.5)$$

Эта особенность инжекции в гетеропереходе делает его уникальным по эффективности инжектором, что очень важно для полупроводниковых лазеров.

4. **Волноводный эффект.** Вследствие разницы показателей преломления материалов, составляющих гетеропереход, будет наблюдаться отражение света от гетерограницы. Как правило, показатель преломления узкозонного материала больше, чем широкозонного (см. § 4.5). Поэтому световые лучи, распространяющи-

ся в узкозонной части под малыми углами к гетерогранице, будут испытывать полное внутреннее отражение. Если узкозонная активная область расположена между двумя широкозонными областями, то световое излучение в ней может распространяться так же, как в волноводе (§ 3.4).

Возбуждение люминесценции электрическим полем может быть осуществлено в неоднородных структурах не только путем инжекции неосновных носителей заряда с приложением к гомо- или гетеро- p - n -переходу напряжения в прямом направлении (этот метод наиболее эффективен), но и другими методами, в том числе с помощью процессов туннелирования через слой изолятора или ударной ионизации в сильных полях. Один из возможных процессов, приводящих к возбуждению электролюминесценции за счет эффектов *туннелирования через слой изолятора*, показан на рис. 9.6. Здесь в качестве примера представлена энергетическая диаграмма для структуры металл — диэлектрик — полупроводник n -типа. Слой диэлектрика, который является достаточно тонким, чтобы через него могли туннелировать электроны, разделяет металлический электрод и полупроводник n -типа (аналогичная ситуация может быть и для полупроводника p -типа). Если к металлическому электроду приложить положительное смещение, достаточное для создания условий, изображенных на рис. 9.6, то электроны будут туннелировать из валентной зоны полупроводника через тонкий слой диэлектрика в металл. В результате в полупроводнике появится дырка, которая может вызвать излучательную рекомбинацию, например, с электроном из зоны проводимости. Процессы ударной ионизации в сильных электрических полях могут приводить к возникновению неравновесных неосновных носителей заряда и тем самым обеспечивать возбуждение люминесценции в полупроводниковых структурах. В частности, при лавинном пробое p - n -перехода, смещенном в обратном направлении, происходит генерация неосновных носителей заряда, которые в последующем могут рекомбинировать с основными носителями излучательным

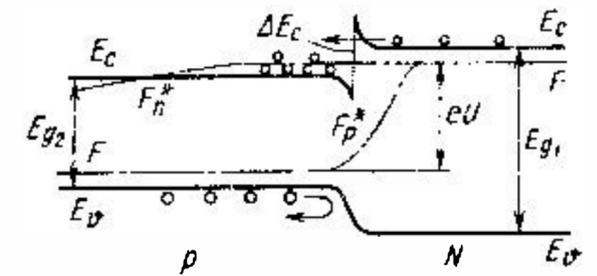


Рис. 9.5. Эффект сверхинжекции в p - N -гетеропереходе

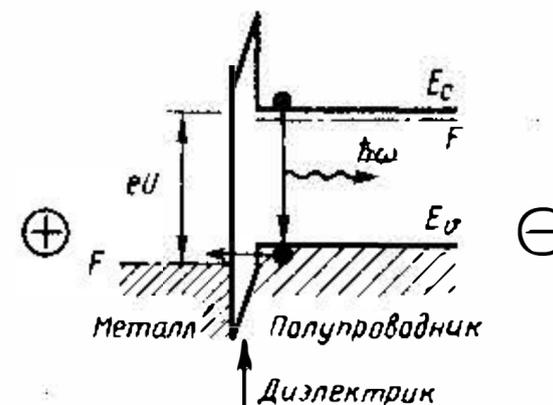


Рис. 9.6. Возбуждение электролюминесценции в структуре металл-диэлектрик-полупроводник n -типа с использованием процессов туннелирования электрона из полупроводника в металл через тонкий слой изолятора

путем. Такая люминесценция называется *предпробойной*. Она может возникать не только в *p-n*-переходах, но также на границах зерен и других неоднородностях, где возможно локальное повышение напряженности электрического поля и развитие процесса локального микропробоя.

§ 9.2. СВЕТОДИОДЫ

Светодиоды — полупроводниковые источники некогерентного оптического излучения, принцип действия которых основан на явлении электролюминесценции при инжекции неосновных носителей заряда через гомо- или гетеро- *p-n*-переход. Как правило, светодиоды работают в спектральном диапазоне 0,4...1,6 мкм. Приборы, излучающие в видимом диапазоне, принято называть *светоизлучающими диодами* — СИД (в этом названии слово «свет» употребляется в узком смысле). В своем большинстве они используются как индикаторы для отображения информации, а также как малоинерционные источники света для генерации световых импульсов малой длительности. Приборы, излучающие в ближней инфракрасной области спектра, принято называть *ИК-светодиодами*. Они, как правило, предназначены для работы в качестве источников излучения в различного рода оптоэлектронных устройствах, в системах автоматического контроля, в датчиках, в системах накачки, ИК-подсветки т. п. Более высокая по сравнению с лазерами надежность и стабильность характеристик, а также сравнительно простая конструкция светодиодов делают их особенно подходящими для систем связи на короткие расстояния при невысокой информационной пропускной способности.

Светодиоды работают при пропускании через них тока в прямом направлении.* За счет инжекции электронов в *p*-область, а дырок в *n*-область вблизи *p-n*-перехода создается неравновесное распределение носителей заряда, как это показано на рис. 9.1. В светодиоде важно обеспечить такие условия, чтобы рекомбинация инжектированных неосновных носителей заряда происходила излучательным путем. Рабочее напряжение, которое необходимо приложить к *p-n*-переходу, определяется шириной запрещенной зоны используемого полупроводникового материала и уровнем его легирования. Как нетрудно убедиться из рис. 9.1 и 9.2, $eU \approx E_g$, типичные значения $U_{\text{раб}} \approx 1...4$ В. При этом необходимо учитывать, что при протекании тока часть напряжения падает в базовых областях и на контактах.

*В специальных случаях для генерации коротких световых импульсов в счет предпробойной люминесценции используют обратное включение светодиодов. Их эффективность при этом мала.

Типичные значения рабочего тока светодиодов составляют 0,1...100 мА. Они зависят от площади *p-n*-перехода и ограничены нагревом.

Основные достоинства светодиодов как видимого, так и ИК-диапазонов обусловлены возможностью непосредственного преобразования электрической энергии в световую с высокой эффективностью. Поэтому важными характеристиками светодиодов являются эффективность и спектральный состав излучения.

Эффективность. Эффективность светодиода η представляет собой его КПД и связана с внешним квантовым выходом электролюминесценции η_e соотношением

$$\eta \approx \frac{h\omega}{eU_{\text{раб}}} \eta_e \quad (9.6)$$

где $h\omega$ — энергия фотона, соответствующая максимуму спектра излучения, $U_{\text{раб}}$ — приложенное внешнее напряжение.

Рассмотрим сначала светодиод в виде прямоугольного параллелепипеда с плоским *p-n*-переходом, как показано на рис. 9.7. Спонтанное излучение генерируется в активной области вблизи *p-n*-перехода и исходит изотропно во все направления. Значение внешнего квантового выхода η_e будет определяться внутренним квантовым выходом люминесценции η_i , коэффициентом инжекции η_j и оптической эффективностью вывода света η_o :

$$\eta_e = \eta_i \eta_j \eta_o \quad (9.7)$$

Проанализируем вклад в общую эффективность каждого из входящих в (9.7) сомножителей.

1. *Внутренний квантовый выход* люминесценции η_i определяется соотношением вероятностей излучательной и безызлучательной рекомбинации [см. (4.89)] и зависит от особенностей зонной структуры полупроводника, типа легирующих примесей и их концентрации, а также от степени совершенства материала. Оказалось, что к качеству полупроводниковых материалов, предназначенных для изготовления светодиодов, предъявляются еще более жесткие требования, чем к материалам для «обычных» полупроводниковых приборов типа диодов и транзисторов. В первую очередь такие материалы должны содержать минимум дефектов, в том числе — глубоких центров, на которых происходит эффективная безызлучательная рекомбинация. Важно, чтобы скорость излучательной рекомбинации превышала скорость

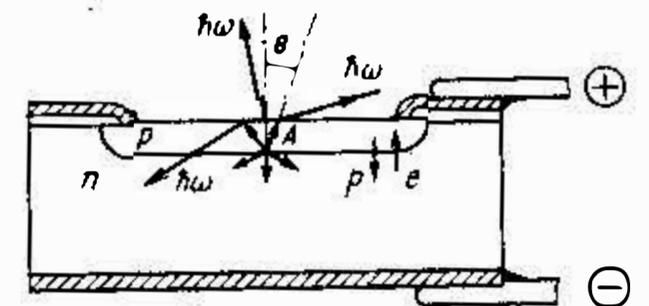


Рис. 9.7. Схематическое изображение плоского светодиода

безызлучательной. Это условие значительно проще выполнить в прямозонных полупроводниках, поскольку вероятности излучательных переходов там существенно выше, чем в непрямозонных материалах (см. § 4.6). По этой причине для изготовления светодиодов предпочтительней использовать полупроводники с прямой структурой энергетических зон. Однако круг материалов, на основе которых могут быть изготовлены светодиоды, весьма ограничен. Классические материалы полупроводниковой электроники — кремний и германий — для этих целей не подходят как по ширине запрещенной зоны, так и по зонной структуре. В полупроводниковых соединениях $A^{III}B^{VI}$, обладающих прямой структурой энергетических зон, в большинстве случаев невозможно инвертировать тип их электропроводности. Если бы в материалах типа ZnS , ZnO , $ZnSe$, CdS и им подобных с помощью надежных и экономичных методов удалось создать гомо- или гетеро- $p-n$ -переходы высокого качества, то они могли бы стать базовыми материалами для светодиодов видимого диапазона (СИД). Пока эти благие намерения не реализованы, основными материалами для производства светодиодов остаются полупроводниковые соединения $A^{III}B^V$.

К сожалению, как видно из рис. 4.3, среди полупроводников $A^{III}B^V$ прямой структурой энергетических зон обладают лишь соединения с $\sum z_i \geq 56$ и $E_g \leq 2$ эВ*. Поэтому для ИК-светодиодов и СИД, излучающих в красной области видимого спектра эти полупроводники по своим свойствам близки к идеальным. Базовым материалом, на основе которого изготавливается большинство таких светодиодов, является арсенид галлия [(GaAs, $E_g = E_g^I = 1,424$ эВ при 300 К)]. На его основе выпускаются наиболее эффективные ИК-светодиоды, работающие в области длин волн 0,86...0,95 мкм. Используется межзонная или квазимежзонная излучательная рекомбинация, а также излучательные переходы на акцепторные уровни (или в примесную зону), созданные амфотерной примесью — кремнием ($E_A = E_{\Sigma} = 0,035$ эВ). Для этих переходов в материале p -типа реализован внутренний квантовый выход излучательной рекомбинации, близкий единице ($\eta_i = 0,95...0,99$) при комнатной температуре.

Светоизлучающие диоды, работающие в видимом диапазоне спектра, должны изготавливаться на основе широкозонных полупроводников с $E_g \geq 1,8$ эВ. Из рис. 4.3 видно, что такие соединения $A^{III}B^V$, за исключением нитридов, обладают не прямой структурой зон. Получение высоких значений η_i в таких материалах является непростой задачей. Вероятность собственной излучательной рекомбинации в непрямозонных полупроводниках мала.

*Как отмечалось в § 4.1, исключения составляют нитриды — GaN и AlN, обладающие прямой структурой энергетических зон с $E_g(\text{GaN}) = 3,6$ эВ и $E_g(\text{AlN}) = 6,2$ эВ. Но для этих материалов существуют те же проблемы создания $p-n$ -переходов, которые, по-видимому, удастся преодолеть.

Поэтому для увеличения η_i здесь используют переходы с участием примесных состояний. Наилучшие результаты по причинам, изложенным в § 4.6, достигнуты при использовании излучательной рекомбинации экситонов, связанных на изоэлектронных ловушках. Базовым материалом здесь служит фосфид галлия (GaP, $E_g = E_g^I = 2,27$ эВ при 300 К) и твердые растворы на его основе $\text{GaAs}_{1-x}\text{P}_x$, $\text{Ga}_x\text{In}_{1-x}\text{P}$, $\text{Ga}_x\text{In}_{1-x}\text{As}_{1-y}\text{P}_y$. Для повышения эффективности люминесценции GaP и $\text{GaAs}_{1-x}\text{P}_x$ ($x \geq 0,4$) легируют изоэлектронной примесью — азотом.

Внутренний квантовый выход люминесценции зависит не только от типа легирующей примеси, но и от ее концентрации и условий роста активного материала. Поскольку к качеству активного слоя, в котором собственно и происходит излучательная рекомбинация, предъявляются очень высокие требования, то для изготовления светодиодов, как правило, используют эпитаксиальные методы выращивания. Наиболее часто применяют газофазовую и жидкофазовую эпитаксии.

Прогресс в технологии выращивания эпитаксиальных структур за последние 25 лет привел к увеличению эффективности светодиодов более чем в 100 раз. Это проиллюстрировано на рис. 9.8, где в качестве характеристики эффективности СИД по вертикальной оси отложена светоотдача, выраженная в люменах на единицу затраченной мощности. Переход к эпитаксиальной технологии позволил улучшить качество материала, увеличить η_i и тем самым увеличить эффективность СИД (первая ступень на рис. 9.8 в середине семидесятых годов). Разработка технологии

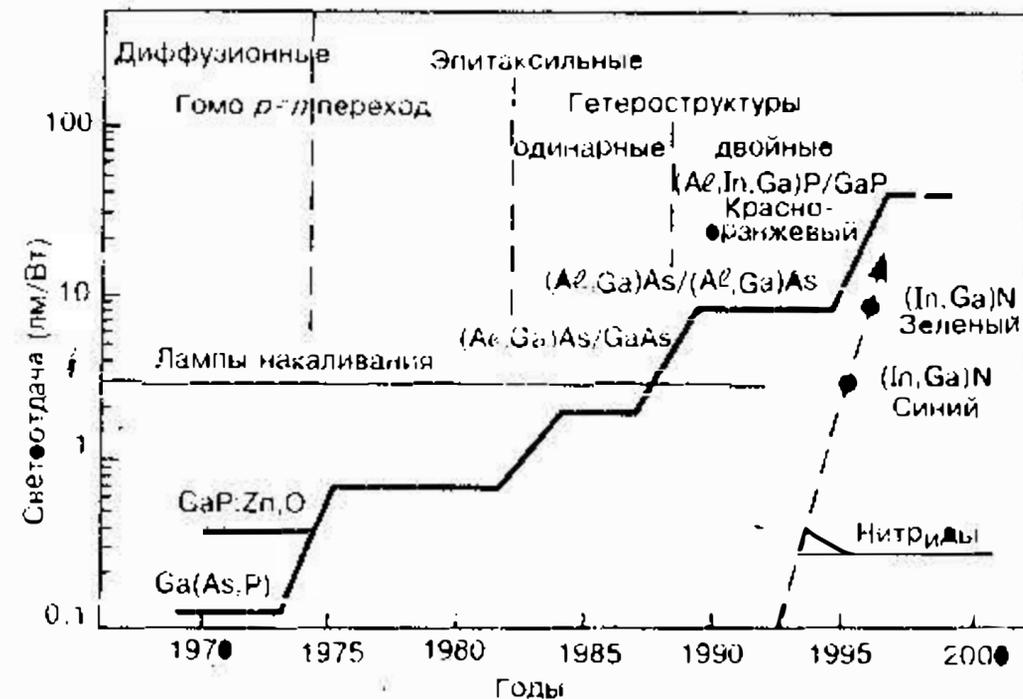


Рис. 9.8. Эволюция эффективности красных светодиодов: Пунктиром изображен возможный прогресс синих и зеленых СИД на основе нитридов.

гетерозпитаксии и применение гетероструктур — сначала оди-
нарных, а затем двойных — позволило еще на порядок увеличить
эффективность СИД, доводя ее до 10 лм/Вт в лучших промыш-
ленных образцах. Это стало возможным благодаря использова-
нию широкозонного твердого раствора $Ga_{1-x}Al_xAs$ с прямой
структурой энергетических зон при $x < 0,36$ (см. § 4.2). В четырех-
компонентных твердых растворах (Ga, In, Al) P можно получить
прямозонные материалы с $E_g \approx 2,2$ эВ.

Применение нитридов в системе InN — GaN — AlN, облада-
ющих прямой структурой энергетических зон (§ 4.1), позволяет
надеяться на быстрый прогресс в достижении высоких значений
 η_i в зеленой, синей и фиолетовой областях спектра и получить
предельно высокую светоотдачу. При этом нет необходимости
проходить все ступени «лестницы» рис. 9.8. Надо сразу приме-
нять эпитаксиальные двойные гетероструктуры.

2. Коэффициент инжекции η_i . Как правило, излучательная
рекомбинация преобладает в одной из областей (p или n), приле-
гающих к p - n -переходу. Поэтому p - n -переход в светодиоде дол-
жен обеспечить преимущественную инжекцию неосновных носи-
телей заряда в ту область, где η_i максимален. Предположим для
определенности, что активной областью с высоким η_i является
 p -область материала. Тогда коэффициент инжекции η_i есть от-
ношение электронного тока I_e к полному току $I = I_e + I_p$, протека-
ющему через p - n -переход в прямом направлении:

$$\eta_i = I_e / (I_e + I_p).$$

Из теории p - n -перехода известно, что электронная и дырочная
составляющие полного тока определяются коэффициентами диф-
фузии электронов и дырок D_n , D_p , диффузионными длинами
неосновных носителей заряда L_p и L_n и уровнем легирования n -
и p -областей — N_D , N_A :

$$\frac{I_e}{I_e + I_p} = \frac{D_n / L_n N_A}{D_n / L_n N_A + D_p / L_p N_D} \quad (9.8)$$

Полагая, что доноры и акцепторы полностью ионизованы, т. е.
в n -области $n \approx N_D$, а в p -области $p \approx N_A$, получаем

$$\eta_i = \frac{nL_p}{nL_p + (D_p/D_n)pL_n} \quad (9.9)$$

откуда видно, что в гомо- p - n -переходе для обеспечения преиму-
щественной инжекции электронов в p -область необходимо иметь
большую L_p и $N_D > N_A$.

Как указывалось в § 9.1, наилучшие условия для односторон-
ней инжекции обеспечиваются в гетеропереходе, где осуществля-

ется преимущественная инжекция из широкозонной области в уз-
козонную. Даже при $L_n \approx L_p$ и $n \approx p$

$$\frac{I_e}{I_p} \approx \exp\left(\frac{\Delta E_g}{kT}\right), \quad (9.10)$$

так что при $\Delta E_g = E_{g1} - E_{g2} > kT$ легко реализуется $\eta_i \approx 1$.

Очевидно, что вблизи границы как гомо-, так и гетеро- p - n -пе-
рехода не должно быть дефектов решетки, механических напря-
жений и нежелательных примесей, которые увеличивают ско-
рость безызлучательной рекомбинации.

3. Коэффициент вывода света η_0 . Он определяется процес-
сами распространения света в активном материале, его отраже-
нием и поглощением на границах раздела, в том числе процес-
сами полного внутреннего отражения. Величина η_0 представляет
собой оптическую эффективность вывода наружу излучения, гене-
рируемого в активной области светодиода. Для оценки η_0 об-
ратимся к рис. 9.7. Свет, генерируемый в точке A вблизи плоско-
сти p - n -перехода, равномерно излучается в телесный угол 4π
стерадиан. Из-за большого значения показателя преломления
полупроводника из диода наружу может выйти только малая
часть излучения, падающего на верхнюю грань в пределах конуса
с критическим углом

$$\theta_{кр} = \arcsin(1/n), \quad (9.11)$$

где n — показатель преломления среды (точнее, n есть относи-
тельный показатель преломления, равный отношению показате-
лей преломления среды, из которой выводится излучение,
и среды, в которую оно вводится). Излучение, распространяюще-
еся вне этого конуса, испытывает полное внутреннее отражение.
Но даже те лучи, которые распространяются в пределах этого
конуса, при прохождении границы раздела частично отражаются.
Для нормального падения коэффициент отражения определяется
соотношением (3.23) при $\alpha = 0$, так что коэффициент пропускания
 T для света, падающего нормально к верхней грани, равен

$$T = (1 - R) = 4n/(1 + n)^2. \quad (9.12)$$

Вблизи критического угла коэффициент отражения растет, а ко-
эффициент пропускания падает, так что среднее значение T по
всему выходному конусу на (3...5)% ниже даваемого соотношени-
ем (9.12).

Таким образом, если пренебречь поглощением света внутри
диода, то доля излучения, которое может быть выведено из
светодиода с плоской геометрией через его верхнюю грань при
первом падении на нее световой волны, составляет

$$\eta_0 = \frac{\sin^2 \theta_{\text{кр}}}{4} T \approx \frac{1}{n(n+1)^2} \quad (9.13)$$

Для GaP ($n=3,4$) имеем $\theta_{\text{кр}}=17^\circ$; $T=0,7$ и $\eta \approx 0,0152$. Это означает, что только 1,5% излучения, генерируемого в точке A такого светодиода, может быть выведено непосредственно на воздух через его выходную грань. Поэтому оптимизация конструкции светодиода с целью повышения η_0 имеет очень важное значение для повышения его полной эффективности.

Коэффициент вывода света η_0 можно повысить, увеличивая критический угол $\theta_{\text{кр}}$ и уменьшая отражение в пределах выходного конуса. На рис. 9.9 показаны поперечные разрезы вариантов светоизлучающих диодов, имеющих полусферическую, усеченную сферическую и эллипсоидальную геометрию. Основное отличие таких структур от структур с плоской геометрией состоит в увеличении телесного угла $\theta_{\text{кр}}$. Их применение приводит к увеличению η_0 почти на порядок, однако значительно усложняет технологию изготовления светодиодов, что приводит к их удорожанию. Естественно, что структуры, изображенные на рис. 9.9, применимы только в том случае, когда полупроводниковый материал (подложка), на основе которого изготовлен СИД (например, эпитаксиальными методами), является прозрачным для излучаемого света. Применение такой геометрии позволяет повысить η_0 более чем в 10 раз, но все равно коэффициент вывода света не превышает 10...30%, т. е. остается низким.

Оптическая эффективность вывода света η_0 ограничена не только трудностью вывода излучения наружу, но и процессами поглощения в материале и на контактах. Прежде чем дойти до выходной поверхности, свет, испущенный вблизи p - n -перехода, проходит через толщу материала, где он может поглотиться. Этот процесс особенно существен в ИК-светодиодах, изготовленных на основе прямозонных полупроводников и использующих

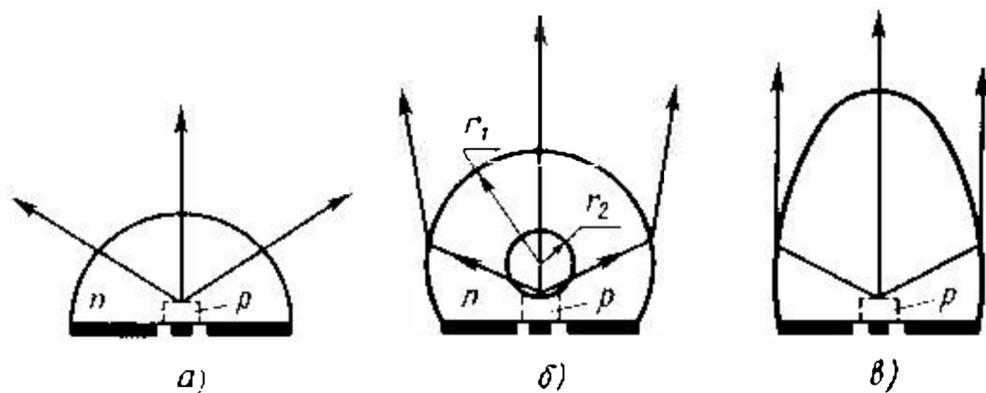


Рис. 9.9. Три конфигурации светодиодов:
а — полусфера; б — усеченная сфера (сфера Вейерштрассе); в — эллипсоид

межзонную или квазимежзонную излучательную рекомбинацию. В этом случае толщина области, через которую выводится излучение в гомо- p - n -переходе, должна быть меньше $1/k_\omega$, т. е. не превышать несколько микрон. Наилучший вывод излучения может быть обеспечен в гетеропереходе через широкозонный эмиттер, используя эффект широкозонного окна (см. § 9.1).

Спектр излучения и яркость. Спектр излучения светодиода определяется шириной запрещенной зоны используемого полупроводникового материала, типом легирующих примесей, уровнем легирования и механизмом излучательной рекомбинации (см. § 4.6). Как указывалось выше, основными материалами для изготовления эффективных светодиодов являются бинарные полупроводниковые соединения $A^{\text{III}}B^{\text{V}}$ и их твердые растворы. На рис. 9.10 в относительных единицах представлены спектры излучения при комнатной температуре некоторых типичных светодиодов, выпускаемых промышленностью.

Наибольшей эффективностью, как говорилось выше, обладают светодиоды на основе арсенида галлия. Из всех прямозонных полупроводников GaAs является технологически наиболее освоенным. Активную p -область создают легированием или классическим для полупроводников $A^{\text{III}}B^{\text{V}}$ акцептором — цинком ($E_A=30,7$ мэВ), или амфотерной примесью — кремнием ($E_A=35$ мэВ). Излучательная рекомбинация определяется переходами электронов из зоны проводимости (или из слившейся с ней примесной зоны, образованной мелкими донорами Te, S или Se с $E_D=6$ мэВ) на акцепторные состояния. Так как энергия ионизации цинка мала, то при концентрации $N_{Zn} \geq 5 \cdot 10^{17} \text{ см}^{-3}$ его примесные уровни сливаются с валентной зоной и излучательная рекомбинация определяется квазимежзонными переходами. Максимум спектра излучения таких светодиодов расположен при энергии, примерно соответствующей ширине запрещенной зоны GaAs, а полуширина спектра примерно равна 2 кТ (см. § 4.6). В таких светодиодах на основе гомопереходов вывод излучения усложняется большим самопоглощением, поскольку $\hbar\omega \approx E_g$, поэтому здесь желательно использовать преимущества гетеропереходов.

В арсенид-галлиевых светодиодах, p -область которых легирована кремнием, спектр излучения сдвинут в длинноволновую область и его максимум расположен при энергии примерно на E_A ниже ширины запрещенной зоны GaAs, что соответствует при комнатной температуре $\lambda \approx 0,95 \text{ мкм}$. GaAs(Si)-светодиоды обладают высокой эффективностью, однако они уступают GaAs(Zn)-светодиодам по быстродействию.

Важно, что спектр излучения GaAs-светодиодов очень хорошо соответствует спектру фоточувствительности наиболее распространенных Si-фотодиодов (см. гл. II).

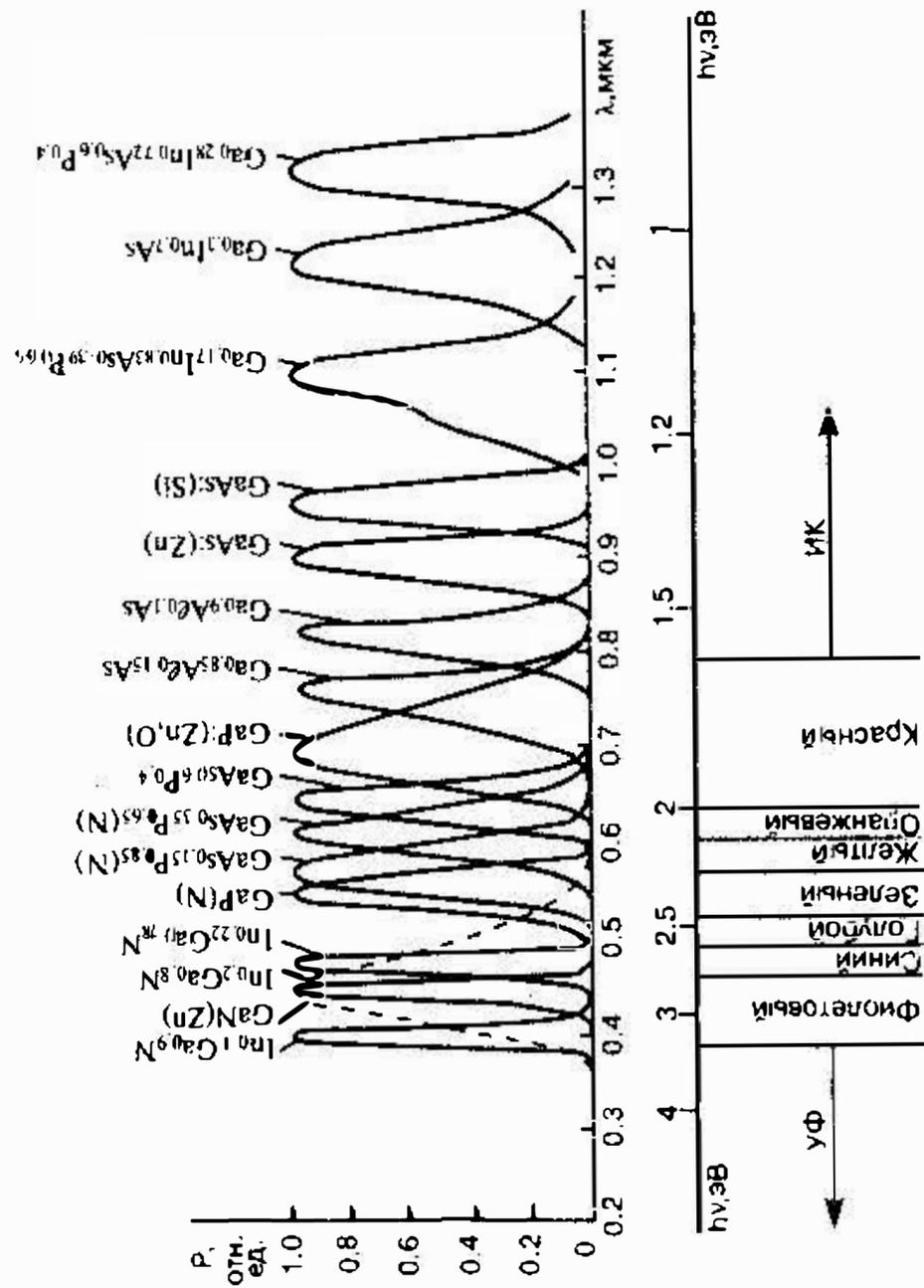


Рис. 9.10. Спектры излучения светодиодов

Светодиоды на более длинноволновую область изготавливаются на основе прямозонных твердых растворов $Ga_xIn_{1-x}As$ и $Ga_xIn_{1-x}As_{1-y}P_y$. Для них преобладающей является квазизонная излучательная рекомбинация. Важно, что максимум спектра излучения таких светодиодов задается составом твердого раствора. Изменяя x и y , можно изготовить светодиод на заданную область спектра, например, совпадающую с минимумом потерь в оптическом волокне или с максимумом спектра поглощения какого-либо вещества, концентрацию которого предстоит контролировать.

Светодиоды на область спектра $\lambda > 5$ мкм могут быть изготовлены на основе халькогенидов свинца: $Pb_xSn_{1-x}Te$, $Pb_xSn_{1-x}Se$ и ртути: $Cd_xHg_{1-x}Te$.

Светоизлучающие диоды на красную область спектра изготавливают из прямозонных твердых растворов $Ga_{1-x}Al_xAs$ и $GaAs_{1-x}P_x$ ($x < 0,4$) (используются квазизонные переходы), а также из непрямозонного GaP, легированного (Zn + O) [используется излучательная рекомбинация экситонов, связанных на изоэлектронном комплексе (Zn — O) (см. рис. 4.39, б)].

Светодиоды, излучающие в оранжевой, желтой и зеленой областях спектра, создаются на основе непрямозонного твердого раствора $GaAs_{1-x}P_x$ ($x > 0,6$), легированного азотом. Как указывалось в § 4.6, азот в этом материале является эффективной изоэлектронной ловушкой и существенно повышает вероятность излучательной рекомбинации. В светодиодах на основе GaP : N кроме основной полосы с максимумом $\lambda = 0,565$ мкм в зеленой области спектра, как правило, присутствует дополнительная красная полоса $\lambda_{\text{длин}} = 0,7$ мкм, обусловленная остаточным кислородом в активном эпитаксиальном слое диода.

Светоизлучающие диоды на коротковолновую область видимого спектра, работающие в голубом, синем и фиолетовом диапазонах, могут быть созданы на основе нитрида галлия GaN и гетеропереходов с использованием твердых растворов $Ga_xIn_{1-x}N$ и $Ga_{1-x}Al_xN$. Успехи в технологии нитридов позволяют надеяться на быстрый прогресс в этом направлении.

Для решения проблемы создания светодиодов коротковолновой области видимого спектра было предложено использовать GaAs — светодиод с нанесенным на его поверхность специальным люминофором, осуществляющим преобразование ИК-излучения в видимый свет. Схема такого диода с преобразователем представлена на рис. 9.11. Преобразование осуществляется за счет последовательного двухступенчатого поглощения фотонов, излучаемых светодиодом, с дальнейшим испусканием одного фотона с уровня E_2 в основное состояние, как показано на рис.

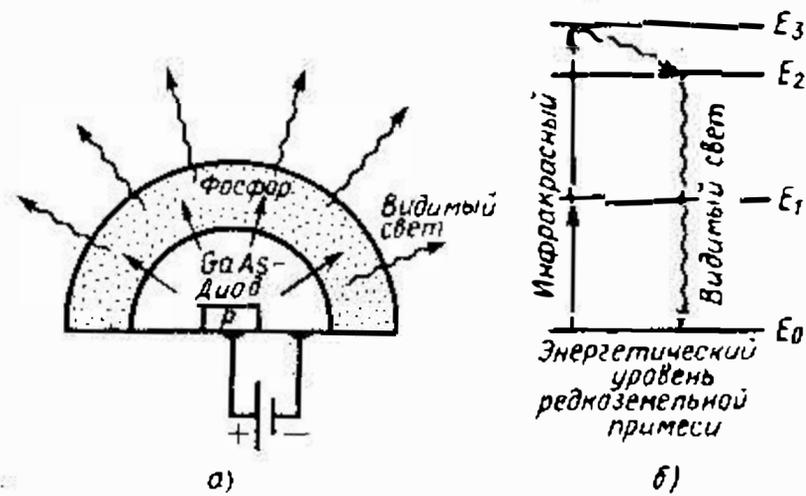


Рис. 9.11. Светодиод с преобразованием излучения:
а) общая схема; б) энергетическая диаграмма

9.11, б. В качестве преобразователя используется диэлектрик, активированный редкоземельными ионами. Необходимо подобрать такие активаторы, чтобы структура их энергетических уровней удовлетворяла требованиям, вытекающим из рис. 9.11, б. Очевидно, что они в значительной степени аналогичны тем требованиям, которые предъявляются к активированным лазерным диэлектрикам (см. § 8.1). В качестве ионов-активаторов применяются иттербий (Yb^{3+}) и эрбий (Er^{3+}). К сожалению, суммарная эффективность такого светодиода с преобразователем пока что остается невысокой.

Для светоизлучающих диодов, используемых в качестве индикаторов, кроме их эффективности важное значение приобретает яркость как оценка эффективности визуального восприятия их излучения. Определение фотометрической яркости V_v светодиода нетрудно провести, зная его спектр излучения и кривую видности

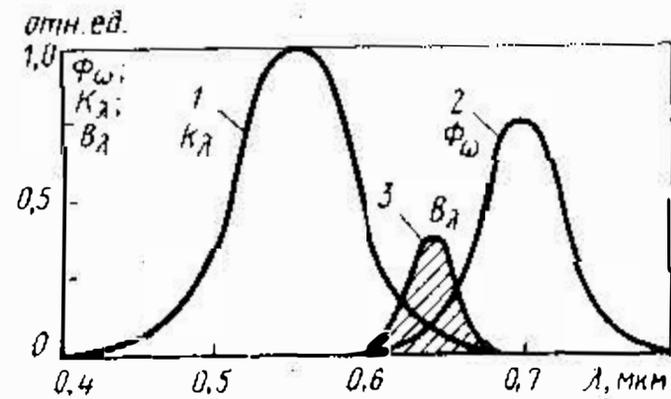


Рис. 9.12. Оценка яркостного эквивалента излучения красного GaP (Zn-O) светодиода:
1 — кривая видности; 2 — спектр излучения;
3 — кривая яркости

глаза (рис. 2.34) по соотношению (2.79). В качестве примера на рис. 9.12 приведены спектр излучения GaP (Zn, O)-светодиода и кривая видности. В то время как максимум излучения диода приходится на $\lambda = 0,69$ мкм, максимум кривой, соответствующей суперпозиции спектра излучения с кривой видности, смещен к $\lambda = 0,65$ мкм, что соответствует яркостному эквиваленту 15 лм/Вт.

Конструкции и характеристики. При кажущейся внешней простоте конструирование светодиодов представляет сложную задачу, поскольку для ее решения надо выполнить ряд противоречивых требований, обсуждавшихся выше. Форма и размеры активного элемента определяются как конкретными целями применения светодиода, так и требованием обеспечения максимального η_0 . Необходимо выполнить и другие требования: обеспечение хорошего теплоотвода, технологичность конструкции, экономические факторы и т. п. Для СИД-индикаторов оптимальные размеры излучающего кристалла близки к $0,3 \times 0,3 \times 0,2$ мм³. Изготовление на таких кристаллах конфигураций, изображенных на рис. 9.9, технологически сложно и дорого. Поэтому в промышленных светодиодах для увеличения η_0 излучающий элемент прямоугольной формы или в виде мезаструктуры заливают оптически прозрачным компаундом с большим показателем преломления (например, на основе эпоксидной смолы) или запрессовывают в специальную пластмассовую линзу. Часто пластмасса окрашивается красителем для повышения контрастности свечения. Для изготовления сигнальных СИД иногда применяют компаунд, диспергированный светорассеивающим наполнителем, позволяющим увеличивать размер светящегося пятна и расширить диаграмму направленности излучения, увеличив тем самым угол обзора.

В качестве примера на рис. 9.13 приведены некоторые типы конструкций светодиодов, выпускаемых отечественной промышленностью. Светодиод типа АЛ102 и аналогичные ему по конструкции (ЗЛ102, ЗЛ341, АЛ112 и др.) в металло-стеклянном корпусе представляют собой герметичную конструкцию с центральным расположением кристалла и гибкими

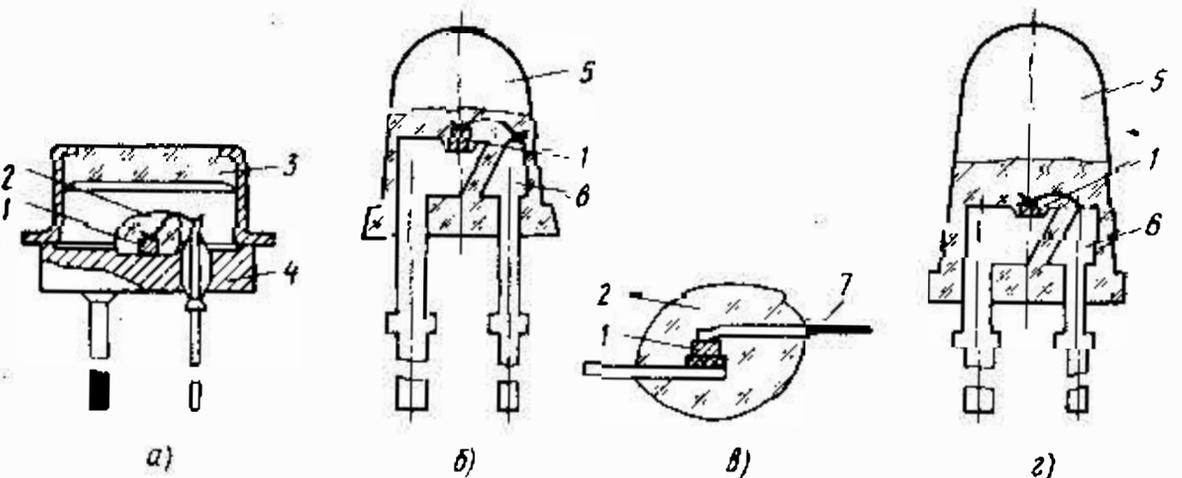


Рис. 9.13. Конструктивные оформления некоторых типов светодиодов:
а) АЛ-102; б) АЛ-307; в) АЛ-301; г) АЛ-336:

1 — светоизлучающий диод; 2 — полимерная защита; 3 — баллон со стеклянным окном; 4 — металлостеклянная вожка; 5 — полимерная линза; 6 — держатель; 7 — гибкий вывод

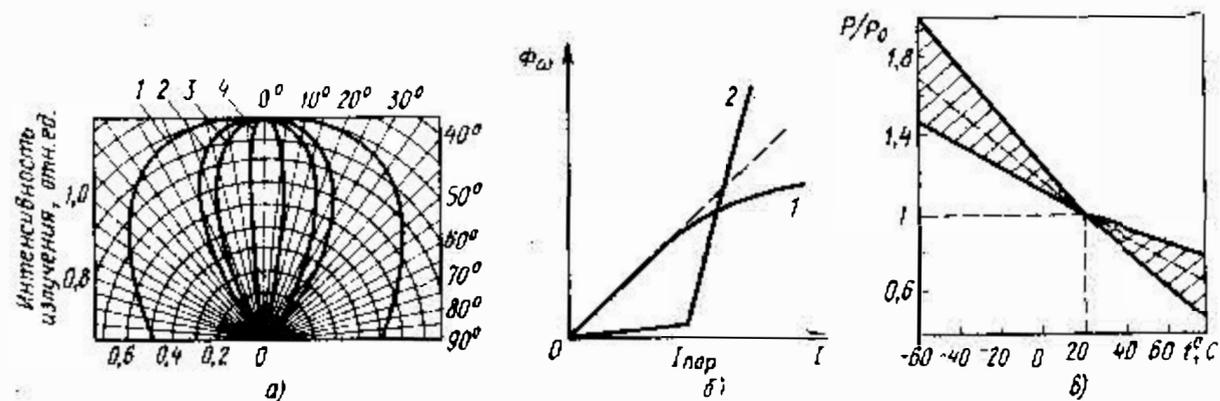


Рис. 9.14. Рабочие характеристики светодиодов: а) диаграммы направленности излучения: 1 — АЛ-301; 2 — АЛ-102; 3 — АЛ-307; 4 — АЛ-336; б) зависимость мощности излучения от тока: 1 — светодиод; 2 — инжекционный лазер; в) температурная зависимость мощности излучения ИК-светодиодов

выводами. Кристалл на ножке защищен прозрачным эпоксидным компаундом.

Светодиод типа АЛ-301 на основе фосфида галлия и карбид-кремниевые светодиоды КЛ101 и 2Л101 являются сигнальными миниатюрными бескорпусными СИД. Он имеет полимерную защиту и гибкие выводы, удобные для монтажа в микросхемах.

Светодиоды типа АЛ307, АЛ310, АЛ316, АЛ336 представляют собой типичные сигнальные СИД с полимерной герметизацией и рамочным держателем, обеспечивающим эффективный теплоотвод от кристалла. В некоторых случаях для увеличения эффективности используют рамочный держатель с встроенным отражателем света. Такая конструкция отличается высокой технологичностью и устойчивостью к механическим и климатическим воздействиям. Типичные рабочие характеристики светодиодов представлены на рис. 9.14.

Конструкции ИК-светодиодов определяются их назначением и могут выполняться в полимерном корпусе с полусферической линзой (АЛ107, АЛ108, АЛ115, АЛ118) аналогично рис. 9.13, а, в бескорпусном варианте с плоской конфигурацией кристалла (АЛ103, 3Л103, АЛ109), в металлоглазном корпусе (АЛ106, АЛ119) и на основе гетеропереходов в системе $Al_xGa_{1-x}As - GaAs$ (АЛ402). Светодиоды, предназначенные для работы в волоконно-оптических линиях связи, должны обеспечивать хорошее согласование излучателя с волокном. Для этих целей служат конструкции, изображенные на рис. 9.15.

Спектр излучения светодиода определяется материалом, на основе которого он изготовлен. В $GaAs$ -, $GaAs_{1-x}P_x$ -, $Al_xGa_{1-x}As$ и $Ga_xIn_{1-x}As_{1-y}P_y$ -светодиодах λ_{max} , соответствующая максимуму спектра излучения, растет с увеличением температуры приблизительно по линейному закону с $d\lambda_{max}/dT \approx 0,3$ нм/град в соответствии с изменением ширины запрещенной зоны этих материалов. Мощность излучения, напротив, падает примерно в 2 —

раза при изменении температуры от -60 до $+70$ °С (исключение составляют SiC-светодиоды). Диаграмма направленности излучения определяется особенностями конструкции прибора.

Ватт-амперные (называемые иногда люкс-амперными и ток-яркостными) характеристики представляют зависимость излучаемой светодiodом мощности от плотности прямого тока. Для светодиодов, изготовленных из прямозонных полупроводников типа GaAs с преобладающей межзонной излучательной рекомбинацией, они линейны вплоть до $j_{пр} \approx 10^3$ А/см². Отклонение от линейности при больших значениях средней подаваемой мощности связано с перегревом кристалла. Для светодиодов на основе непрямозонных материалов типа GaP линейность ватт-амперной характеристики может нарушаться при $j > 10$ А/см². При малых $j_{пр}$ порядка 0,1 А/см² наблюдается резкое уменьшение $P_{изл}$.

При использовании светодиодов в системах передачи информации важной характеристикой является их **быстродействие**. Постоянная времени определяется как скоростью (вероятностью) излучательных переходов, так и электрическими характеристиками диода, в том числе постоянной RC-цепочки. Предельная частота работы GaAs-светодиодов зависит от уровня легирования активного слоя, как показано на рис. 9.16. Аналогичная ситуация наблюдается и в других ИК-светодиодах. Эта зависимость обусловлена уменьшением $\tau_{изл}$ при межзонной (или квазимежзонной) излучательной рекомбинации, вызванным увеличением n или p (см. § 4.6).

Характеристики светодиодов могут изменяться во времени, что вызвано их **деградацией**. Как правило, деградация светодиодов проявляется в постепенном уменьшении мощности излучения при длительном пропускании через них прямого тока. Она может быть вызвана несколькими причинами: а) миграцией

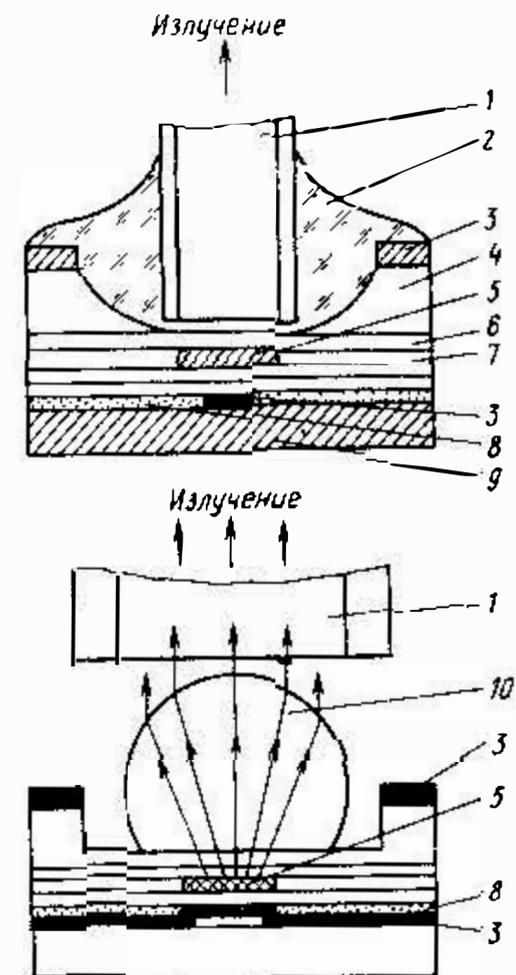


Рис. 9.15. Светодиоды для волоконно-оптических линий связи, состыкованные с оптическим волокном: 1 — оптическое волокно; 2 — эпоксидный компаунд; 3 — омический контакт; 4 — подложка $n-GaAs$; 5 — область рекомбинации; 6 — N -эмиттер $Al_xGa_{1-x}As$; 7 — активный слой $p-Al_xGa_{1-y}As$; 8 — диэлектрик; 9 — теплоотвод; 10 — микролинза

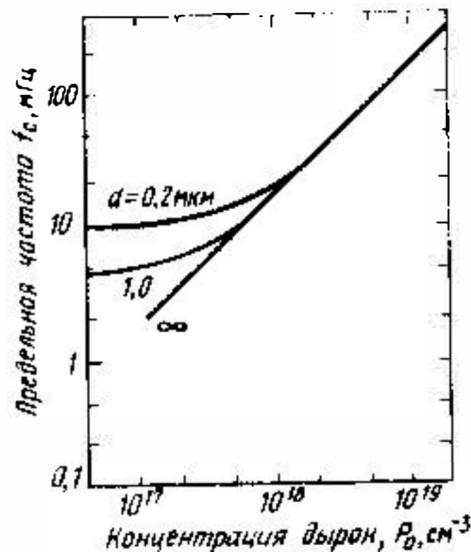


Рис. 9.16. Зависимость предельной рабочей частоты GaAs-светодиодов от концентрации дырок в активном слое

в электрическом поле неконтролируемых примесей (например, Cu, Au, Na) и связанным с этим увеличением концентрации центров безызлучательной рекомбинации в активной области; б) переходом центров излучательной рекомбинации в центры безызлучательной рекомбинации, например, путем их перехода из узлов в междоузлия; в) внутренними напряжениями вблизи *p-n*-перехода; г) поверхностными эффектами, связанными с химическими реакциями на поверхности. Влияние всех этих факторов может быть сведено до минимума за счет совершенствования технологии, так что во многих промышленных светодиодах срок службы превышает $10^5 \dots 10^6$ часов.

При наличии деградации было установлено, что спустя некоторый период «приработки», называемый *периодом катастрофической деградации* (он может отсутствовать), изменение мощности подчиняется экспоненциальному закону

$$P(t) = P(0) \exp(-t/\tau_{дет}). \quad (9.14)$$

Время деградации

$$\tau_{дет} \approx A I_{пр}^{-m} \exp(-E_a/kT) \quad (9.15)$$

зависит от конструкции прибора, его технологии, используемых материалов и т. п., так что константы *A* и *m* различны не только для разных типов светодиодов, но могут различаться для разных партий. Обычно $m \approx 1 \dots 2$. С ростом $I_{пр}$ $\tau_{дет}$ уменьшается. Активационная энергия деградации E_a такова, что при возрастании температуры на $10 \dots 30$ °C от комнатной срок службы снижается вдвое. Для большинства светодиодов $\tau_{дет} \approx 10^4$ ч.

§ 9.3. ОБЩАЯ ХАРАКТЕРИСТИКА И ОСОБЕННОСТИ ПОЛУПРОВОДНИКОВЫХ ЛАЗЕРОВ

Во всех лазерах и мазерах, рассмотренных ранее в гл. 6 — 8, рабочие переходы осуществлялись между индивидуальными энергетическими уровнями, принадлежащими определенным активным частицам. Отличительной особенностью полупроводниковых лазеров является то, что в создании инверсной населенности в них участвуют состояния, определенные по всему кристаллу

и описываемые блоховскими волновыми функциями вида (4.3). Поскольку свободный электрон в зоне проводимости и дырку в валентной зоне нельзя «привязать» к какому-то определенному узлу решетки как активному центру, то обычное условие инверсии в полупроводниках трансформируется в условие (4.76). В этом отношении к полупроводниковым лазерам наиболее близки жидкостные лазеры на органических красителях и лазеры на центрах окраски, где связь между отдельными активными атомами очень эффективно осуществляется за счет интенсивного взаимодействия с колебаниями кристаллической решетки.

Рассмотрим для простоты собственный полупроводник с прямой структурой энергетических зон, законы дисперсии в которых параболичны и определяются соотношениями (4.7), (4.9). Спектральная зависимость показателя поглощения для межзонных переходов без учета экситонных эффектов определена в § 4.4 как

$$k_{\omega}^0 = A (\hbar\omega - E_g)^{1/2}, \quad (9.16)$$

где коэффициент пропорциональности *A* зависит от свойств конкретного материала и для полупроводников $A^{III}V^V$ в первом приближении определяется соотношением (4.43). Напомним, что при выводе спектральной зависимости (9.16) мы предполагали, что полупроводник находится в состоянии термодинамического равновесия и все начальные состояния *j* в валентной зоне заняты электронами, а конечные состояния *f* в зоне проводимости свободны (на это указывает верхний индекс 0 у k_{ω}). Очевидно, что теперь нас будет интересовать другой случай: для выполнения условия (4.76) и получения усиления необходимо путем внешнего воздействия (накачки) создать достаточно высокие концентрации неравновесных электронов в зоне проводимости и дырок в валентной зоне.

Напомним, что усиление есть отрицательное поглощение и вычисление спектральной зависимости показателя усиления $\alpha(\hbar\omega) = -k_{\omega}$ сводится просто к вычислению спектральной зависимости показателя поглощения с учетом заселенности начальных и конечных состояний, а также с учетом вынужденного испускания. По причинам, изложенным в § 4.6, в каждой из зон за время порядка времени внутризонной релаксации ($10^{-12} - 10^{-13}$ с) устанавливаются квазиравновесные распределения свободных носителей заряда, характеризующиеся функциями (4.72). Если все начальные состояния заняты, а конечные свободны, то вероятность поглощения фотона $\hbar\omega$ в единицу времени равна $k_{\omega}^0 v$, где *v* — групповая скорость потока фотонов в веществе. Если вероятность заселения нижнего состояния дыркой равна f_p , а верхнего — электроном — f_n , то вероятность поглощения фотона будет равна

$$k_{\omega}^0 v (1 - f_p) (1 - f_n).$$

Аналогичным образом, вероятность вынужденного испускания такого же фотона есть

$$k_{\omega}^0 v f_p f_n.$$

Истинная скорость поглощения определяется разностью этих двух процессов, откуда получаем выражение для показателя поглощения:

$$k_{\omega} = k_{\omega}^0 (1 - f_n - f_p). \quad (9.17)$$

С учетом (9.16) и (4.72) имеем

$$\begin{aligned} k_{\omega} &= A (\hbar\omega - E_g)^{1/2} (1 - f_n - f_p) = \\ &= A (\hbar\omega - E_g)^{1/2} \left(e^{\frac{\hbar\omega - (F_n^* - F_p^*)}{kT}} - 1 \right) f_n f_p. \end{aligned} \quad (9.18)$$

Зависимости (9.18) для двух температур $T_1 \approx 0$ и $T_2 > T_1$ проиллюстрированы на рис. 9.17. При высоких уровнях накачки в области энергий фотонов $\hbar\omega < F_n^* - F_p^*$ в полном соответствии с (9.18) показатель поглощения k_{ω} становится отрицательным. Этот диапазон энергий с отрицательным показателем поглощения как раз является спектральным диапазоном, в котором возможно усиление. Очевидно, что с ростом интенсивности накачки увеличивается расстояние между квазиуровнями Ферми, показатель усиления $\alpha_{\omega} = -k_{\omega}$ в области низких энергий насыщается, а его максимум сдвигается в сторону больших $\hbar\omega$. Оценить показатель усиления весьма просто из примера, рассмотренного в § 4.4, где мы получили, что в полупроводнике с прямой структурой энергетических зон при энергии фотонов на 0,01 эВ больше ширины запрещенной зоны показатель поглощения равен $\approx 1000 \text{ см}^{-1}$. Очевидно, что если в этом же материале с помощью накачки обеспечить такую концентрацию неравновесных носителей, при которой $F_n^* - F_p^*$ превышает E_g на ≈ 10 мэВ, то показатель усиления в максимуме будет иметь порядок $\alpha_{\omega} \approx 10^3 \text{ см}^{-1}$.

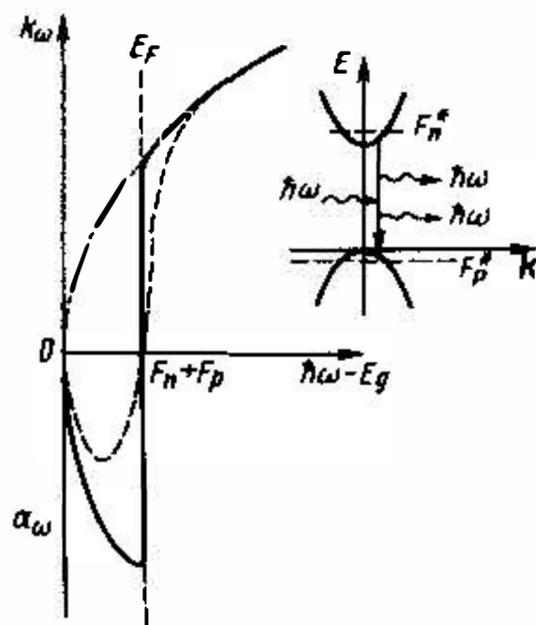


Рис. 9.17. Спектральные зависимости показателя поглощения (усиления) для прямых межзонных переходов при $T=0$ (сплошная линия) и $T>0$ (пунктирная линия). На вставке показаны переходы, определяющие усиление. Уровень накачки характеризуется положением F_n^* и F_p^* .

Для выполнения условия (4.76) необходимо, чтобы концентрации неравновесных носителей заряда в активной области были достаточно высоки. Это означает, что активная область по своим свойствам будет напоминать сильнолегированный полупроводник, для которого характерно образование хвостов плотности состояний, возможны нарушения правил отбора по волновому вектору за счет рассеяния на свободных электронах, дырках и т. п. (см. § 4.3). Если в качестве активной области используется легированный материал (например, p -типа), то указанные эффекты будут проявляться еще сильнее и к ним будут добавляться переходы на примесные состояния или в примесную зону. Это приведет к тому, что спектральная зависимость показателя поглощения будет отклоняться от (9.16). В частности, порог вблизи $\hbar\omega = E_g$ не будет столь резким, а в спектре поглощения (и усиления) в этой области будет наблюдаться экспоненциальный «хвост», характерный для сильнолегированного полупроводника (см. рис. 4.18).

Рисунок 9.17 не учитывает эти особенности и потому отражает ситуацию лишь качественно. Количественные результаты могут быть получены для конкретного полупроводника с учетом указанных обстоятельств. Спектральная зависимость показателя усиления рассчитывается по соотношению (9.17). Квазиуровни Ферми F_n^* и F_p^* для электронов и дырок определяются по заданным концентрациям носителей n и p в зонах. Поскольку соотношения (4.73) для вырожденных случаев не применимы, то положения квазиуровней Ферми надо находить через интегралы:

$$n = \int_{-\infty}^{\infty} \rho_c(E - E_c) f_n dE; \quad p = \int_{-\infty}^{\infty} \rho_v(E - E_v) f_p dE, \quad (9.19)$$

где f_n и f_p определяются выражениями (4.72), а $E_c - E_v = E_g$. В этих выражениях для плотности состояний в зоне проводимости $\rho_c(E - E_c)$ и в валентной зоне $\rho_v(E - E_v)$ можно использовать удобные параболические аппроксимации (4.9), поскольку хвосты зон не оказывают значительного влияния на F_n^* и F_p^* при столь высоких уровнях возбуждения. В то же время на спектральные зависимости поглощения и излучения хвосты плотности состояний оказывают существенное влияние. Поэтому при нахождении k_{ω}^0 в (9.17) их надо учитывать.

Очевидно, что при высоких концентрациях неосновных носителей концентрация основных носителей также должна увеличиваться, чтобы выполнялось условие электронейтральности

$$n + N_A^- = p + N_D^+. \quad (9.20)$$

В качестве примера на рис. 9.18 показаны распределения свободных носителей заряда в арсениде галлия p -типа

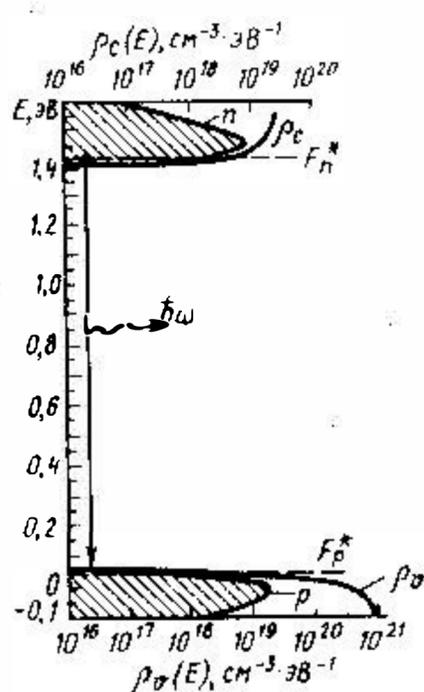


Рис. 9.18. Распределение неравновесных электронов и дырок в арсениде галлия при накачке до концентрации $n = 1 \cdot 10^{18} \text{ см}^{-3}$ и $p = 2,2 \cdot 10^{18} \text{ см}^{-3}$; $T = 300 \text{ K}$; $N_A - N_D = 1,2 \cdot 10^{18} \text{ см}^{-3}$

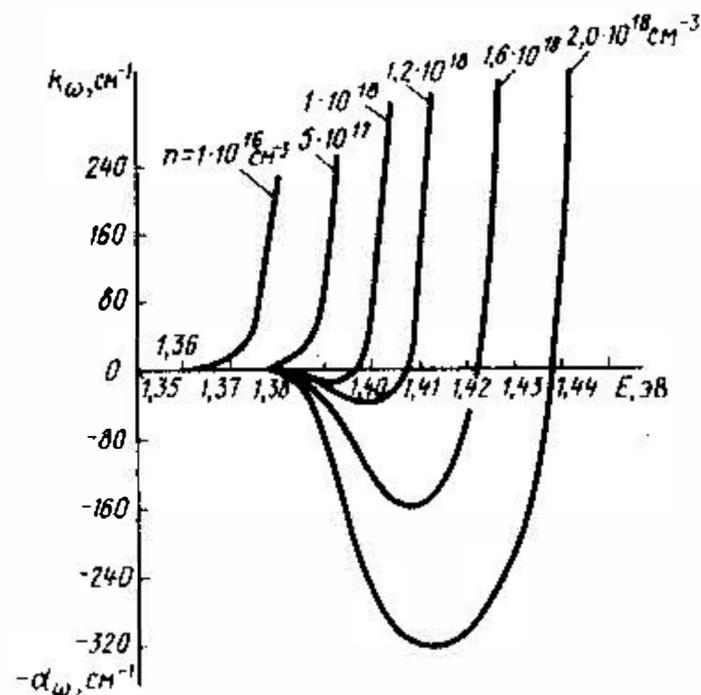


Рис. 9.19. Спектральные зависимости показателя поглощения (усиления), рассчитанные для GaAs при различных уровнях накачки; $T = 300 \text{ K}$; $N_A - N_D = p_0 = 1,2 \cdot 10^{18} \text{ см}^{-3}$

($p_0 = 1,2 \cdot 10^{18} \text{ см}^{-3}$) при высоком (лазерном) уровне возбуждения. (Обратите внимание, что в отличие от рис. 4.35 на рис. 9.18 применен логарифмический масштаб по оси плотности состояний.) На рис. 9.19 для этого же материала приведено изменение спектров поглощения при изменении концентрации неосновных носителей n . Видно, что при накачке до концентрации электронов $n = 1 \cdot 10^{18} \text{ см}^{-3}$ (при этом концентрация дырок становится равной $p = 2,2 \cdot 10^{18} \text{ см}^{-3}$) усиление при комнатной температуре практически отсутствует. Дальнейшее небольшое увеличение n приводит к появлению области отрицательного поглощения, так что при $n = 2 \cdot 10^{18} \text{ см}^{-3}$ показатель усиления в максимуме превышает 300 см^{-1} . Использование квантоворазмерных структур, рассмотренных в § 4.8, позволяет повысить показатель усиления вплоть до значений 10^4 см^{-1} для КТ за счет группировки плотности состояний вблизи уровней размерного квантования.

Таким образом, первой важной особенностью полупроводников как лазерных активных материалов является возможность получения исключительно высоких значений коэффициентов усиления с единицы длины. Следствием этого является возможность, а часто — необходимость применения активных элементов исключительно малых размеров. Действительно, при $\alpha_0 \approx 10^2 \dots 10^4 \text{ см}^{-1}$ длина активной области должна составлять $20 \dots 2000 \text{ мкм}$.

а ее размеры в поперечнике не должны превышать нескольких микрон, чтобы исключить развитие усиления в поперечных направлениях.

Как и в других генераторах, в полупроводниковых лазерах недостаточно получить только усиление внутри активной области. Необходимо, чтобы это усиление превысило все потери. Напомним (см. §§ 4.5, 4.7), что в полупроводниках с непрямой структурой зон выполнить эти условия весьма сложно, а иногда и вовсе невозможно. Поэтому полупроводниковые лазеры (в отличие от светодиодов) изготавливают только на основе прямозонных материалов. Излучательные времена жизни в них малы и имеют порядок 1 нс и менее (см. табл. 4.2), уменьшаясь с ростом концентрации свободных носителей. Поэтому для получения концентрации неравновесных носителей на уровне 10^{18} см^{-3} необходимо в активном объеме осуществлять генерацию электронно-дырочных пар со скоростью не менее $10^{27} \text{ см}^{-3} \text{ с}^{-1}$, т. е. при толщине активного слоя всего 1 мкм плотность потока возбуждения должна быть не менее $10^{23} \text{ см}^{-2} \text{ с}^{-1}$. При возбуждении такого активного слоя (например, фотонами) необходима мощность возбуждения более 10^{23} фотонов на см^2 в секунду, что для $\lambda \approx 1 \text{ мкм}$ составляет $P_{\text{возб}} \geq 20 \text{ кВт/см}^2$. При возбуждении путем инжекции неравновесных носителей заряда через p - n -переход этому уровню возбуждения соответствует плотность тока порядка $2 \cdot 10^4 \text{ А/см}^2$. Эти значения хотя и велики, во реально достижимы.

Второй важной особенностью полупроводников как лазерных материалов является возможность непосредственного преобразования электрической энергии в световую, происходящего при инжекционной электролюминесценции (§ 9.1). Эта особенность реализуется только в инжекционных лазерах.

Указанные две важнейшие особенности полупроводников как активных лазерных материалов как раз и определяют те преимущества полупроводниковых лазеров, которые были перечислены в начале главы.

По механизму возбуждения полупроводниковые лазеры разделяют на лазеры с электронной или оптической накачкой и на инжекционные лазеры.

§ 9.4. ПОЛУПРОВОДНИКОВЫЕ ЛАЗЕРЫ С ЭЛЕКТРОННОЙ И ОПТИЧЕСКОЙ НАКАЧКОЙ

Возбуждение люминесценции в полупроводниках при воздействии на них быстрыми электронами (катодолуминесценция) хорошо известно и широко используется в телевизионных электронно-лучевых трубках: электроны, ускоренные электрическим полем, бомбардируют экран (мишень), на который нанесен слой люминофора. В качестве люминофора обычно применяют полу-

проводниковый сульфид цинка (ZnS) с различными добавками (активаторами), определяющими цвет свечения. При соблюдении определенных условий этот же метод возбуждения может быть использован для генерации лазерного излучения.

Накачка быстрыми электронами является наиболее универсальным способом возбуждения лазерной генерации в полупроводниках, поскольку она применима к материалам с любой шириной запрещенной зоны и типом электропроводности и не требует создания специальных структур типа $p-n$ -переходов, т. е. применима к однородным кристаллам и слоям. Благодаря этому с помощью электронной накачки получена лазерная генерация почти во всех прямозонных полупроводниках, в том числе в полупроводниках $A^{\text{III}}B^{\text{V}}$, $A^{\text{II}}B^{\text{VI}}$ и $A^{\text{IV}}B^{\text{VI}}$ и их твердых растворах.

При попадании в полупроводник быстрые электроны тормозятся, отдавая часть своей энергии на возбуждение неравновесных электронно-дырочных пар. За счет процессов многоступенчатой ионизации образуются «горячие» электроны и дырки, которые быстро термализуются, приходя в квазиравновесие отдельно в зоне проводимости и в валентной зоне, как это обсуждалось в § 4.6. Накачка должна быть настолько интенсивной, чтобы охлажденные (термализованные) носители были вырождены в каждой из зон и выполнялось условие инверсии (4.76). При остывании горячих носителей их энергия передается решетке, приводя к разогреву кристалла. Теоретические расчеты, подтвержденные многочисленными экспериментальными результатами, показывают, что минимальная энергия быстрого электрона, необходимая для создания одной электронно-дырочной пары, равна $3 E_g$. Это означает, что только $1/3$ энергии электронного пучка расходуется полезно, а $2/3$ идет на разогрев полупроводника. Это требует специальных мер по осуществлению теплоотвода и охлаждению активного элемента.

Полупроводниковые лазеры с электронной накачкой могут быть реализованы в двух геометрических разновидностях: с поперечной и с продольной накачкой. Первая из них представлена на рис. 9.20, а. В этом случае ось оптического резонатора и направление лазерного излучения примерно перпендикулярны направлению электронного пучка накачки. Резонатором служит сам кристалл, боковые грани которого строго параллельны друг другу. Вследствие большого значения показателя преломления полупроводника (см. § 4.5) коэффициент зеркального отражения от этих граней составляет 30...40%. При высоких значениях коэффициента усиления этого достаточно для реализации положительной обратной связи. В первых полупроводниковых лазерах резонатор изготовлялся путем тщательной оптической полировки торцов активного элемента. В настоящее время резонатор полупроводниковых лазеров изготовляют путем скола боковых граней кристалла, из которого изготовлен активный элемент.

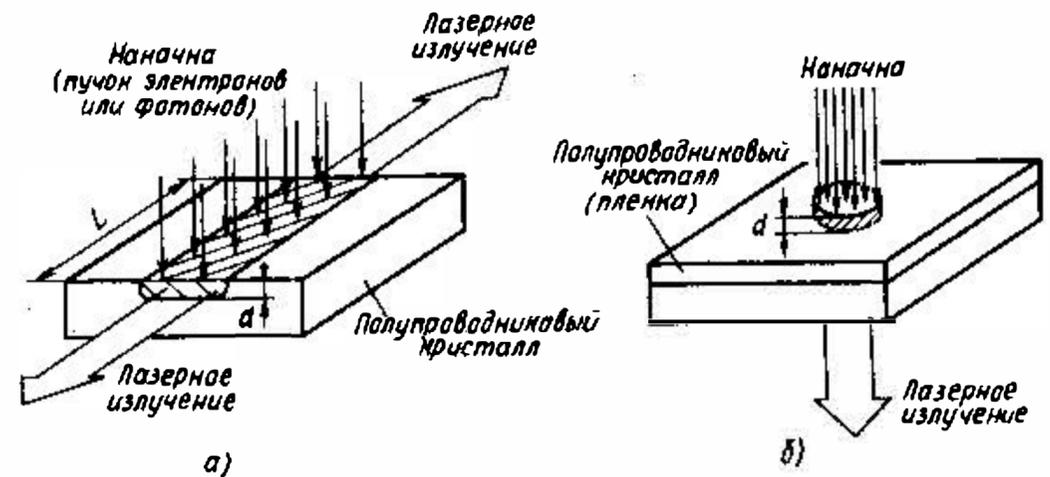


Рис. 9.20. Полупроводниковый лазер с поперечной (а) и продольной (б) электронной или оптической накачкой. Активная область заштрихована

У большинства полупроводников $A^{\text{III}}B^{\text{V}}$ и $A^{\text{II}}B^{\text{VI}}$, кристаллизующихся в решетке сфалерита, плоскостями спайности являются плоскости (110). При высоком качестве монокристалла скалывание по этим плоскостям дает возможность реализовать плоский резонатор.

Длина активной области L должна быть порядка нескольких величин $1/\alpha_\omega$. Поскольку показатель усиления α_ω в полупроводниках по оценкам, выполненным в § 9.2, достигает нескольких сотен см^{-1} , то для возникновения генерации достаточно L порядка десятков микрон. Толщина активной области d определяется глубиной проникновения быстрых электронов. Последняя связана с энергией электронов E_e соотношением

$$d \approx KE_e^\gamma, \quad (9.21)$$

где параметры K и γ зависят от свойств материала. Обычно $\gamma = 1,3 \dots 2,0$.

Для ZnS и GaAs при $E_e \approx 10$ кэВ $d \approx 1$ мкм.

В варианте с продольным возбуждением (см. рис. 9.20, б) ось резонатора совпадает с направлением электронного пучка. В этом случае длина активной области определяется глубиной проникновения быстрых электронов. Поэтому для достижения требуемой $L_{\text{акт}}$ необходимы большие E_e , достигающие 100 кэВ и более. Верхняя граница E_e определяется возникновением дефектов, ведущих к деградации лазера, а также появлением характеристического рентгеновского излучения.

Основной интерес к полупроводниковым лазерам с продольной электронной накачкой вызван простотой двумерного сканирования остро сфокусированным электронным лучом. На этом принципе работает лазерная электронно-лучевая трубка, схема которой приведена на рис. 9.21. В такой трубке ускоренные до

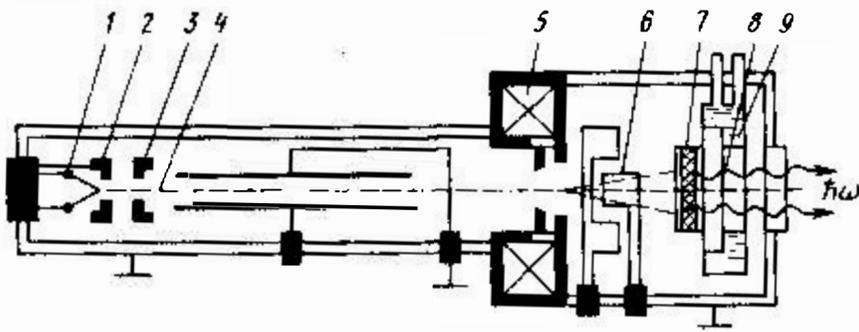


Рис. 9.21. Схема лазерной электронно-лучевой трубки:
 1 — катод; 2 — управляющий электрод; 3 — анод; 4 — электронный пучок; 5 — магнитная линза; 6 — отклоняющая система; 7 — лазерная мишень (экран); 8 — сапфировая подложка; 9 — охлаждающая жидкость

энергий $eU_{\text{раб}}$ электроны осуществляют продольную накачку плоскопараллельной полупроводниковой мишени — пластинки диаметром несколько сантиметров и толщиной несколько десятков микрон. Со стороны падения электронов на мишень наносится тонкое металлическое зеркальное покрытие, легко проникаемое для быстрых электронов и выполняющее функцию одного из зеркал резонатора. Второе (выходное) зеркало наносится на противоположную сторону кристалла или подложки. Пятно лазерного излучения диаметром десятки микрон может перемещаться непрерывно или дискретно по пластине. В расстрове режиме пятно построчно проходит весь кадр. Благодаря направленности ($\theta \approx 20^\circ$) и высокой плотности излучения лазерного луча яркость изображения в лазерной трубке в сотни и тысячи раз превышает яркость стандартного телевизионного изображения. Это дает возможность путем простой оптической системы спроектировать полученное изображение на площадь в несколько квадратных метров, оставляя яркость, приемлемую для его последующего восприятия. К сожалению, ряд технических и технологических трудностей, а также необходимость применения больших рабочих напряжений пока что ограничивают применение таких трубок как лабораторных образцов.

Полупроводниковые лазеры с поперечной схемой электронной накачки выпускаются промышленностью. Как правило, они работают в импульсном режиме с пиковой мощностью до десятков кВт в сине-зеленой ($\lambda_{\text{раб}} = 0,5$ мкм, рабочее вещество — сульфид кадмия) и оранжевой ($\lambda_{\text{раб}} = 0,62 \dots 0,63$ мкм, рабочее вещество — твердый раствор сульфида-селенида кадмия) областях видимого спектра, т. е. в тех областях, которые пока что не освоены инжекционными лазерами.

Очевидно, что вместо электронной накачки может быть применена оптическая накачка фотонами с энергией $\hbar\omega \geq E_g$ и сказанное выше в своей основе справедливо для полупроводникового лазера с оптической накачкой (рис. 9.20). Необходимость острого

фокусировки и большой интенсивности возбуждения требует применения для накачки лазера с $\hbar\omega \approx E_g$ (или $2\hbar\omega \geq E_g$ при двухфотонной накачке), так что полупроводниковый лазер с оптической накачкой выполняет скорее функции конвертора лазерного излучения и применяется редко.

§ 9.5. ИНЖЕКЦИОННЫЕ ДГС-ЛАЗЕРЫ

В инжекционных лазерах накачка производится путем инжекции неосновных носителей заряда через p - n -переход при пропускании через него тока в прямом направлении. Эти лазеры являются наиболее важными из всех полупроводниковых лазеров.

Общая схема инжекционного лазера приведена на рис. 9.22. При подаче смещения в прямом направлении электроны инжектируются в p -область, а дырки — в n -область, создавая вблизи p - n -перехода активный слой, в котором происходит излучательная рекомбинация. При большой плотности тока (большое смещение) в вырожденном переходе в этом слое может быть реализовано условие инверсии (4.76), как показано на рис. 9.2. Функцию резонатора, как указывалось в § 9.4, могут выполнять сколотые по плоскостям спайности грани активного кристалла. Очевидно, что плоскость p - n -перехода должна быть строго перпендикулярна сколотым граням резонатора, как показано на рис. 9.22. Для обеспечения эффективного взаимодействия света с активной средой необходимо совместить область, в которой создана инверсия, с областью распространения светового излучения, т. е. в одном и том же активном слое локализовать неравновесные носители заряда и фотоны. В обычном p - n -переходе,

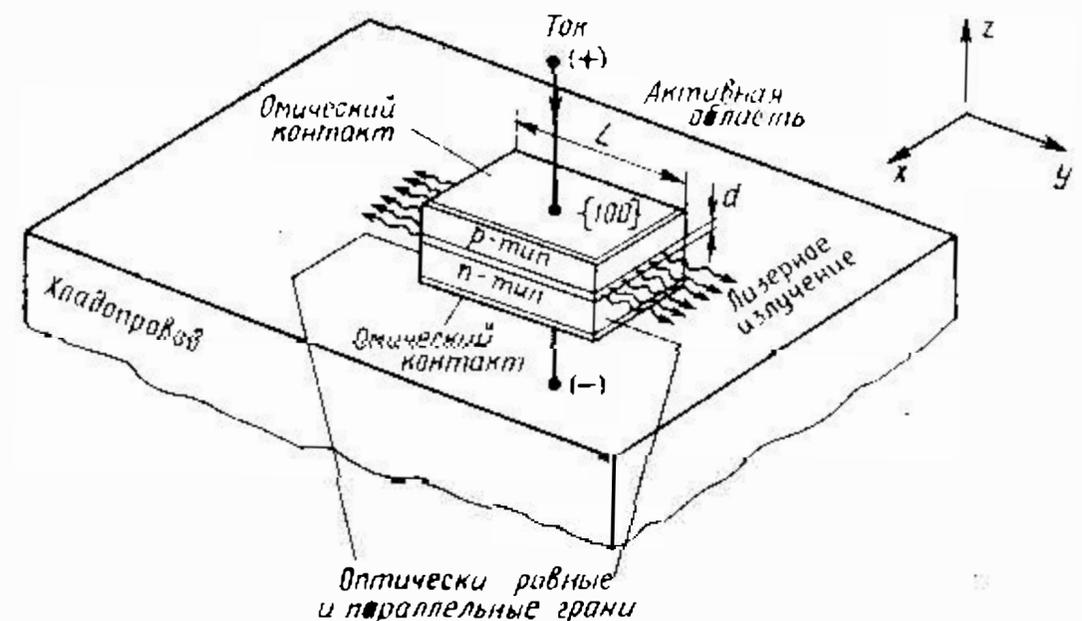


Рис. 9.22. Упрощенная структура инжекционного лазера

изготовленном, например, путем диффузии акцептора Zn в n-GaAs, это требование реализуется далеко не оптимальным образом. Поэтому такие лазеры обладают плохими рабочими характеристиками, в частности — очень высокими значениями пороговой плотности тока (выше 50 кА/см²). Из-за большого тепловыделения такие лазеры могут работать лишь в импульсном режиме при низких температурах.

Осуществить локализацию электронного и светового потоков и их совмещение возможно в двойных гетероструктурах (ДГС), используя свойства гетеропереходов, рассмотренные в § 9.1. Общая схема инжекционного ДГС-лазера такая же, как на рис. 9.22, но активная область толщиной d представляет собой тонкий слой узкозонного полупроводника, расположенного между двумя широкозонными n - и p -областями, выполняющими функции эмиттеров. На рис. 9.23 проведено сравнение свойств лазерных структур на основе p - n -гомоперехода (рис. 9.23, а) и двойной гетероструктуры (рис. 9.23, а'). В p - n -гомопереходе (левая часть рис. 9.23) электроны из сильно легированной n^+ -области инжектируются в сильно легированную p^+ -область GaAs- p - n -перехода, рассасываясь там с характерной длиной, равной диффузионной длине электронов в p -области L_n^p . Аналогично для дырок, инжектируемых в n -область. Но обычно $L_n^p \gg L_p^n$, и потому на рис. 9.23, б L_p^n не указана, а спад концентрации дырок на рис. 9.23, в происходит значительно быстрее, чем электронов.

Свойства гетеропереходов, рассмотренные в § 9.1 (односторонняя инжекция, сверхинжекция и волноводный эффект), позволяют существенно облегчить достижение условия инверсии в инжекционных ДГС-лазерах. В качестве примера рассмотрим двойную гетероструктуру $(n\text{-Al}_{x_1}\text{Ga}_{1-x_1}\text{As}) = (p\text{-GaAs}) = (p^+\text{-Al}_{x_2}\text{Ga}_{1-x_2}\text{As})$. В правой части рис. 9.23 для этой структуры приведены энергетическая диаграмма при прямом смещении (б'), распределение концентрации электронов и дырок (в'), профиль показателя преломления (г') и распределение интенсивности основной аксиальной моды излучения лазера (д').

Центральная область (p -GaAs) является активной. Из широкозонной n -области в нее инжектируются электроны. За счет эффекта сверхинжекции концентрация электронов в активной области оказывается больше, чем в N -эмиттере (скачок концентрации на диаграмме рис. 9.23, в'). Одновременно потенциальный барьер на гетерогранице p - p^+ препятствует движению инжектированных электронов, запирая (локализуя) их в пределах узкозонной активной области. Таким образом в ДГС-лазере осуществляется электронное ограничение.

Аналогичная ситуация наблюдается для дырок: потенциальный барьер в валентной зоне на гетерогранице $(N\text{-Al}_{x_1}\text{Ga}_{1-x_1}\text{As})$ - $(p\text{-GaAs})$ ограничивает распространение дырок из p -области, запирая их в пределах активной области. Но так как $\Delta E_v < \Delta E_c$, то

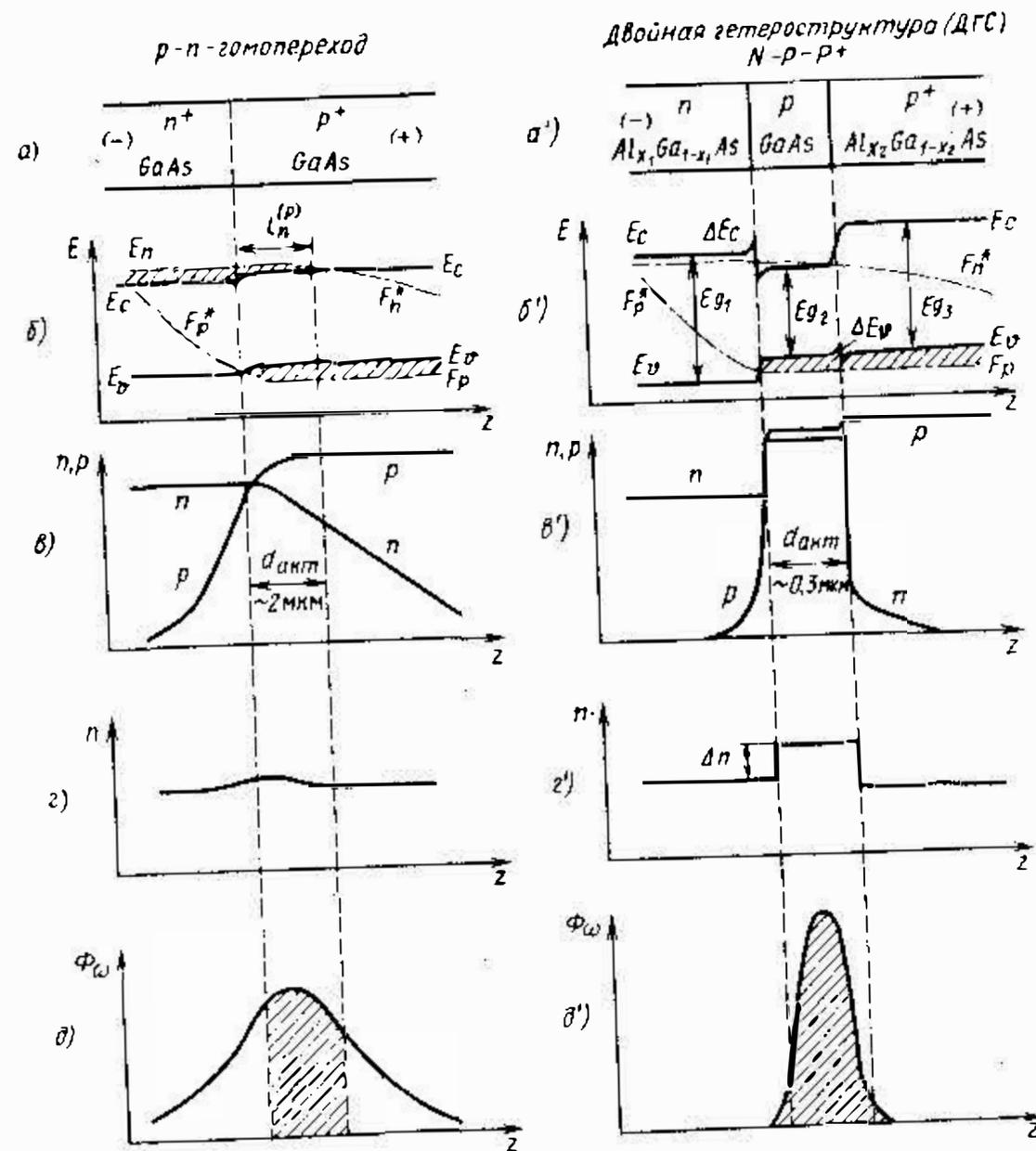


Рис. 9.23. Сравнительные характеристики лазерных структур на гомо- и гетеро- p - n -переходах

для дырок этот эффект выражен слабее. По этой же причине эффект сверхинжекции из p - $\text{Al}_{x_1}\text{Ga}_{1-x_1}\text{As}$ в n -GaAs мал и в качестве эмиттера дырок применяется легированный до вырождения p^+ -материал.

Кроме электронного ограничения в двойной гетероструктуре осуществляется эффективное оптическое ограничение. Показатель преломления узкозонного активного слоя больше, чем показатель преломления прилегающих широкозонных областей. На гетерограницах наблюдаются скачки показателя преломления (рис. 9.23, г') и активный слой ведет себя как волновод, локализуя излучение, распространяющееся в плоскости p - n -перехода (рис. 9.23, д'). Ситуация здесь аналогична рассмотренной в § 3.4 для световой волны, распространяющейся в плоском волноводе (рис. 3.7 — 3.9). В обычном p - n -переходе (рис. 9.23, г, д) эффекты

выражены крайне слабо лишь за счет разницы показателей преломления, возникающей вследствие различного легирования p и n -областей.

Таким образом, в активной области двойной гетероструктуры возможно создать благоприятные условия для получения инверсной населенности, локализовать электронный и световой потоки и совместить их. Свет и носители, удерживаемые в активном слое, эффективно взаимодействуют, благодаря чему пороговый уровень накачки (пороговая плотность тока) существенно уменьшается.

В полупроводниковых лазерах, как и в лазерах других типов, условие инверсии (4.76) является необходимым, но не достаточным для возникновения генерации. Как и ранее (см. § 2.4), достаточным условием является преобладание усиления над потерями. Выведем формулу для пороговой плотности тока инжекционного ДГС-лазера.

Будем считать, что в активном слое толщиной d в направлении y (рис. 9.22) как в диэлектрическом волноводе распространяется основная $TE_{M_{00}}$ -мода. Длина резонатора L , а коэффициент отражения зеркальных граней R . Аналогично тому, как это было сделано в § 2.4, условие самовозбуждения будет выполнено, когда усиление скомпенсирует все потери, т. е.

$$R \exp[(\alpha_{\omega}^0 - \beta_{\omega})L] \geq 1, \quad (9.22)$$

откуда

$$\alpha_{\omega}^0 \geq \beta_{\omega} + 1/L \ln(1/R), \quad (9.23)$$

где $\alpha_{\omega}^0 = -k_{\omega}$ — ненасыщенный показатель усиления вещества активного слоя, а β_{ω} — показатель объемных потерь; как следует из § 4.4, он в основном определяется поглощением на свободных носителях заряда на частоте генерации ω .

Как видно из диаграмм рис. 9.23, δ и δ' , световая волна частично просачивается в неактивные области, что не учтено в соотношении (9.23). Учесть это обстоятельство можно, домножив α_{ω}^0 на безразмерный коэффициент ξ , называемый *коэффициентом удержания света* или *параметром оптического ограничения* и представляющий собой отношение энергии световой моды, заключенной внутри активного слоя, к ее полной энергии. Поскольку вне активного слоя усиление не происходит, то условие самовозбуждения запишется в виде

$$\alpha_{\omega}^0 \xi \geq \beta_{\omega} + 1/L \ln(1/R). \quad (9.24)$$

Произведение $\alpha_{\omega}^0 \xi$ называется *показателем усиления моды*. Нетрудно показать, что в случае разных коэффициентов отражения R_1 и R_2 граней или зеркал резонатора вместо $1/R$ в (9.24) надо поставить $1/\sqrt{R_1 R_2}$.

Как обсуждалось в § 2.4 и следует из соотношения (2.47), ненасыщенный показатель усиления пропорционален инверсии ΔN . Аналогично в нашем случае показатель усиления α_{ω}^0 возрастает пропорционально концентрациям неравновесных носителей в активном слое. Обозначив коэффициент пропорциональности через A_0 , запишем

$$\alpha_{\omega}^0 = A_0 (n - n_0), \quad (9.25)$$

где n_0 имеет смысл концентрации носителей, при которой поглощение между зонами гасится и возникает усиление. В случае, для которого приведены расчеты на рис. 9.19, нетрудно оценить $n_0 \approx 5 \cdot 10^{17} \text{ см}^{-3}$.

Концентрация носителей заряда (например, электронов n) в активном слое ДГС-лазера толщиной d связана с временем их жизни τ_n и плотностью тока инжекции J соотношением

$$n = \frac{\tau_n}{e d} J. \quad (9.26)$$

Предположим, что время жизни в активной области определяется излучательными переходами зона — зона с квантовым выходом, равным единице. Тогда, аналогично (4.84), излучательное время жизни обратно пропорционально концентрации носителей:

$$\tau_n = \tau_{изл} = \frac{1}{n B} \quad (9.27)$$

с той разницей, что в знаменателе (4.84) стоит концентрация собственных носителей n_i или p_0 , а в (9.27) — концентрация инжектированных носителей. В этом приближении получаем

$$n = \sqrt{\frac{J}{e B d}} \quad (9.28)$$

т. е. концентрация неравновесных носителей в активном слое пропорциональна корню из плотности тока инжекции J и обратно пропорциональна корню из толщины слоя d и эффективному коэффициенту рекомбинации B , смысл которого вытекает из соотношений (4.77), (4.83) и (4.84).

Зависимость показателя усиления от плотности тока инжекции получаем из (9.25), (9.28):

$$\alpha_{\omega}^0 = \frac{A_0}{\sqrt{e B}} \sqrt{\frac{J}{d}} - \alpha_0, \quad (9.29)$$

где $\alpha_0 = A_0 n_0$.

Пороговые условия соответствуют знаку равенства в (9.24), откуда с учетом (9.29) находим пороговую плотность тока по генерации:

$$J_{\text{пор}} = \frac{eBd}{A_0^2 \xi^2} \left[\alpha_0 \xi + \beta_\omega + \frac{1}{L} \ln \frac{1}{R} \right]^2 \quad (9.30)$$

Из этой формулы видно, что пороговая плотность тока падает с уменьшением толщины d активного слоя. Это имеет ясный физический смысл, поскольку усиление определяется объемной концентрацией носителей в активном слое. Но прямая пропорциональность $J_{\text{пор}} \sim d$ будет соблюдаться лишь тогда, когда все входящие в (9.30) величины от d не зависят. Обратимся к рис. 9.23, d и 3.9. Если толщина волноводного слоя d будет меньше половины длины волны света в веществе, то волна будет просачиваться за стенки волновода, так что при $d < \frac{\lambda}{2n}$ (n — показатель преломления) резко уменьшится коэффициент удержания света ξ и возрастет $J_{\text{пор}}$. На рис. 9.24 приведены типичные зависимости пороговой плотности тока от толщины активной области при комнатной температуре для ДГС-лазеров, работающих на длинах волны $\lambda = 1,3$ мкм (GaInAsP/InP-лазер) и $\lambda = 0,86$ мкм (Al_xGa_{1-x}As/GaAs-лазер).

Поскольку в определенной области $J_{\text{пор}} \approx d$, то для оценки качества лазера применяют понятие *номинальной пороговой плотности тока* $J_{\text{пор.ном.}}$, которая определяется как отношение $J_{\text{пор.}}/d$ и измеряется в [кА/(см² мкм)]. Оценим $J_{\text{пор.ном.}}$ для ДГС-лазера

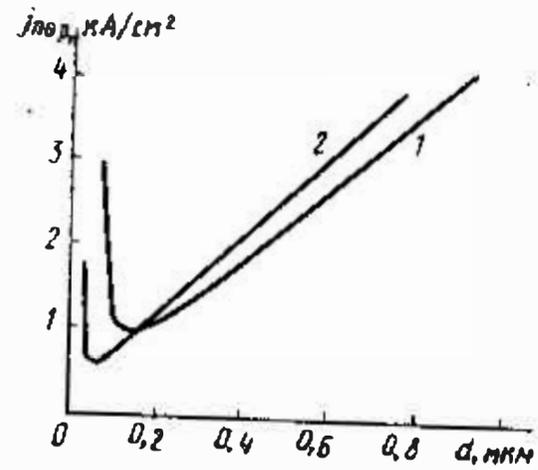


Рис. 9.24. Зависимости пороговой плотности тока при комнатной температуре от толщины активной области ДГС-лазеров:
1 — GaInAsP/InP, $\lambda = 1,3$ мкм; 2 — Al_{0,3}Ga_{0,7}As/GaAs, $\lambda = 0,86$ мкм

длиной $L \approx 300$ мкм с активной областью из арсенида галлия. Примем $\xi = 1$; $B = 7 \cdot 10^{-10}$ см² с⁻¹ (табл. 4.2); $\beta_\omega \approx 10$ см⁻¹ (поглощение на свободных носителях заряда при $n, p \sim 10^{18}$ см⁻³, $\lambda \approx 0,9$ мкм); R определим из (3.23) и (4.69), а величины A_0 и α_0 найдем по данным рис. 9.19: при $n = 2 \cdot 10^{18}$ см⁻³, $\alpha_\omega \approx 300$ см⁻¹, а $n_0 \approx 5 \cdot 10^{17}$ см⁻³, так что из (9.25) имеем $A_0 \approx 2 \cdot 10^{16}$ см² и $\alpha_0 = A_0 n_0 \approx 100$ см⁻¹. Подставляя эти значения в (9.30), получаем $J_{\text{пор.ном.}} \approx 5$ кА/(см² мкм). Вообще, если в ДГС-лазерах при комнатной температуре $J_{\text{пор.ном.}} \approx (3 \dots 5)$ кА/(см² мкм), то такой лазер считается хорошего качества.

Помимо $J_{\text{пор}}$ важным параметром, характеризующим качество полупроводникового лазера, является *дифференциальная эффективность* или *дифференциальный квантовый выход* η_D , который задает крутизну линейной части $\Phi_\omega(I)$ характеристики и определяется как

$$\eta_D = \frac{\Delta D_\omega / h\nu}{\Delta I / e} = \frac{\Delta \Phi_\omega}{\Delta I / e} \quad (9.31)$$

т. е. как отношение приращения числа фотонов Φ_ω , излучаемых лазером, к приращению числа носителей, инжектируемых в единицу времени в активную область (рис. 9.14, б).

Напомним, что выражения (9.29) и (9.30) для показателя усиления и пороговой плотности тока были получены в предположении, что внутренний квантовый выход люминесценции в активном слое равен единице, а время жизни определяется межзонными или квазимежзонными переходами по (9.27). Предоставляем читателю возможность самому получить выражения для α_ω^0 и $J_{\text{пор}}$ при $\eta < 1$ и $\tau_{n,p} = \text{const}$.

Рассматривая свойства ДГС-лазера, мы предполагали, что в активном слое распространяется только одна основная TEM₀₀-мода. На самом деле возможно возбуждение колебаний более высокого порядка, например TEM₁₀ и TEM₂₀, распределение поля для которых приведено на рис. 2.20 и 3.9. Возможность распространения волноводных мод более высокого порядка в активном слое как в планарном волноводе определяется отношением характерного геометрического размера волновода к длине волны. В частности, при $d \approx \lambda / (2n)$, как видно из рис. 3.9 и 9.25, коэффициент удержания света ξ для мод с $m \geq 1$ будет мал и они будут затухать. При увеличении d возможно появление мод более высокого порядка.

Если в направлении z модовое ограничение достигается просто уменьшением d , то в поперечном направлении x в ДГС-лазере ограничения нет, поскольку ширина активной области $a \gg \lambda$. Это приводит к появлению большого числа волноводных мод типа TEM_{0m}. Эти моды взаимодействуют друг с другом, создавая биения и неустойчивости, зависящие от плотности тока накачки. Кроме того, в геометрии ДГС-лазера (см. рис. 9.22) при больших значениях плотности тока возможно появление токовых неустойчивостей, связанных с эффектами шнурования, локального нагрева и т. п. Набор всех этих причин

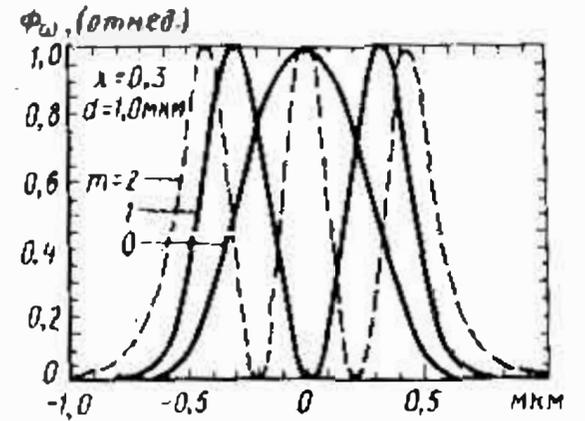


Рис. 9.25. Распределение интенсивности излучения в волноводе на основе двойной гетероструктуры для трех первых мод

приводит к появлению характерных изломов или «колен» на ватт-амперной характеристике лазера $\Phi_0(I)$, к возрастанию шума и к неустойчивой его работе. Избежать этих явлений можно в полосковых ДГС-лазерах.

§ 9.6. ПОЛОСКОВЫЕ ГЕТЕРОЛАЗЕРЫ

Полосковым лазером называют инжекционный гетеролазер в котором активная область выполнена в виде узкой полоски. Такая конструкция обеспечивает уменьшение рабочего тока (при неизменном $J_{пор}$), эффективную селекцию поперечных мод в направлении, параллельном $p-n$ -переходу, и устойчивую работу лазера.

Простейшим способом ограничения ширины активной области в плоскости $p-n$ -перехода является создание омического контакта не по всей площади, а в виде узкой полоски. Эта полоска шириной от единиц до десятков микрон может быть изготовлена обычными методами фотолитографии. В качестве примера на рис. 9.26 показана рабочая структура контактного полоскового лазера на основе гетероструктуры GaInAsP/InP. В отличие от рис. 9.22 реальный элемент содержит пять слоев, из которых три центральных образуют непосредственно рабочую ДГС-структуру. Лазер изготавливается следующим образом. На подложку низкоомного n -InP толщиной около 150 мкм эпитаксиальным методом наращивают слой n -InP, служащий широкозонным N -эмиттером. На него наращивается тонкий слой твердого раствора $Ga_xIn_{1-x}As_{1-y}P_y$ с составами x и y , отвечающими условию изопериодичности с InP. Этот слой слабого p - (иногда n -) типа образует активную область. На него наращивается широкозонный P -InP-эмиттер, который покрывается тонким слоем низкоомного p^+ -InGaAsP для обеспечения хорошего омического контакта. На этот слой наносится слой диэлектрика (например, SiO_2), в котором методом фотолитографии протравливается уз-

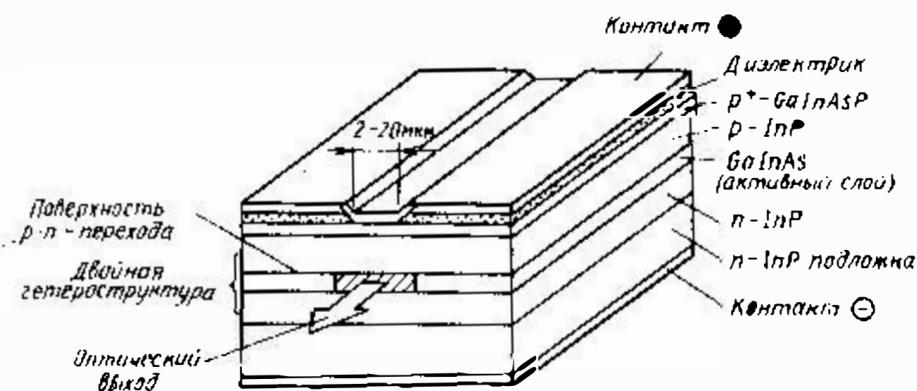


Рис. 9.26. Структура контактного ДГС-полоскового лазера $Ga_xIn_{1-x}As_{1-y}P_y/InP$

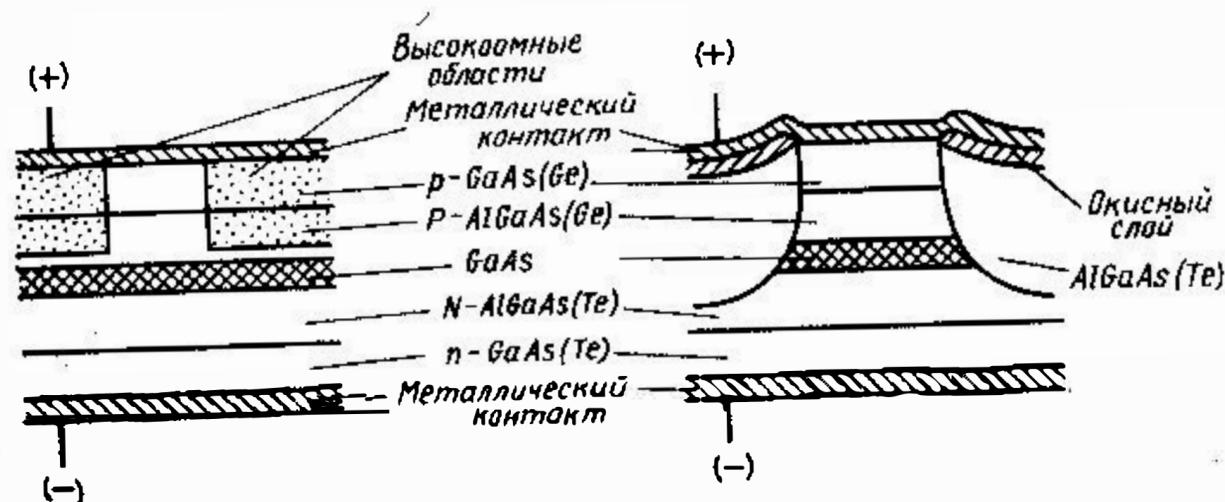


Рис. 9.27. Варианты структуры ДГС-полосковых лазеров

кое окно. Верхний омический контакт напыляется на всю структуру. Такая многослойная структура создается на всей площади InP-подложки. Затем из этой структуры параллельно полосковым контактам нарезаются полосы шириной 200...300 мкм, из которых затем методом скола по плоскостям {110} выкалываются рабочие лазерные структуры. Эти структуры напаяются на массивный держатель, выполняющий роль теплоотвода и служащий омическим контактом к n -области. Верхний контакт также делается достаточно массивным для улучшения теплоотвода. Иногда всю структуру помещают на термоэлектрический холодильник.

Для достижения наилучших характеристик лазера необходимо обеспечить электронное и оптическое ограничения не только в направлении z , но и в плоскости $p-n$ -перехода в направлении x , т. е. создать нитевидную структуру. Очевидно, что структура контактного полоскового лазера не обеспечивает хорошего ограничения в плоскости $p-n$ -перехода. Поэтому были предложены несколько вариантов более сложных структур полосковых лазеров, две из которых приведены на рис. 9.27 (для разнообразия взята система $GaAs-Al_xGa_{1-x}As$). В первом случае полосковая геометрия создается путем бомбардировки протонами всей площади, за исключением контактной полости. Бомбардировка производится на ускорителе на глубину, немного не доходящую до активного слоя. Облучение протонами создает дефекты в материале, в результате чего эта область становится высокоомной и рабочий ток идет только через необлученную полосу.

Во втором случае узкая полоска формируется вытравливанием мезаструктуры и ее последующим наращиванием широкозонным твердым раствором $Al_xGa_{1-x}As$ n -типа. В этом случае, как нетрудно убедиться из рис. 9.27, б, обеспечивается как электронное, так и оптическое ограничение по двум координатам.

§ 9.7. ГЕТЕРОЛАЗЕРЫ С РАЗДЕЛЬНЫМ ЭЛЕКТРОННЫМ И ОПТИЧЕСКИМ ОГРАНИЧЕНИЕМ

Для уменьшения рабочего тока ДГС-лазера необходимо уменьшать толщину активной области d (см. § 9.5 и соотношение 9.30). Однако при $d < \lambda/n$ этому препятствует увеличение потерь, вызванных просачиванием света в низкоомные P^+ - и N -области эмиттеров, и уменьшение коэффициента удержания света ξ (рис. 9.24). Чтобы уменьшить оба этих эффекта, необходимо разделить области электронного и оптического ограничения, поместив сверхтонкий активный слой толщиной $d \ll \lambda$ внутри более толстого слоя толщиной $D \approx \lambda/n$. Ширина запрещенной зоны области, осуществляющей оптическое ограничение, должна быть больше, чем активного слоя. Свойства такой структуры проиллюстрированы на рис. 9.28, а, б. При пропускании тока в прямом направлении электроны и дырки из широкозонных N - и P -эмиттеров инжектируются в прилегающие к ним области аналогично тому, как это происходит в ДГС-лазере (рис. 9.23). Толщина слоя D примерно такая же, как активной области ДГС-лазера, т. е. $D \approx \lambda/(2n)$. Этот слой играет двойную роль. Во-первых, он является оптическим волноводом и осуществляет оптическое ограничение (рис. 9.28, б). Во-вторых, он служит своего рода резервуаром для неравновесных носителей, откуда они поступают в активную область толщиной $d \ll D$. Учитывая, что при генерации время жизни неравновесных носителей в активной области резко уменьшается за счет вынужденных переходов, последнее обстоятельство является достаточно важным. Локализация носителей в слое D увеличивает вероятность их захвата в активную область d .

Как видно из рис. 9.28, б, параметр оптического ограничения ξ уменьшается с уменьшением d , однако это уменьшение происходит медленнее, чем в ДГС-лазере. Действительно, при $d \ll \lambda$ в ДГС-лазере без отдельного ограничения величина ξ уменьшается примерно как $\xi \approx d^{-2}$, что определяется двумя причинами: 1) увеличением ширины распределения $\Phi_{\omega}(z)$, вызванным ослаблением волноводного эффекта при $d \ll \lambda$, т. е. расширением области оптической локализации, и 2) чисто геометрическим уменьшением d . В лазерах с отдельным электронным и оптическим ограничением (РОДГС-лазерах) первая причина аннулируется, а вторая остается. На рис. 9.29 представлены зависимости величины параметра оптического ограничения от толщины активного слоя d , рассчитанные для волноводной структуры, профиль показателя преломления которой приведен на вставке к этому рисунку. Расчеты проведены для $\lambda = 1,55$ мкм.

Для успешной работы РОДГС-лазера уменьшение ξ при уменьшении d должно компенсироваться ростом показателя усиления α_0 в активном слое. В современных лазерах с отдельным

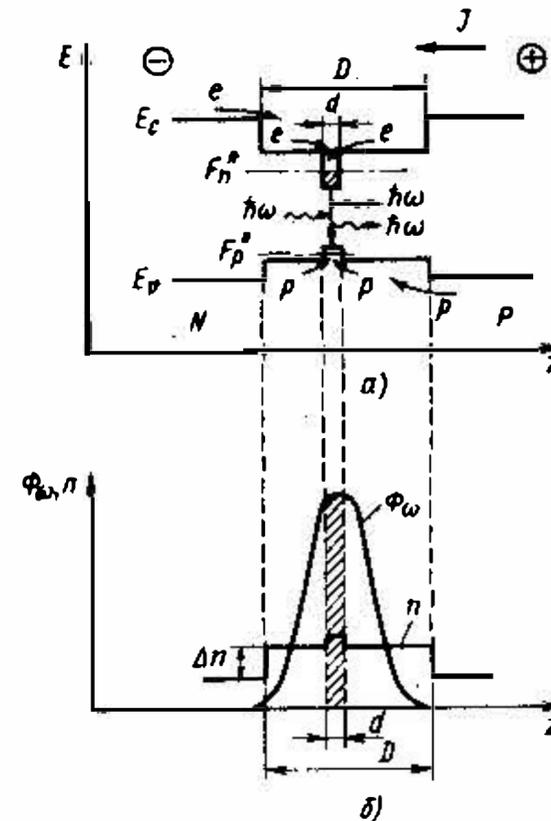


Рис. 9.28. Гетеролазер с отдельным электронным и оптическим ограничением (РО ДГС-лазер): а — энергетическая диаграмма; б — распределение показателя преломления n и поле излучения Φ_{ω}

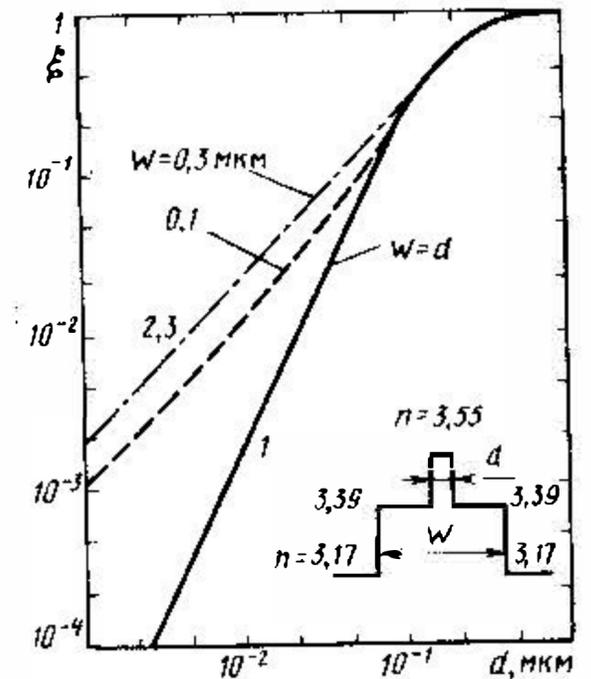


Рис. 9.29. Зависимость параметра оптического ограничения от толщины активного слоя d для ДГС-лазера (1) и РО ДГС-лазеров с различной шириной области оптического ограничения w (2, 3)

электронным и оптическим ограничением в качестве активной области, где происходит накопление и последующая излучательная рекомбинация неравновесных электронов и дырок, используются квантоворазмерные структуры, рассмотренные в § 4.8: квантовые ямы (КЯ) и квантовые точки (КТ). Их применение дает целый ряд существенных преимуществ, связанных прежде всего с особенностями электронного спектра и функции плотности состояний. Ступенчатый вид функции плотности состояний для двумерного электронного газа в квантовой яме (см. рис. 4.57, б) уменьшает тепловое «размытие» в распределении носителей по энергии и облегчает достижения условий инверсии. Коэффициент усиления в максимуме возрастает, а температурная зависимость пороговой плотности тока ослабевает. В еще большей мере это проявляется для квантовых точек, где электронный спектр и функция плотности состояний имеют атомноподобный вид (см. рис. 4.57, г). Если расстояние между дискретными уровнями размерного квантования в квантовых точках меньше kT , то температура вообще не должна влиять на пороговую плотность тока накачки

(при условии, что с увеличением температуры не уменьшается квантовый выход люминесценции). Расчеты показывают, что при выполнении последнего условия температурная зависимость пороговой плотности тока ДГС и РО ДГС-лазеров в области температур $-60 < T < 60$ °С может быть аппроксимирована экспоненциальной функцией

$$I_{\text{пор}} = I_{\text{пор}}(0) \cdot \exp(T/T_0), \quad (9.32)$$

где T_0 — характеристическая температура, равная: $T_0 = 104$ °С для объемной (3D) активной области; $T_0 = 285$ °С для активной области в виде квантовой ямы (КЯ, 2D — электронный газ); $T_0 = 481$ °С для квантовой проволоки (КП, 1D) и $T_0 = \infty$ для квантовых точек (КТ, 0D). Эти зависимости показаны на рис. 9.30.

Таким образом, применение квантоворазмерных структур в активной области РО ДГС-лазеров позволяет: 1) уменьшить пороговую плотность тока накачки; 2) ослабить температурные зависимости; 3) увеличить коэффициент усиления с единицы длины активной области; 4) улучшить спектральные характеристики РО ДГС-лазеров с активной областью в виде одной или нескольких квантовых ям превосходят характеристики обычных ДГС-лазеров. В частности, пороговая плотность тока накачки $J_{\text{пор}}$ при комнатной температуре снижается до значений 100...300 А/см², уменьшаясь почти на порядок по сравнению с ДГС-лазерами.

По ряду причин в активной области гетеролазеров иногда целесообразно использовать квантоворазмерные слои с внутренними напряжениями сжатия или растяжения. При толщинах таких слоев меньше некоторой критической величины в них не образуются дислокации несоответствия и другие дефекты как в самом слое, так и на гетерограницах. Использование напряженных слоев позволит, во-первых, значительно расширить круг материалов, поскольку требования к величине рассогласования параметров решетки $\Delta a/a$, определяемые соотношением (9.1), значительно смягчатся. Во-вторых, варьируя величиной рассогласования Δa , возможно изменять внутренние напряжения в слое и тем самым менять параметры зонной структуры материала, в частности — параметры валентной зоны. Это дает возможность, например, подавить процессы безызлучательной Оже-рекомбинации дырок.

По тем же причинам, которые обсуждались в § 9.6, для улучшения рабочих характеристик лазера необходимо ограничивать размеры активной области в поперечном направлении, т. е. создавать полосковые структуры. Полосковые РО ДГС-лазеры

обладают наименьшими значениями порогового тока накачки, который может составлять всего несколько мА, т. е. быть даже меньше рабочего тока стандартного светодиода.

Для повышения мощности излучения РО ДГС-лазера возможно использовать не один слой, а серию квантоворазмерных структур в пределах одной области D, определяемой условием оптического ограничения. Если число слоев не превышает нескольких (< 6), то состояния в них примерно соответствуют состояниям в одиночной квантовой яме. При больших количествах слоев из разрешенных состояний начинают образовываться минизоны, характерные для одномерных сверхрешеток (сверхструктур). (см. § 4.8 и рис. 4.50). В таких структурах происходит эффективный обмен носителями между отдельными квантовыми ямами за счет туннельного механизма, что облегчает их захват на нижние квантовые уровни E_1 .

В качестве примера на рис. 9.31 приведена лазерная структура, активная область которой содержит GaAs-квантовую яму шириной $d = 100$ Å (10 нм). Волновод имеет сложный профиль и состоит из расположенных по краям ямы короткопериодных сверхрешеток общей толщиной по 0,1 мкм и прилегающих к ним широкозонных слоев $\text{Al}_x\text{Ga}_{1-x}\text{As}$ переменного состава, обеспечивающих заданный профиль волноводного слоя. Применение чередующихся слоев, образующих сверхрешетку, с одной стороны, ускоряет поступление носителей в квантовую яму и, с другой стороны, дает возможность избежать возможных деформаций и связанных с ними дефектов на резкой гетерогранице. В лазерах с такой структурой пороговая плотность тока снижается до 40 А/см².

Правила отбора для межзонных излучательных переходов в квантовых ямах, как говорилось в § 4.8, устанавливают $\Delta l = 0$.

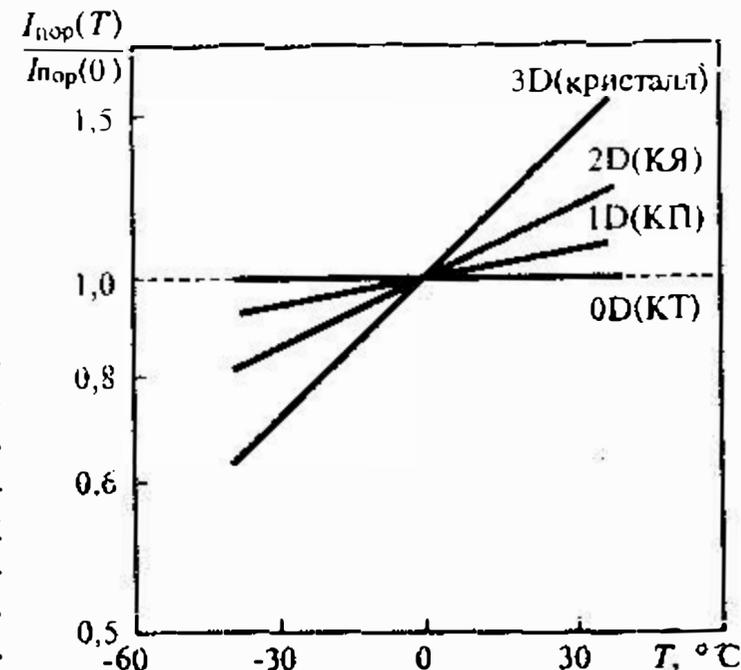


Рис. 9.30. Температурная зависимость пороговой плотности тока, нормированной к 0 °С, для ДГС-лазеров с активной областью в виде: толстого эпитаксиального слоя 3D; квантовой ямы 2D; квантовой проволоки 1D и квантовой точки 0D. По шкале ординат применен логарифмический масштаб

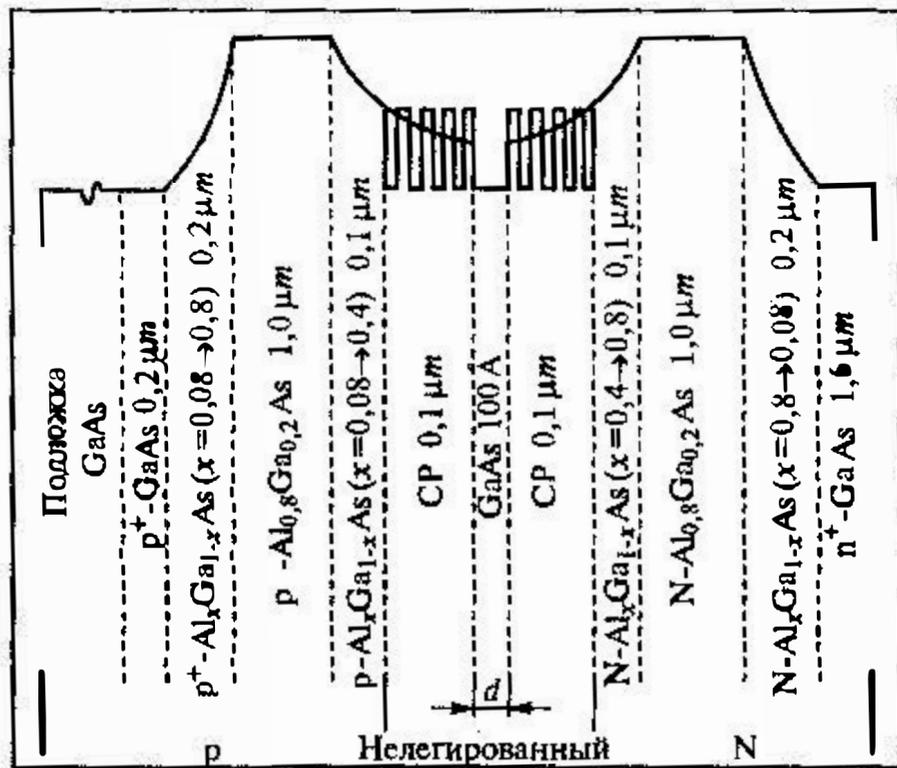


Рис. 9.31. Лазерная РО ДГС-структура с GaAs-квантовой ямой и параболлическим профилем волновода, сформированным короткопериодной сверхрешеткой и твердым раствором $\text{Al}_x\text{Ga}_{1-x}\text{As}$ переменного состава

Это означает, что с испусканием фотона могут рекомбинировать электроны из состояния E_1 ($n=1$) с дырками $E_1^{V_1}$ ($1hh$) и $E_1^{V_2}$ ($1lh$) как показано на рис. 9.32 или электроны из состояния E_2 с дырками $E_2^{V_1}$ ($2hh$), $E_2^{V_2}$ ($2lh$) и т. п. Строгость выполнения правил отбора нарушается как дефектами и напряжениями в самом материале активного слоя, так и неидеальностью структуры.

Необходимость изготовления сверхтонких слоев для создания активных структур РО ДГС-лазеров предъявляет исключительно высокие требования как к технологической, так и к измерительной культуре при их изготовлении. По существу необходимо выращивать и контролировать структуры с точностью до одного атомного слоя. Для осуществления этих фантастических задач были разработаны в высшей степени уточненные методы выращивания полупроводниковых твердых растворов группы $A^{III}B^V$. Для этого используются три метода: 1) молекулярно-пучковая эпитаксия (МПЭ), 2) газофазная эпитаксия из металлоорганических соединений (МО ГФЭ) и 3) жидкофазная эпитаксия (ЖФЭ). Первые два метода дают возможность получать гетероструктуры, в которых толщина переходной области на гетерогранице находится в пределах одного монослоя, а толщина самой активной — несколько молекулярных слоев, т. е. несколько постоянных решетки. Типичные размеры активной области РО ДГС-лазеров $d=(5...20)$ нм.

Необходимость изготовления сверхтонких слоев для создания активных структур РО ДГС-лазеров предъявляет исключительно высокие требования как к технологической, так и к измерительной культуре при их изготовлении. По существу необходимо выращивать и контролировать структуры с точностью до одного атомного слоя. Для осуществления этих фантастических задач были разработаны в высшей степени уточненные методы выращивания полупроводниковых твердых растворов группы $A^{III}B^V$.

Для этого используются три метода: 1) молекулярно-пучковая эпитаксия (МПЭ), 2) газофазная эпитаксия из металлоорганических соединений (МО ГФЭ) и 3) жидкофазная эпитаксия (ЖФЭ). Первые два метода дают возможность получать гетероструктуры, в которых толщина переходной области на гетерогранице находится в пределах одного монослоя, а толщина самой активной области — нескольких молекулярных слоев, т. е. нескольких постоянных решетки. Типичные размеры активной области РО ДГС-лазеров $d=(5...20)$ нм.

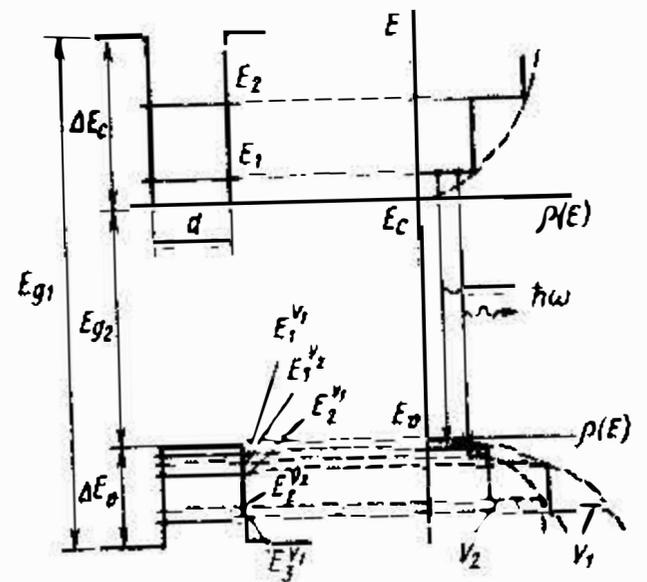


Рис. 9.32. Межзонные оптические переходы в структуре с квантовой ямой

§ 9.8. ГЕТЕРОЛАЗЕРЫ С РАСПРЕДЕЛЕННОЙ ОБРАТНОЙ СВЯЗЬЮ

Во всех рассмотренных выше полупроводниковых лазерах обратная связь обеспечивалась за счет отражения на торцевых гранях резонатора, сформированного скалыванием или оптической полировкой. Рассматривая различные типы резонаторов в § 2.3, мы указали на возможность использования распределенной обратной связи (рис. 2.26). Упрощенная схема ДГС-лазерного диода с распределенной обратной связью показана на рис. 9.33. Одна из гетерограниц делается гофрированной, что приводит к интерференционным эффектам. Шаг гофрировки b для решетки, работающей в первом порядке, должен быть $b=\lambda/(2n)$, т. е. около 0,13 мкм для арсенид-галлиевого лазера, изображенного на рис. 9.33. Решетки с таким малым периодом создаются в пленке фоторезиста при помощи интерференции лазерных пучков, излучаемых одним из рассмотренных в гл. 7 ко-

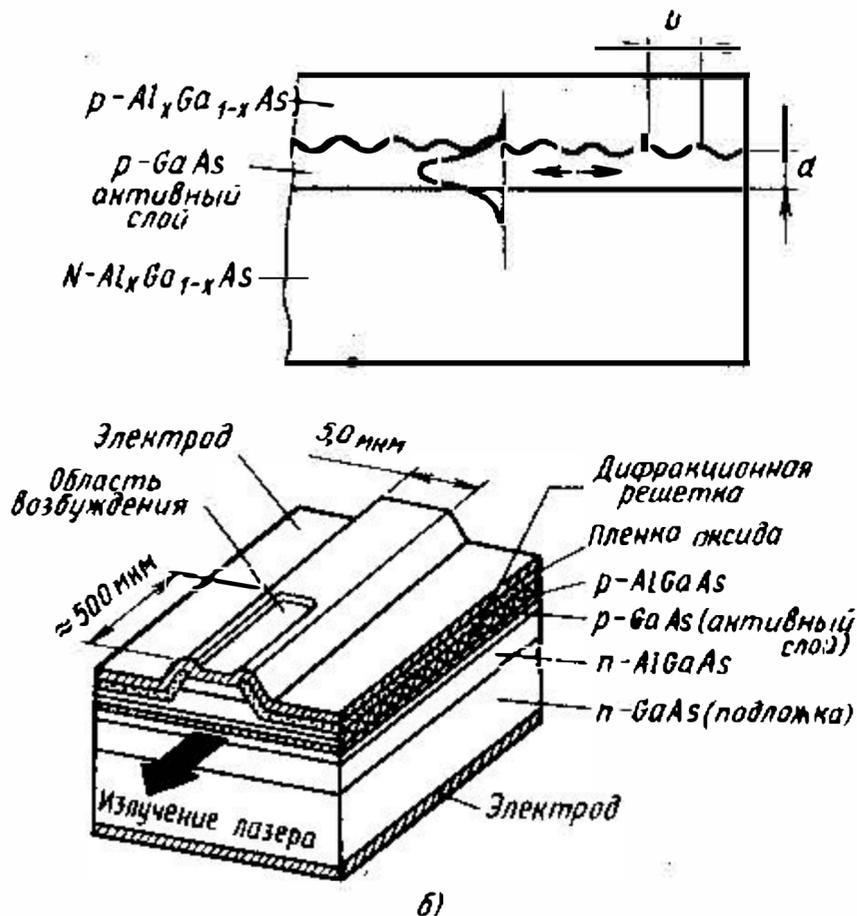


Рис. 9.33. Инжекционный лазер с распределенной обратной связью:
 а — распространение электромагнитной волны; б — общий вид

ротковолновых газовых лазеров. Затем эти решетки могут быть перенесены на рабочую поверхность с помощью химического или ионного травления.

Если в структуре, изображенной на рис. 9.33, а, шаг гофрировки сделать не 0,13, а 0,39 мкм, то вывод излучения, генерируемого в активном слое, может быть осуществлен под углом 30° к плоскости волновода, как это показано на рис. 2.26.

Одним из достоинств лазера с распределенной обратной связью является температурная стабильность его спектральных характеристик и простота работы в одномодовом режиме. Слабая температурная зависимость объясняется тем, что в обыч-

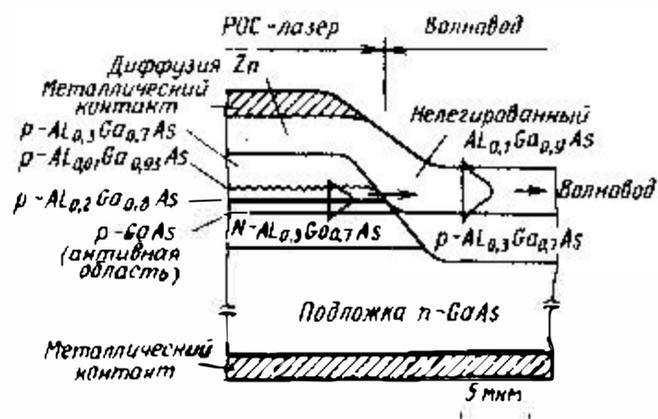


Рис. 9.34. Ввод излучения в оптический волновод в лазере с распределенной обратной связью

ном гетеролазере основным фактором, приводящим к сдвигу спектра излучения, является температурная зависимость ширины запрещенной зоны, а в лазере с распределенной обратной связью — более слабая температурная зависимость показателя преломления.

Естественно, что для получения оптимальной связи должна быть подобрана толщина активного слоя d . Для улучшения характеристик лазера активная область может быть выполнена из нескольких слоев, в том числе с использованием отдельного электронного и оптического ограничений. РОДГС-лазеры с распределенной обратной связью (РОДГС РОС-лазер) дают возможность осуществить эффективный ввод излучения в оптический волновод, как это показано на рис. 9.34, что важно для устройств интегральной оптики.

§ 9.9. РАБОЧИЕ ХАРАКТЕРИСТИКИ ИНЖЕКЦИОННЫХ ЛАЗЕРОВ

Высокая эффективность, малые габариты, возможность непосредственного питания от низковольтных источников тока, мгновенная готовность к работе, высокое быстродействие, экономичность, надежность, долговечность и сравнительно низкая стоимость — все эти качества потенциально заложены в инжекционных лазерах и в своем большинстве уже реализованы в промышленных образцах. Это определяет их широкое практическое применение в таких областях, как волоконно-оптическая связь, оптические системы записи и считывания информации, включая видеодисковую аппаратуру, знакопечатающие устройства, датчики и системы контроля окружающей среды, интегральная оптика, спектроскопия и т. п. Успехи в разработке мощных инжекционных лазеров, излучающих в непрерывном режиме при комнатной температуре мощности от долей до единиц Вт, дополнительно расширяют их области применения, так что в ряде случаев инжекционные лазеры начинают вытеснять лазеры других типов, в частности — гелий-неоновые газовые лазеры. Возможность варьирования характеристик инжекционных лазеров в широких пределах и многообразие областей их использования привели к необходимости специализации лазерных конструкций и материалов для определенных их применений. Освоен выпуск не только целого ряда разновидностей полупроводниковых лазеров, но и более сложных специализированных модулей, включающих такие лазеры.

Наибольшая доля промышленного выпуска приходится на инжекционные лазеры на основе гетероструктур в системах $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ и $\text{Ga}_x\text{In}_{1-x}\text{As}_{1-y}\text{P}_y/\text{InP}$, перекрывающих диапазон 0,67...1,67 мкм. Из них подавляющее большинство приходится на три диапазона длин волн, соответствующие трем областям

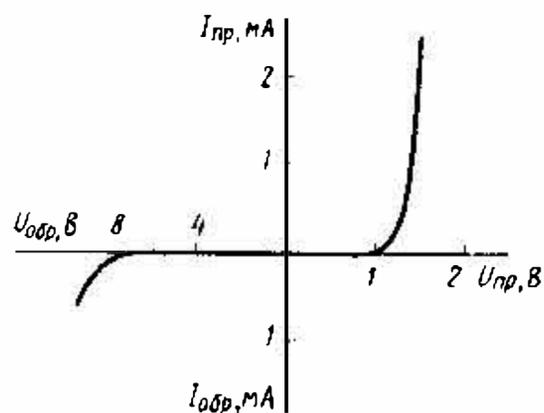


Рис. 9.35. Типичная вольт-амперная характеристика $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}/\text{GaAs}$ -ДГС-лазера с широким контактом

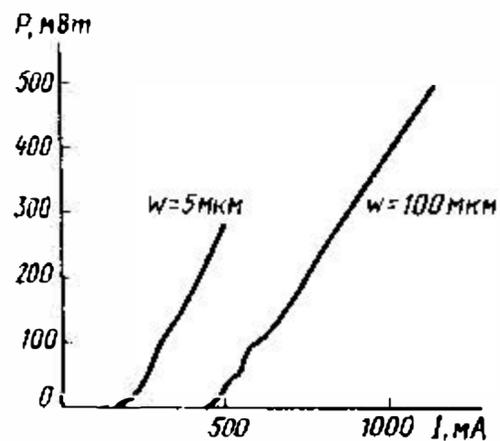


Рис. 9.36. Зависимость выходной мощности излучения $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}/\text{GaAs}$ -полоскового ДГС-лазера от тока накачки для двух значений ширины полоска

пропускания оптических волокон (см. § 12.3): 0,84 мкм (активная среда GaAs или $\text{Al}_{0,05}\text{Ga}_{0,95}\text{As}$); 1,3 мкм (активная среда $\text{In}_{0,72}\text{Ga}_{0,28}\text{As}_{0,6}\text{P}_{0,4}$) и 1,55 мкм (активная среда $\text{In}_{0,60}\text{Ga}_{0,40}\text{As}_{0,88}\text{P}_{0,12}$). Большинство лазерных структур выполняются в виде полосковых ДГС или полосковых РОДГС.

Суммируем типичные рабочие характеристики лазеров, рассмотренных в предыдущих параграфах.

Электрические характеристики, выходная мощность и спектры излучения определяются особенностями конструкции и материалом активной области. Типичная вольт-амперная характеристика ДГС-лазера с широким контактом на основе $\text{GaAs} - \text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ приведена на рис. 9.35. Напряжение, при котором ток равен 1 мА, обычно составляет 1,1...1,2 В. Если при таком напряжении ток прямого смещения значительно превосходит 1 мА, то это свидетельствует о наличии дополнительных паразитных каналов избыточного тока, приводящих к увеличению $J_{\text{пр}}$. Напряжение пробоя при обратном смещении зависит от уровня легирования. Характерное значение — около 6 В. Ток в прямом направлении при его изменении от 10^{-10} до 10^{-3} А хорошо описывается известным соотношением $I_{\text{пр}} \sim \exp(eU/2kT)$. При больших токах на вольт-амперную характеристику начинают влиять разогрев и падение напряжения на контактах и в приконтактных областях.

Ватт-амперные характеристики и типичные спектры излучения полоскового $\text{GaAs} - \text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ -ДГС-лазера приведены на рис. 9.36 и 9.37. В лазерах с широким контактом или с большой шириной полоски, как обсуждалось в § 9.6, в зависимости выходной мощности от тока накачки могут наблюдаться характерные изломы, вызванные нестабильностью работы. В полосковых лазерах с узким контактом эти неприятности отсутствуют.

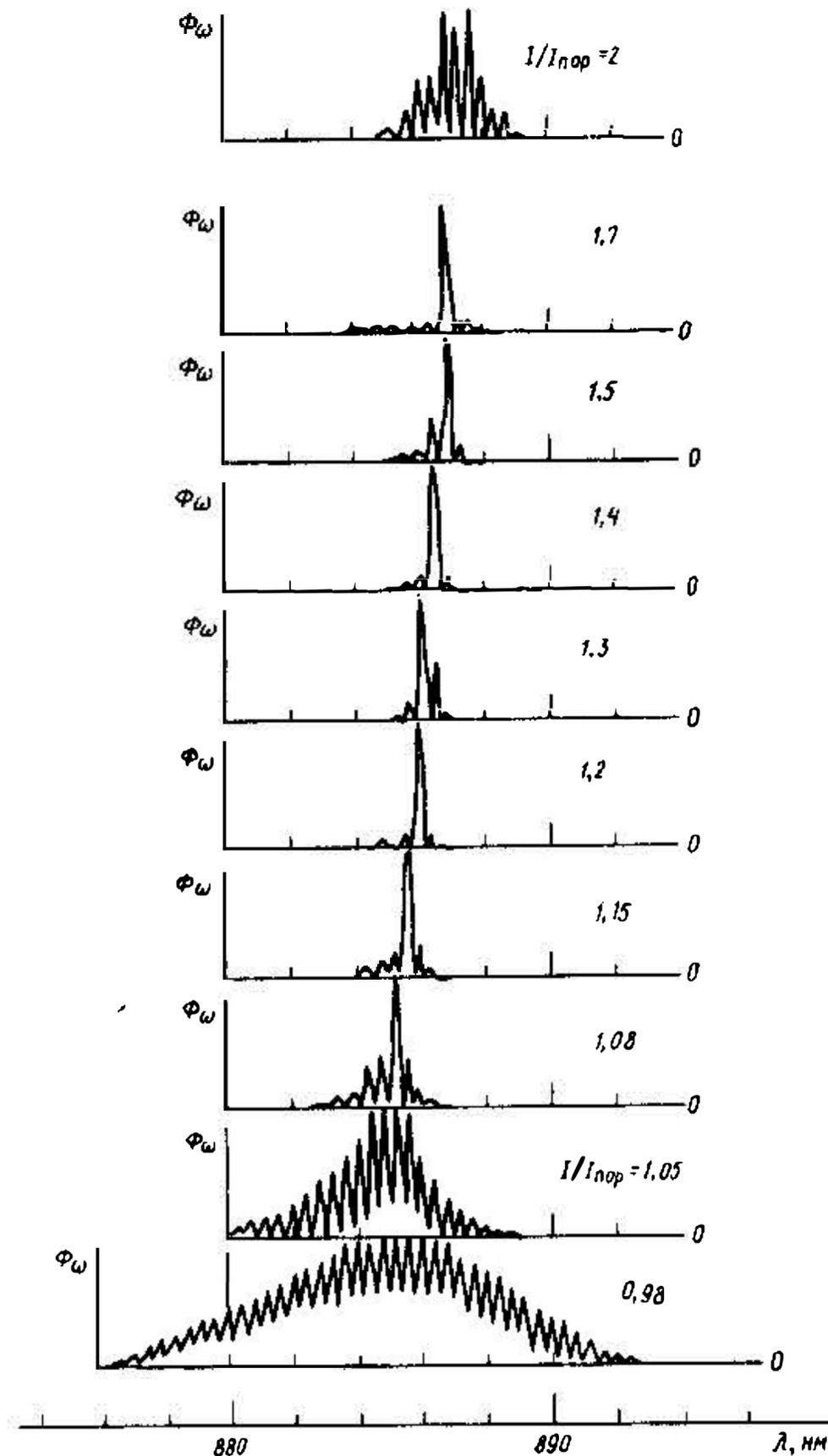


Рис. 9.37. Изменение спектра излучения $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}/\text{GaAs}$ — полоскового ДГС-лазера с изменением тока накачки. Ширина полоска 0,3 мкм; $I_{\text{пор}} = 50$ мА

Спектр излучения при токах ниже порогового представляет собой сравнительно широкую полосу, обусловленную межзонными или квазимежзонными спонтанными переходами в активной области гетеролазера. При приближении к порогу в спектре излучения возникает модовая структура, которая при токе вблизи порогового проявляется очень отчетливо (рис. 9.37). Этот режим работы называется *сверхлюминесценцией* или *суперлюминесценцией*: в активной области создана инверсия населенности, но ее еще не хватает для начала генерации, т. е. еще не выполнено условие самовозбуждения (9.23). При $J > J_{\text{пор}}$ спектр излучения сужается, увеличивается направленность и когерентность излучения в соответствии с общими закономерностями, обсуждавшимися в гл. 2. При изменении плотности тока накачки возможен перескок с одной моды на другую. Из рис. 9.19 видно, что при увеличении тока накачки максимум спектральной зависимости показателя усиле-

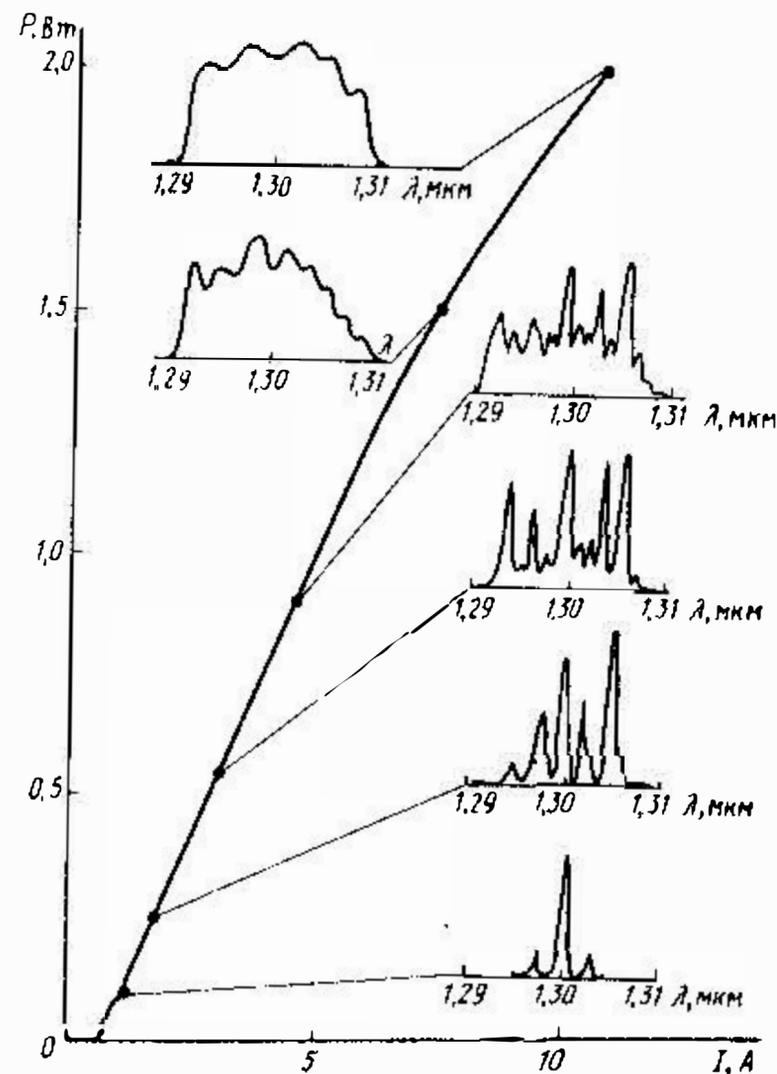


Рис. 9.38. Ватт-амперная характеристика полоскового РО ДГС-лазера на GaInAsP/InP. В разных точках характеристики приведены спектры излучения

ния сдвигается в коротковолновую область. Одновременно при больших J происходит нагрев кристалла, за счет чего спектр сдвигается в длинноволновую область. Часто второй эффект преобладает над первым, что видно из рис. 9.37.

На рис. 9.38 приведена ватт-амперная характеристика полоскового лазера с отдельным электронным и оптическим ограничением на гетероструктуре InGaAsP/InP и спектры излучения в разных ее точках. Если в активной области лазера имеются факторы, приводящие к неоднородному уширению спектральной линии люминесценции (например, неоднородность состава по площади активного слоя), то даже в полосковом РОДГС-лазере велика вероятность генерации в многомодовом режиме с участием нескольких продольных мод.

Наибольшей чистотой спектра среди инжекционных лазеров обладает РОДГС РОС-лазер. Дифракционная решетка, обеспечивающая распределенную обратную связь, одновременно проводит эффективную селекцию по частоте. Изменяя в незначительных пределах шаг гофра, можно менять рабочую частоту лазера.

Направленность излучения. Диаграмма направленности инжекционного лазера, как и лазеров других типов, определяется факторами, обсуждавшимися в § 2.6. Для основной моды угол расходимости дается соотношением (2.77). Так как поперечные размеры активной области в полупроводниковых лазерах очень малы, то дифракционная расходимость будет велика. В плоскости, перпендикулярной активному слою, она будет определяться толщиной области оптического ограничения D , а в плоскости p - n -перехода — шириной области генерации или шириной полоски W в полосковом лазере. Так как $D < W$, то угол расходимости излучения в плоскости p - n -перехода $\theta_{\perp} \approx \lambda/W$ будет меньше, чем в перпендикулярной плоскости, где $\theta_{\parallel} \approx \lambda/D$, как показано на рис. 9.39. Картина излучения полоскового гетеролазера, работающего в одномодовом режиме, в дальней зоне представляет собой эллипс, расширяющийся в направлении, перпендикулярном поверхности слоя (рис. 9.39). Типичные значения $\theta_{\perp} = 30 \dots 60^\circ$ и $\theta_{\parallel} = 10 \dots 30^\circ$. Для его фокусировки могут быть использованы цилиндрические линзы.

Увеличение ширины полоски не всегда приводит к сужению диаграммы направленности θ_{\perp} . Как отмечалось в § 9.6, увеличение ширины полоски больше критического значения способствует возникновению многомодовой структуры излучения. Это иллю-

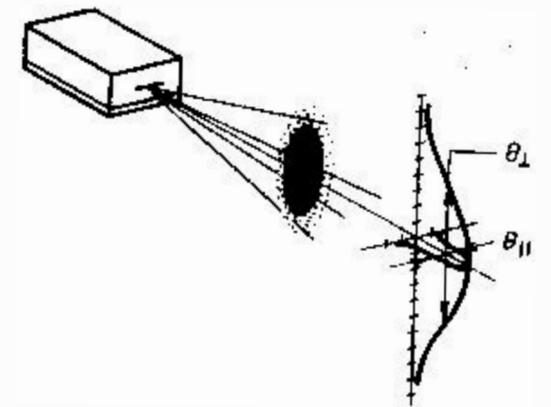


Рис. 9.39. Диаграмма направленности и картина дальнего поля полоскового ДГС-лазера

стрирует рис. 9.40, где представлены экспериментальные данные, наглядно свидетельствующие об изменении модовой структуры излучения лазера при увеличении ширины полоски. Эти моды характерны для распределения (2.39). Весьма полезно сравнить эти данные с приведенными на рис. 2.16 и 2.17. Регулярная модовая структура вдоль плоскости p - n -перехода при широких W встречается не всегда. Часто на нее накладываются эффекты, связанные со шнурованием и генерацией в каналах.

Температурные зависимости. Наиболее сильно температура влияет на пороговый ток и мощность излучения, что вызвано в основном двумя причинами. Во-первых, с увеличением температуры происходит «размазывание» носителей по энергии. Для РОДГС-лазеров с квантовыми ямами этот фактор уменьшен за счет ступенчатой функции плотности состояний (рис. 9.31, б), благодаря чему эти лазеры способны работать при повышенных температурах. Во-вторых, с увеличением температуры обычно происходит снижение эффективности люминесценции в активной области (температурное гашение люминесценции), приводящее к уменьшению квантового выхода и времени жизни неравновесных носителей заряда в активном слое. Поэтому в полупроводниковых лазерах следует уделять самое серьезное внимание теплоотводу.

Обычно температурная зависимость пороговой плотности тока в области $T > 250$ К приближенно описывается экспоненциальной функцией вида (9.32)

Температура, при которой получена непрерывная генерация на длине волны 0,84 мкм, достигает 200 °С, а на длине волны 1,3 мкм — 100 °С.

Изменение ширины запрещенной зоны с температурой вызывает сдвиг спектра лазерного излучения. Если лазер работает в многомодовом режиме, то увеличение температуры приводит к плавному смещению огибающей спектра в длинноволновую область в соответствии с величиной $\partial E_g / \partial T$ материала активного слоя. Если лазер работает в одномодовом режиме, то увеличение температуры будет приводить к скачкообразным изменениям $\lambda_{\text{ген}}$, определяемым перескоком с одной моды на другую при изменении E_g . Наилуч-

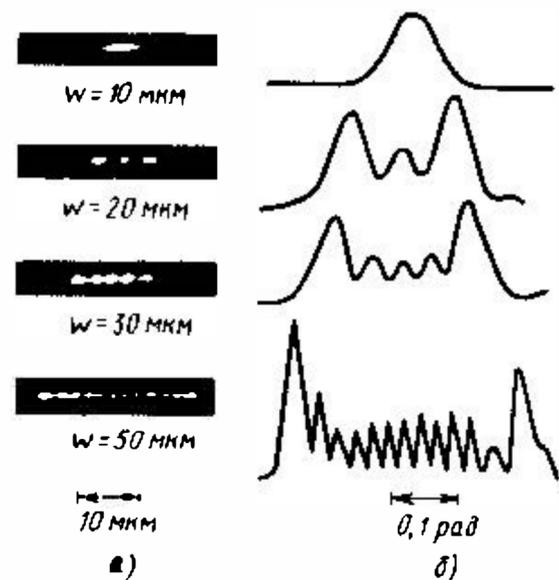


Рис. 9.40. Изменение модовой структуры излучения при изменении ширины полоски ДГС-лазера: а) фотография структуры поля на выходном зеркале; б) — распределение интенсивности

шей температурной стабильностью спектра обладают РОДГС РОС-лазеры.

Модуляционные и частотные характеристики. Одним из существенных достоинств инжекционных лазеров для применения в линиях связи и передачи информации является простота их амплитудной модуляции, которая осуществляется простой модуляцией рабочего тока.

Если на лазер подать прямоугольный токовый импульс с временем нарастания меньше времени жизни τ носителей в активном слое, то лазерный импульс будет задержан на время $t_{\text{зд}} = t_2 - t_1$ и на нем будут наблюдаться экспоненциально затухающие релаксационные колебания, как это показано на рис. 9.41. Причина их возникновения та же, что обсуждалась в § 2.5 (кривая 3 на рис. 2.27). Время задержки светового импульса связано с временем жизни τ_n неравновесных носителей и током накачки как

$$t_{\text{зд}} = t_2 - t_1 = \tau_n \ln \left(\frac{I}{I - I_{\text{пор}}} \right), \quad (9.33)$$

что нетрудно получить из решения уравнения непрерывности, например, для электронов в активной p -области. Частота релаксационных колебаний примерно совпадает с характерной резонансной частотой f_p и имеет порядок ГГц. В первом приближении f_p может быть оценена из

$$f_p = \frac{1}{2\pi} \left[\frac{1}{\tau_\phi \tau_n} \left(\frac{I}{I_{\text{пор}}} - 1 \right) \right]^{1/2}, \quad (9.34)$$

где τ_ϕ — время жизни фотона в резонаторе.

Релаксационные колебания представляют нежелательное явление для высокочастотной амплитудной модуляции излучения инжекционного лазера. Уменьшить это явление, как и время задержки t_3 , можно, осуществляя подпитку лазера постоянным током, равным $I_{\text{пор}}$. Но наилучший способ борьбы — применение плосковых лазеров с узким полоском. При ширине полоски порядка диффузионной длины ($L_n \approx 2$ мкм) эти колебания полностью исчезают, что объясняется компенсацией флуктуаций концентрации носителей в середине

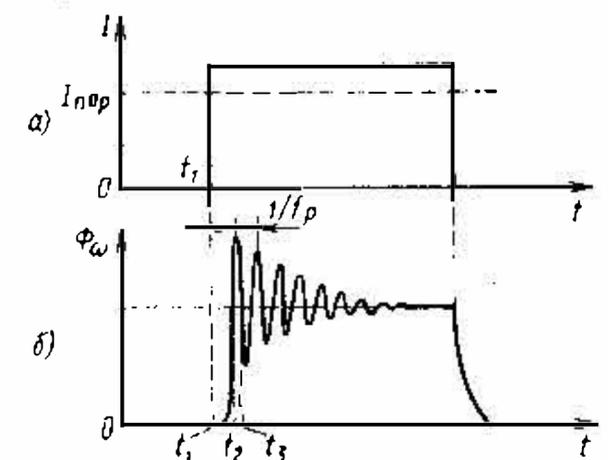


Рис. 9.41. Импульсная модуляция излучения инжекционного лазера: а) токовый импульс; б) оптический отклик (лазерное излучение)

активной области за счет диффузии из периферийных областей полосы.

Типичные частотные зависимости мощности излучения полоскового инжекционного лазера показаны на рис. 9.42. На частоте $f \approx f_p$ наблюдается резонансный максимум, после чего при $f > f_p$ выход резко падает. Величина f_p , как видно из рис. 9.42 и соотношения (9.34), зависит от тока. Значение f_p служит пределом верхней частоты токовой модуляции инжекционного лазера.

При высокочастотной модуляции возможно так называемое динамическое уширение спектра излучения, вызванное возрастанием числа генерируемых продольных мод. Это уширение $\Delta\lambda$ в AlGaAs/GaAs полосковых лазерах имеет порядок 2 нм.

Деградация. Полупроводниковые лазеры работают в экстремальном режиме. В активной области плотность энергии электромагнитного поля очень велика. Например, для лазера, излучающего в непрерывном режиме 10 мВт, в активной области с типичными размерами $(300 \times 0,2 \times 3)$ мкм³ плотность световой энергии ρ_E достигает 10^8 Вт/см³. Естественно, что в таких условиях возможны различные необратимые процессы, приводящие к ухудшению рабочих характеристик лазера во время его работы, т. е. к деградации. Деградация в основном приводит к увеличению порогового тока накачки и снижению мощности излучения.

Причины, приводящие к деградации инжекционного лазера, могут быть те же, что и светодиода (§ 9.2). Но в силу экстремальных условий работы лазерный диод еще более чувствителен как к качеству материалов активной и прилегающей к ней областей, так и к тщательности изготовления прибора и его эксплуатации.

Выделяют три основных типа деградации инжекционных лазеров:

1. **Катастрофическое разрушение.** Оно может быть вызвано либо наличием дефекта в активной области и связанным с ним лавинообразным разрушением материала, либо повреждением зеркал, либо оптической прочностью материала (см. § 8.3). Последняя ограничивает пиковую мощность, которую можно получить от инжекционного лазера. В GaAs-лазерах предельные световые потоки составляют $(1...2)$ МВт/см². Для GaInAsP/InP-лазеров они

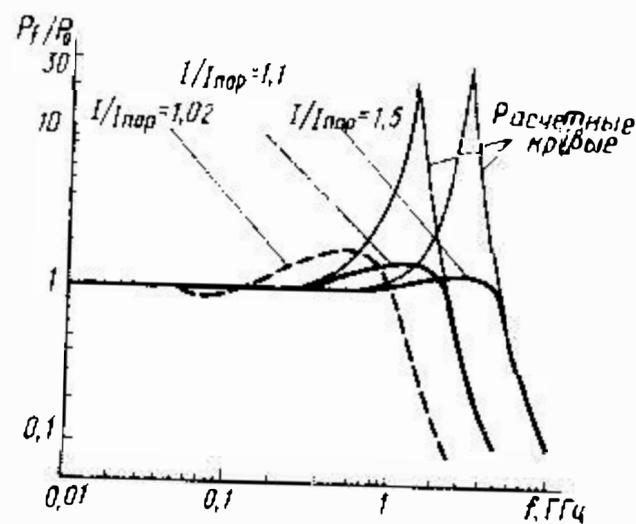


Рис. 9.42. Типичные зависимости нормированной мощности излучения от частоты модуляции ДГС-лазера для различных уровней возбуждения

в 3...5 раз выше в связи с более высокой оптической прочностью материала.

2. **Образование так называемых «дефектов темных линий».** Обычно этот тип деградации связан с присутствием исходных дислокаций в материале и гетероструктуре, в частности дислокаций несоответствия на гетерогранице. В процессе работы из этих единичных дислокаций могут формироваться дефекты темных линий, представляющие собой сетку дислокаций. Этот вид деградации сильнее выражен в GaAlAs/GaAs-лазерах, чем в GaInAsP/InP-приборах, в силу рассмотренных в §§ 4.2 и 9.1 преимуществ этого четырехкомпонентного твердого раствора.

3. **Постепенная деградация.** Если устранить факторы, приводящие к первым двум типам деградации, то время деградации $\tau_{дет}$ в (9.14) для некоторых лабораторных образцов инжекционных лазеров, подвергнутых испытанию, достигает 10^5 часов непрерывной работы и более. Экстраполяция данных, полученных для повышенных температур работы, к комнатной температуре по соотношению (9.15) показывает, что $\tau_{дет}$ может превышать 10^6 часов (более 100 лет). Это доказывает, что принципиальных ограничений срока службы инжекционных лазеров нет, а возможные отказы вызваны несовершенством технологии, дефектами конструкции, неправильной эксплуатацией и т. п.

§ 9.10. МОЩНЫЕ ИНЖЕКЦИОННЫЕ ЛАЗЕРЫ, ЛАЗЕРНЫЕ ЛИНЕЙКИ И РЕШЕТКИ

Разделение полупроводниковых лазеров на маломощные и мощные весьма условно и зависит от многих факторов, в том числе — от области применения, типа лазера, спектрального диапазона и режима работы. Принято называть мощными полупроводниковые лазеры, генерирующие в одночастотном (одномодовом) непрерывном режиме мощности более 50 — 100 мВт, а в многомодовом непрерывном режиме более 0,5...1 Вт.

Фундаментальным фактором, ограничивающим возможность получения больших мощностей в инжекционных гетеролазерах, является предельно допустимая плотность электромагнитной энергии внутри активной области. Если напряженность электрического поля, создаваемого световой волной, будет сравнима с внутриатомными полями, то наступит разрыв межатомных и межмолекулярных связей, что неминуемо приведет к разрушению — явлению, обсуждавшемуся в гл. 5 и известному как «оптический пробой». В конечном итоге именно этот фактор является тем пределом, обойти который вряд ли удастся. Оценим объемную плотность энергии электромагнитного излучения ρ , которая создает в световой волне напряженность электрического поля E порядка 10^7 В/м. Эта величина выбрана из соображений,

что она не должна превышать одной десятой от среднего значения внутрикристаллического поля $E_{кр}$ в активной области. Оценка, сделанная для арсенида галлия, дает $E_{кр} = 2 \cdot 10^8$ В/м, что согласуется с величиной критического поля в кремнии, при котором наступает туннельный пробой.

Напряженность поля E в световой волне связана с объемной плотностью электромагнитной энергии ρ и величиной вектора Пойнтинга S , определяющего среднюю по времени плотность мощности излучения, соотношениями (1.12), (1.25), которые с учетом ϵ и n перепишутся в виду

$$|S| = \rho \cdot v = \epsilon \epsilon_0 E^2 \frac{c}{n}, \quad (9.35)$$

откуда для арсенида галлия при $E \approx 10^7$ В/м получается $I = |S| \approx \approx 10^{12}$ Вт/м² ≈ 1 Вт/мкм². Это означает, что предельно допустимый съем энергии с полоскового ДГС-лазера с поперечными размерами активной области (волноводного слоя) $H \times W = 0,2$ мкм \times 5 мкм, ограниченный лучевой объемной стойкостью, составляет ~ 1 Вт.

Сделанная оценка соответствует идеально однородному распределению поля внутри активной области, что, естественно, никогда не реализуется. Неоднородности могут быть весьма существенны даже в идеально однородном активном слое, что связано с гауссовым характером распределения поля даже в идеальном случае ТЕМ₀₀-моды (см. 2.39, а); эффектами пространственного «выжигания дыр» (см. § 2.4); интерференционными эффектами, связанными с взаимодействием мод высокого порядка, что характерно для работы лазера при токах (уровнях) накачки, существенно превышающих пороговые (см. § 9.9); эффектами «шнурования» тока накачки; неоднородным распределением поля в плоскости $H \times W$, особенно по направлению W вдоль «полоска» и эффектами «оптического шнурования».

Наконец, на самих гетерограницах, формирующих активную область, уже существует встроенное электрическое поле, которое в разных структурах может составлять $10^3 \dots 10^8$ В/м. В неблагоприятных случаях это поле, складываясь с полем световой волны, может привести к локальному оптическому пробую. По этой причине желательно использовать однородный рабочий объем, в котором отсутствуют внутренние локальные поля, хотя это может привести к возрастанию пороговой плотности тока. Впрочем, надо учитывать и другой фактор — потери в активной области, в том числе — поглощение на свободных носителях заряда. Необходимость уменьшения этих потерь приводит к необходимости применения структур с отдельным оптическим

и электронным ограничением, содержащих в активной области одну или несколько квантовых ям (см. §§ 4.8 и 9.7).

Перечисленные выше факторы были учтены в коэффициенте 0,1 при оценке предельного значения E , вызывающего оптический пробой в гетеролазере с идеальной активной областью, не содержащей никаких дефектов и неоднородностей. В реальном лазере, работающем в многомодовом режиме, необходимо ввести коэффициент «запаса прочности» как минимум 2, так что напряженность поля в световой волне не должна превышать $(0,4 \dots 0,6) \cdot 10^7$ В/м, что соответствует средней плотности световой мощности на выходе лазерного диода на уровне 0,2 Вт/мкм². Это означает, что для лазера с активной областью $W \times H = = 100 \times 1$ мкм² выходная мощность ограничивается 20 Вт как разумный предел, превышение которого нецелесообразно в связи с неизбежными необратимыми изменениями в объеме активного материала и на гетерограницах. Отметим, что самой «опасной» гетерограницей, на которой, как правило, сосредоточены самые сильные внутренние поля, является торцевая поверхность, через которую в обычном лазере выводится излучение, или на которое нанесено глухое зеркало. По этой причине торцевые поверхности в мощных лазерах подвергаются специальной обработке — пассивации, а на выходную грань, через которую выводится лазерное излучение, наносится антиотражающее покрытие, уменьшающее ее коэффициент отражения до $R \approx 0,05$ и пропускающее 95% световой мощности. Это, естественно, увеличивает пороговую плотность тока, но при $I > 2 I_{пор}$ позволяет получить максимальный съем энергии, что в конечном итоге и надо в мощном лазере.

В AlGaAs/GaAs-лазерах с открытыми сколотыми гранями критическая плотность оптической мощности составляет $(1 \dots 5)$ МВт/см². Пассивация с последующей защитой поверхности и нанесением антиотражающих покрытий приводит к ее возрастанию почти на порядок, достигая $(10 \dots 30)$ МВт/см², т. е. $(0,1 \dots 0,3)$ Вт/мкм². Эта величина всего в полтора-два раза ниже предельных значений, приведенных выше. Она близка к лучевой стойкости иттрий-алюминиевого граната и превышает лучевую стойкость стекла с ниодимом (см. § 8.3).

Многомодовая генерация значительно снижает среднюю плотность энергии, которую можно получить во всех типах лазеров. Такой режим работы характерен для инжекционных лазеров с шириной активной области $W \gg \lambda/n$ при большом превышении тока накачки над пороговым (см. рис. 9.37, 9.38 и 9.40). Возникающие при многомодовой генерации локальные области с большой концентрацией энергии способствует развитию необратимых процессов в рабочих структурах. В мощных инжекционных лазерах с шириной полоска более 50...100 мкм,

работающих при многократном превышении порогового тока накачки, избавиться от многочастотной и многомодовой генерации чрезвычайно сложно.

Важной причиной, ограничивающей мощность излучения инжекционных лазеров, является тепловой нагрев. Особенно опасен и нежелателен локальный нагрев, вызванный различного рода неоднородностями и приводящий к быстрой деградации лазера. Единственным механизмом потерь, от которого принципиально не избавиться, является поглощение на свободных носителях заряда с последующей внутризонной релаксацией, сопровождающейся передачей энергии решетке, т. е. нагревом. Все остальные виды потерь, включая омические потери на контактах и прилегающих областях должны быть сведены к минимуму. В современных мощных лазерных диодах последовательное омическое сопротивление не превышает 0,01...0,05 Ом. Поскольку рабочие токи мощных лазеров составляют единицы и десятки ампер через активную область площадью доли мм², то очень жесткие требования предъявляются к теплоотводу. Активный лазерный элемент напаяется на массивный держатель, выполняющий роль теплообменника. При генерации мощности более 2 Вт в непрерывном режиме, как правило, проводится дополнительное охлаждение с помощью термоэлементов Пельтье. Признаком того, что именно тепловой нагрев ограничивает мощность излучения, является насыщение и даже спад выходной мощности с возрастанием тока накачки. Причин такого поведения может быть несколько, но в большинстве случаев это связано с тепловым выбросом носителей заряда из активной области или с падением квантового выхода излучательной рекомбинацией.

В мощном полупроводниковом лазере с шириной активной области W , существенно превышающей длину волны, электрическое поле в световой волне $E(x, y, z, t)$ формируется как суперпозиция многих боковых (латеральных, зависящих от координаты x), поперечных (зависящих от y) и продольных (зависящих от z) мод, каждая из которых имеет различные собственные частоты. Полный количественный анализ ситуации в этом случае невозможен и обычно ограничиваются рассмотрением генерации на одной продольной моде с собственной частотой ω_0 . При таком приближении генерация осуществляется преимущественно в ТЕ-поляризации. Для этой поляризации максимальны как коэффициент усиления, так и коэффициент отражения. Для ТЕ-моды x -компонента электрического поля в световой волне, распространяющейся вдоль направления z , записывается в виде:

$$E_x(x, y, z, t) = E(x) F(x, y) \exp i(\beta z - \omega_0 t), \quad (9.36)$$

где β — комплексная постоянная распространения, связанная с комплексным эффективным показателем преломления η и волновым вектором $k_0 = 2\pi/\lambda_0$ как $\beta = k_0 \cdot \eta$. С точки зрения получения излучения высокого качества в соотношении (9.36) нас интересует зависимость $E(x)$, которая определяет латеральную моду, и зависимость $F(x, y)$, которая определяет поперечную моду и обычно является слабо меняющейся функцией координаты x . Эти функции для полоскового лазера с $W \gg \lambda$ и $H \approx \lambda$ представлены на рис. 9.43 и весьма полезны для анализа. В частности, приведенные результаты расчета показывают, что даже в идеальном, модельном случае при генерации на одной моде распределение поля внутри активной области мощного лазера с широким полоском далеко неоднородно.

Представленные на рис. 9.43 данные иллюстрируют те предельные возможности по пространственной однородности и когерентности излучения, которые можно получить от полупроводникового лазера с широкой активной областью. На практике эти неоднородности существенно больше и определяются в основном эффектами шнурования, которые проявляются в расщеплении высокоапертурного луча на ряд узких шнуров. Это неприятное явление связано с нелинейными процессами насыщения усиления, рассмотренными в § 2.4, и их влиянием на эффективный показатель преломления n_{eff} . В результате в активной области происходит самофокусировка лазерного излучения, которая сопровождается эффектами «пространственного выжигания дыр».

При заметном превышении порога, когда проявляются эффекты насыщения, оптический пучок большой мощности, генерируемый в слое толщиной больше 10 мкм, склонен разрываться на отдельные шнуры с характерными размерами 5 — 10 мкм по ширине волноводного слоя. В результате наблюдается модуляция (флуктуация) интенсивности, которая сопровождается модуляцией (флуктуацией) волнового фронта и диаграммы направленности. Для борьбы с этим явлением разработаны специальные методы подавления самопроизвольного шнурования, сводящиеся к разделению широкого (100 мкм и более) полоска на отдельные нитевидные участки.

В мощных лазерах, когда излучательное время жизни в активной области (квантовой яме на рис. 9.43, б), становится очень мало за счет интенсивно протекающих процессов вынужденного испускания, важным становится ее подпитка неравновесными носителями заряда из волноводного слоя, выполняющего одновременно функции резервуара. Для ускорения поступления носителей в активную область энергетический профиль этого слоя делают не плоским, как на рис. 9.28, а спадающим к центру квантовой ямы, как показано на рис. 9.43, б. Это легко осуществ-

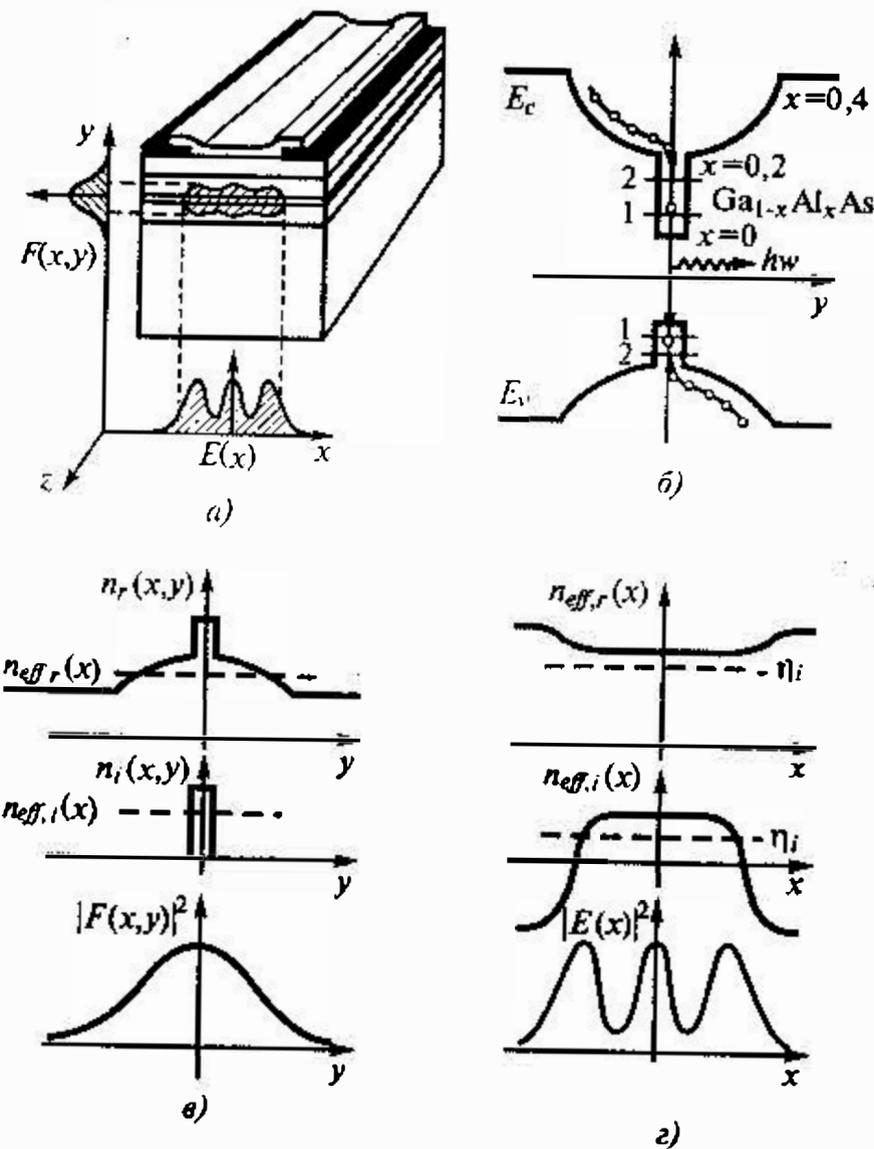


Рис. 9.43. Особенности распределения поля в мощном лазере с широкой активной областью:

а — пространственные распределения, указывающие на поперечные и латеральные оптические моды; б — зонаная диаграмма в поперечном направлении y ; в — профили для поперечного усиления $n_i(x, y)$ и показателя преломления $n_r(x, y)$ для заданной координаты x в поперечной моде $F(x, y)$; г — профили для латерального (горизонтального) усиления $n_{eff,i}(x)$ и показателя преломления $n_{eff,r}(x)$ для латеральной моды $E(x)$

вить, например, путем изменения состава твердого раствора по координате y при его эпитаксиальном наращивании.

В качестве основных активных материалов мощных инжекционных лазеров применяются системы на основе твердых растворов $(Al, Ga)As/(In, Ga)As/GaAs$ и $(Ga, In)(As, P)/InP$ или $(Ga, In)(As, P)/GaAs$. Они перекрывают спектральный диапазон $0,78...1,55$ мкм, исключительно важный как для волоконно-оптических линий связи (см. § 12.3), так и для накачки твердотельных лазеров (см. § 8.3) и лазерных усилителей (см. § 8.5). Гетероструктуры на основе твердого раствора $(Ga, Al)As$ более просты в технологическом отношении, чем структуры на основе четырех-

компонентной системы $(Ga, In)(As, P)$ (см. § 4.2), однако первые по ряду параметров проигрывают вторым. В частности, лучевая (оптическая) стойкость первых в 2 — 3 раза меньше, чем вторых. Объясняется это, по-видимому, наличием в составе первых алюминия. С точки зрения эффективности теплоотвода система $GaAs/(Al, Ga)As$ также проигрывает системе $InP/(Ga, In)(As, P)$ как по характеристической температуре T_0 в выражении (9.32), так и по теплопроводности. Для $(In, Ga)As/(Al, Ga)As$ ($\lambda = 780...980$ нм) значения T_0 лежат в пределах $100^\circ C < T_0 < 200^\circ C$, в то время как для $InP/(Ga, In)(As, P)$ -лазеров ($\lambda = 1,3...1,55$ мкм) $50^\circ C < T_0 < 90^\circ C$. Плохая теплопроводность, особенно $(Al, Ga)As$, приводит к сильным градиентам температуры и, как следствие, к появлению характерных «дефектов темных линий» и термической деградации. Особенно опасен становится циклический характер работы мощных лазеров.

Эффективность преобразования (КПД), определяемая как отношение оптической выходной мощности лазерного излучения к электрической мощности, подаваемой на диод, в выпускаемых промышленностью мощных инжекционных лазерах достигает 50 — 60%. Несмотря на внушительность этой цифры, остается реальная перспектива ее повышения до 80 — 90%. Увеличение эффективности для мощных инжекционных лазеров имеет принципиально важное значение не только и не столько как увеличение КПД, но, прежде всего как снижение тепловыделения и связанных с ним негативных явлений. Достигнутая в лабораторных условиях дифференциальная квантовая эффективность для лазеров, работающих в области $0,98$ мкм, составила 86%. Это говорит о реальном повышении КПД инжекционных лазеров до фантастической величины 80 — 90%.

Мощности, генерируемые одиночными лазерными диодами, составляют $(0,2...0,5)$ Вт в одночастотном (одномодовом) режиме и достигают $5...10$ Вт в многомодовом непрерывном режиме генерации. Для ряда важных применений (медицина, технология и др.) требуются еще большие мощности излучения. Дальнейшее увеличение энергии и мощности оптического излучения, генерируемого полупроводниковыми лазерами, достигается суммированием световых потоков от отдельных лазерных элементов. Здесь развитие идет в трех направлениях.

1. Излучение отдельного лазерного элемента вводится в оптическое волокно, а затем методами волоконной оптики оно складывается с излучением других лазерных диодов в одно общее волокно, которое обеспечивает суммарный вывод мощности. Вместо отдельных элементов могут применяться готовые приборы или чипы с волоконным выводом.

2. На одном кристалле методами интегральной технологии создается набор однотипных полосковых лазерных структур, раз-

деленных друг от друга и не связанных или слабо связанных между собой оптически, но имеющих общую базу и общие рабочие эпитаксиальные слои. Такая структура образует лазерную линейку, и излучение всех лазерных элементов складывается в дальнем поле.

3. Суммирование излучения отдельных лазерных элементов, созданных на одном кристалле, проводится методами интегральной оптики за счет создания специальных волноводных структур на этом же кристалле.

Последнее направление развивается пока что лишь в исследовательских лабораториях.

В рамках первого направления созданы и промышленно выпускаются лазерные модули, включающие в себя до восьми мощных лазерных диодов, излучение которых вводится в общее волокно с диаметром сердцевины порядка 200 мкм. Выходная мощность излучения при числовой апертуре 0,22 составляет 40 Вт при термоэлектрическом охлаждении. Плотность оптической мощности на выходе волокна достигает миллиона Ватт на квадратный сантиметр.

В рамках второго направления разработаны и выпускаются монолитные конструкции в виде лазерных линеек и собранные из них блоки в виде объемных лазерных решеток.

Типовая схема лазерной линейки представлена на рис. 9.44. Такие линейки часто применяются в различного рода блочных конструкциях, что очень удобно. В качестве базового размера блока выбрана его ширина, равная 1 см, и кратные ей величины. Каждая линейка включает в себя ряд активных элементов (областей), через которые протекает ток накачки. В свою очередь каждый активный элемент линейки может состоять из многих узких параллельных полосков, как показано точками на рис. 9.44. Периодически расположенные активные элементы занимают до-

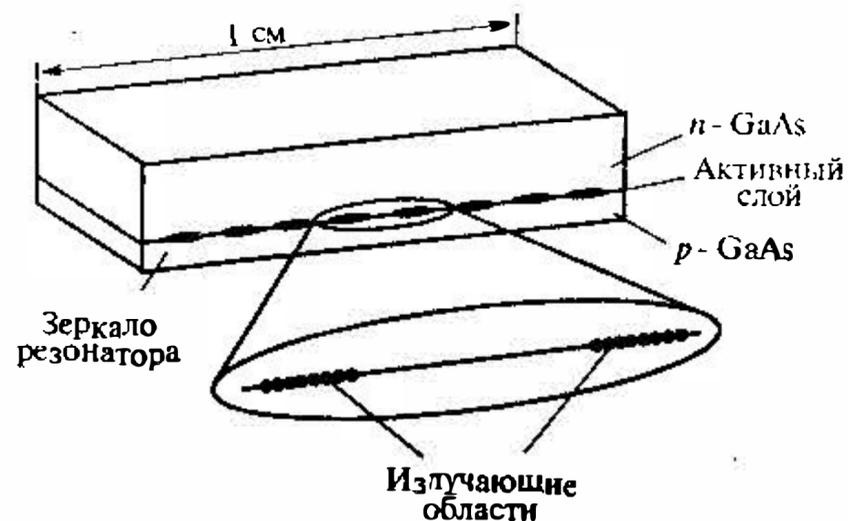


Рис. 9.44. Схема типовой лазерной линейки. Указаны отдельные излучающие области, представляющие собой многополосковые лазеры

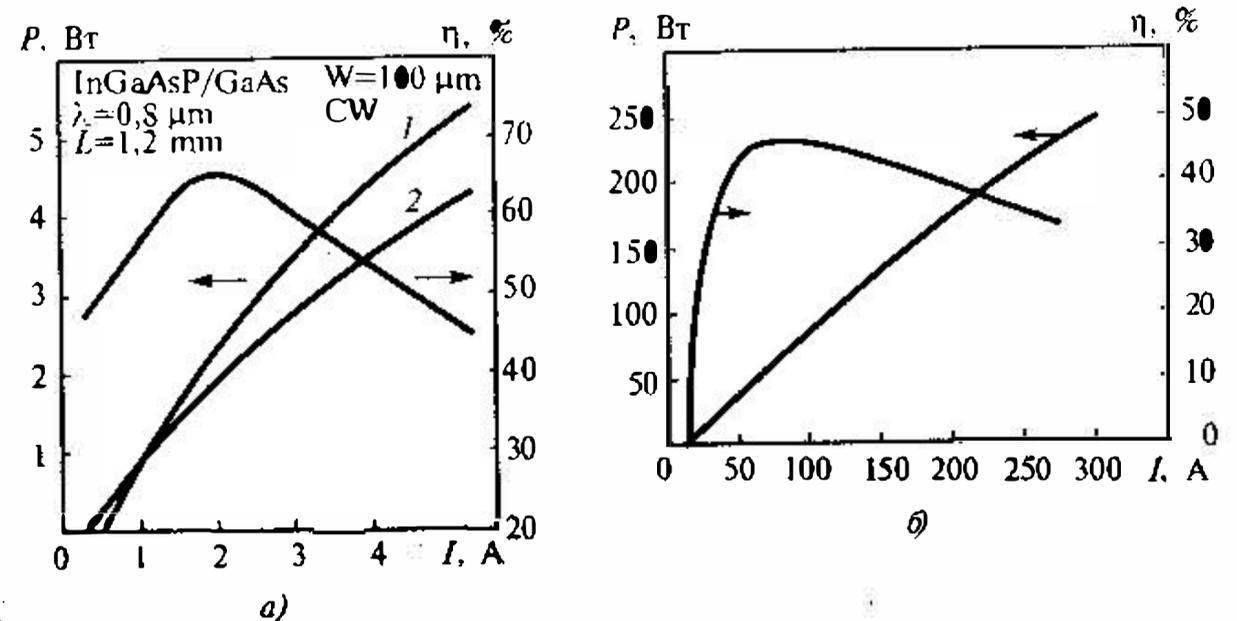


Рис. 9.45. Ватт-амперные характеристики и КПД одиночного лазерного диода (а) и лазерной линейки шириной 1 см (б). Для одиночного диода Ватт-амперные характеристики 1 и 2 относятся к диодам, у которых на выходную грань нанесено (1) и не нанесено (2) антиотражающее покрытие

лю площади активной области чипа, которая выражается в процентах и называется «фактором заполнения». В зависимости от целей применения фактор заполнения в современных линейках может изменяться от величин меньше 10 — 20% до почти 100%. Такие линейки могут работать как в непрерывном режиме, так и в импульсном и квазинепрерывном режимах. Эффективность лазерных линеек возросла за последние несколько лет с 30 до более 50% преимущественно за счет применения структур с раздельным оптическим и электронным ограничением (см. § 9.7). Типичные характеристики таких линеек представлены на рис. 9.45.

Предельные мощности, которые могут быть получены от лазерных линеек, ограничены теми же причинами, которые обсуждались выше для отдельных лазерных диодов.

Стандартные лазерные линейки шириной 1 см, излучающие до 50 Вт в непрерывном режиме при сроке службы (5...10) тыс. ч, служат модулями для строительства наиболее мощных лазерных излучателей — лазерных решеток. Их конфигурация и число элементов (линеек) в решетке могут быть весьма разнообразны. Например, наборная решетка из двадцати лазерных линеек, смонтированных на держателе — хладопроводе с поперечным размером 1 см × 1 см при проточном водяном охлаждении генерирует пиковую оптическую мощность 1,5 кВт в квазинепрерывном режиме при скважности импульсов 3%. Применяя «наборную» технику, возможно получать средние выходные уровни мощности до 200 Вт/см² и пиковые плотности мощности 5 кВт/см² и более.

§ 9.11. ПОВЕРХНОСТНО-ИЗЛУЧАЮЩИЕ ИНЖЕКЦИОННЫЕ МИКРОЛАЗЕРЫ

Особенности электронного спектра квантоворазмерных гетероструктур, рассмотренные в § 4.8, и возможность получения больших значений показателя усиления α_w позволяют реализовать лазеры, активная область которых составляет всего несколько микрон. Весьма привлекательно создать инжекционный лазер, излучающий не с торца p - n перехода, а с его поверхности аналогично полупроводниковому лазеру с продольной электронной накачкой, изображенному на рис. 9.20, б. Для его реализации прежде всего необходима среда с очень большим показателем усиления α_w , так, чтобы $\alpha_w > 1/d$, где d — толщина активной области. В отличие от геометрии традиционной лазерной структуры, изображенной на рис. 9.22, в поверхностно-излучающем лазере усиление электромагнитной волны осуществляется в направлении z , совпадающем с направлением электрического тока. Максимальная длина активной области (на рис. 9.20, б, рис. 9.22 это толщина d) в такой геометрии ограничена диффузионной длиной неосновных носителей заряда в активной области и составляет единицы мкм. Зеркала резонатора должны быть нанесены на излучающие поверхности, через которые пропускается электрический ток. Оптические потери при этом должны быть минимальны. Современная полупроводниковая технология позволяет создавать такие структуры в едином технологическом цикле. На рис. 9.46 приведен пример реализации такого лазера, активная область которого содержит вертикально-связанные квантовые точки на основе узкозонного InGaAs в матрице GaAs или $Ga_{1-x}Al_xAs$ эпитаксиального слоя, как показано справа на рис. 9.46. Квантовые точки, как это следует из § 4.8, за счет стягивания плотности состояний к уровням размерного квантования обеспечивают исключительно большую силу осциллятора для резонансных оптических переходов и как следствие — большую величину показателя поглощения k_w , а при реализации условий инверсии (4.76, а) — показателя усиления α_w . Зеркала резонатора выполнены в виде чередующихся четвертьволновых слоев с разными показателями преломления так, чтобы их оптическая толщина удовлетворяла условию (7.20): $n_1 d_1 = n_2 d_2 = \lambda/4$. Такие многослойные структуры работают как интерференционные зеркала аналогично изображенным на рис. 7.9. Они называются распределенными брэгговскими отражателями (РБО) и создаются в процессе эпитаксии, например слоев GaAs — $Al_xGa_{1-x}As$. Величина коэффициента отражения R задается числом слоев и для нижнего «глухого» зеркала $R=1$. При необходимости эти же зеркала могут выполнять функцию электрических контактов.

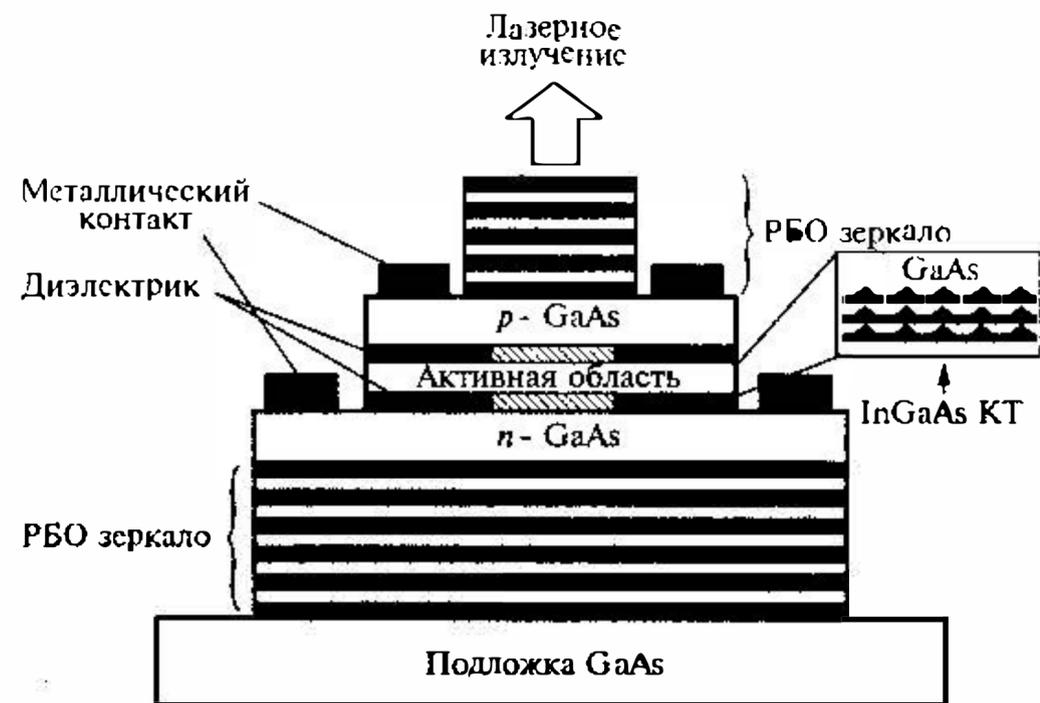


Рис. 9.46. Поверхностно-излучающий инжекционный микролазер с активной областью, содержащей вертикально-связанные квантовые точки

Отметим два принципиально важных обстоятельства, которые необходимо учитывать при создании рассматриваемых лазерных структур.

1. В поверхностно-излучающих лазерах максимальное усиление должно быть обеспечено в направлении, перпендикулярном p - n переходу, а в плоскости p - n перехода оно должно быть подавлено не только в направлении x , как в полосковом лазере на рис. 9.26 и 9.27, но и в направлении y . С этой целью площадь рабочей области ограничивается, например, с помощью тонких диэлектрических (или высокоомных полупроводниковых) слоев, как показано на рис. 9.46. Электрический ток проходит только через отверстие в диэлектрической маске. Его размеры в поперечнике составляют несколько мкм. Поскольку эффективный показатель усиления активной среды должен превышать 1000 см^{-1} , то в активной области используется двойная гетероструктура, содержащая набор квантовых ям, квантовых нитей или квантовых точек (см. § 4.8). Применение квантовых нитей может обеспечить преимущественное усиление в направлении их осей, а особенности в функции плотности состояний (см. рис. 4.57) — большой показатель усиления в узком спектральном интервале. Однако их технология получения пока что не обеспечивает требуемых параметров и потому активная область микролазеров, как правило, содержит или набор квантовых ям или квантовые точки, которые с помощью специальной технологии, использующей эффекты самоорганизации, «складируются» друг на друга, образуя верти-

кальво-связанные КТ, например InGaAs в эпитаксиальном слое GaAs, как показано на рис. 4.46.

2. Длина резонатора в рассматриваемых лазерах составляет несколько длин волн. Расстояние между собственными типами колебаний велико, а спектр усиления — узкий. С одной стороны, это облегчает работу лазера в одночастотном режиме, обеспечивая тем самым высокое качество излучения. С другой стороны, это накладывает жесткие требования к согласованию размеров резонатора с параметрами активной среды (рабочей длиной волны и максимумом коэффициента усиления).

Поверхностно-излучающие полупроводниковые лазеры создаются с использованием современной групповой полупроводниковой технологии в едином технологическом цикле. На одной подложке могут быть выращены структуры, содержащие сотни тысяч активных элементов. Каждый из этих элементов может работать или отдельно, независимо, или в комплексе с другими элементами, обеспечивая работу всей матрицы в целом. Рабочий ток отдельного микролазера не превышает единиц мА при пороговом токе десятые и сотые доли мА. При необходимости такой микролазер удобно стыкуется с оптическим волокном с минимальными потерями излучения.

Обладая малой инерционностью и высоким качеством излучения, поверхностно-излучающие микролазеры используются в системах передачи и обработки оптических сигналов со скоростью до 100 Гбит/с.

§ 9.12. КАСКАДНЫЕ ЛАЗЕРЫ

Во всех рассмотренных ранее типах полупроводниковых лазеров использовалась излучательная рекомбинация электрона зоны проводимости с дыркой валентной зоны. Она могла происходить непосредственно (зона-зона), либо с участием экситонных и мелких примесных состояний (квазимежзонные переходы). Правила отбора по волновому вектору (4.33) разрешают только прямые, вертикальные переходы и потому в активной области полупроводниковых лазеров применяют только прямозонные материалы (см. § 9.3).

В каскадных лазерах в качестве рабочих переходов используют переходы между уровнями размерного квантования, принадлежащими одной зоне, т. е. внутризонные переходы. Это могут быть, например, переходы между подзонами размерного квантования электронов в квантовой яме или электронными уровнями квантовой точки (см. § 4.8). Этим каскадные лазеры принципиально отличаются от обычных полупроводниковых лазеров.

Основная трудность в лазере, использующем внутризонные переходы, связана с созданием инверсной населенности между рабочими уровнями энергии. Как говорилось в § 4.6, время внутризонной релаксации в полупроводниках чрезвычайно мало и имеет порядок величины 10^{-12} с. Поэтому в объемных 3D-кристаллах получить инверсию населенности между энергетическими состояниями внутри зоны не представляется возможным. Применение квантоворазмерных структур, рассмотренных в § 4.8, открывает в этом отношении новые перспективы, поскольку в таких структурах энергетическими состояниями и волновыми функциями электронов можно управлять, меняя параметры структуры и тем самым конструировать электронную систему с желаемыми свойствами.

Принцип работы каскадного лазера на квантовых ямах пояснен на энергетической диаграмме (рис. 9.47). Активная область содержит три квантовые ямы толщиной d_1 , d_2 и d_3 , разделенные барьерами шириной b_1 и b_2 . Это связанные ямы, аналогичные изображенным на рис. 4.49. Связь между ними задается шириной барьеров b_1 и b_2 , а положения уровней размерного квантования определяются в основном высотой барьера (которая в принципе может быть разной для каждой ямы как на рис. 4.49), и толщиной ям d . В электрическом поле за счет эффектов резонансного туннелирования (рис. 4.53 и 4.54) электроны из эмиттера будут заселять в основном уровень 3, как показано на рис. 9.47. Пара-

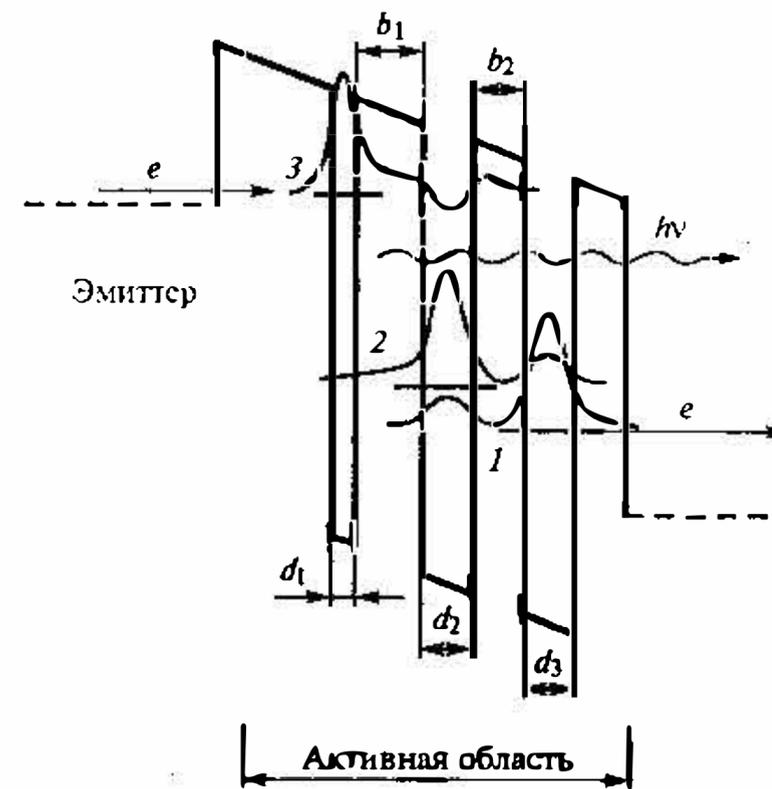


Рис. 9.47. Принцип работы каскадного лазера на внутризонных переходах между подзонами размерного квантования связанных квантовых ям

метры первой ямы d_1 подбираются таким образом, что этот уровень для нее является наименьшим и поэтому релаксация внутри ямы невозможна, а квантовые переходы $3 \rightarrow 2$ идут с испусканием фотона $h\nu = E_2 - E_1$. Нижний рабочий лазерный уровень E_2 быстро опустошается за счет переходов $2 \rightarrow 1$, передавая электроны в третью яму, откуда они электрическим полем выбрасываются в коллектор.

Нетрудно заметить, что по принципу работы каскадный полупроводниковый лазер близок к обычным газовым лазерам, рассмотренным в гл. 7. Действительно, заселение верхнего рабочего уровня осуществляется селективно за счет резонансных процессов аналогично тому, как это происходит при резонансной передаче возбуждения (7.15) в газовых лазерах (см. рис. 7.4), в том числе — в He-Ne (см. рис. 7.8) и CO₂ (см. рис. 7.13) лазерах. Это обеспечивает инверсию населенности. Как обычные атомарные или молекулярные лазеры, каскадный лазер, работает по трех- или четырехуровневой схеме аналогично представленной на рис. 7.4. Но в отличие от газовых лазеров, параметры активных частиц в которых задаются Природой, в каскадных полупроводниковых лазерах мы можем сами конструировать квантовые ямы, придавая им нужные нам свойства, в частности, реализуя условия $\tau_{32} > \tau_{21}$. Рабочая длина волны определяется расстоянием между уровнями размерного квантования и может изменяться от нескольких десятых долей эВ почти до 0, так что каскадные лазеры могут перекрывать весь ИК — диапазон оптического спектра, реально работая в области примерно от 3 до 100 мкм.

ГЛАВА 10

ПРИБОРЫ УПРАВЛЕНИЯ ЛАЗЕРНЫМ ИЗЛУЧЕНИЕМ

Приборы управления лазерным излучением призваны осуществлять те преобразования лазерных пучков, о которых говорилось в § 3.6. Кроме традиционных элементов классической оптики — линз, зеркал, поляризаторов, призм, оптических фильтров, клиньев и т. п., развитие лазерной техники стимулировало создание новых устройств, которые и рассматриваются в настоящей главе. Их работа основана на использовании электро-, магнито- и пьезооптических эффектов в кристаллах.

§ 10.1. ЭЛЕКТРООПТИЧЕСКИЕ, МАГНИТООПТИЧЕСКИЕ И ПЬЕЗООПТИЧЕСКИЕ ЭФФЕКТЫ

Под *электрооптическими, магнитооптическими и пьезооптическими эффектами* понимают эффекты, связанные с изменением оптических свойств вещества под воздействием внешних электрических и магнитных полей или упругих механических деформаций. Так как в общем случае диэлектрическая проницаемость вещества комплексная, то все эти эффекты могут быть подразделены на две большие группы, связанные с изменением показателей преломления и поглощения. Здесь будем рассматривать только явления, обусловленные изменением вещественной части показателя преломления.

Общим для указанных эффектов является то, что внешние воздействия приводят к изменению симметрии кристалла, следствием чего является изменение его свойств. Коэффициенты оптической индикатрисы (3.41) для данного кристалла являются постоянными лишь при определенных условиях. Воздействие внешних механических напряжений, электрических и магнитных полей приводит к их изменению. Оптически изотропные кристаллы становятся анизотропными, а оптическая индикатриса одноосных и двухосных кристаллов будет поворачиваться и деформироваться. В этом состоит сущность рассматриваемых эффектов.

Линейный и квадратичный электрооптические эффекты. Электрооптические свойства кристаллов удобно рассматривать, введя вместо тензора (3.37) диэлектрической проницаемости ϵ тензор диэлектрической непроницаемости a , компоненты a_{ij} которого обратны компонентам тензора ϵ_{ij} :

$$a_{ij} = 1/\epsilon_{ij} = 1/n^2. \quad (10.1)$$

Тензор a , так же как и тензор ϵ , является симметричным тензором второго ранга, а его компоненты — компонентами тензора диэлектрической непроницаемости, или *поляризационными константами*. В главной системе координат при $E = 0$

$$\begin{aligned} a_{11} &= 1/\epsilon_{11} = 1/\epsilon_x = 1/n_x^2, \\ a_{22} &= 1/\epsilon_{22} = 1/\epsilon_y = 1/n_y^2, \\ a_{33} &= 1/\epsilon_{33} = 1/\epsilon_z = 1/n_z^2. \end{aligned} \quad (10.2)$$

При наложении электрического поля эллипсоид оптической индикатрисы (3.41) поворачивается и деформируется. Главные оси этого эллипсоида в общем случае не будут совпадать с главными осями исходного эллипсоида, в системе координат x, y, z которого уравнение оптической индикатрисы принимает вид

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{31}zx = 1. \quad (10.3)$$

Изменение коэффициента оптической индикатрисы под воздействием внешнего электрического поля в общем случае будет описываться соотношением

$$\Delta a_{ij} = r_{ijk}E_k + R_{ijkl}E_kE_l + \dots, \quad (10.4)$$

где E_k и E_l — компоненты вектора напряженности внешнего поля E .

Первое слагаемое в правой части выражает *линейный электрооптический эффект*, второе — *квадратичный электрооптический эффект*, остальные слагаемые соответствуют эффектам более высокого порядка.

Коэффициенты r_{ijk} составляют тензор третьего ранга, компоненты которого называются *линейными электрооптическими коэффициентами*. Этот тензор симметричен по первым двум индексам и по симметрии аналогичен тензору пьезоэлектрических модулей. В центросимметричных кристаллах все компоненты тензора коэффициентов линейного электрооптического эффекта обращаются в нуль. Поэтому линейный электрооптический эффект возможен только в кристаллах, не обладающих центром инверсии. Изменение показателя преломления Δn в слабых полях пропорционально приложенному полю.

Для сокращения записи часто вместо трехиндексной системы обозначений r_{ijk} используют двухиндексную (матричную) систему, заменяя два индекса, по которым тензор симметричен, одним со следующим соответствием: 11→1; 22→2; 33→3; 23→4; 31→5; 12→6. Тогда линейное изменение поляризационных констант под воздействием внешнего электрического поля E будет определяться выражением

$$\Delta a_i = \sum_{j=1}^3 r_{ij}E_j, \quad (10.5)$$

где индекс суммирования j означает $x=1$; $y=2$ и $z=3$. Уравнение (10.5) можно записать в матричной форме:

$$\begin{pmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta a_3 \\ \Delta a_4 \\ \Delta a_5 \\ \Delta a_6 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}. \quad (10.6)$$

Матрица 6×3 с элементами r_{ij} есть матрица электрооптических коэффициентов. Зная правила перемножения матриц (10.6) и (10.5), находим, например, $\Delta a_4 = r_{41}E_1 + r_{42}E_2 + r_{43}E_3$. Вид матрицы r_{ij} , но не числовые значения ее коэффициентов можно определить на основе анализа симметрии кристалла, который показывает, какие из 18 коэффициентов r_{ij} равны нулю, и характеризует соотношения между остальными коэффициентами. Например, для кристаллов с центром инверсии все коэффициенты $r_{ij} = 0$, т. е. линейный электрооптический эффект отсутствует. Для кристаллов со структурой сфалерита (GaAs, GaP, CdTe и др.) отличны от нуля только три коэффициента, причем все они равны: $r_{41} = r_{52} = r_{63}$. Для кристаллов типа титаната бария (BaTiO_3) отличны от нуля $r_{13} = r_{23}$ и $r_{11} = r_{32}$. Для кристаллов типа КДР (дигидрофосфат калия KH_2PO_4) $r_{41} = r_{52} \neq r_{63}$, а остальные коэффициенты $r_{ij} = 0$ и т. д.

Отметим, что линейный электрооптический эффект проявляется в тех же классах кристаллов, в которых существует пьезоэффект, т. е. в пьезоэлектриках.

Таким образом, с приложением электрического поля к кристаллу в уравнении эллипсоида показателя преломления появляются «смешанные» члены xy , yz , xz . Это означает, что главные оси эллипсоида при наличии поля не параллельны осям кристалла x , y , z . Для того чтобы найти направления и длины главных полуосей нового эллипсоида, необходимо диагонализировать мат-

рицу, элементами которой являются константы эллипсоида (10.3).

Рассмотренный линейный электрооптический эффект называется *первичным или истинным*. При воздействии электрического поля на механически не закрепленные кристаллы они деформируются за счет обратного пьезоэффекта. Эти деформации из-за упругооптического эффекта приводят к дополнительному изменению оптической индикатрисы (*вторичный электрооптический эффект*). Так как деформация кристалла зависит от способа его закрепления и частоты внешнего поля, то вторичный электрооптический эффект будет максимален на частоте собственных колебаний кристалла в держателе. В механически зажатом кристалле, когда он не может деформироваться под воздействием поля E , или на высоких частотах, когда деформация не успевает следовать за изменением поля, будет проявляться только истинный электрооптический эффект.

Квадратичный электрооптический эффект характеризуется тензором четвертого ранга R_{ijkl} . Он проявляется во всех материалах, в том числе в газах, жидкостях, аморфных и кристаллических твердых телах. Квадратичный электрооптический эффект известен также как *эффект Керра*. Изменение показателя преломления, вызванное квадратичным эффектом, обычно значительно меньше, чем изменение от линейного эффекта (за исключением случаев, когда последний отсутствует).

Рассмотренные электрооптические эффекты — линейный (Поккельса) и квадратичный (Керра) — обладают малой инерционностью и используются для модуляции и отклонения света. Постоянная времени истинного электрооптического эффекта определяется молекулярным временем релаксации и может составлять 10^{-10} с и менее.

Магнитооптические эффекты. Внешнее магнитное поле также может приводить к изменению оптических характеристик материала. Как и электрооптические эффекты, магнитооптические эффекты подразделяют на эффекты первого порядка (линейные), пропорциональные первой степени напряженности магнитного поля, и эффекты второго и более высоких порядков. Наиболее сильным эффектом первого порядка является *эффект Фарадея*, заключающийся во вращении плоскости поляризации света, распространяющегося вдоль направления магнитного поля. Угол поворота плоскости поляризации φ пропорционален длине пути света в веществе и *индукции* магнитного поля:

$$\varphi = VIB, \quad (10.7)$$

где V — коэффициент пропорциональности, называемый *постоянной Верде*.

Типичное значение постоянной Верде для стекол 10 град/(Тл·см) указывает, что для наблюдения эффекта Фарадея

обычно нужны сильные магнитные поля. Поскольку управлять магнитным полем сложнее, чем электрическим, эффект Фарадея используют для модуляции света реже, чем электрооптический эффект.

Пьезооптические эффекты. В твердых телах под действием упругих механических напряжений изменяется показатель преломления вещества. Это явление называется *фотоупругостью*, а связанные с ним эффекты — пьезооптическими или упругооптическими эффектами.

Линейный упругооптический (фотоупругий) эффект заключается в том, что показатель преломления среды изменяется пропорционально механической деформации. Если к кристаллу приложить механическое напряжение σ_i , то в нем появятся деформации, что в свою очередь приведет к изменению характеристик эллипсоида показателя преломления. При линейном эффекте изменение коэффициентов оптической индикатрисы может быть записано в виде

$$\Delta a_i = \sum_j \pi_{ij} \sigma_j \quad (10.8)$$

Здесь π_{ij} — *пьезооптические коэффициенты*, которые являются компонентами соответствующего тензора четвертого ранга и хорошо известны из кристаллографии.

Пьезооптические эффекты можно рассматривать аналогично электрооптическим. Однако в отличие от линейного электрооптического эффекта, который имеет место лишь в кристаллах, не обладающих инверсией, т. е. в пьезокристаллах, линейный пьезооптический эффект наблюдается во всех кристаллах и аморфных средах. В частности, он заметно проявляется в стеклах, плавленном кварце, кремнии и германии. Под действием одномерной деформации изотропная среда становится одноосной с оптической осью, совпадающей с направлением механической деформации. При распространении света в плоскости, перпендикулярной наведенной оптической оси, наблюдается двулучепреломление. Разность показателей преломления для обыкновенного и необыкновенного лучей при этом пропорциональна приложенному напряжению:

$$\Delta n = n_o - n_e = K\sigma, \quad (10.9)$$

§ 10.2. ОПТИЧЕСКИЕ МОДУЛЯТОРЫ

Модуляция света — это изменение его параметров в зависимости от управляющего (модулирующего) сигнала. С ее помощью производят наложение информации на световую волну или световой поток, осуществляющие перенос этой информации. Из

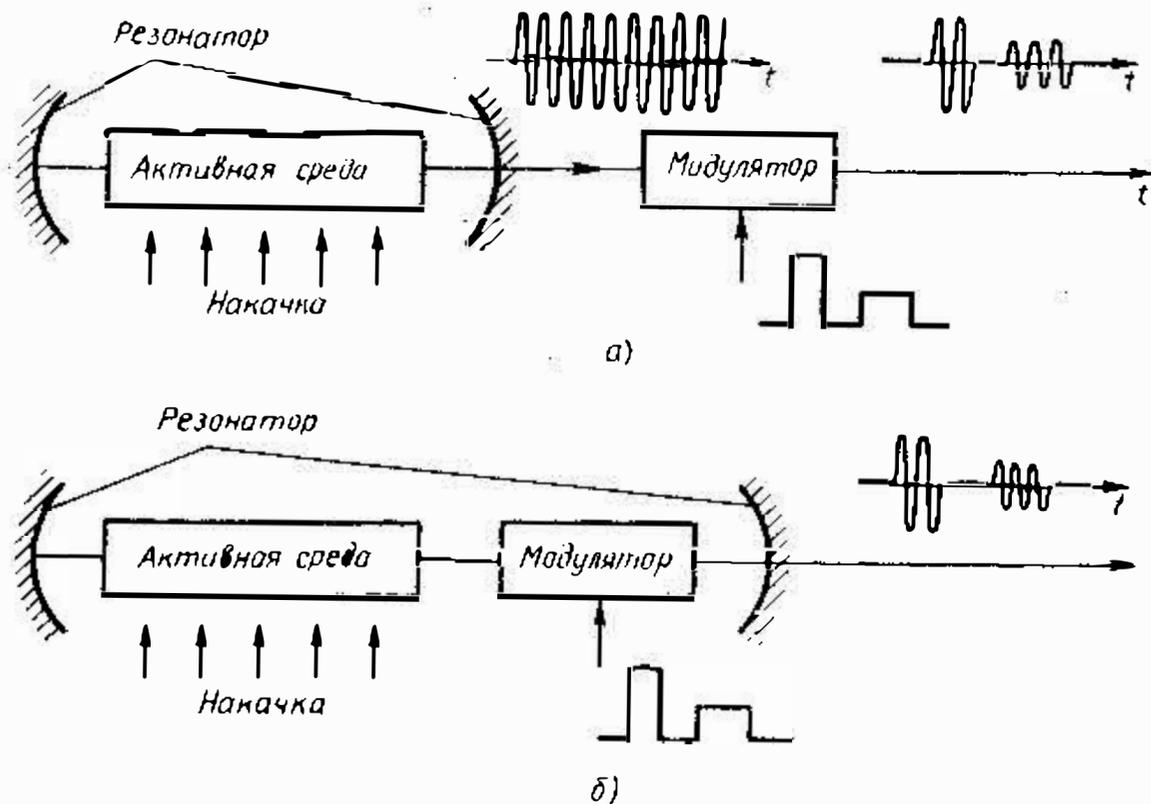


Рис. 10.1. Внешняя (а) и внутренняя (б) модуляции лазерного излучения. Временные диаграммы показывают пример амплитудной модуляции

характеристики световой волны (1.23) видно, что модулировать, т. е. целенаправленно изменять, можно: амплитуду (интенсивность), частоту, фазу, поляризацию, направление распространения и пространственное распределение волны или светового потока. Как и в радиотехнике, в оптической электронике наибольшее распространение как наиболее эффективные получили амплитудная и фазовая модуляция.

Различают внешнюю и внутреннюю модуляцию. В первом случае (рис. 10.1, а) модулятор находится вне резонатора и осуществляет модуляцию излучения, генерируемого лазером. Во втором случае (рис. 10.1, б) модулятор находится внутри резонатора, изменяя его свойства (например, добротность Q) и осуществляя модуляцию генерируемого излучения. Пример внутренней модуляции был приведен в § 2.5 при рассмотрении генерации коротких лазерных импульсов в режиме модуляции добротности (см. рис. 2.27) и в режиме синхронизации мод (см. рис. 2.30). Частным и весьма важным случаем внутренней модуляции является так называемая *прямая модуляция*, при которой сигнал модуляции воздействует на свойства самой активной среды. Типичный пример рассмотрен в § 9.9, где модуляция излучения полупроводникового лазера осуществляется прямым изменением тока накачки.

Важнейшими характеристиками оптических модуляторов являются: а) *глубина модуляции*

$$\eta = (I_{\max} - I_{\min}) / I_{\max}, \quad (10.10)$$

где I_{\max} и I_{\min} — интенсивности света при полностью открытом и закрытом состоянии модулятора; б) *ширина полосы пропускания* или диапазон модулирующих частот $\Delta\nu$, которые определяются как разность между верхней и нижней частотами, при которых глубина модуляции уменьшается на 50% от максимального значения; полоса частот $\Delta\nu$ определяет предельный объем информации, который можно передать с помощью данного модулятора; в) *рабочая апертура* — тот угол, измеряемый в градусах или стерadiansах, в пределах которого оптическое излучение может быть введено в модулятор; г) *спектральная область* — область длин волн, в которой модулятор способен работать; д) *рабочее напряжение* или *напряжение полуволнового сдвига* — те величины сигнала, которые необходимо подать на вход модулятора, чтобы перевести его из «открытого» состояния в «закрытое»; е) *потери*, вносимые модулятором, выражаемые, как правило, в децибеллах:

$$\beta = 10 \lg I_0 / I_{\max}, \quad (10.11)$$

где I_0 — интенсивность света в отсутствие модулятора, I_{\max} — интенсивность света, прошедшего через модулятор в открытом состоянии; ж) *потребляемая мощность* на единицу ширины полосы пропускания модулятора $P/\Delta f$, выражаемая обычно в Вт/Гц.

Для осуществления модуляции необходимо, чтобы управляющий сигнал воздействовал на свет. Стало быть, работа оптического модулятора должна основываться на процессах взаимодействия света с веществом. Возможно использование эффектов, связанных с внешним воздействием на вещественную часть n комплексного показателя преломления n^* или комплексной диэлектрической проницаемости. Это рассмотренные выше электрооптические, магнитооптические и пьезооптические эффекты. Соответствующие модуляторы называются *электрооптическими*, *магнитооптическими* и *пьезооптическими* или *акустооптическими*, если деформация в кристалле создается с помощью акустической волны. Если принцип работы модулятора основан на использовании эффектов, связанных с внешним воздействием на мнимую часть k комплексного показателя преломления, т. е. связан с процессами оптического поглощения, то такие модуляторы называются *абсорбционными*.

По соотношению характерных геометрических размеров рабочей области модулятора к длине световой волны и вытекающим отсюда отличиям в системах ввода излучения различают

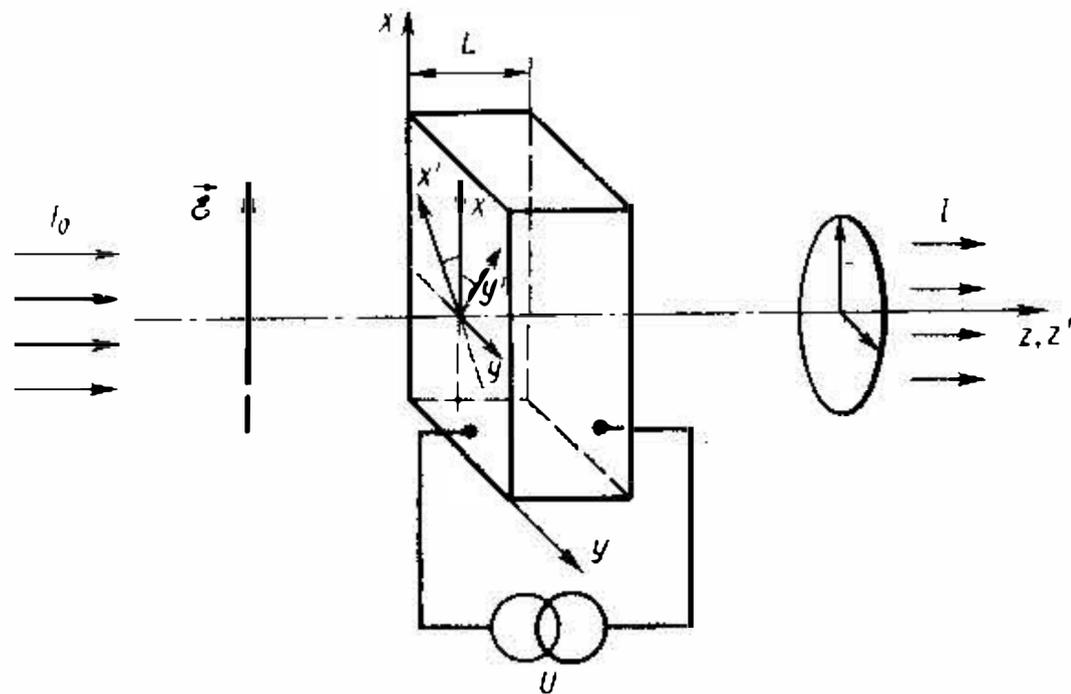


Рис. 10.2. Проложение плоскополяризованного луча через электрооптический кристалл

объемные и волноводные модуляторы. Первые работают с плоскими волнами и гауссовыми пучками. Вторые являются составными элементами оптических интегральных схем и волоконной оптики.

Наибольшее значение как отдельные элементы оптической электроники получили электрооптические и акустооптические модуляторы. Рассмотрим их свойства.

Электрооптические модуляторы. Это модуляторы, принцип действия которых основан на использовании рассмотренных в § 10.1 электрооптических эффектов в кристаллах. Рассмотрим модулятор, работающий на основе линейного электрооптического эффекта. Пусть плоскополяризованный свет $E = E_x$, распространяющийся в направлении z , падает на электрооптический кристалл, как это показано на рис. 10.2.

Для определенности возьмем кристалл дигидрофосфата калия KN_2PO_4 , известный как КДР. Это кристалл тетрагональной сингонии с группой симметрии $42m$. Его ось симметрии четвертого порядка совпадает с оптической осью кристалла и направлена по оси z на рис. 10.2. Две взаимно перпендикулярные оси второго порядка направлены по x и y . В одноосном кристалле, каким является КДР, $n_x = n_y = n_0$ и $n_z = n_e$. Для кристаллов симметрии $42m$ в матрице электрооптических коэффициентов (10.6) отличны от нуля только три элемента r_{ij} , а именно $r_{41} = r_{52}$ и r_{63} . При приложении поля E_z вдоль направления z уравнение эллипсоида показателя преломления (3.44) запишется в виде

$$(a_x^2 + a_y^2)/n_0^2 + a_z^2/n_e^2 + 2r_{63}a_x a_y E_z = 1. \quad (10.12)$$

Чтобы это уравнение привести к диагональному виду, необходимо выбрать систему координат $x'y'z'$ с осью z' , совпадающей с z и осями x' и y' , повернутыми на 45° относительно x и y , как показано на рис. 10.2. В этой системе координат уравнение (10.12) преобразуется к диагональному виду

$$a_x^2/n_x^2 + a_y^2/n_y^2 + a_z^2/n_z^2 = 1. \quad (10.13)$$

Осуществив математические преобразования, связанные с переходом из системы координат xy в систему $x'y'$, находим

$$\begin{aligned} 1/n_x^2 &= 1/n_0^2 + r_{63}E_z, \\ 1/n_y^2 &= 1/n_0^2 - r_{63}E_z. \end{aligned} \quad (10.14)$$

Учитывая, что $r_{63}E_z \ll 1/n_0^2$, получаем

$$\begin{aligned} n_x &= n_0 - \frac{n_0^3}{2} r_{63}E_z, \\ n_y &= n_0 + \frac{n_0^3}{2} r_{63}E_z, \\ n_z &= n_e. \end{aligned} \quad (10.15)$$

Аналогичные соотношения можно получить из (10.1) и (10.4):

$$\Delta a_{ij} = \Delta \left(\frac{1}{n^2} \right) = \frac{-2\Delta n_x}{n^3} = r_{ijk}E_z,$$

откуда находим изменение показателя преломления Δn , вызванное электрическим полем

$$\Delta n = n(E) - n_0(0) = \pm \frac{1}{2} r_{ijk} n_0^3 E_z,$$

в точном соответствии с (10.15), где $r_{ijk} = r_{63}$.

Таким образом, если на электрооптический кристалл падает световая волна, поляризованная в плоскости xz , то, раскладывая ее на две компоненты E_x и E_y , как показано на рис. 10.2, получаем для них изменение показателя преломления

$$\Delta n_y = -\Delta n_x = \frac{1}{2} r_{63} n_0^3 E_z \quad (10.16)$$

и вызванные этим изменением фазовые сдвиги для E_y - и E_x -компонент световой волны:

$$\Delta \varphi_y = -\Delta \varphi_x = \frac{2\pi}{\lambda} L \Delta n, \quad (10.17)$$

так что общий фазовый сдвиг

$$\Delta\varphi = \Delta\varphi_y - \Delta\varphi_x = \frac{2\pi}{\lambda} r_{63} n_0^3 E_z L = \frac{2\pi}{\lambda} r_{63} n_0^3 U \quad (10.18)$$

пропорционален приложенному напряжению U . Если на вход такого модулятора падает плоская монохроматическая волна

$$E = E_0 \cos \omega t,$$

то по прохождении электрооптического кристалла x' - и y' -компоненты этой волны

$$\begin{aligned} E_x &= \frac{E_0}{\sqrt{2}} \cos(\omega t + \varphi_0 + \Delta\varphi_x), \\ E_y &= \frac{E_0}{\sqrt{2}} \cos(\omega t + \varphi_0 + \Delta\varphi_y) \end{aligned} \quad (10.19)$$

имеют фазовый сдвиг, непосредственно зависящий от приложенного напряжения, так что по схеме рис. 10.2 можно осуществлять фазовую модуляцию световой волны.

Типичная схема амплитудного электрооптического модулятора приведена на рис. 10.3. Электрооптический кристалл 2 с таким расположением, как это приведено на рис. 10.2, помещен между двумя скрещенными друг относительно друга поляризаторами 1 и 4. Поляризатор 1 пропускает лишь волну, поляризованную в вертикальной плоскости xz , а анализатор 4 — волну, поляризованную в горизонтальной плоскости yz . Будем сначала считать, что элемент 3 в схеме рис. 10.3 отсутствует. Тогда через анализатор будут проходить $E_y/\sqrt{2}$ и $(-E_x/\sqrt{2})$ — составляющие, как

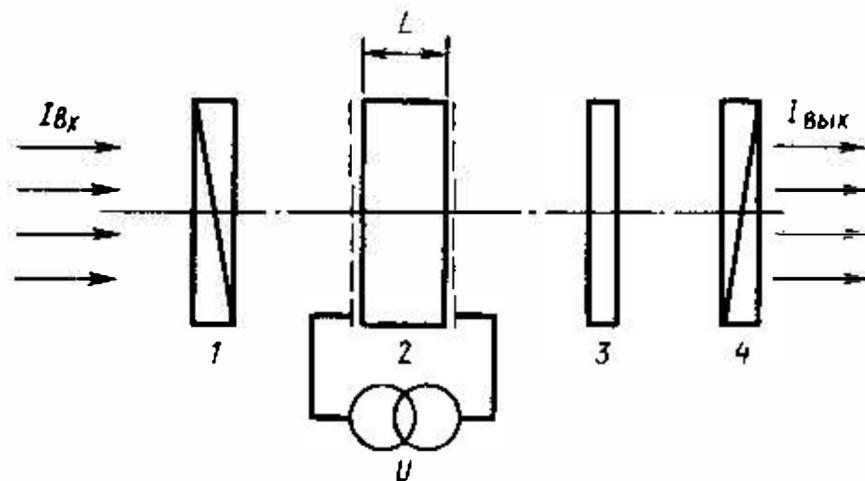


Рис. 10.3. Схема электрооптического амплитудного модулятора:

1 — входной поляризатор; 2 — электрооптический кристалл; 3 — четвертьволновая пластинка; 4 — выходной поляризатор (анализатор)

нетрудно убедиться из рис. 10.2. Тогда с учетом (10.19) прошедшая через анализатор волна запишется в виде

$$E_y = \frac{E_0}{2} [\cos(\omega t + \varphi_0 + \Delta\varphi_y) - \cos(\omega t + \varphi_0 + \Delta\varphi_x)]. \quad (10.20)$$

При $\Delta\varphi_y = -\Delta\varphi_x = \Delta\varphi/2$ из (10.20) имеем

$$E_y = -E_0 \sin \frac{\Delta\varphi}{2} \sin(\omega t + \varphi_0). \quad (10.21)$$

Интенсивность прошедшего излучения, которое определяется усреднением E_y^2 за период $T = 2\pi/\omega$, будет

$$I = \frac{1}{T} \int_0^T E_y^2 dt = I_0 \sin^2 \frac{\Delta\varphi}{2}, \quad (10.22)$$

где I_0 — интенсивность света, падающего на кристалл (отражением и поглощением в кристалле и оптических элементах схемы пренебрегаем). С учетом (10.18), (10.22) получаем зависимость оптического пропускания модулятора от приложенного напряжения

$$\frac{I}{I_0} = \sin^2 \left(\frac{\pi}{\lambda} r_{63} n_0^3 U \right), \quad (10.23)$$

которое можно переписать в виде

$$\frac{I}{I_0} = \sin^2 \left(\frac{\pi}{2} \frac{U}{U_{\lambda/2}} \right), \quad (10.24)$$

где

$$U_{\lambda/2} = \lambda / (2r_{63} n_0^3) \quad (10.25)$$

называется напряжением полуволинового смещения. При $U = U_{\lambda/2}$ сдвиг фаз соответствует $\lambda/2$, плоскость поляризации падающей волны после прохождения электрооптического кристалла поворачивается на 90° и пропускание модулятора максимально. Как видно из (10.25), $U_{\lambda/2}$ определяется свойствами материала, из которого изготовлен модулятор. Параметры некоторых электрооптических кристаллов приведены в табл. 10.1. Из предыдущего изложения ясно, что величина $U_{\lambda/2}$ должна зависеть также от ориентации кристалла относительно напряженностей E светового и внешнего электрического полей. Оптимальная конфигурация показана на рис. 10.2.

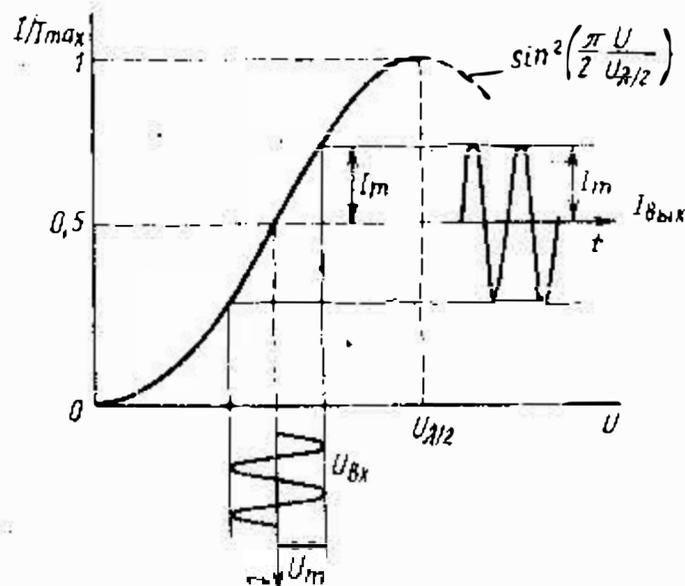


Рис. 10.4. Зависимость оптического пропускания электрооптического модулятора от приложенного напряжения.

Рабочая точка сдвинута за счет введения четвертьволновой пластинки так, что при $U=0$ $T=I/I_{\max}=0,5$. Показано, как внешнее синусоидальное напряжение модулирует интенсивность прошедшего света

оптическим кристаллом и одним из поляризаторов ставится четвертьволновая пластинка (см. § 3.6), как показано на рис. 10.3. Она осуществляет фазовый сдвиг между E_x и E_y — компонентами на $\pi/2$, что эквивалентно сдвигу рабочей точки модулятора на $U_{\lambda/4}$, как показано на рис. 10.4. В этом случае фазовый сдвиг между E_x и E_y — компонентами будет

$$\Delta\varphi_y - \Delta\varphi_x = \frac{\pi}{2} + \Delta\varphi = \frac{\pi}{2} + \pi \frac{U}{U_{\lambda/2}}$$

а оптическое пропускание

$$T = \frac{I}{I_0} = \sin^2 \left(\frac{\pi}{4} + \frac{\pi U}{2 U_{\lambda/2}} \right) = \frac{1}{2} \left(1 + \sin \frac{\pi U}{U_{\lambda/2}} \right) \approx 0,5 + \frac{\pi U}{2 U_{\lambda/2}}, \quad (10.26)$$

т. е. при $U/U_{\lambda/2} \ll 1$ электрооптический модулятор по схеме рис. 10.3 осуществляет амплитудную модуляцию, пропорциональную приложенному напряжению U .

Модулятор, схема которого приведена на рис. 10.3, называется *модулятором с продольным полем*. Он применяется для широких световых пучков. Очевидны два недостатка такого модулятора. Во-первых, электроды, с помощью которых прикладывается электрическое напряжение к кристаллу, должны пропускать свет,

Поскольку оптическое пропускание системы зависит от приложенного напряжения, то она может осуществлять амплитудную модуляцию света. При малых $U/U_{\lambda/2}$ из (10.24) имеем $I/I_0 \sim (U/U_{\lambda/2})^2$, т. е. модуляция не линейна. Функция (10.24) представлена на рис. 10.4. Видно, что для осуществления эффективной модуляции рабочую точку надо сдвинуть на $0,5U_{\lambda/2}$. Для этого можно подать на модулятор постоянное смещение, на которое накладывают переменный сигнал модуляции. Гораздо удобнее другой путь: между электро-

т. е. быть полупрозрачными, что вызывает нежелательные потери. Во-вторых, для работы такого модулятора требуются большие напряжения. Например, для модулятора на основе кристалла КДР, как нетрудно получить из (10.25) и табл. 10.1, для $\lambda = 1,06$ мкм $U_{\lambda/2} = 14,5$ кВ.

Избавиться от этих недостатков можно в *модуляторах с поперечным полем*, типичная схема которого приведена на рис. 10.5. В этом случае электрическое поле прикладывается перпендикулярно направлению распространения света, электроды не препятствуют его прохождению, а фазовая задержка, пропорциональная произведению поля на длину кристалла, зависит от отношения L/D и может быть увеличена при использовании длинных кристаллов.

В случае кристалла КДР его ориентация относительно направления распространения света и напряженности поля E показана на рис. 10.5. Внешнее поле прикладывается, как и ранее, вдоль оси z , совпадающей с оптической осью кристалла. Свет распространяется вдоль оси y' , вектор его поляризации находится в плоскости $x'z$ под углом 45° к оси z . В этом случае из (10.15) находим фазовую задержку

$$\Delta\varphi = \varphi_z - \varphi_x = \frac{2\pi}{\lambda} (n_o - n_e) L - \frac{2\pi}{\lambda} r_{63} n_o^3 \frac{UL}{D}. \quad (10.27)$$

Обратим внимание, что $\Delta\varphi$ содержит слагаемое, не зависящее от приложенного напряжения U . Сравнение (10.18) с (10.27) показывает, что напряжение полуволнового смещения в модуляторе с поперечным полем может быть уменьшено в L/D раз. К сожалению, эти модуляторы обладают очень малой апертурой и работают лишь с параллельными пучками.

Важным параметром модулятора, как говорилось выше, является требуемая управляемая мощность. Ее можно оценить, представив модулятор в виде конденсатора емкостью C , включенного в колебательный контур, ширина полосы пропускания которого $2\pi\Delta\nu = \Delta\omega \approx 2/RC$. Тогда требуемая управляемая мощность будет равна

$$P = \frac{U^2}{2R} = \frac{\pi}{2} U^2 C \Delta\nu. \quad (10.28)$$

Учитывая, что емкость конденсатора $C = \epsilon_0 \epsilon_r \frac{aL}{D}$, а переменная составляющая фазового сдвига

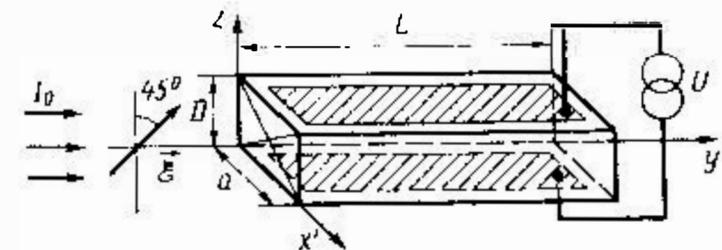


Рис. 10.5. Схема электрооптического модулятора с поперечным полем. Для кристалла КДР обозначения осей $x'y'z$ соответствуют рис. 10.2

Кристалл	Симметрия	Диапазон оптической прозрачности, мкм	Электрооптические параметры				Диэлектрическая проницаемость, ϵ/ϵ_0	
			Линейный электрооптический коэффициент r , 10^{-12} м/В	Длина волны λ , мкм	Показатель преломления			$n_0 \cdot r_i$, 10^{-12} м/В
					n_o	n_e		
KH ₂ PO ₄ (KDP)	Тетрагональная 42m	0,2 — 1,2	$r_{41} = 8,6$	0,55	1,51	1,47	29	20 (E c) 45 (E⊥c) 50 (E c) 12 (E⊥c)
			$r_{63} = 10,6$				36	
			$r_{63} = 23,6$				81	
NH ₄ H ₂ PO ₄ (ADP)	Тетрагональная 42m	0,13 — 1,2	$r_{41} = 28$	0,55	1,52	1,48	98	12 (E c)
			$r_{63} = 8,5$				29	
GaAs GaP ZnTe ZnSe β -ZnS CdTe Bi ₄ Ge ₃ O ₁₂	Кубическая 43m	1 — 15 0,56 — 10 0,57 — 25	$r_{41} = 1,6$	10,6	3,34	3,34	59	12,9 11,1 8,7 8,9 8,3 10,2 16
			$r_{41} = 1,06$				39	
			$r_{41} = 4,55$				135	
			$r_{41} = 3,9$				85	
			$r_{41} = 2,0$				38	
			$r_{41} = 2,0$				27	
LiNbO ₃	Тригональная 3m	0,4 — 4,5	$r_{33} = 30,8$	0,63	2,286	2,200	368	38 (E c) 80 (E⊥c)
			$r_{13} = 8,6$				103	
			$r_{42} = 28$				334	
LiTaO ₃	Тригональная 3m	0,4 — 4,5	$r_{33} = 30,3$	0,63	2,176	2,180	312	43 (E c)
			$r_{13} = 5,7$				59	
			$r_{42} = 20$				206	
BaTiO ₃	Тетрагональная 4mm	0,4 — 4,5	$r_{13} = 8,0$	0,55	2,437	2,365	116	106 (E c) 4300 (E⊥c)
			$r_{42} = 820$ (30° С)					

из (10.27) с учетом (10.25) равна $\Delta\varphi = \pi \frac{U}{U_{\lambda/2}} \frac{L}{D}$, получаем

$$P = \frac{\pi}{2} \epsilon_0 \epsilon_r V_{\lambda/2}^2 \Delta v \frac{aD}{L} \left(\frac{\Delta\varphi}{\pi} \right)^2. \quad (10.29)$$

Отсюда следует, что потребляемая модулятором мощность пропорциональна ширине полосы пропускания и квадрату глубины модуляции $\Delta\varphi$. Поэтому можно ввести критерий качества модулятора

$$q = \frac{P}{\Delta v (\Delta\varphi)^2}. \quad (10.30)$$

Из (10.27) видно, что для понижения рабочего напряжения целесообразно использовать тонкие и длинные кристаллы. Это же, согласно (10.29), приводит к уменьшению потребляемой мощности. В пределе толщина D ограничена длиной волны и мы приходим к волноводным модуляторам.

Рабочим элементом волноводного электрооптического модулятора является тонкопленочный оптический волновод. Наибольшее распространение получили модуляторы на основе кристаллов сегнетоэлектриков и полупроводниковых соединений A^{III}B^V. Волноводные модуляторы на сегнетоэлектриках, как правило, изготавливаются на основе кристаллов ниобата или танталата лития — LiNbO₃ и LiTaO₃, обладающих хорошим сочетанием электрооптических свойств (см. табл. 10.1). Для волноводных модуляторов с планарным волноводом достигнуты следующие параметры: рабочий диапазон длин волн от 0,4 до 1,5 мкм; полоса частот $\Delta\nu = 1...5$ ГГц; напряжение полуволнового смещения $U_{\lambda/2} = 10...20$ В, потребляемая мощность на единицу полосы частот $P/\Delta f = 0,2...2$ Вт/ГГц. К сожалению, в таких модуляторах потери света достигают 20 дБ и более.

Поскольку волноводные модуляторы наиболее часто применяются как элементы оптических интегральных схем, то особое значение приобретают полупроводниковые модуляторы, технология изготовления которых совместима с интегральной технологией электронных микросхем. Так как в кристаллах с центром инверсии электрооптический эффект отсутствует, то модуляторы не могут быть изготовлены на основе кремния или германия. Наиболее эффективны модуляторы на основе двойных гетероструктур, подобных тем, которые применяются в ДГС-лазерах. Как указывалось в § 9.5 и видно из рис. 9.23, z' и d' , рабочий слой обладает хорошими волноводными свойствами. Если этот слой сделать высокоомным, а прилегающие широкозонные области будут выполнять роль омических контактов, то такая структура

при соответствующей ориентации кристалла будет работать как модулятор с поперечным полем. Наиболее часто такие модуляторы изготавливают на основе арсенида галлия в системе $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$. Они работают в диапазоне от 1,0 до 10,6 мкм. Длина рабочего элемента 1...20 мм при толщине активной области 10...100 мкм. В интегрально-оптическом исполнении толщина может быть уменьшена до единиц и долей мкм. На длине волны $\lambda=1,15$ мкм достигнуты следующие характеристики: $U_{1/2}=10$ В при $\eta \approx 95\%$, $\Delta\nu \approx 4$ ГГц, $P/\Delta\nu=0,2$ Вт/ГГц. Как и в волноводных модуляторах на основе сегнетоэлектриков, потери в полупроводниковых модуляторах составляют около 20 дБ. В основном они вызваны поглощением на свободных носителях заряда.

В видимой области GaAs использовать нельзя, поэтому для аналогичных целей может быть применен фосфид галлия (GaP) и гетероструктуры в системе $\text{Al}_x\text{Ga}_{1-x}\text{P}/\text{GaP}$.

Возможны и другие типы полупроводниковых модуляторов, основанные, например, на использовании сильного электрического поля в p - n -переходе при приложении к нему напряжения в обратном направлении, на инжекции носителей заряда через p - n -переход, на смещении края собственного поглощения под действием электрического поля (эффект Франца — Келдыша) и др. У каждого из них имеются определенные достоинства и недостатки. Универсального модулятора, пригодного для всех потенциально возможных применений, не существует.

Магнитооптические модуляторы. Эти модуляторы работают на рассмотренном в § 10.1 эффекте Фарадея. Значения постоянной Верде в (10.7) изменяются от 4,0 рад/мТл для кварца до 82 рад/мТл для ZnS ($\lambda=589,3$ нм). Поэтому для ощутимого поворота плоскости поляризации ϕ необходимы сильные магнитные поля, коммутировать которые сложнее, чем электрические.

Схема построения амплитудного магнитооптического модулятора выглядит аналогично приведенной на рис. 10.3 с той разницей, что электрооптический кристалл должен быть заменен на магнитооптический элемент, помещенный в магнитное поле, направление которого совпадает с осью модулятора. Линейность эффекта Фарадея не требует применения четвертьволновой пластинки.

Акустооптические модуляторы. Их принцип действия основан на акустооптическом эффекте, связанном с изменением показателя преломления оптической среды под влиянием механических напряжений, сопровождающих прохождение акустической волны через эту среду. Акустическая волна длиной Λ вызывает пространственное изменение показателя преломления, обусловленное пьезооптическими эффектами (§ 10.1). Обычно используется линейный упругооптический (фотоупругий) эффект. С помощью

акустической волны, генерируемой, например, каким-либо пьезоэлектрическим устройством, в оптическом элементе модулятора создается заданный профиль показателя преломления. Проще всего осуществить периодическое изменение Δn , создавая для света своеобразную дифракционную решетку. Проходя через эту решетку или отражаясь от нее, световая волна будет испытывать дифракцию и отклоняться. Это отклонение с помощью системы линз и диафрагм может быть преобразовано в амплитудную модуляцию. Изменяя шаг решетки и ее глубину путем изменения частоты и амплитуды акустической волны, возможно осуществлять модуляцию света. Очевидно, что на этом же принципе могут работать системы отклонения и сканирования света. Поэтому более подробно мы их рассмотрим в следующем параграфе.

§ 10.3. ДЕФЛЕКТОРЫ

Оптические дефлекторы предназначены для управления направлением распространения светового луча в пространстве (*сканирования*). Простейший дефлектор представляет собой зеркало, угловым поворотом которого можно осуществлять угловое перемещение (отклонение) отраженного от него светового луча, реализуя, например, строчную развертку оптического изображения. В сочетании с другими методами пространственного преобразования лазерных пучков (см. § 3.6) такие системы могут быть довольно эффективны до частот порядка единиц килогерц. При больших частотах применяют электрооптические и акустооптические дефлекторы.

Как указывалось в § 3.3, практически все способы управления распространением световых потоков основаны на использовании явлений рефракции света в неоднородных структурах. Чтобы изменить направление светового пучка в пространстве, необходимо управлять изменением пространственного распределения показателя преломления. Если взять прямоугольную пластинку длиной L , показатель преломления которой в поперечном направлении x линейно изменяется по закону $n(x)=n_0+ax$, то плоская световая волна, распространяющаяся вдоль направления L , будет отклоняться на угол

$$\varphi = -L \, dn/dx. \quad (10.31)$$

Эту пластину можно изготовить из двух склеенных по диагонали призм (рис. 10.6), сделанных из электрооптических кристаллов, оптические оси z которых направлены навстречу друг другу и перпендикулярны плоскости xu . Кристаллы ориентированы так, что оси x' и y' на рис. 10.2 совпадают с осями x и y на рис. 10.6. Прикладывая электрическое поле в направлении z , согласно (10.16) в такой системе можно изменять показатель преломления

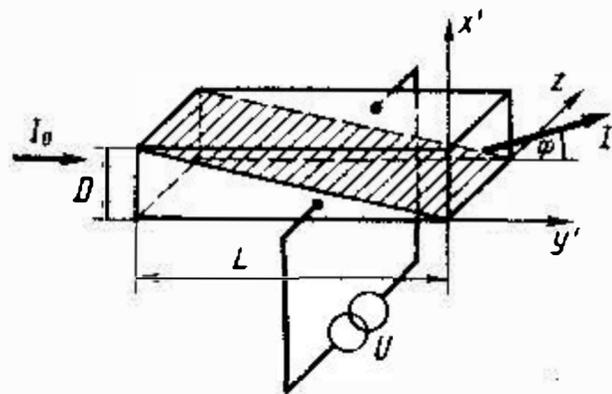


Рис. 10.6. Схема электрооптического дефлектора на двойной призме из кристаллов КДР. Обозначения осей x' и y' соответствуют рис. 10.2, а ось z у верхней и нижней призм направлены противоположно

для лучей, распространяющихся в нижней и верхней частях пластины, на величину

$$\Delta n = n_0^3 r_{63} E_z.$$

Это вызовет отклонение светового пучка на угол

$$\varphi = \frac{L}{D} n_0^3 r_{63} E_z, \quad (10.32)$$

где D — толщина пластины в направлении x . Так работает один из типов электрооптических дефлекторов.

Если по формуле (10.32)

найти угол φ , то он окажется довольно мал. Казалось бы, что этот угол отклонения в дальнейшем можно увеличить, например, с помощью короткофокусной линзы, выпуклого зеркала малого радиуса или какого-нибудь другого хитроумного приспособления. На самом деле каждый оптический пучок имеет конечный угол расходимости θ и основной характеристикой дефлектора является не угол отклонения φ , а параметр N , показывающий, во сколько раз φ превышает θ . Оценим наилучшую ситуацию, когда рассмотренный выше электрооптический дефлектор помещен в каустику гауссова пучка (рис. 2.21) с размером пятна $r_0 = D/2$. Тогда из (10.32) и (3.3) получаем

$$N = \frac{\varphi}{\theta} = \frac{\pi L n_0^3 r_{63}}{2\lambda} E_z. \quad (10.33)$$

Из сопоставления этого выражения с (10.24), (10.25) получаем, что электрическое поле, которое на расстоянии L индуцирует разность фаз, равную π , позволяет получить $N \approx 1$, что эквивалентно отклонению пучка на диаметр его пятна.

Значительно более хорошими характеристиками обладают *акустооптические дефлекторы*. Их принцип работы основан на взаимодействии света со звуковыми волнами. Напомним, что звуковая волна представляет собой периодические изменения плотности вещества (или напряжения деформации), которые распространяются со скоростью звука v_s . Считая, что одинаковые атомы (молекулы) вносят одинаковый вклад в показатель преломления, получаем, что изменение плотности среды приводит к прямо пропорциональному изменению ее показателя преломления n . На самом деле в кристаллах вследствие анизотропии их свойств фотоупругие эффекты описываются тензором пьезооп-

тических коэффициентов π_{ij} (10.8), но в первом приближении можно считать, что изменение показателя преломления Δn пропорционально корню квадратному из акустической мощности, поскольку последняя пропорциональна квадрату амплитуды волны.

Рассмотрим акустическую волну, распространяющуюся в направлении z в среде со скоростью v_s . Пусть она приводит к изменению показателя преломления

$$\Delta n(z, t) = \Delta n \sin(\Omega_s t - K_s z), \quad (10.34)$$

где $\Omega_s/K_s = v_s$, а $K_s = 2\pi/\Lambda$. На эту среду под небольшим углом φ_1 к направлению x падает плоская световая волна (рис. 10.7). Для акустической волны в образце может быть создан как режим бегущей волны, так и режим стоячей волны. Но поскольку $v_s \ll c$, то свет не будет замечать перемещения акустической волны. Поэтому мы можем оперировать «замороженной» акустической волной в момент времени t_0 . Области максимального сжатия на рис. 10.7 показаны горизонтальными сплошными линиями. В них наблюдается увеличение n . В областях разрежения, обозначенных горизонтальными пунктирными линиями, показатель преломления уменьшается. В результате мы имеем дифракционную решетку, при падении на которую световая волна будет отклоняться, если выполнены условия дифракции. Получим эти условия, используя для разнообразия не волновой подход, как это делается обычно, а корпускулярный. В этом случае падающую и диф-

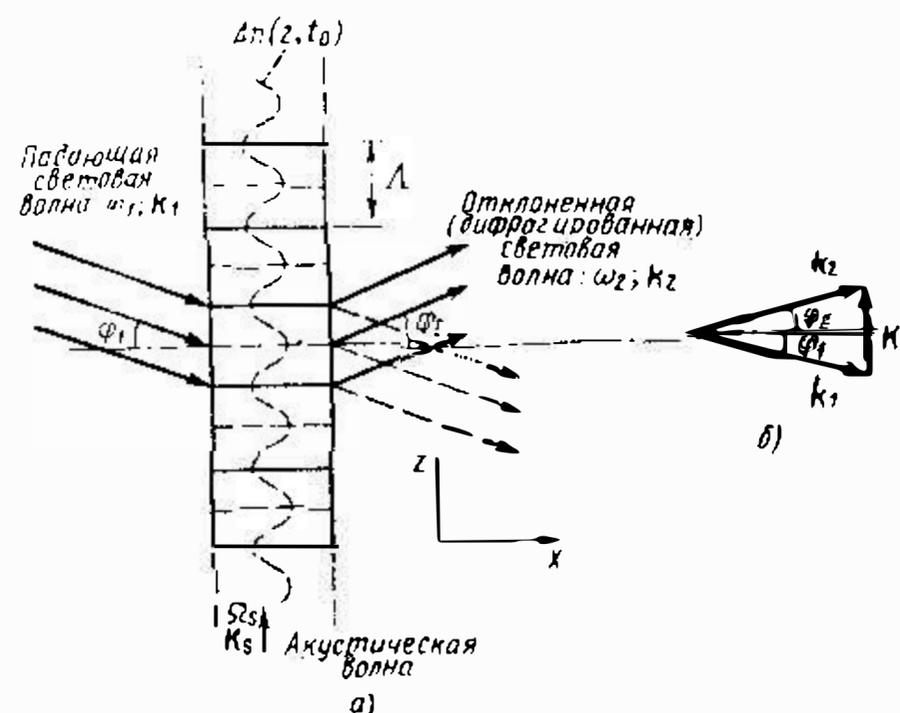


Рис. 10.7. Взаимодействие света с акустической волной и принцип работы акустооптического дефлектора

рагированную световые волны представим как поток фотонов с энергиями $\hbar\omega_1$ и $\hbar\omega_2$ и волновыми векторами \mathbf{k}_1 и \mathbf{k}_2 , а акустическую волну — как поток фононов с энергией $\hbar\Omega_s$ и волновым вектором \mathbf{K}_s . Тогда дифракция света на акустической волне может быть рассмотрена как ряд столкновений, приводящих с определенной вероятностью к исчезновению одного падающего фотона ($\omega_1; \mathbf{k}_1$) и одного фонона ($\Omega_s; \mathbf{K}_s$) с одновременным испусканием нового (дифрагированного) фотона ($\hbar\omega_2; \mathbf{k}_2$). Этот процесс будет возможен, если выполняются законы сохранения энергии (4.37) и волнового вектора (4.38), т. е. если выполнены условия:

$$\omega_2 = \omega_1 + \Omega_s, \quad (10.35)$$

$$\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{K}_s. \quad (10.36)$$

Поскольку $\omega_1 \gg \Omega_s$, то

$$\omega_2 \approx \omega_1; \quad \lambda_2 \approx \lambda_1,$$

и так как $\Lambda \gg \lambda_1 \approx \lambda_2$, то волновые векторы (но не их направления):

$$|\mathbf{k}_2| \approx |\mathbf{k}_1| \gg |\mathbf{K}_s|. \quad (10.37)$$

Эти соотношения показывают, как сделать *акустооптический дефлектор*. Световая волна должна быть почти перпендикулярна акустической волне, возбуждаемой в материале с большими значениями пьезооптических коэффициентов. Угол падения и угол выхода световой волны определяются из векторной диаграммы в правой части рис. 10.7. Поскольку $|\mathbf{k}_1| \approx |\mathbf{k}_2|$, то $\varphi_1 \approx \varphi_2$ и эти углы малы. Из рис. 10.7, б находим

$$\sin \varphi_2 \approx \frac{K_s}{2|\mathbf{k}_1|} = \frac{\lambda}{2n\Lambda}, \quad (10.38)$$

где n — показатель преломления, а λ/n — длина световой волны в материале. Это условие можно записать

$$2\Lambda \sin \varphi = \lambda/n,$$

которое совпадает с условием дифракции Брегга на решетке с периодом Λ . Такая дифракция называется *дифракцией Брегга*. Она подобна дифракции рентгеновских лучей на кристаллах. Для оценки φ рассмотрим дифракцию света с $\lambda = 1$ мкм на ультразвуковой волне с частотой 500 МГц. При скорости звука $v_s = 3000$ м/с имеем $\Lambda = v_s/v_s = 6$ мкм. Для $n = 1,7$ получаем $\varphi_2 = 5 \cdot 10^{-2}$ рад $\approx 2,8^\circ$.

Таким образом, если условие Брегга (10.38) выполнено, то диаграмма волновых векторов на рис. 10.7, б замкнута и $\varphi_1 = \varphi_2 = \varphi$. Будем теперь изменять частоту ультразвуковой

волны от v_s до $v_s + \Delta v_s$. Это приведет к изменению волнового вектора звуковой волны на $|\Delta \mathbf{K}_s| = \frac{2\pi \Delta v_s}{v_s}$. Оставляем угол падения

$\varphi_1 = \varphi$ неизменным. Модуль волнового вектора дифрагированной световой волны в силу (10.37) также остается неизменным. Поэтому конец этого вектора движется по окружности, как показано пунктирной линией на рис. 10.8. Акустическая волна не является идеально плоской хотя бы в силу наличия границ раздела. Поэтому пучок дифрагирует в направлении, соответствующем наименьшему отклонению на угол $\Delta\varphi$, причем, как видно из рис. 10.8,

$$\Delta\varphi = \frac{\Delta K_s}{k} = \frac{\lambda}{\pi v_s} \Delta v_s. \quad (10.39)$$

Следовательно, мы реализовали акустооптический дефлектор, который осуществляет отклонение световой волны путем изменения частоты ультразвуковой волны, причем $\Delta\varphi \sim \Delta v$.

Как говорилось выше, для оптического дефлектора важное значение имеет не сам угол отклонения φ , а его превышение над углом дифракционной расходимости пучка θ , которое показывает число возможных неперекрывающихся наложений отклоненного луча. Принимая угол дифракционной расходимости $\theta_D \approx \lambda/(\pi D)$, где диаметр пучка D определяется поперечными размерами модулятора, получаем

$$N = \frac{\Delta\varphi}{\theta_D} = \frac{\lambda \Delta v_s D}{v_s \lambda} = \Delta v_s \left(\frac{D}{v_s} \right) \quad (10.40)$$

или

$$N = \Delta v_s \tau, \quad (10.41)$$

где τ — время установления акустического режима в объеме материала; оно равно времени прохождения звука через диаметр оптического пучка $\tau = D/v_s$. Очевидно, что это и есть его постоянная времени.

Выражение (10.41) хорошо известно в электронике. Оно дает максимальное число степеней свободы сигнала. В нашем случае значения τ и Δv ограничены как физическими, так и техническими причинами. Увеличение τ ограничено не только необходимым быстродействием детек-

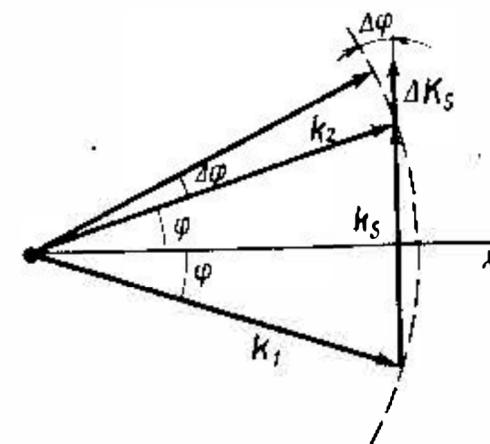


Рис. 10.8. Векторная диаграмма, поясняющая отклонение дифрагированного светового пучка на угол $\Delta\varphi$ при изменении частоты ультразвука от v_s до $v_s + \Delta v_s$

ктора, но и длиной акустооптического взаимодействия D , которое определяется поглощением акустических волн в материале. Расширение $\Delta\nu$ также ограничено, ибо пьезоэлектрический преобразователь обычно имеет частотную характеристику с ярко выраженным резонансным пиком.

Оценим возможности акустооптического дефлектора, изготовленного из стекла (тяжелый флинт, $v_s = 3100$ м/с; $n = 1,9$). Пусть диаметр оптического пучка $D = 1$ см, а ν_s изменяется от 100 до 150 МГц. Тогда $\tau = D/v_s = 3,2$ мкс и число разрешимых элементов $N = 160$, что существенно лучше, чем в электрооптическом дефлекторе, изображенном на рис. 10.6.

В выпускаемых промышленностью акустооптических дефлекторах число разрешаемых направлений достигает 1000 и более при постоянной времени $\tau \approx 1$ мкс и эффективности $\eta = (50...80)\%$.

Для снижения акустической мощности, требуемой для работы дефлектора, как и в модуляторах, используют акустооптическое взаимодействие в тонких волноводных слоях.

§ 10.4. ПРЕОБРАЗОВАТЕЛИ ЧАСТОТЫ

Преобразователи частоты представляют собой устройства управления лазерным излучением, предназначенные для изменения его частоты (длины волны). Различают *дискретные* и *непрерывные* преобразователи частоты. Среди дискретных преобразователей выделяют *умножители частоты*, осуществляющие генерирование частот, кратных основной частоте лазерного излучения. Они называются *генераторами гармоник*.

Принцип действия большинства преобразователей частоты основан на использовании нелинейных оптических эффектов (см. гл. 5). Наибольшее распространение среди дискретных преобразователей получили: умножители частоты (генераторы гармоник); преобразователи частоты на основе вынужденного комбинационного рассеяния света (комбинационные лазеры) и преобразователи частоты вверх (генераторы суммарных частот). Из непрерывных преобразователей частоты интерес представляют прежде всего параметрические генераторы света, а также акустооптические и электрооптические модуляторы частоты. Остановимся на некоторых из этих приборов.

Генератор второй гармоники. Его принцип действия рассмотрен в § 5.2. На нелинейный кристалл, ориентированный специальным образом для соблюдения условия волнового синхронизма (5.22), подается плоская линейно поляризованная световая волна, генерируемая лазером на частоте ω . За счет процессов нелинейной поляризации кристалл преобразует эту волну во вторую гармонику, так что на выходе появляется излучение удвоенной частоты. В силу причин, рассмотренных в § 5.2, этот процесс

будет эффективен лишь для достаточно интенсивного излучения при жестком соблюдении условия (5.22). Поэтому нелинейный кристалл-удвоитель частоты желательно термостабилизировать и помещать внутрь лазерного резонатора. Схема лазера с генератором второй гармоники, помещенном внутри резонатора, показана на рис. 10.9. Зеркала лазера R_1 и R_2 делаются непрозрачными для излучения на основной частоте ω , т. е. $R_1(\omega) = R_2(\omega) \approx 1$. Коэффициент пропускания выходного зеркала на удвоенной частоте $R_2(2\omega)$ подбирается так, чтобы обеспечить оптимальную связь. При выполнении этих условий коэффициент преобразования излучения во вторую гармонику может быть очень близок к 100%.

По аналогичной схеме могут быть построены умножители частоты — генераторы третьей, четвертой и т. д. гармоник, но их эффективность существенно ниже.

Наиболее часто удвоители частоты используются совместно с твердотельными неодимовыми лазерами, излучающими на длине волны 1,06 мкм (см. § 8.3). Промышленность выпускает твердотельные лазеры на иттрий-алюминиевом гранате со встроенным внутри резонатора удвоителем частоты. Они излучают в зеленой области спектра на длине волны 0,53 мкм, т. е. вблизи максимума чувствительности глаза (см. рис. 3.5) и в области прозрачности морской воды.

Параметрические преобразователи частоты. Принцип действия параметрических усилителей и генераторов света рассмотрен в § 5.3. Так же как и для удвоения частоты, для эффективного параметрического преобразования необходимо выполнить условие волнового синхронизма (5.27) и обеспечить высокие плотности световой энергии на частоте накачки ω_p . Плавная перестройка частоты осуществляется или поворотом нелинейного кристалла, как это показано на рис. 5.4, или изменением его температуры. Наиболее часто в качестве задающего лазера, осуществляющего накачку параметрического усилителя или генератора, используют лазер на ИАГ: Nd^{3+} ($\lambda_{\text{ген}} = 1,06$ мкм). Для получения большего нелинейного эффекта излучение лазера накачки фокусируется в нелинейный кристалл, помещенный в термостат. При применении в качестве нелинейного элемента кри-

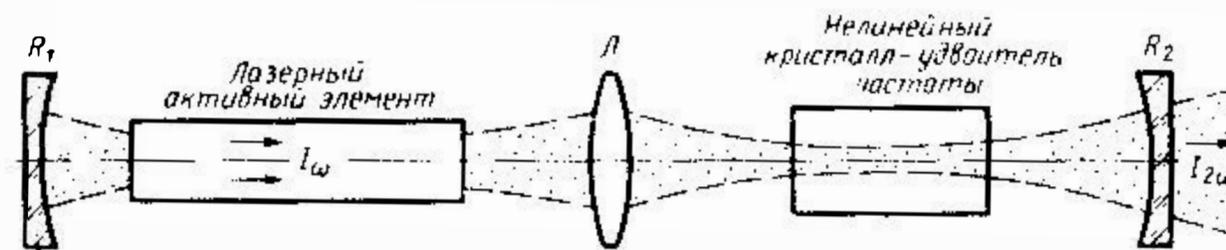


Рис. 10.9. Схема генератора второй гармоники с внутррезонаторным расположением нелинейного элемента

сталла ниобата лития диапазон перестройки может быть от 0,55 до 3,6 мкм при изменении температуры кристалла от 50 до 450°С. Коэффициент преобразования достигает 40%.

Достижения в области создания мощных полупроводниковых гетеролазеров открывают новые возможности создания миниатюрных и эффективных параметрических генераторов света с плавной перестройкой частоты.

Акустооптический модулятор частоты. Его принцип действия вытекает из соотношения (10.35) и пояснен на рис. 10.10. Пьезоэлектрический элемент создает в оптической среде, выполняющей роль звукопровода, бегущую ультразвуковую волну с частотой Ω и длиной волны Λ . Эта бегущая волна образует дифракционную решетку. При выполнении условия Вульфа — Брега

$$\sin \theta_B = \lambda/2\Lambda$$

происходит дифракция световой волны λ на ультразвуке. После прохождения такой решетки появляется несмещенная волна нулевого порядка дифракции и волна первого порядка дифракции, частота которой отличается от падающей ω_1 на Ω_s согласно (10.35). В зависимости от направления волновых векторов k_1 и k_s она может быть $\omega_1 + \Omega_s$ или $\omega_1 - \Omega_s$.

Основным элементом акустооптического модулятора частоты света является звукопровод, изготовленный из высококачественного оптического материала с хорошими акустооптическими характеристиками. Для света видимого, ближнего УФ- и ИК-диапазонов применяют стекло (тяжелый флинт), плавленый кварц, молибдат свинца ($PbMoO_4$), парателлурит (GeO_2), в ИК-диапазоне — германий и др. Рабочие частоты модуляторов — от десятков МГц до ГГц. Относительная эффективность модулятора, определяемая как отношение интенсивности светового пучка в первом порядке дифракции к интенсивности падающего света, для промышленных устройств составляет 60...70%, достигая иногда 90% и выше. Мощность, затрачиваемая на управление модулятором, составляет единицы ватт. Понятно, что перестройка частоты осуществляется в очень узком диапазоне. Как правило, он используется в различного рода лазерных измерительных устройствах типа гетеродинных интерферометров и доплеровских измерителей скорости.

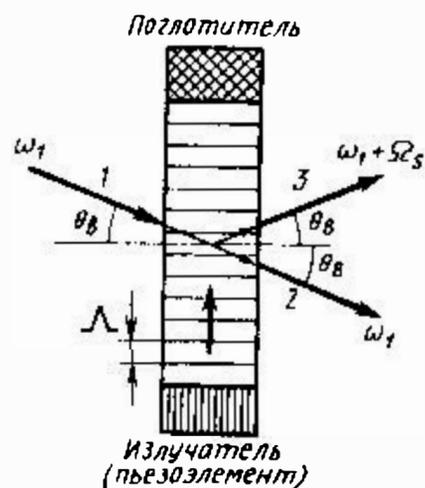


Рис. 10.10. Акустооптический модулятор частоты.
1 — падающий световой пучок;
2 — прошедший пучок (нулевой порядок дифракции);
3 — дифрагированный пучок, смещенный по частоте (первый порядок дифракции)

Люминесцентный преобразователь частоты. Это преобразователь частоты, действие которого основано на вынужденном излучении, возникающем в люминесцирующем веществе при лазерной накачке. Примеры таких лазеров-преобразователей мы рассматривали ранее. Наибольшее значение здесь имеют жидкостные лазеры на органических красителях, способные перестраивать длину волны излучения в широких пределах (§ 8.5).

§ 10.5. УПРАВЛЯЕМЫЕ ОПТИЧЕСКИЕ СИСТЕМЫ

Простейшие оптические элементы, такие, как зеркала, линзы, призмы и т. п., осуществляют хорошо известные из классической оптики преобразования световых пучков. Профиль оптической поверхности (отражающей или преломляющей) задается функциональным назначением элемента. Развитие вычислительной техники привело к появлению оптических систем, управляемых компьютерами. Простейшим примером такой системы является управляемое оптическое зеркало, выполненное в виде гибкой мембраны, поверхность которой может принимать ту или иную форму в зависимости от управляющих напряжений, приложенных к зеркалу. Такое гибкое зеркало, управляемое ЭВМ, позволяет по заданной программе сформировать необходимый волновой фронт светового пучка, изменять его во времени, менять интенсивность пучка и профиль его распределения в зоне фокусировки и т. п.

Оптические системы, управляемые от ЭВМ, способны решать новый класс задач лазерной техники, связанных с необходимостью пространственной модуляции фазы. Прежде всего — это задачи компенсации фазовых искажений, возникающих вследствие каких-либо дефектов системы.

Изменяя профиль фазы когерентного светового поля, можно практически без потерь энергии сформировать заданное распределение интенсивности и, в частности, сфокусировать его в линию, в том числе в двумерную кривую. Предназначенные для этого устройства называются *фокусаторами*. Формирование световых пучков с требуемым пространственным распределением интенсивности является одной из важных задач лазерной технологии.

Разработка управляемых оптических систем находится лишь в начальной стадии развития. Но уже в настоящее время на основе методов нелинейной оптики созданы системы, способные управлять длительностью и формой сверхкоротких лазерных импульсов. С помощью специальных методов *временной компрессии* (сжатия) световых импульсов достигнуты предельно малые длительности световых импульсов, равные одному периоду световых колебаний и составляющие $\tau \approx 10^{-14}$ с. Этот метод основан на самовоздействии света при его распространении в нелинейной среде, показатель преломления которой зависит от интенсивности.

§ 11.1. КЛАССИФИКАЦИЯ И ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ ФОТОДЕТЕКТОРОВ

Все фотоприемники по принципу действия можно разделить на две большие группы: *тепловые* и *фотонные*. В свою очередь фотонные приемники подразделяют на фотодетекторы, основанные на а) внешнем фотоэффекте (фотоэлектронные умножители и вакуумные фотоэлементы, электронно-оптические преобразователи) и б) внутреннем фотоэффекте (фоторезисторы, фотодиоды, фототранзисторы, фототиристоры и т. п.).

Для определения технических возможностей конкретного фотоприемника используются следующие характеристики:

1. Спектральная (монохроматическая) чувствительность S_λ — мера реакции фотоприемника на оптическое излучение с длиной волны λ :

$$S_\lambda = dU/d\Phi_\lambda \text{ [В/Вт]}$$

или

$$S_\lambda = dI/d\Phi_\lambda \text{ [А/Вт]}. \quad (11.1)$$

Для тепловых приемников S_λ не зависит от длины волны, а для фотонных приемников существует максимальная (пороговая) длина волны λ_m , выше которой энергии фотона $h\omega = hc/\lambda$ недостаточно для возникновения фотоэффекта.

На рис. 11.1 представлены спектральные характеристики идеализированного теплового и фотонного приемников. Для фотонных детекторов наряду с S_λ применяют понятие квантового выхода фотоответа β как отношение числа носителей заряда, генерируемых за счет внешнего или внутреннего фотоэффекта, к числу падающих фотонов. В идеальном фотонном детекторе $\beta = 1$ при $\lambda < \lambda_m$ и $\beta = 0$ при $\lambda > \lambda_m$. Обратите внимание, что при $\beta = \text{const}$ в коротковолновой области S_λ линейно уменьшается с уменьшением λ , так как уменьшается число фотонов при $\Phi_\lambda = \text{const}$.

2. Интегральная чувствительность S — мера реакции фотоприемника на световой поток Φ заданного спектрального состава

$$S = \frac{dU}{d\Phi} \text{ [В/Вт]}; \quad S = \frac{dI}{d\Phi} \text{ [А/Вт]}. \quad (11.2)$$

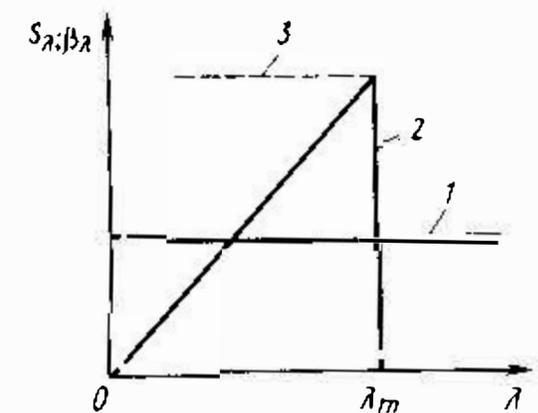


Рис. 11.1. Идеальные спектры фоточувствительности тепловых (1) и фотонных (2) детекторов. Пунктирная кривая (3) — квантовый выход идеального фотонного фотодетектора

ГЛАВА 11 ФОТОПРИЕМНИКИ

Фотоприемники предназначены для преобразования оптического сигнала в электрический. Они называются также *фотодетекторами*, так как осуществляют детектирование оптического сигнала, т. е. его демодуляцию. Фотоприемник как оптический детектор должен обладать: а) большим откликом на входное воздействие оптического сигнала; б) низким уровнем собственных шумов; в) широкой полосой пропускания, согласованной со спектром входного сигнала; г) линейностью характеристик. Эти требования противоречивы и их невозможно в максимальной степени реализовать в одном приборе, поэтому невозможно создать один универсальный фотоприемник, пригодный на все случаи жизни.

К фотоприемникам также относят *фотопреобразователи*, основной целью которых является эффективное преобразование световой энергии заданного спектрального состава в электрическую. Типичным примером здесь служат солнечные батареи.

Фотоприемник является первым и основным элементом системы демодуляции и обработки оптического сигнала. Как и в радиоэлектронике, системы *детектирования* в оптической электронике разделяют на две группы: а) непосредственного детектирования и б) детектирования с преобразованием. Последняя аналогична гетеродинному приему. Она применима для детектирования лишь когерентного светового потока, который предварительно смешивается с когерентным излучением опорного сигнала, генерируемого лазером и выполняющим функцию местного гетеродина. Оптическое детектирование с преобразованием, как правило, осуществляется в длинноволновом оптическом диапазоне с $\lambda \geq 10 \text{ мкм}$. Наиболее часто применяют схемы прямого детектирования.

Для идеального теплового приемника $S=S_1$ и не зависит от спектра Φ . Поэтому такие приемники используются для спектральной градуировки. Для фотонных приемников величина S зависит как от спектра фоточувствительности приемника, так и от спектра регистрируемого светового потока. Наиболее часто в качестве эталонного светового потока для определения S используют излучение абсолютно черного тела (1.107) с заданной температурой T или излучение эталонной лампы накаливания с вольфрамовой нитью. Если температура нити накала равна 2850 К, то такой режим принято называть «режим А».

3. Минимально различимый сигнал Φ_{\min} — та величина светового потока, измеряемая в [Вт], которая на выходе фотоприемника создает сигнал, равный шуму. Так как интенсивность белого шума пропорциональна корню квадратному из полосы пропускания Δf усилительного тракта, то вводят следующую характеристику.

4. Эквивалентная мощность шума NEP* — та величина светового потока, которая на выходе фотоприемника в единичной полосе частот вызывает сигнал, равный шуму,

$$\text{NEP} = \frac{\Phi}{\Delta f} \frac{U_{\text{ш}}}{U_c} [\text{Вт/Гц}^{1/2}]. \quad (11.3)$$

5. Обнаружительная способность D

$$D = \frac{1}{\text{NEP}} [\text{Гц}^{1/2}/\text{Вт}].$$

Эта величина зависит от площади приемника A , так как шум пропорционален \sqrt{A} .

6. Детектирующая способность D^* , называемая также *нормированной обнаружительной способностью*

$$D^* = \sqrt{A} D [\text{см} \cdot \text{Гц}^{1/2}/\text{Вт}]. \quad (11.4)$$

Это наиболее объективная и важная характеристика фотодетектора данного типа, поскольку она не зависит от его площади и полосы частот усилителя.

7. Инерционность — способность фотоприемника без искажения регистрировать быстрые изменения интенсивности светового потока. Она характеризуется или граничной частотой f_{\max} , при которой чувствительность фотоприемника падает в заданное число раз (обычно в 2 или e раз), или постоянной времени τ (для линейных процессов, см. рис. 4.47, а).

Для детекторов, фотоответ которых имеет экспоненциальные законы нарастания и спада с постоянной времени τ подобно (4.102), (4.103), зависимость чувствительности S или S_1 от частоты модуляции света f имеет вид

$$S(f) = \frac{S(0)}{\sqrt{1 + 4\pi^2 f^2 \tau^2}}. \quad (11.5)$$

Принцип действия *тепловых фотоприемников* основан на регистрации изменения свойств материала при изменении его температуры вследствие поглощения оптического излучения. Существуют различные типы тепловых фотоприемников, основанных на различных эффектах. Среди них наиболее распространены: а) *болометры*, использующие изменение сопротивления тонкой металлической, полупроводниковой или сверхпроводящей пленки; б) *термоэлектрические* детекторы типа термопар или термостолбиков, использующие эффект возникновения термоЭДС на контактах двух металлов; в) *пирозлектрические* приемники, основанные на пирозлектрическом эффекте в пирозлектрических, в том числе в ферроэлектрических кристаллах вблизи температуры Кюри; г) *оптико-акустические* приемники (ОАП), называемые иногда шевматическими ИК-детекторами или элементами Голя, использующие периодическое расширение и сжатие газа при его нагреве от промодулированного по амплитуде оптического излучения, поглощаемого тонкой мембраной.

Инерционность тепловых приемников велика (> 10 мс), а чувствительность сравнительно низка ($D^* \approx 10^8 \dots 10^{10}$ см \cdot Гц^{1/2} / Вт). Поэтому в системах передачи информации они не используются. Тепловые приемники применяются там, где необходимо обеспечить постоянство спектральной чувствительности, а также в далекой ИК-области спектра.

Фотонные приемники эффективно работают в той области спектра, где энергия фотона существенно превышает kT . В случае, когда тепловая энергия сравнима или превышает энергию фотона $\hbar\omega$, тепловое возбуждение действует активнее оптического и эффективность фотонного приемника резко падает. Поэтому фотонные приемники, предназначенные для работы в области $\lambda \geq 3$ мкм, как правило, требуют охлаждения тем более глубокого, чем больше рабочая длина волны.

Фотонные приемники, принцип действия которых основан на использовании внешнего или внутреннего фотоэффектов, обладают малой инерционностью, большой чувствительностью и высокой обнаружительной способностью. В ряде современных приборов достигнуты значения этих величин, близкие к своему теоретическому пределу. Поэтому в оптической электронике применяются в основном фотонные приемники, на свойствах которых мы остановимся ниже.

*NEP — от англ. Noise Equivalent Power — мощность, эквивалентная шуму.

Оптическую информацию, передаваемую или принимаемую в оптической электронике, можно разделить на два вида: 1) оптические сигналы, дискретные во времени и пространстве и 2) оптические образы или картины. Соответственно все фотоприемники можно разбить на две группы.

1. *Дискретные*, как правило, одноэлементные фотоприемники с малой рабочей площадью, предназначенные для приема коротких оптических импульсов, обладающие высокой спектральной чувствительностью S_λ в заданной области спектра, большой дегактирующей способностью D^* и малой инерционностью τ . Лучшим сочетанием параметров в этой группе приборов обладают фотодиоды, особенно *p-i-n* и лавинные фотодиоды.

2. Фотоприемники, предназначенные для восприятия световых образов. Как правило, это *многоэлементные фотоприемники* с самосканированием и высокой пространственной разрешающей способностью, обладающие хорошей чувствительностью в сравнительно широком спектральном интервале. Лучшими характеристиками из этой группы приемников обладают фоточувствительные приборы с зарядовой связью.

§ 11.2. ФОТОЭЛЕКТРОННЫЕ УМНОЖИТЕЛИ

Принцип действия *фотоэлектронного умножителя (ФЭУ)* основан на выбивании электронов с поверхности световыми квантами $h\omega$ (внешний фотоэффект) и последующем усилении электронного потока с помощью вторичной электронной эмиссии. Фотоумножитель обладает низкими шумами и высоким внутренним сопротивлением по току. Он является самым чувствительным фотоприемником и одним из наиболее чувствительных приборов, созданных человеком. При оптимальных режимах работы он может регистрировать отдельные фотоны и световые мощности на уровне 10^{-19} Вт.

На рис. 11.2 показана схема ФЭУ с боковым оптическим входом. Он состоит из *фотокатода ФК* и нескольких электродов, называемых *динодами*, которые обозначены цифрами с 1 по 8. На них через штыревые выводы *B* с помощью делителя подается

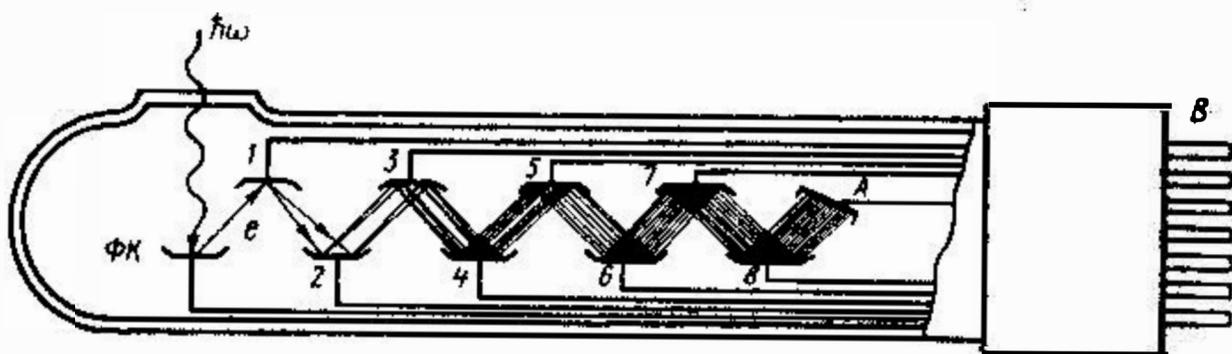


Рис. 11.2. Схема фотоэлектронного умножителя с боковым оптическим входом

~~внешнее напряжение~~ таким образом, что по мере удаления от катода потенциал каждого последующего динода увеличивается примерно на 100 В. Последний электрод *A* — *анод* предназначен для сбора электронов. Вся система помещена в вакуумную колбу. Вакуум должен быть достаточно высоким, чтобы исключить столкновение электронов с молекулами газа. Форма динодов подобрана так, что электроны, испускаемые фотокатодом, фокусируются электростатическим способом и ускоряются к первому диноду, достигая его с энергией порядка 100 эВ и выбивая с его поверхности вторичные электроны (вторичная электронная эмиссия). Этот процесс повторяется на каждом диноде и приводит к значительному усилению начального фототока. Если коэффициент вторичного электронного умножения принять равным q , т. е. считать, что на каждый первичный электрон возникает q вторичных электронов, а число динодов N , то полное усиление по току между катодом и анодом составит $K = q^N$. Типичное значение $q \approx 5$, так что при $N = 9$ получаем $K \approx 2 \cdot 10^6$.

Главной частью ФЭУ является фотокатод. Он преобразует оптическое излучение в поток электронов за счет внешнего фотоэффекта, как это показано на энергетических диаграммах рис. 11.3. Если материалом фотокатода является металлическая плен-

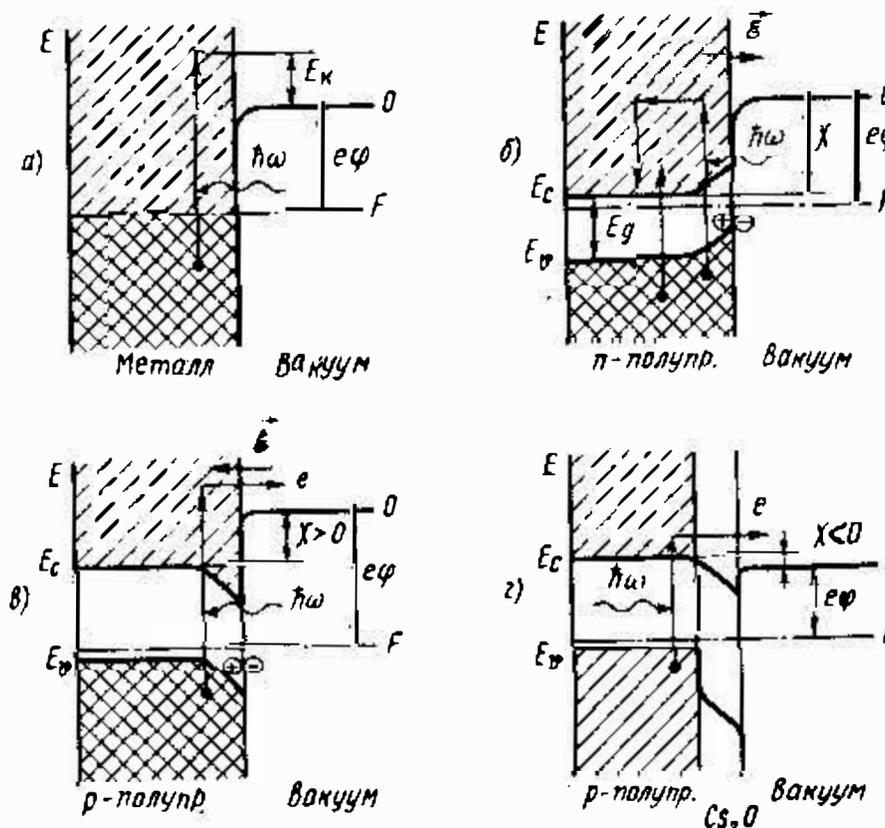


Рис. 11.3. Упрощенные энергетические диаграммы фотокатодов:

а) металлического; б) полупроводникового *n*-типа; в, г) полупроводникового *p*-типа с положительным (в) и отрицательным (г) зарядом

ка, то электрон с энергией вблизи уровня Ферми F , взаимодействуя с фотоном $\hbar\omega$, переходит в состояние с энергией $F + \hbar\omega$. Если эта энергия превышает работу выхода $e\phi$ электрона из металла в вакуум, то электрон может выйти наружу, обладая кинетической энергией

$$E_k = \hbar\omega - e\phi. \quad (11.6)$$

Отношение числа испущенных электронов к числу поглощенных фотонов называется квантовой эффективностью фотокатода. Очевидно, что существует минимальная энергия фотонов $\hbar\omega_0 = e\phi$ и максимальная длина волны $\lambda_0 = hc/e\phi$, определяющие «красную границу» фоточувствительности.

Из всех металлов наименьшей работой выхода $e\phi \approx 2$ эВ обладает цезий (Cs). Еще меньше работа выхода у окиси цезия ($e\phi \approx 1,1$ эВ), являющейся полупроводником. Как видно из сравнения диаграмм рис. 11.3, б, в, для фотокатодов желательнее использовать полупроводники p -типа, где благоприятный изгиб зон при закреплении уровня Ферми у поверхности понижает электронное сродство χ , а приповерхностное поле E способствует повышению квантового выхода фотоэмиссии.

При большой работе выхода $e\phi > 3E_g$, велика вероятность потери энергии возбужденного электроном фотона за счет процесса ударной ионизации, как показано на рис. 11.3, б. Поэтому для фотокатодов используются материалы с низкой работой выхода электронов с поверхности. Обычно это полупроводниковые соединения на основе Ag-O-Cs или Sb-Cs с работой выхода около 1,1...1,8 эВ (4,5 эВ для обычных металлов). Длинноволновая граница фоточувствительности ФЭУ с такими фотокатодами

лежит в районе (0,7...1,0) мкм. Спектральные зависимости фоточувствительности ряда промышленных фотокатодов, имеющих стандартные обозначения S1 (Ag-O-Cs), S11 (Cs₂Sb) и S20 (Cs-Na₂KSb), приведены на рис. 11.4. Пунктирными кривыми там же приведены спектральные чувствительности для идеальных фотокатодов с квантовыми эффективностями 0,1; 1,0 и 10%.

Работа выхода электрона очень сильно зависит от состояния поверхности ма-

териала. Поэтому в создании фотокатодов важную роль играет формирование поверхностных покрытий с малой работой выхода. Часто используют пленки цезия и его окисла толщиной несколько атомарных слоев. В результате специальной обработки (активирования) формируется фотокатод, представляющий собой гетероструктуру, образованную низкоомным полупроводником p -типа и тонким слоем типа Cs₂O с низким электронным сродством. При применении полупроводников $A^{III}B^V$ (GaAs, Ga_xIn_{1-x}As и Ga_xIn_{1-x}As_{1-y}Py) может быть достигнута ситуация, изображенная в упрощенном виде на рис. 11.3, г, когда $e\phi < E_g$ и $\chi < 0$. Такие фотокатоды называются *фотокатодами с отрицательным электронным сродством*. Они обладают наибольшей эффективностью. В качестве примера на рис. 11.4 приведена спектральная зависимость S_λ для арсенид-галлиевого фотокатода GaAs-Cs₂O с отрицательным электронным сродством.

Часто фотокатоды делают полупрозрачными, так что выход фотоэлектронов может быть осуществлен как бы «на просвет», а не «на отражение», как это показано на рис. 11.2. Такие фотокатоды используются в ФЭУ с торцевым оптическим входом.

Быстродействие обычных ФЭУ составляет 10...30 нс. За счет применения диодной системы со специальной формой электродов это значение может быть улучшено до 1 нс и менее. При использовании скрещенных магнитных и электрических полей в системе фокусировки достигается $\tau \approx 0,1$ нс, т. е. полоса частот на уровне 10 ГГц.

При всех своих положительных качествах фотоэлектронные умножители как приборы оптической электроники обладают недостатками, присущими большинству вакуумных приборов: высокие рабочие напряжения ($U_{раб.} \approx 500...1500$ В) и низкая прочность.

§ 11.3. ФОТОРЕЗИСТОРЫ

Фоторезистор представляет собой простейший тип фотонного фотоприемника, принцип действия которого основан на рассмотренном в § 4.7 фоторезистивном эффекте в однородном полупроводнике. Поскольку в разных спектральных диапазонах фотопроводимость полупроводника определяется различными по своей природе процессами собственного или примесного поглощения, то различают *собственные* и *примесные фоторезисторы*. Первые, как правило, предназначены для работы в видимой и ближней ИК-областях спектра. Их спектральный диапазон определяется шириной запрещенной зоны используемого материала. Вторые предназначены для работы в средней и дальней ИК-областях спектра и их спектральный диапазон определяется энергией ионизации примеси в полупроводнике. Примесные

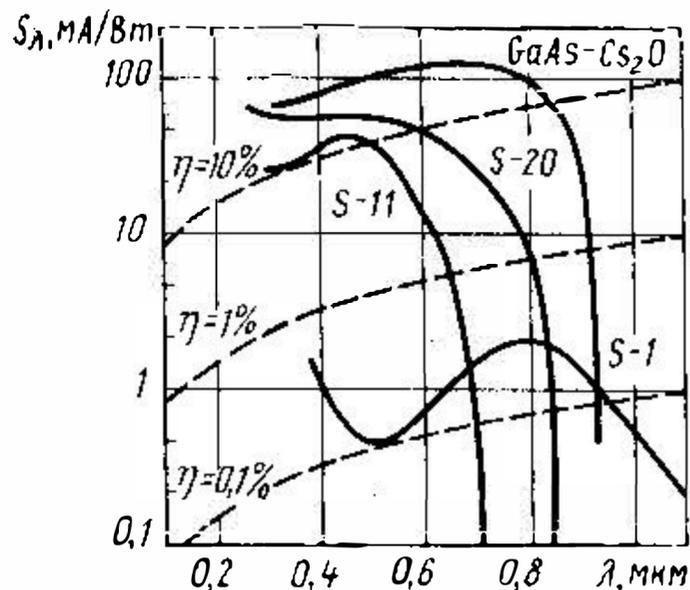


Рис. 11.4. Спектральные чувствительности S_λ наиболее распространенных типов фотокатодов

фоторезисторы, как правило, требуют глубокого охлаждения.

Фоторезистор выполняется в виде тонкой пластинки полупроводника, изготовленного из монокристалла или монокристаллической (иногда — поликристаллической) пленки с омическими контактами. Эти контакты наносятся или на противоположных концах пластинки, или в виде «гребенки», как показано на рис. 11.5. Фоточувствительный элемент закрепляется на специальном держателе.

Фоторезистор как фотоприемник характеризуется параметрами, рассмотренными в § 11.1. Кроме того, его специфическими характеристиками являются темновое R_T и световое $R_{св}$ сопротивления и их отношение $R_T/R_{св}$, определяемое как кратность изменения сопротивления фоторезистора в темноте и при заданной засветке, а также коэффициент внутреннего усиления фототока K_T .

Типовая схема включения фоторезистора как детектора амплитудно-модулированных оптических сигналов приведена на рис. 11.6. Эта схема, хотя и кажется отличной от схемы рис. 4.45, на самом деле ей аналогична, если принять во внимание внутреннее сопротивление измерительных приборов. Сопротивление нагрузки R_H , с которого снимается сигнал, необходимо согласовывать с сопротивлением фоторезистора и выбирать, исходя из технических требований. Разделительная емкость C при измерении постоянного или низкочастотного сигнала может отсутствовать.

Как следует из § 4.7, рассчитывать фототок в общем случае довольно затруднительно, поскольку в фоторезисторе, как и в любом другом фотонном фотоприемнике, необходимо учитывать три основных процесса: 1) генерацию носителей светом, которая происходит неравномерно по толщине фоточувствительного слоя; 2) перенос носителей заряда путем диффузии и дрейфа с возможным их умножением за счет внутреннего механизма усиления тока; 3) взаимодействие тока с внешней цепью, обес-

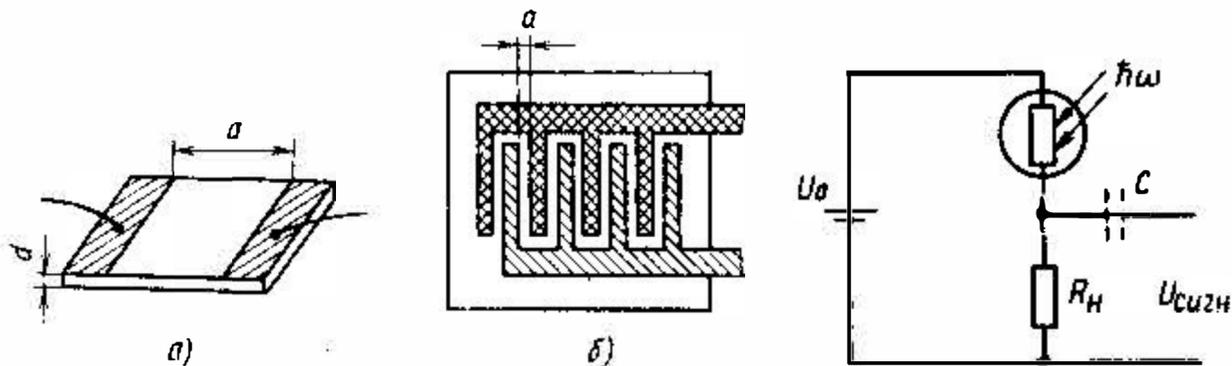


Рис. 11.5. Пленочные фоторезисторы: а) с двумя параллельными электродами, б) с гребенчатыми электродами

Рис. 11.6. Типовая схема включения фоторезистора как фотоприемника

печаивающее получение выходного сигнала. Поэтому для оценки параметров фоторезисторов в большинстве случаев допускают упрощения, предполагая, что:

фоторезистор представляет собой параллелепипед (см. рис. 4.45);

рабочая область фоторезистора освещена равномерно по площади, излучение полностью поглощается в полупроводнике и генерируемые носители равномерно распределяются по толщине d рабочего слоя;

дрейф и рекомбинация носителей заряда в объеме характеризуются некоторыми усредненными постоянными значениями подвижности (μ_n ; μ_p) и временами жизни (τ_n ; τ_p);

полупроводник обладает монополярной фотопроводимостью, для конкретности электронной; это означает, что генерируемые дырки быстро захватываются центрами рекомбинации, так что $\tau_n \gg \tau_p$ и $\Delta\sigma = e\mu_n \delta n$;

реализуется высокий уровень возбуждения, т. е. $\delta n \gg n_0$ (или $\delta p \gg p_0$).

При этих упрощениях для рис. 4.45 в соотношении (4.98) $k_{\omega} I = \Phi_{\omega} / V$, где Φ_{ω} — мощность светового излучения, поглощенного в фоторезисторе, а $V = abd$ — его объем. Учитывая, что фототок $I_{\phi} = U_R / R_{св}$, где U_R — напряжение на фоторезисторе, получаем для статических параметров:

$$\frac{R_T}{R_{св}} = \frac{\delta\sigma}{\sigma_T} = \frac{e\mu_n \tau_n \Phi_{\omega}}{\sigma_T V \hbar\omega} = \beta \frac{\Phi_{\omega}}{V \hbar\omega}; \quad (11.7)$$

$$S_{\lambda} = \frac{\delta I_{\phi}}{\delta \Phi_{\omega}} = \frac{U_R \delta\sigma bd/a}{\delta \Phi_{\omega}} = \beta \frac{e\mu_n \tau_n U_R}{\hbar\omega a^2}; \quad (11.8)$$

$$K_T = \frac{I_{\phi}/e}{\Phi_{\omega}/\hbar\omega} = S_{\lambda} \frac{\hbar\omega}{e} = \beta \mu_n \tau_n \frac{U_R}{a^2} = \beta \frac{\tau_n}{t_{пр}}. \quad (11.9)$$

В последнем соотношении, определяющем коэффициент внутреннего усиления фототока как отношение числа электронов I_{ϕ}/e , перенесенных фототоком, к числу поглощенных фотонов $\Phi_{\omega}/\hbar\omega$, величина

$$t_{пр} = \frac{a}{v_{ср}} = \frac{a}{\mu_n E} = \frac{a^2}{\mu_n U_R} \quad (11.10)$$

есть среднее время пролета электрона через фоторезистор. Из (11.9) вытекает физический смысл коэффициента K_T . Для образцов с большим временем жизни τ_n и малым расстоянием a между

контактами коэффициент усиления K_I может быть существенно больше единицы, достигая 10^5 .

Анализ выражений (11.7) — (11.9) показывает, что для достижения высоких значений R_T/R_{cb} , S_I и K_I необходимо применять полупроводники с высокими значениями подвижности μ и времени жизни носителей заряда τ . Поверхностная рекомбинация вследствие (4.99) должна быть уменьшена до минимума. Спектральная зависимость S_I при соблюдении указанных выше допущений будет следовать кривой 2 на рис. 11.1, т. е. линейно уменьшаться с уменьшением λ до λ_m , а при λ_m иметь резкий порог.

Из (11.8) и (11.9) видно, что коэффициент усиления K_I и фоточувствительность S_I растут пропорционально U_R/a^2 . Это справедливо до той поры, пока время пролета носителей t_{np} не станет соизмеримо со временем диэлектрической (максвелловской) релаксации τ_M , когда оно не будет зависеть ни от U_R , ни от a . Поэтому предельное значение коэффициента усиления ограничено величиной $K_{I_{max}} = \tau_n/\tau_M$. В низкоомных полупроводниках, кроме того, ограничение K_I может быть обусловлено насыщением дрейфовой скорости и уменьшением μ_n при увеличении E .

Для достижения большой кратности изменения сопротивления фоторезистора R_T/R_{cb} , как видно из (11.7), желательно уменьшать V при $\Phi_{\omega} = \text{const}$ или при постоянной интенсивности излучения $I_{\omega} = \Phi_{\omega}/(ab)$ надо уменьшать толщину d . Однако это справедливо только до $d \geq 1/k_{\omega}$, поскольку в противном случае не будут выполняться сделанные выше допущения. Поэтому оптимальная толщина фоторезистора d выбирается как $d \approx 2/k_{\omega}$. В области собственного поглощения в полупроводниках с прямой структурой зон $k_{\omega} \approx 10^3 \dots 10^5 \text{ см}^{-1}$, в то время как в примесной области показатель поглощения k_{ω} существенно меньше (см. § 4.4).

Соотношения (11.7) — (11.9) записаны для стационарной фотопроводимости. В динамическом режиме быстрое изменение Φ_{ω} вызовет изменение δn и фототока I_{ϕ} с некоторым запаздыванием τ , определяемым в случае линейной рекомбинации соотношениями типа (4.102) и (4.103). Для светового потока, промодулированного по амплитуде по синусоидальному закону с частотой f , это запаздывание вызовет фазовый сдвиг, так что зависимость коэффициента усиления K_I от частоты модуляции f светового потока Φ_{ω} можно записать в виде

$$K_I(f) = K_I(0) \frac{1}{\sqrt{1 + 4\pi^2 f^2 \tau^2}} = \frac{\tau}{t_{np}} \frac{1}{\sqrt{1 + 4\pi^2 f^2 \tau^2}} \quad (11.11)$$

где τ — время жизни носителей заряда. Отсюда видно, что произведение коэффициента усиления тока на ширину полосы Δf при больших f

$$K_I(f) \Delta f = \frac{1}{2\pi t_{np}} \quad (11.12)$$

не зависит от времени жизни носителей заряда и определяется только временем их пролета. В режиме насыщенного дрейфа при $v \leq v_{max} \approx 10^5 \text{ м/с}$ и $a \approx 10 \text{ мкм}$ из (11.10) и (11.12) получаем $K_I \Delta f \leq 2 \text{ ГГц}$. Это означает, что даже при оптимальных условиях фоторезисторы не могут обеспечить детектирования оптических сигналов с высокой частотой модуляции (для сравнения из § 11.2 нетрудно определить, что в ФЭУ величина $K_I \Delta f$ на 6...7 порядков выше). Поэтому фоторезисторы в основном применяются для регистрации сравнительно медленно протекающих процессов.

Если фоторезистор применяется для детектирования слабых оптических сигналов, то важнейшей его характеристикой становится детектирующая способность D^* (11.4) и определяющую роль начинают играть шумовые характеристики (естественно, что шумы усилителя и всех электрических контактов должны быть исключены). На рис. 11.7 в двойном логарифмическом масштабе представлена качественная зависимость от частоты f величины шумового тока, представляющего собой корень квадратный из среднеквадратичной флуктуации тока $\sqrt{\Delta I^2}$. В низкочастотной области до частоты f_1 порядка 1 кГц преобладает так называемый «1/f-шум» или «фликер-шум». Этот шум характерен для большинства полупроводниковых приборов, хотя природа его возникновения до сих пор не вполне ясна. Его можно оценить как

$$\Delta I_{ш}^{(1/f)} = I (B \Delta f / f)^{1/2}, \quad (11.13)$$

где постоянная B имеет порядок 10^{-11} , а I — полный ток через фоторезистор.

Основным источником шума в полупроводниковых фотодетекторах являются процессы генерации и рекомбинации электронно-дырочных пар. Этот шум называется генерационно-рекомбинационным. Он не зависит от частоты и является преобладающим в области от f_1 до $f_2 \approx 1/2\pi\tau$, где τ — время жизни неосновных носителей заряда. Вклад в генерационно-рекомбинационный шум вносят как оптические, так и тепловые процессы возбуждения. Для уменьшения последних рабочая температура фоторези-

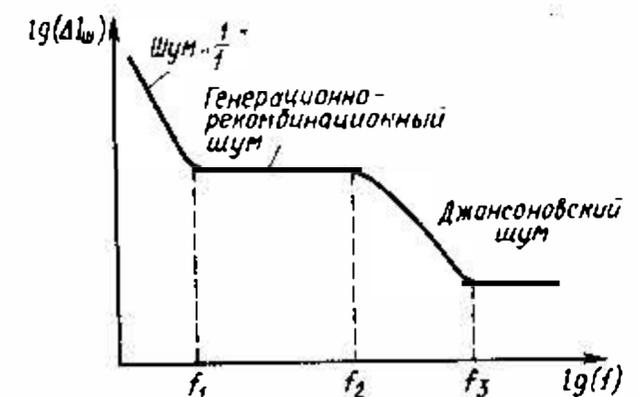


Рис. 11.7. Схематическая диаграмма спектра шумов в фотодетекторе

сторы с шириной запрещенной зоны E_g должна быть $T < E_g/25k$. В этом случае генерационно-рекомбинационный шум можно записать

$$\Delta I_{ш}^{(2-p)} = \left(\frac{4I_e K_I \Delta f}{1 + 4\pi^2 f^2 \tau^2} \right)^{1/2} \quad (11.14)$$

Для высоких частот $f > f_2 \approx (1/2\pi\tau)$ генерационно-рекомбинационный шум уменьшается с уменьшением частоты. При очень больших частотах $f > f_3$ преобладает так называемый джонсоновский шум.

На рис. 11.8 представлены спектральные зависимости обнаружительной способности для различных типов фотодетекторов, в том числе фоторезисторов. В видимой области спектра применяют пленочные фоторезисторы на основе CdS и CdSe. Для компенсации собственных мелких донорных уровней эти матери-

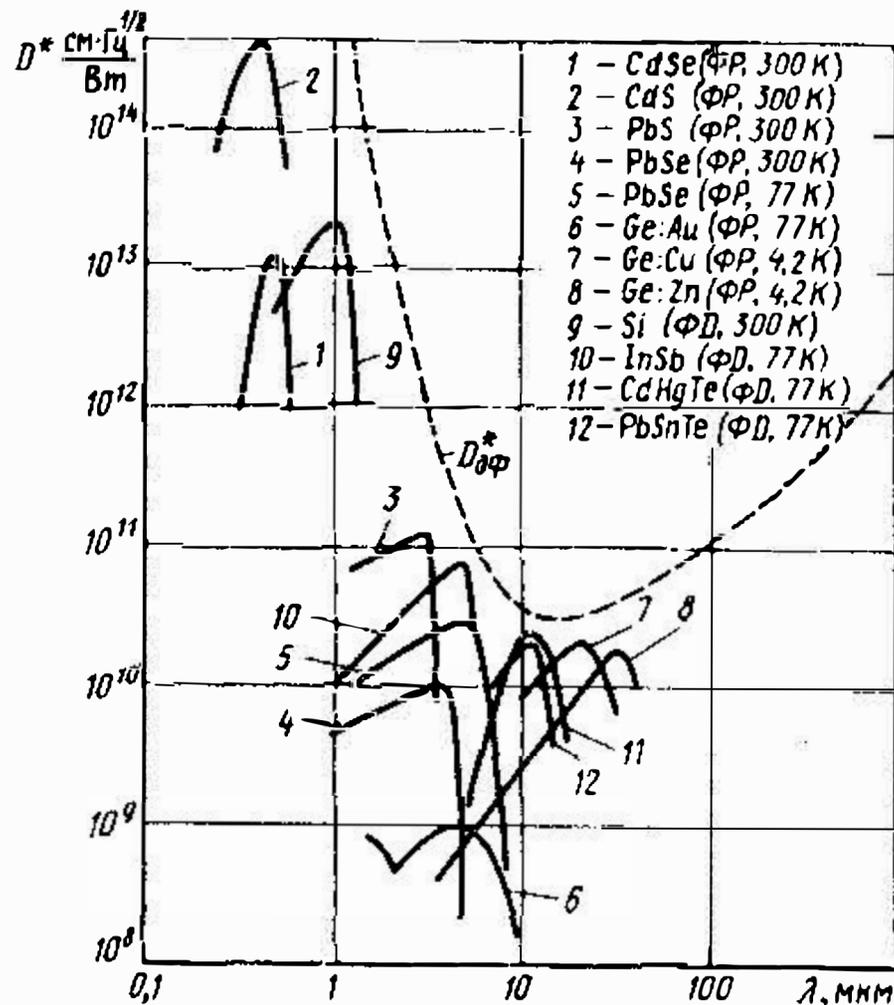


Рис. 11.8. Спектральные зависимости обнаружительной способности D^* для различных типов фоторезисторов (ФР) и фотодиодов (ФД). Штриховой кривой представлен теоретический предел чувствительности, ограниченный фоновым излучением при 300 К и угле обзора 2π

алы легируют медью, что приводит к их «очувствлению» — резкому возрастанию τ_n , снижению n_0 и повышению фоточувствительности. Для фоторезисторов этого типа возможно достижение $R_T/R_{c0} \approx 10^5 \dots 10^6$ при $R_T \approx 10^7 \dots 10^{14}$ Ом. Их инерционность очень велика (≈ 100 мс при 300 К). В области спектра 1...5 мкм, важной для прикладных целей, широкое распространение получили пленочные фоторезисторы на основе сульфида и селенида свинца PbS и PbSe, обладающие хорошей чувствительностью и удовлетворительной инерционностью. В области 5...12 мкм применяют фоторезисторы на основе узкозонных полупроводниковых твердых растворов кадмий-ртуть-теллурид ($Cd_xHg_{1-x}Te$), называемых КРТ, или халькогениды свинца-олова ($Pb_xSn_{1-x}Te$) (см. § 4.2 и рис. 4.12). В длинноволновом ИК-диапазоне применяют фоторезисторы, работающие на примесном фотоэффекте. Их делают на основе монокристаллов германия, легированных различными примесями: Ge: Au; Ge: Hg; Ge: Cu; Ge: Zn. При $\lambda > 100$ мкм возможно использование примесной фотопроводимости в л-GaAs, где энергия ионизации мелких доноров составляет ≈ 5 мэВ. Основные характеристики промышленных фоторезисторов приведены в табл. 11.1.

Таблица 11.1

Параметры различных типов фоторезисторов

Материал	$T_{\text{раб.}}$ К	Спектральный диапазон, мкм	$\lambda_{\text{макс.}}$ мкм	$D^*_{\text{макс.}}$ см·Гц ^{1/2} Вт ⁻¹	τ , мс	$R_{\text{Темп.}}$ Мом	Примеры промышленных типов фоторезисторов
CdS	295	0,4...0,8	$0,6 \pm 0,05$	10^{14}	10...200	0,1...100	ФСК-1...ФСК-7, ФР-765, СФ2-1...СФ2-19
CdSe	295	0,4...1	$0,7 \pm 0,05$	10^{13}	10...100	2...1000	ФСД-1, ФСД-Г1, ФР-764, СФ3-1...СФ3-16
PbS	295	1...3,0	$2,2 \pm 0,2$	10^{12}	0,4	0,02...5	ФСА-0, ФСА-1, ФСА-4, ФСА-6, ФСА-Г1, ФСА-Г2, ФР1-3, ФР1-4, ФР-СС-138, ФСВ-16АН, ФСВ-18АА
	193 80	1...3,5 1...4,0	2,7 3,2	$5 \cdot 10^{11}$ $2 \cdot 10^{11}$	5 3		
PbSe	295	0,6...4,5	$3,7 \pm 0,2$	10^{10}	0,01	0,2..1	СФ4-1Д...СФ4-3
	193 80	0,6...5,1 1...6,5	4,4 5,0	$3 \cdot 10^{10}$ $2 \cdot 10^{10}$	0,05 0,04		

Материал	$T_{\text{раб}}$, К	Соединительный диапазон, мкм	λ_{max} , мкм	$D_{\text{фот}}^*$, см · Гц ^{1/2} · Вт ⁻¹	τ , мс	$R_{\text{ТЭМ}}$, МОм	Примеры промышленных типов фоторезисторов
InSb	80	2,0...5,6	$5,2 \pm 0,2$	$5 \cdot 10^{10}$	0,005	0,002...0,02	ФРО-146, ФРО-148, ФС-17ДА
Au	80	2,4...7,6	$5,9 \pm 0,3$	$3 \cdot 10^9$	30 нс	0,03...1	ФСГ-22
Ge	40	4...14	11 ± 1	$3 \cdot 10^{10}$	100 нс	1	ФСГ-23
Cu	15	2...28	24	$2 \cdot 10^{10}$	10 нс	0,02	ФСГ-28РТА
Hg _{0,8} Cd _{0,2} Te	80	8...13	12 ± 1	$5 \cdot 10^{10}$	0,001	50...2000 Ом	ФРОХ1-142
Pb _{0,8} Sn _{0,2} Te	80	6...11	10	$2 \cdot 10^{10}$	0,01	50 Ом	—
Болометр	295	Определяется пропусканием окна	—	$2 \cdot 10^8$	3	0,6...3	БКМ

На рис. 11.8 пунктирной линией изображен теоретический предел чувствительности, ограниченный фоновым излучением для $T=300$ К при угле обзора 2π стерадиан, определенный для фотовольтаических приемников типа фотодиодов. Для фоторезисторов аналогичная зависимость идет примерно в 2 раза ниже.

В связи с большой инерционностью применение фоторезисторов как демодуляторов сигналов в оптической электронике ограничено. Если на основе одного и того же материала могут быть изготовлены и фотодиод, и фоторезистор, то при одинаковых условиях возбуждения параметры первого, как правило, оказываются выше.

Возможности существенного улучшения характеристик фоторезисторов связаны в первую очередь с использованием квантово-размерных слоев и сверхрешеток (см. § 9.7), благодаря которым могут быть реализованы коэффициенты внутреннего усиления $K_I > 10^4$ при очень низких рабочих напряжениях ($U_R < 1$ В) и эквивалентной мощности шума на уровне 10^{-14} (Вт/Гц^{1/2}).

§ 11.4. ФОТОВОЛЬТАИЧЕСКИЕ ЭФФЕКТЫ В НЕОДНОРОДНЫХ СТРУКТУРАХ

Для возникновения фотоЭДС в полупроводнике при возбуждении его светом должны существовать причины, приводящие к разделению в пространстве неравновесных носителей зарядов разных знаков. В однородном кристалле, как говорилось в § 4.7, такой причиной может быть разная подвижность носителей раз-

ного знака, что приводит к фотодиффузионному эффекту. Однако этот эффект очень слаб. Значительно более сильные эффекты проявляются в неоднородных структурах, наличие потенциальных барьеров и внутренних электрических полей в которых создает благоприятные условия для пространственного разделения носителей заряда, созданных в результате фотоактивного поглощения света. При освещении такой неоднородной структуры в ней генерируется фотоЭДС. Этот эффект называется *фотовольтаическим* или *вентильным фотозффектом*.

В качестве примера рассмотрим возникновение фотоЭДС в структуре с изменяющейся шириной запрещенной зоны, как показано на рис. 11.9. Это изменение может быть вызвано различными причинами, например изменением химического состава твердого раствора (варизонный полупроводник) или наличием локальной механической деформации. В последнем случае наблюдается фотопьезоэлектрический эффект.

В структуре с плавно изменяющейся шириной запрещенной зоны возникает локальное поле. В примере, приведенном на рис. 11.9, это поле ускоряет неосновные носители заряда (электроны) и они будут «скатываться» по потенциальному барьеру. Заметим, что стрелка на рис. 11.9, а условно отображает движение электрона. На самом деле траектория его движения в координатах ($E; z$) представляет собой ломаную линию (рис. 11.9, б). При свободном движении электрона в кристалле под воздействием электрического поля E его полная энергия не изменяется. В электрическом поле электрон ускоряется, и его потенциальная энергия переходит в кинетическую. Такому движению на диаграмме $E(z)$ будет отвечать горизонтальная линия (участки 1 — 2, 3 — 4, 5 — 6 на рис. 11.9, б). Электрическое поле не может изменить полную энергию электрона. Приобретая кинетическую энергию, электрон при своем движении может рассеивать ее, например, передавая решетке при столкновении с фононами. При этом полная энергия электронной системы уменьшается. Этим процес-

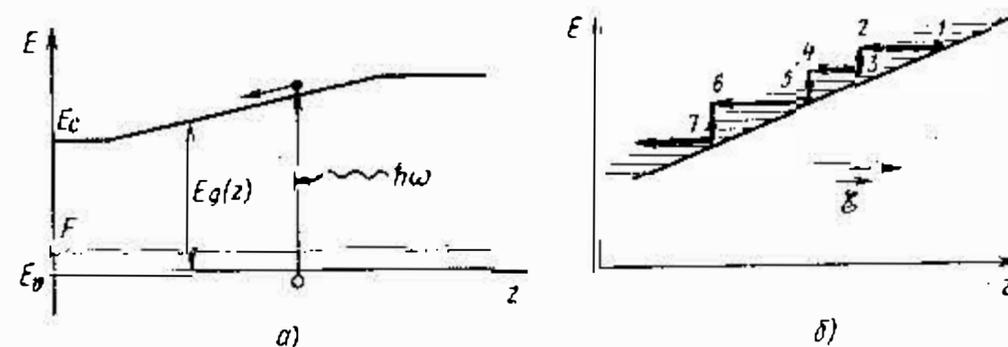


Рис. 11.9. Возникновение фотоЭДС в структуре с изменяющейся шириной запрещенной зоны: а — общий вид энергетической диаграммы; б — траектория движения электрона на диаграмме $E(z)$

сом рассеяния соответствуют участки 2 — 3, 4 — 5 и 6 — 7 на рис. 11.9, б.

Наиболее эффективное разделение носителей заряда происходит в слое объемного заряда p - n -перехода. Поэтому в таких структурах, называемых *фотодиодами*, фотовольтаический эффект проявляется наиболее сильно. Рассмотрим p - n -переход, на который падает оптическое излучение с энергией фотонов $\hbar\omega > E_g$, приводящее к образованию электронно-дырочных пар. Будем считать для определенности, что p - n -переход освещается с p -стороны, как показано на рис. 11.10. Если поглощение фотона происходит в области объемного заряда p - n -перехода (процесс 1 на рис. 11.10), то под воздействием внутреннего поля перехода носители заряда будут перемещаться в противоположных направлениях: электроны — в n -область, а дырки — в p -область. На самом деле толщина слоя объемного заряда обычно очень мала, поэтому вероятность поглощения фотона в этой области также незначительна. Генерация светом избыточных носителей заряда происходит в основном в областях, непосредственно примыкающих к p - n -переходу (процессы 2 и 3). Эти избыточные носители заряда диффундируют к области p - n -перехода. Если генерация произошла на расстоянии, меньшем диффузионной длины L_n или L_p для неосновных носителей заряда, то они успеют дойти до p - n -перехода, не рекомбинируя с основными носителями. В области объемного заряда неосновные носители подхватываются полем и выбрасываются в противоположную область p - n -перехода. Созданные светом и разделенные p - n -переходом избыточные носители заряда накапливаются в разных областях, что и приводит к возникновению фототока и фотоЭДС. При этом p -область будет заряжаться положительно, а n -область — отрицательно.

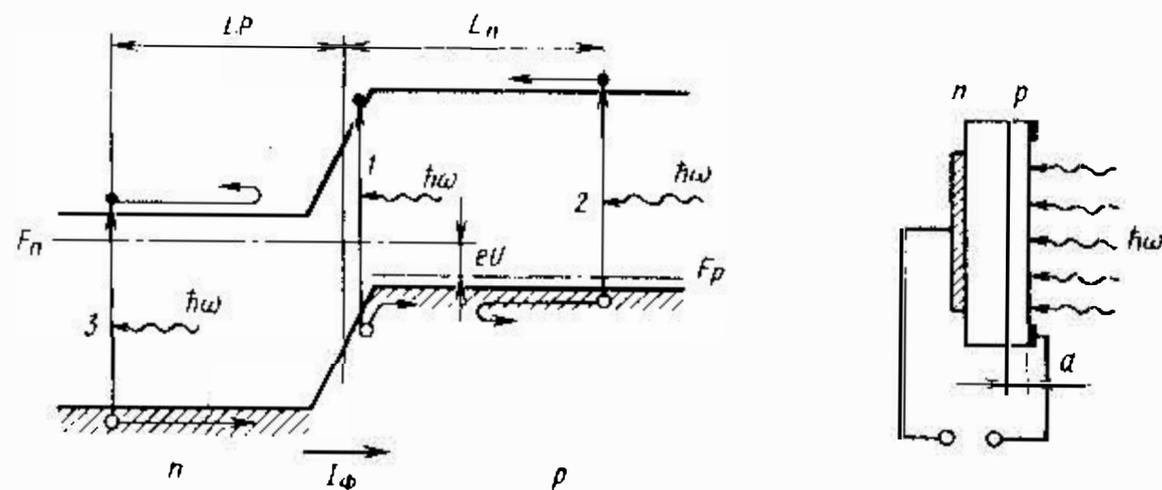


Рис. 11.10. Фотовольтаический эффект в p - n -переходе

Таким образом, под действием света в p - n -переходе возникает фототок I_ϕ , пропорциональный скорости генерации G избыточных электронно-дырочных пар в области, ограниченной диффузионными длинами неосновных носителей заряда. Этот фототок, проходящий в обратном направлении, будет изменять (уменьшать) контактную разность потенциалов, вследствие чего через переход начнет проходить ток в прямом направлении, значение которого

$$I_y = I_s \left(e^{\frac{eU}{kT}} - 1 \right). \quad (11.15)$$

Здесь I_s — ток насыщения, который создается свободными носителями заряда, генерируемыми за счет теплового возбуждения.

Можно считать, что фототок через p - n -переход протекает независимо от приложенного напряжения. Тогда выражение для полного тока будет иметь вид

$$I = I_s \left(e^{\frac{eU}{kT}} - 1 \right) - I_\phi. \quad (11.16)$$

Это общее уравнение фотодиода. Вольт-амперная характеристика p - n -перехода в отсутствие освещения ($G = 0$) и при воздействии света ($G > 0$) представлена на рис. 11.11. В случае разомкнутой цепи (режим холостого хода) фотоЭДС определяется из (11.16) при $I = 0$:

$$U_x = \frac{kT}{e} \ln \left(\frac{I_\phi}{I_s + I} \right). \quad (11.17)$$

Из (11.16) при $U = 0$ находим, что ток короткого замыкания равен фототоку:

$$I_x = I_\phi.$$

В свою очередь, фототок для p - n -перехода с освещаемой поверхностью A определяется скоростью генерации избыточных носителей заряда G и диффузионными длинами L_n и L_p :

$$I_\phi = I_x = Ae(L_n + L_p)G. \quad (11.18)$$

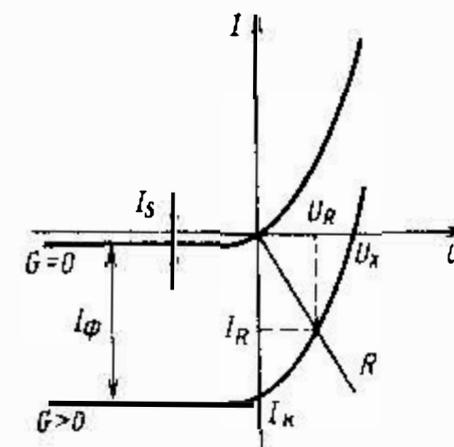


Рис. 11.11. Вольт-амперные характеристики p - n -перехода в отсутствие освещения ($G = 0$) и при воздействии света ($G > 0$)

В промежуточном случае, когда p - n -переход нагружен на сопротивление R , ток и напряжение находят из общего уравнения фотодиода (11.16):

$$U_R = \frac{kT}{e} \ln \left(\frac{I_R + I_\phi}{I_S} + 1 \right), \quad (11.19)$$

где $I_R = U_R/R$ (рис. 11.11).

При увеличении интенсивности света, т. е. скорости генерации G , возрастают ток короткого замыкания и напряжение холостого хода. Ток I_x возрастает линейно с интенсивностью света, а напряжение U_x — по логарифмическому закону. Предельное значение U_x приблизительно соответствует ширине запрещенной зоны.

Заряды, определяющие возникновение фотоЭДС в p - n -переходе (или в других неоднородных структурах), генерируются в процессе поглощения фотона в слое толщиной $L_n + L_p$. К появлению фотоЭДС приводят только те процессы поглощения, в результате которых создаются неосновные носители заряда. Доминирует здесь собственное поглощение. Процессы фотонионизации мелких примесей не дают вклада в фотоЭДС, а процессы их фотонейтрализации могут вызвать фотоЭДС.

Рассмотрим спектральную зависимость фоточувствительности p - n -перехода. Для определенности освещение будем проводить с p -стороны, как на рис. 11.10. Если толщина p -области меньше диффузионной длины для электронов ($d < L_n$), то все фотогенерированные в p -слое носители заряда дойдут до p - n -перехода и дадут вклад в фотозффект. При малых значениях ($\hbar\omega - E_g$) фототок пропорционален показателю поглощения, поскольку скорость генерации G пропорциональна k_ω [см. выражение (4.96)]. Поэтому в области длинноволнового края при $k_\omega < 1/d$

спектральная зависимость фототока будет определяться спектральной зависимостью k_ω . При продвижении в коротковолновую область при $\hbar\omega > E_g$ показатель поглощения быстро возрастает, и в случае $k_\omega \gg 1/d$ практически все излучение будет поглощаться в p -области. Дальнейшее увеличение k_ω не приводит к росту фототока (рис. 11.12).

Если толщина p -области, через которую производят освещение, больше диффузионной длины L_n ($d > L_n$), то при увеличении k_ω все большая часть излучения

будет поглощаться вдали от p - n -перехода и все меньшая часть генерированных неосновных носителей заряда достигать p - n -перехода. Поэтому при $d > L_n$ в спектрах фотоЭДС и фототока будет наблюдаться максимум, как показано на рис. 11.12.

Отметим, что изображенные на рис. 11.12 спектры фототока отражают спектральные зависимости квантового выхода фотоэффекта. Они сняты при постоянной плотности падающих фотонов. Если спектр фототока снимается при постоянной интенсивности засветки, то вместо выхода на насыщение при $d < L_n$ при увеличении $\hbar\omega$ будет наблюдаться спад в соответствии с рис. 11.1.

Из приведенного выше качественного рассмотрения ясно, что для получения высокой фоточувствительности, высокого быстродействия, исключения потерь на поверхностную и объемную рекомбинацию желательно, чтобы генерация электронно-дырочных пар осуществлялась преимущественно в области объемного заряда перехода. Этого можно достигнуть в p - i - n -диодах, которые рассматриваются в следующем параграфе, и в гетеропереходах, используя широкозонную часть в качестве «окна», через которое производится освещение. Такое «окно» прозрачно для света, поглощаемого в узкозонной части p - n -перехода (рис. 11.13,

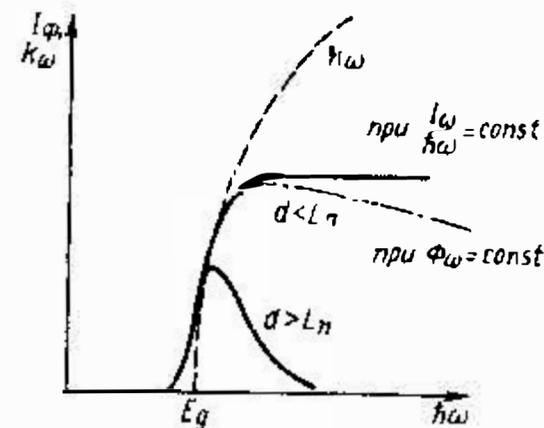


Рис. 11.12. Спектральные зависимости фототока короткого замыкания $I_\phi(\hbar\omega)$ p - n -перехода (пунктиром изображен спектр собственного поглощения (4.42))

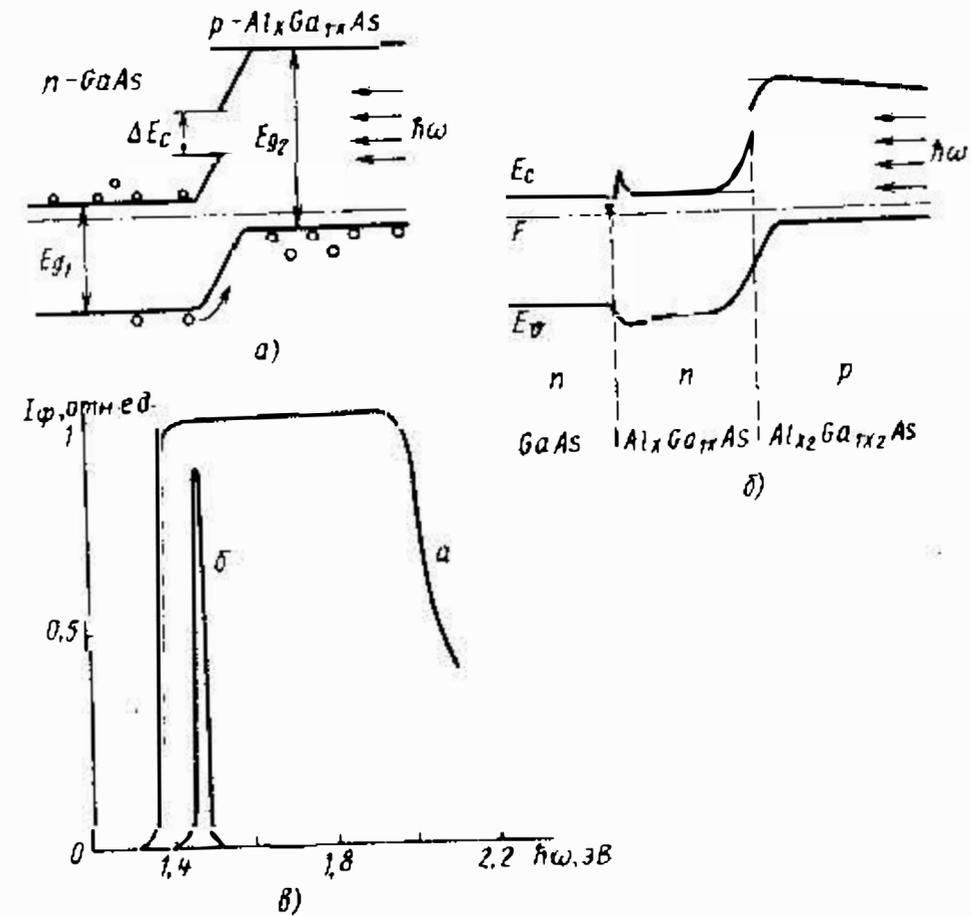


Рис. 11.13. Энергетические диаграммы (а, б) и спектральные характеристики (в) p - n -гетеропереходов при освещении через широкозонное «окно»

а). Спектральная характеристика фоточувствительности при этом близка к П-образной и ограничена с длинноволновой стороны шириной запрещенной зоны узкозонного, а с коротковолновой — широкозонного материала (кривая 1 на рис. 11.13, а). Используя твердые растворы, можно в заданных пределах изменять спектральную характеристику. В частности, используя гетеропереход, энергетическая диаграмма которого приведена на рис. 11.13, б, можно получить очень узкую область спектральной чувствительности (селективный фотоэлемент), характеристика которого соответствует кривой 2 на рис. 11.13, а.

Фотовольтаические эффекты наблюдаются и на контакте полупроводника с металлом. В месте такого контакта происходит перераспределение зарядов, в результате чего в полупроводнике возникает обедненный слой и появляется энергетический барьер, называемый *барьером Шотки*. Высота барьера равна разности работ выхода для металла и полупроводника. В такой структуре возможно также эффективное разделение носителей заряда противоположных знаков, как показано на рис. 11.14, а. Металлический слой, через который производят освещение, служащий электрическим контактом, для уменьшения оптических потерь делают очень тонким (~ 10 нм). Если энергия фотона $\hbar\omega$ больше ширины запрещенной зоны, то поглощение происходит в основном в слое объемного заряда полупроводника и диод Шотки работает как фотодиод с *p-n*-переходом. Кроме того, на барьере Шотки возможно наблюдать фотовольтаический эффект для $\hbar\omega_2 < E_g$, но $\hbar\omega_2 > e\phi$ (рис. 11.14, б). В этом случае электроны, возбужденные в металлической пленке, могут преодолеть потенциальный барьер и перейти в полупроводник.

Фотодиоды как фотоприемники могут работать в двух режимах: а) без приложения внешнего электрического поля, т. е. как источники тока (напряжения); такой режим работы называется *вентильным* и б) при приложении напряжения в запирающем

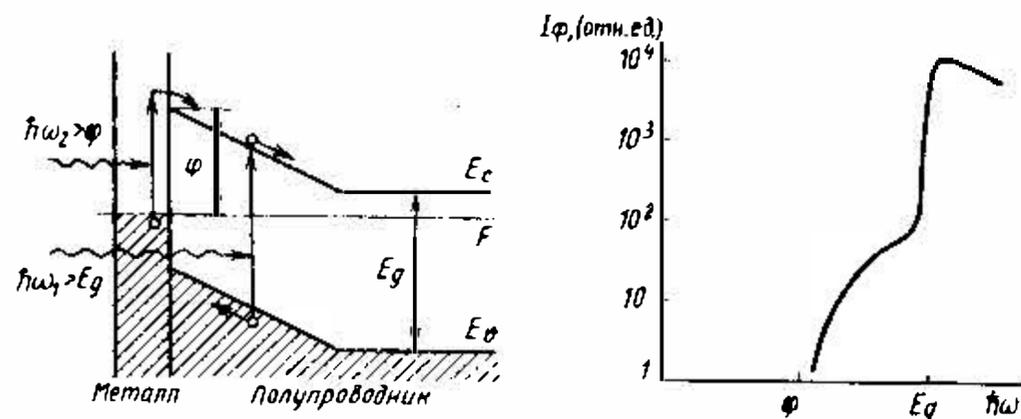


Рис. 11.14. Фотоэффект на контакте металл—полупроводник (на барьере Шотки): а — энергетическая диаграмма; б — спектральная характеристика

направлении; такой режим называется *фотодиодным*. При подаче больших обратных смещений в фотодиодном режиме в *p-n*-переходе или диоде Шотки возможно лавинное умножение носителей заряда, что используется для усиления фототока (§ 11.6).

Помимо рассмотренных фотовольтаических эффектов, при которых генерируемая фотоЭДС не превышает ширины запрещенной зоны полупроводника, в тонких пленках некоторых материалов (Se, PbS, Si, Ge, GaAs, GaP, HgTe, ZnS и др.), прошедших специальную обработку, наблюдался фотовольтаический эффект, при котором генерируемое напряжение существенно превышало ширину запрещенной зоны и достигало нескольких киловольт (до 5 кВ). Этот эффект называют эффектом *аномальных фотонапряжений*. Хотя в деталях природа его возникновения окончательно не ясна, наиболее вероятно, что он является следствием суммирования элементарных фотоЭДС, генерируемых в микропереходах или на границах зерен в микрочайках пленки, имеющей зернистую структуру.

§ 11.5. *p-i-n*-ФОТОДИОДЫ

В фотодиоде необходимо совместить область поглощения света с обедненным слоем, чтобы одновременно удовлетворить требованиям быстродействия и высокого квантового выхода. Это реализуется в фотодиодах с *p-i-n*-структурой, которые являются наиболее распространенным типом фотодетекторов.

Структура и принцип действия *p-i-n*-фотодиода пояснены на рис. 11.15. Он состоит из низкоомной n^+ -подложки, слабо легированного (собственного) *i*-слоя и тонкого низкоомного p^+ -слоя толщиной до 0,3 мкм, через которые производится освещение. Низкоомные n^+ - и p^+ -области выполняют роль контактов (а). Наличие центрального высокоомного *i*-слоя приводит к увеличению ширины слоя объемного заряда (б) по сравнению с обычным *p-n*-переходом. Его толщина d_i подбирается так, чтобы поглощение света происходило в этом *i*-слое (в), совпадающем со слоем объемного заряда. При приложении обратного смещения U обедненный слой распространяется на всю *i*-область. Это приводит к уменьшению емкости перехода, повышению чувствительности и быстродействия. Падающий свет, затухая по экспоненциальному закону с постоянной, определяемой показателем поглощения k_ω для данной длины волны, вызывает генерацию носителей заряда преимущественно в *i*-слое. Фотогенерированные носители ускоряются электрическим полем до скорости насыщения дрейфа ($\sim 10^5$ м/с), поскольку напряженность электрического поля в обедненном слое обычно превышает 1 кВ/см. Эта скорость дрейфа примерно на три порядка превышает скорость диффузии. Поэтому *p-i-n*-фотодиод конструктивно выполняется так, чтобы мак-

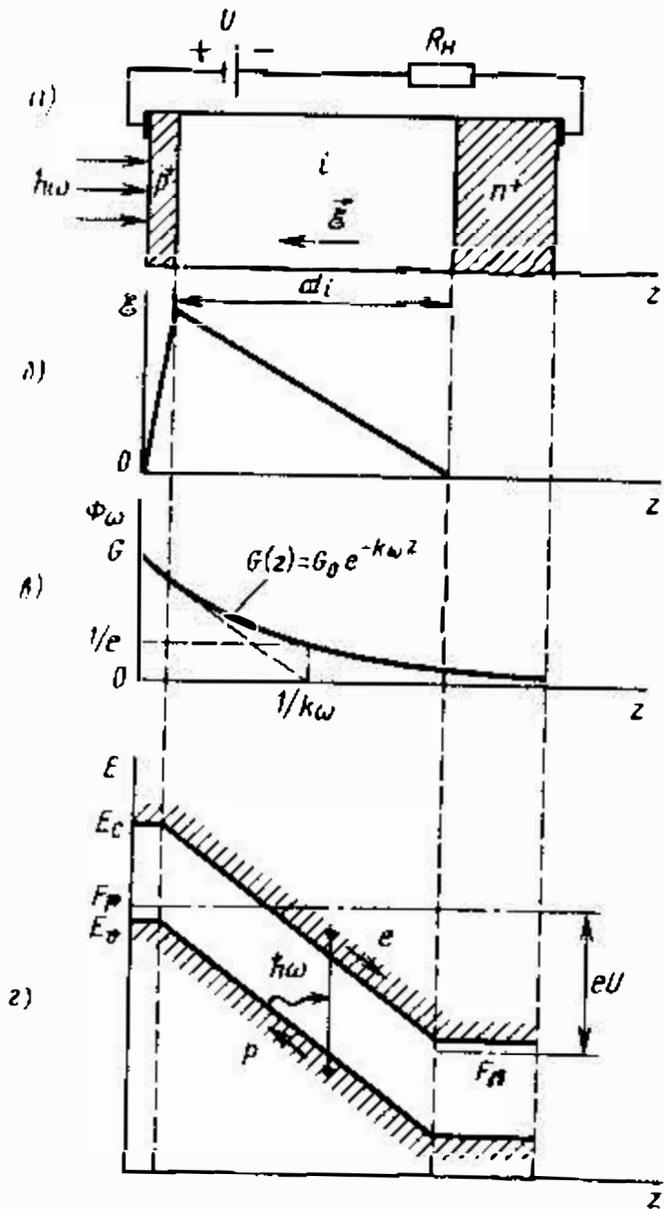


Рис. 11.15. Принципы работы $p-i-n$ фотодиода:

а) структура диода; б) распределение электрического поля; в) распределение интенсивности света Φ_ω и скорость генерации G ; г) зонная диаграмма при обратном смещении U

ческой генерации можно пренебречь. Тогда в соответствии с (1.170) и рис. 11.5, в скорость генерации электронно-дырочных пар

$$G(z) = G_0 e^{-k_\omega z},$$

где G_0 определяется потоком падающих фотонов $\Phi_\omega = I_\omega / \hbar\omega$, коэффициентом оптического отражения R_ω и площадью окна A как

$$G_0 = \Phi_\omega k_\omega (1 - R) / A. \quad (11.21)$$

При этих условиях дрейфовый ток

симально уменьшить долю поглощенного света вне i -слоя. С этой целью переход формируется у самой поверхности кристалла (как это реализуется в кремниевых фотодиодах), или используется эффект широкозонного окна (как в $p-i-n$ -гетерофотодиодах).

В стационарном режиме плотность полного фототока, протекающего через обратносмещенный $p-i-n$ -переход, можно разбить на две части:

$$J_\phi = J_{др} + J_{дифф}, \quad (11.20)$$

где $J_{др}$ — плотность дрейфового тока, обусловленного генерацией носителей в i -слое толщиной d , а $J_{дифф}$ — плотность диффузионного тока, обусловленного генерацией носителей в объеме полупроводника за пределами обедненного слоя и их последующей диффузией к области объемного заряда. Будем считать толщину приповерхностного слоя p^+ -типа существенно меньше $1/k_\omega$. Током термической генерации можно пренебречь. Тогда в соответствии с (1.170) и рис. 11.5, в скорость генерации электронно-дырочных пар

$$J_{др} = -e \int_0^d G(z) dz = e \frac{\Phi_\omega}{A} (1 - R) (1 - e^{-k_\omega d}). \quad (11.22)$$

Плотность диффузионного тока $J_{дифф}$ определяется через концентрацию неосновных носителей (в нашем случае — дырки в n -области) и их коэффициент диффузии D_p , как

$$J_{дифф} = -e D_p \left. \frac{dp_n}{dz} \right|_{z=d}. \quad (11.23)$$

В свою очередь концентрация неосновных носителей (дырок) находится из одномерного диффузионного уравнения, которое в нашем случае имеет вид:

$$D_p \frac{\partial^2 p_n}{\partial z^2} - \frac{(p_n - p_{n0})}{\tau_p} + G(z) = 0. \quad (11.24)$$

Здесь τ_p — время жизни неравновесных носителей, а p_{n0} — равновесная концентрация дырок. Решая это уравнение с граничными условиями $p_n = p_{n0}$ при $z \rightarrow \infty$ и $p_n = 0$ при $z = d$, подставляя это решение в (11.23), находим

$$J_{дифф} = e G_0 \frac{L_p}{1 + k_\omega L_p} e^{-k_\omega d} + e p_{n0} \frac{D_p}{L_p}, \quad (11.25)$$

где $L_p = \sqrt{D_p \tau_p}$ — диффузионная длина.

Полная плотность фототока получается из (11.20), (11.22) и (11.25) как

$$J_\phi = e \frac{\Phi_\omega}{A} (1 - R) \left(1 - \frac{e^{-k_\omega d}}{1 + k_\omega L_p} \right) + e p_{n0} \frac{D_p}{L_p}. \quad (11.26)$$

Обычно второе слагаемое, содержащее p_{n0} , значительно меньше первого и полный фототок J_ϕ пропорционален потоку фотонов Φ_ω . Из (11.26) получаем выражение для квантовой эффективности

$$\eta = \frac{J_\phi / e}{\Phi_\omega / A} = (1 - R) \left(1 - \frac{e^{-k_\omega d}}{1 + k_\omega L_p} \right). \quad (11.27)$$

Для достижения высокой эффективности фотодиода необходимо выполнить $k_\omega d \gg 1$. Однако при увеличении d будет расти время пролета носителей, которое равно времени их дрейфа через

i-область. Это приведет к уменьшению быстродействия диода. При высокочастотной модуляции интенсивности света появится разность фаз между потоком фотонов и фототоком. Разумный компромисс между быстродействием и квантовой эффективностью достигается при значении ширины области поглощения d_i от $1/k_{\omega}$ до $2/k_{\omega}$. В кремниевых *p-i-n*-фотодиодах, предназначенных для приема излучения арсенид-галлиевого лазера ($\lambda \approx 0,85 \dots 0,92$ мкм), это соответствует толщине обедненного слоя $d = 20 \dots 50$ мкм. Те же диоды, предназначенные для приема излучения твердотельных неодимовых лазеров ($\lambda = 1,06$ мкм), должны иметь толщину *i*-слоя $d \approx 500$ мкм.

Предельная рабочая частота, ограниченная временем дрейфа носителей через *i*-область и соответствующая уменьшению фототока на 3 дБ по сравнению со стационарным значением, примерно равна

$$f_{гр} \approx 0,4v_s/d \approx 0,4k_{\omega}v_s, \quad (11.28)$$

где v_s — скорость насыщенного дрейфа.

Наибольшее распространение в оптической электронике получили *p-i-n*-фотодиоды, изготовленные на основе высокоомного кремния *n*-типа. Типичная конструкция кремниевого *p-i-n*-фотодиода приведена на рис. 11.16, а. Отметим, что эти диоды просты по своей структуре, обладают хорошей линейностью в широком динамическом диапазоне от десятков пиковатт до десятков милливольт. Они просты в эксплуатации и дешевы. Их спектр фоточувствительности хорошо согласуется со спектром излучения GaAs-лазеров. Все это, вместе взятое, определило широкое использование *p-i-n*-фотодиодов в оптической электронике для детектирования оптических сигналов, модулированных частотами до гигагерцового диапазона.

Естественно, что для реализации возможностей фотодиода как фотоприемника слабых оптических сигналов его электрические параметры должны быть согласованы с параметрами схемы, обеспечивающей дальнейшее усиление и регистрацию сигнала. Освещаемый фотодиод эквивалентен генератору тока. В простейшей схеме включения, изображенной на рис. 11.16, б, с нагрузочного сопротивления R_n снимается напряжение, пропорциональное фототоку, которое подается на усилитель напряжения. Отношение сигнала к шумовому току на входе будет тем больше, чем больше сопротивление R_n . В сочетании с полевым транзистором такая схема позволяет регистрировать световые мощности на уровне десятков пиковатт. Напомним, что при больших R_n инерционность приемного тракта может быть ограничена постоянной времени RC .

Вместо обычного усилителя напряжения может быть выбран стандартный операционный усилитель с большим коэффициентом

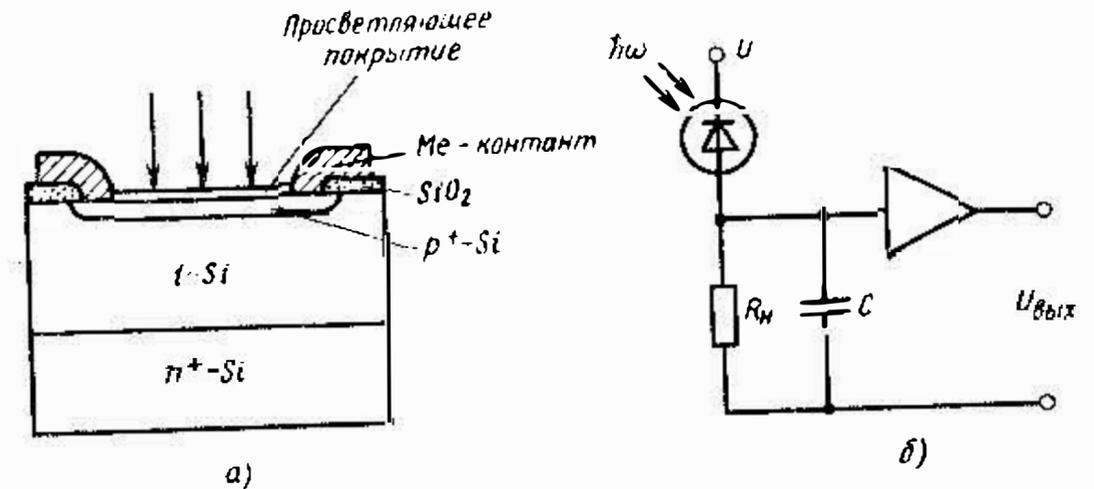


Рис. 11.16. Конструкция (а, б) и схема включения (в) *p-i-n*-фотодиода

том усиления и малым входным током, имеющий отрицательную обратную связь, благодаря чему его выходное напряжение пропорционально входному току.

Кремниевые *p-i-n*-фотодиоды применяются для области спектра $0,4 \dots 1$ мкм. Для более длинноволновой области $1 \dots 1,6$ мкм используют *p-i-n*-фотодиоды на основе гетеропереходов в системе $In_{1-x}Ga_xAs/InP$ или $In_{1-x}Ga_xAs_{1-y}P_y/InP$. Поскольку эти материалы обладают прямой структурой энергетических зон и край собственного поглощения у них очень крутой (см. §§ 4.2, 4.4), то толщина *i*-области у таких фотодиодов существенно меньше, чем у кремниевых диодов, и их быстродействие выше.

§ 11.6. ЛАВИННЫЙ ФОТОДИОД

При регистрации очень малых световых мощностей (≤ 1 нВт) фототоки будут весьма малы (< 1 нА). В этом случае желательно использовать внутреннее усиление в фотоприемнике аналогично тому, как это осуществляется в фотоэлектронном умножителе.

В лавинном фотодиоде (ЛФД), являющемся твердотельным аналогом ФЭУ, усиление фототока происходит за счет лавинного умножения генерированных светом носителей заряда в обратном смещенном *p-n*-переходе (рис. 11.17). При приложении к *p-n*-переходу обратного напряжения U , близкого к напряжению лавинного пробоя $U_{пр}$, энергия носителей заряда, ускоренных электрическим полем, может превысить порог ионизации вещества. Столкновение такого «горячего» носителя с электронами валентной зоны приведет к образованию пары электрон — дырка. Если образовавшиеся вторичные носители тоже ускорятся до энергии, превышающей порог ионизации, то они создадут другие носители и т. д., как показано на рис. 11.17. В результате проводимость нарастает за счет образования лавины носителей заряда

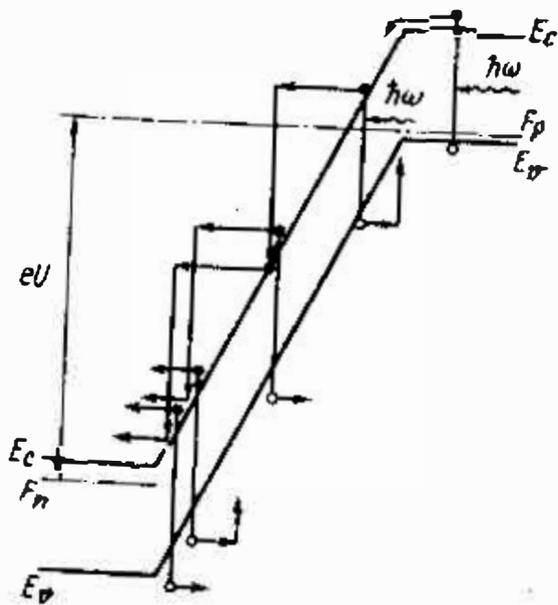


Рис. 11.17. Лавинное умножение генерированных светом носителей заряда в *p-n*-переходе, смещенном в обратном направлении

(рис. 11.18). Поглощение фотона является началом процесса, вызвавшего лавину.

Коэффициент лавинного умножения M сильно зависит от напряжения смещения U . Эта зависимость может быть представлена эмпирической формулой

$$M = \frac{1}{1 - (U/U_{пр})^n}, \quad (11.29)$$

где показатель степени n принимает значения от 2 до 6 в зависимости как от характеристики полупроводникового материала, так и от структуры *p-n*-перехода. При $U \approx U_{пр}$ с повышением напряжения происходит резкое увеличение коэффициента умножения, кото-

рый может достигать 10^3 . Обычно используют рабочие напряжения, при которых $M \approx 10 \dots 100$. Очевидно, что коэффициент лавинного умножения M и характеристики фотодиода сильно изменяются при изменении не только напряжения, но и температуры. Поэтому в электрической схеме смещения ЛФД необходимо предусматривать жесткие меры, устраняющие влияние этих изменений.

Процесс образования лавины носит вероятностный характер. Величина M изменяется случайным образом, флуктуируя около своего среднего значения (11.29). Это создает дополнительный шум, который весьма нежелателен. Для его уменьшения необходимо, чтобы коэффициенты ионизации электронов и дырок различались как можно более сильно, а лавинный пробой стимулировался носителями заряда, обладающими более высокими их значениями. Обычно такими носителями являются электроны.

Отношение коэффициентов ионизации электронов и дырок $K = \alpha_n / \alpha_p$ в кремнии зависит от напряженности электрического поля, изменяясь примерно от 0,1 при $E \approx 3 \cdot 10^5$ В/см до 0,5 при $E \approx 6 \cdot 10^5$ В/см. Поэтому для получения минимальных шумов желате-

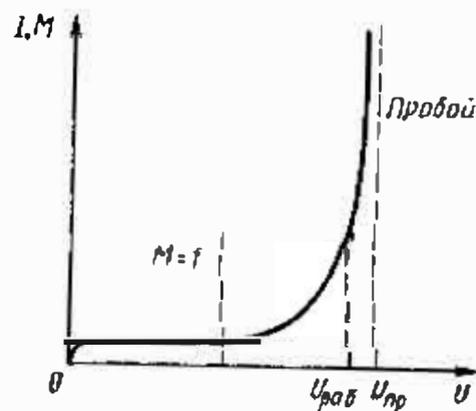


Рис. 11.18. Зависимость тока I и коэффициента умножения M от напряжения для лавинного фотодиода

льно низкое значение напряженности электрического поля лавинного пробоя.

В германии коэффициент ионизации электронов и дырок сравнимы, $K \approx 1$ и среднеквадратичный дробовой шумовой ток изменяется по закону M^3 . Поэтому германиевые ЛФД обладают худшей обнаружительной способностью и меньшей D^* , чем аналогичные диоды из кремния.

Наибольшее различие в коэффициентах ионизации и минимальное K достигается в таких материалах, в которых значение спин-орбитального расщепления валентной зоны $\Delta_{сo}$ примерно равно или немного меньше ширины запрещенной зоны E_g^r . Это условие реализуется, в частности, в твердом растворе $Al_xGa_{1-x}Sb$. В лавинных фотодиодах на основе этого материала при $M \approx 100$ шум-фактор увеличивается всего в 3 раза. Аналогичная ситуация может быть реализована и в некоторых других полупроводниковых твердых растворах, а также в структурах на основе сверхрешеток.

Для получения максимального отношения сигнал/шум в лавинном фотодиоде необходимо подбирать оптимальную величину напряжения обратного смещения, регулируя тем самым коэффициент умножения M . Поясним это с помощью рис. 11.19, где в двойном логарифмическом масштабе представлены зависимости мощности сигнала и мощности различных видов шумов от коэффициента лавинного умножения M . Мощность полезного сигнала растет пропорционально M^2 (поскольку $P \sim I^2$). При небольших значениях M дробовой шум лавинного умножения обычно меньше теплового шума, величина которого остается постоянной. Поэтому увеличение M приводит к росту отношения сигнал/шум до той поры, пока дробовой шум не превысит тепловой. Оптимальный коэффициент умножения M_{opt} , при котором отношение сигнал/шум максимально, достигается тогда, когда дробовой шум примерно в два раза превысит тепловой (рис. 11.19). На практике оптимальный коэффициент лавинного умножения M_{opt} подбирают регулировкой напряжения смещения. Для разных фотодиодов эта величина колеблется от 10 до 150 В.

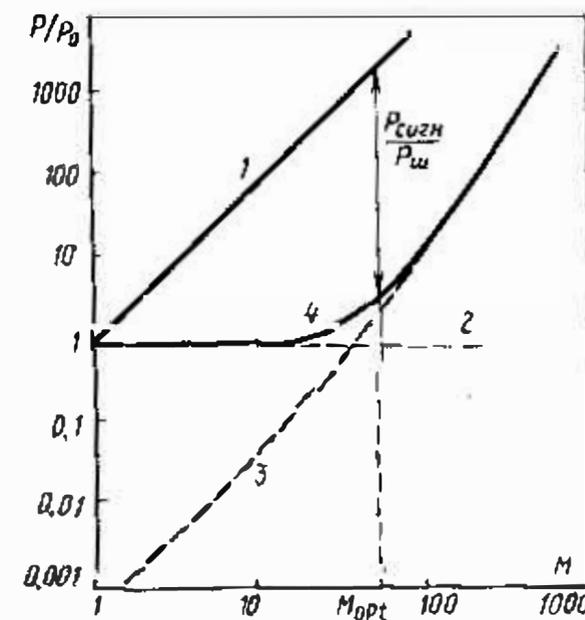


Рис. 11.19. Зависимости мощности выходного сигнала и мощности шумов от коэффициента умножения M в лавинном фотодиоде: 1 — мощность сигнала; 2 — мощность теплового шума; 3 — мощность дробового шума; 4 — суммарная мощность шумов

Параметры различных типов фотодиодов

Материал; тип диода	Т _{р.б.} К	Спектр, диапазон, мкм	λ_{max} мкм	D_{max}^* см·Гц ^{1/2} · Вт ⁻¹	τ , мкс	Примеры промышленных типов фотодиодов
$p-n$ Si $p-i-n$ ЛФД	295	0,4...1 0,4...1,1 0,4...1,1	$0,85 \pm 0,05$ $0,9 \pm 0,05$ 0,9	$3 \cdot 10^{12}$ $3 \cdot 10^{13}$ 10^{14}	1-10 <0,01 (до 2 нс) <0,01 (до 2 нс)	ФДК, КФДМ, ФД-3К — ФД-28КФД-К-155, 1690, ФД-10КП, ФД-20КП, ФД-252, ФД-256
$p-n$ Ge ЛФД	295	0,5...1,8	$1,55 \pm 0,05$	$5 \cdot 10^{10}$ 10^{12}	1-10 <0,1	ФД-1 — ФД-3, ФД-4Г — ФД-10Г 9Э-111, ФТГ-3 — ФТГ-5 (фототранзисто- ры)
InAs	300 80	1...3,6 1...3,1	3,5 2,8	10^{10} $5 \cdot 10^{11}$	RC RC	—
GaAs	300	0,5...0,9	0,8	10^{12}	0,1	—
InSb	80	2,2...5,4	$5,2 \pm 0,2$	$5 \cdot 10^{10}$	5	ФДО-117, ФДО-119, ФДО-257, ФД-268
$Ga_xIn_{1-x}As/InP$	300	1,3...1,6	1,5	10^{11}	0,005	—
$Cd_xHg_{1-x}Te$	80	2...30	8...14	10^{10}	0,01	—
$Cd_{0,2}Hg_{0,8}Te$	80	5...13	12	$5 \cdot 10^9$	RC	—
$Pb_xSn_{1-x}Te$	80	5...30	8...14	$10^9 \dots 10^{10}$	1	—

§ 11.7. ПРИЕМНИКИ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

Рассмотренные выше фотоприемники регистрируют оптическое излучение, попадающее на приемный элемент, как целое. Для регистрации оптического изображения, его последующей обработки и передачи электронными методами необходимо проводить поэлементный прием. Современные методы микроэлектроники позволяют сформировать на одном кристалле планарную структуру в виде сетки фотодиодов с количеством элементов 1000×1000 и более. Если на такой многоэлементный фотоприемник спроектировать оптическое изображение, то сигнал в каждом элементе будет пропорционален освещенности в данной точке. Весь вопрос заключается в том, каким образом свести сигнал с каждого из элементов и как осуществить последовательное сканирование (выборку) этих элементов. В качестве примеров успешного решения этого вопроса рассмотрим:

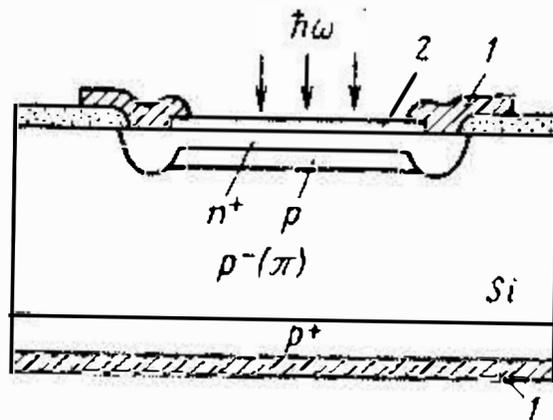


Рис. 11.20. Структура лавинного фотодиода на основе кремния:
1 — омические контакты; 2 — антиотражающее покрытие

Конструкции ЛФД могут меняться в зависимости от свойств применяемых материалов. Поскольку эти приборы работают в предпробойном режиме, то основным требованием является очень высокое качество и однородности как самого материала, так и $p-n$ -структуры. Недопустимы утечки тока и появление самопроизвольных микроплазм в дефектных областях. В качестве примера на рис. 11.20 приведена структура кремниевого лавинного фотодиода. Для уменьшения отражения света рабочая поверхность покрывается просветляющей диэлектрической пленкой. Защитное кольцо по периметру $p-n$ -перехода служит для предупреждения локальных лавинных пробоев и достижения равномерного по площади лавинного усиления. В фотодиодах на основе кремния глубина проникновения света велика вследствие малости показателя поглощения. Поэтому область обедненного слоя по аналогии с $p-i-n$ -фотодиодом формируют в виде слаболегированного высокоомного p^- -слоя (π -слоя). К этой области примыкает p -слой с высокой концентрацией носителей, образующий лавинную область с большой напряженностью электрического поля. В фотодиодах на основе прямозонных полупроводников $A^{III}B^V$ показатель поглощения велик, необходимость в создании широкого обедненного слоя отсутствует и они могут быть выполнены в виде простой p^+-n -структуры.

Лавинные фотодиоды обладают очень высоким быстродействием, достигающим (0,2...0,5) нс. Они имеют максимальное произведение коэффициента усиления на ширину полосы пропускания, составляющее 100 ГГц и более. В то же время ЛФД значительно дороже, требуют специального источника питания, они капризнее в эксплуатации, чем $p-i-n$ -фотодиоды. Они применяются в оптоэлектронике для регистрации слабых оптических потоков, промодулированных высокочастотным сигналом. Спектральные характеристики некоторых типов фотодиодов приведены на рис. 11.8.

Кроме рассмотренных примеров существуют другие разновидности ЛФД, в том числе на основе гетероструктур, с варизонной активной областью, с активной областью на сверхрешетках, каналные ЛФД и др. Характерные параметры наиболее распространенных типов фотодиодов представлены в табл. 11.2.

- а) передающие телевизионные трубки типа «видикон» и
 б) фоточувствительные приборы с переносом заряда.

Видикон, плюмбикон и кремникон. Это электровакуумные приборы, представляющие собой электронно-лучевую трубку, мишень которой изготовлена из фоточувствительного материала. Они предназначены для конверсии оптического изображения в электрические сигналы и наиболее часто применяются как *передающие телевизионные трубки*. Схема трубки типа «видикон», поясняющая принцип ее работы, приведена на рис. 11.21. Оптическое изображение с помощью объектива формируется на тонкой фоточувствительной мишени. Эта мишень со стороны изображения покрыта электропроводящим слоем, прозрачным для оптического излучения (обычно SnO_2). На этот слой через нагрузочное сопротивление R_n подается положительный потенциал $U_{\text{раб}}$ порядка $+50$ В относительно катода. С противоположной стороны мишень сканируется электронным лучом, который управляется с помощью обычных фокусирующих и отклоняющих систем аналогично тому, как это делается в электронно-лучевой трубке. В рабочем режиме мишень действует подобно конденсатору с утечкой. При отсутствии освещения сопротивление рабочего слоя велико и электрический заряд накапливается на противоположных обкладках конденсатора. Со стороны электронного луча потенциал обкладки будет равен потенциалу катода, т. е. 0, в то время как с противоположной стороны он равен $U_{\text{раб}}$. При освещении мишени сопротивление фоточувствительного материала уменьшается и конденсатор будет разряжаться. Сопротивления слоев и их толщина подобраны таким образом,

чтобы за время сканирования одного кадра растекание заряда по площади мишени было невелико. Тогда разрядка будет происходить только в том месте, куда падает свет. Когда электронный луч достигнет «разряженной» области, он будет ее дозаряжать, вызывая ток через конденсатор (рабочую мишень) и через нагрузочное сопротивление R_n . Сумма протекающего заряда будет зависеть от того, насколько разрядился конденсатор, т. е. суммой света, падающего в данном месте на фоточувствительную мишень. Электрический сигнал, снимаемый с нагрузочного сопротивления, пропорционален протека-

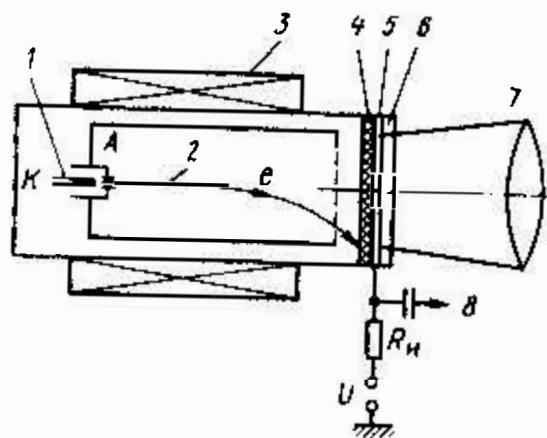


Рис. 11.21. Передающая телевизионная трубка типа «Видикон»:
 1 — катод; 2 — электронный пучок;
 3 — фокусирующие и отклоняющие системы; 4 — фоточувствительная мишень;
 5 — прозрачный электрод (слой SnO_2);
 6 — анод; 7 — объектив;
 8 — выходной сигнал

ющему через мишень току, т. е. освещенности мишени в том месте, куда падает электронный луч. Сканируя электронным лучом по поверхности мишени, мы таким образом преобразуем оптическое изображение в электрический сигнал.

Недостатком описанной выше трубки типа «видикон» является большое значение темнового тока. Этот недостаток отсутствует в трубке, получившей название «плюмбикон». Принцип действия этой трубки такой же, но рабочая мишень «плюмбикона» представляет собой слоистую $p-i-n$ -структуру, изготовленную на основе окиси свинца PbO (отсюда — название трубки), как показано на рис. 11.22, а. Прозрачный слой SnO_2 выполняет роль контакта p -типа. С противоположной поверхности слой PbO обогащен кислородом, что создает p -тип электропроводности. На $n\text{-SnO}_2$ подается положительный потенциал $U_{\text{раб}}$, смещающий $p-i-n$ -структуру в обратном направлении. Поэтому темновой ток мал.

Ширина запрещенной зоны PbO — около 2 эВ. Поэтому к красному свету с $\lambda > 0,6$ мкм этот материал нечувствителен. Для повышения чувствительности в длинноволновом диапазоне добавляют тонкий слой PbS ($E_g \approx 0,4$ эВ) со стороны мишени, обращенной к электронному лучу.

Дальнейшее совершенствование мишени привело к замене сплошных фоточувствительных слоев сеткой из кремниевых фотодиодов, как показано на рис. 11.22, б. Передающая электронно-лучевая трубка с мишенью в виде кремниевой фотодиодной матрицы, предназначенная для преобразования оптического изображения в электрический сигнал, называется *кремниконом*. Мозаика $p-n$ -переходов общим числом 10^6 и более (до 10^8) изготавливается на кремниевой пластине методом диффузии или

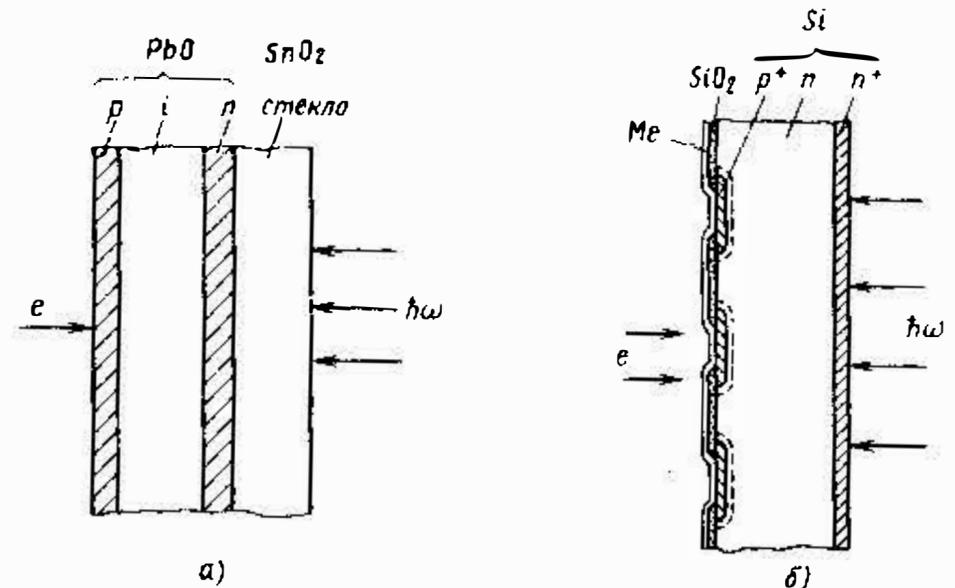


Рис. 11.22. Структура мишеней плюмбикона (а) и кремникона (б)

ионной имплантацией. На поверхность, обращенную к электронному лучу, наносится тонкая проводящая пленка, предохраняющая пленку SiO_2 от накопления заряда.

На n -область мишени через контактный n^+ -слой подается небольшое положительное напряжение $U_{\text{раб}} \approx 5 \dots 10$ В, смещающее p - n -переходы в обратном направлении и заряжающее их емкости. Освещение приводит к появлению фототока и разрядке емкостей p - n -переходов. При сканировании электронным пучком происходит их дозарядка. Протекающий через сопротивление R_n ток формирует видеосигнал, в котором закодировано изображение.

Темновой ток, искажающий видеосигнал, в кремнионе очень мал и находится на уровне 10 нА. Устройство обладает хорошей чувствительностью в спектральном диапазоне 0,4...0,9 мкм.

Отметим, что все рассмотренные выше мишени как приемники оптических изображений работают в режиме накопления. Типичная длительность кадра $\tau_k \approx (1/25)$ с, а длительность цикла опроса τ_0 на 2 — 3 порядка меньше. Это позволяет повысить чувствительность приемника.

Основные недостатки рассмотренных выше приемников изображений типа «видикон» характерны для всех электровакуумных приборов и связаны с необходимостью вакуумирования, а также использования больших ускоряющих напряжений U_A и сравнительно больших мощностей для управления электронным пучком. Этим недостаткам лишены полностью твердотельные приемники оптических изображений, основанные на использовании эффекта переноса заряда в приборах с зарядовой связью (ПЗС).

Приемники изображения на ПЗС. Это растровые безвакуумные приемники оптических изображений. Они осуществляют восприятие изображения, его разложение на элементарные фрагменты, поэлементное электронное считывание (сканирование) и формирование на выходе видеосигнала, адекватного изображению. Поэлементное считывание происходит за счет управляемого перемещения макроскопических зарядовых пакетов вдоль полупроводниковой подложки в приборах с переносом заряда при подаче на них определенной последовательности тактовых импульсов.

Фотоприемник на ПЗС представляет собой специальную фоточувствительную МДП (или МОП)* — микросхему с регулярной системой электродов, расположенных на поверхности диэлектрика настолько близко друг к другу, что за счет перекрытия электрических полей соседних электродов внутри полупроводника становится существенным их взаимодействие. Основу прибора составляет элементарный конденсатор со структурой металл — диэлектрик — полупроводник (МДП-конденсатор). Наиболее часто в качестве полупроводника используется крем-

ний, а функцию диэлектрика выполняет его окисная пленка SiO_2 . На рис. 11.23, а изображен такой МОП-конденсатор: металлический электрод, нанесенный на термически окисленную подложку из p -кремния. Если к металлическому электроду приложить положительное напряжение U_0 относительно p -подложки, то на границе раздела SiO_2 - Si образуется потенциальная яма для неосновных носителей заряда (электронов). Распространение потенциальной ямы вдоль границы раздела, т. е. вдоль поверхности кремния, ограничивается специально созданными p^+ -областями полупроводника, имеющими тот же тип проводимости, что и подложка, но степень легирования на несколько порядков выше. Их называют областями *стоп-диффузии*. Ограничение обеспечивается тем, что в низкоомных областях стоп-диффузии поверхностный потенциал на границе раздела близок нулю (рис. 11.23, б). При воздействии света с $\hbar\omega > E_g$ возникающие в полупроводнике неосновные носители заряда (электроны) собираются в потенциальной яме вблизи границы раздела и образуют инверсионный слой толщиной порядка 10 нм. Это пояснено на рис. 11.23, в. Процесс идет аналогично тому, как это происходит в фотодиоде, с той разницей, что сквозному движению носителей заряда препятствует потенциальный барьер на границе диэлектрик — полупроводник. Поэтому в потенциальной яме накаплива-

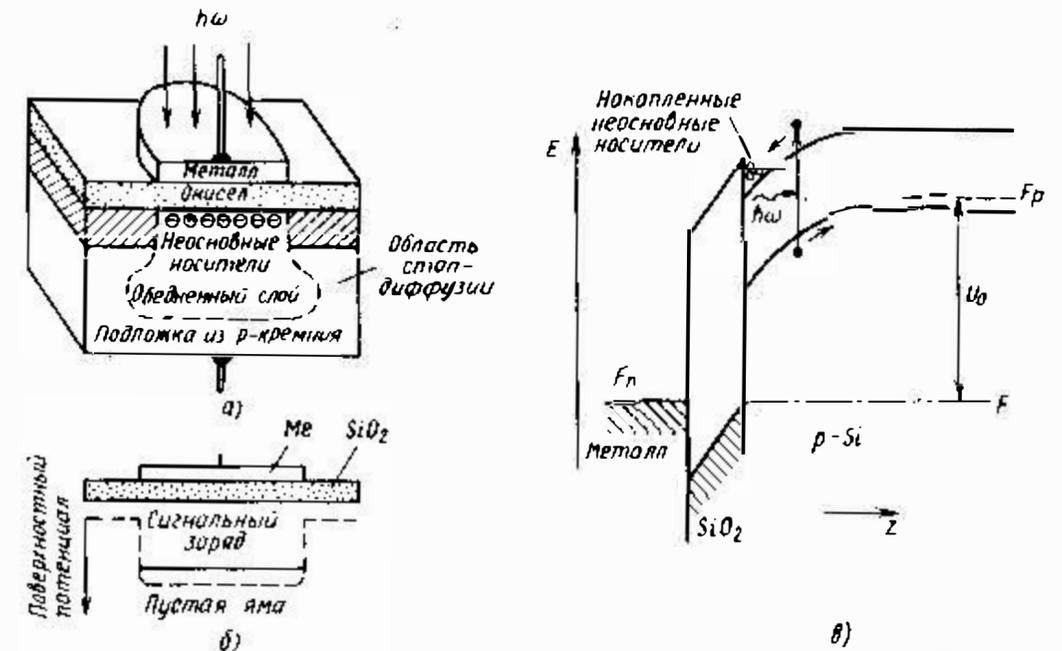


Рис. 11.23. МОП-конденсатор как фотоприемник:

а) поперечное сечение; б) условное изображение потенциальной ямы; в) энергетическая диаграмма МОП-структуры при действии света в ориентации положительного напряжения U_0 к металлическому электроду. Пунктирной линией изображен потенциальный рельеф при отсутствии засветки, сплошной — при ее наличии. Плотность инжектированной области пропорциональна накопленному заряду

*МДП (МОП) — общепринятые обозначения структур металл — диэлектрик — полупроводник (металл — окисел — полупроводник).

ется заряд Q , который пропорционален интенсивности и времени воздействия света. При увеличении заряда Q в яме поверхностный потенциал уменьшается, как это показано сплошной линией на рис. 11.23, б. Обедненная область схлопывается, а емкость электрод — подложка увеличивается.

Таким образом, в течение промежутка времени, меньшего времени релаксации, МДП (МОП)-конденсатор может служить запоминающим элементом для аналоговой информации. Это прибор динамического типа, а носителем информации в нем является зарядовый пакет, который определяется интегралом от светового потока по времени накопления с учетом разрядки за счет процессов рекомбинации и возможных утечек.

Пусть теперь два МДП-конденсатора, изготовленные на общей подложке с общим диэлектрическим слоем, расположены настолько близко друг к другу, что их обедненные области перекрываются и потенциальные ямы взаимодействуют («связываются»). Тогда подвижный заряд неосновных носителей будет накапливаться в том месте, где выше значение поверхностного потенциала, т. е. где глубже яма. Если изменять потенциал во времени, то заряд будет перетекать в наиболее глубокую часть потенциальной ямы. В этом и состоит идея управляемого переноса заряда от одного электрода к другому. Она проиллюстрирована на рис. 11.24 на примере ставшей классической трехкратной (трехфазной) схемы.

Электроды в ПЗС-фотоприемной матрице располагаются в виде линейки (строки) или матрицы, как показано на рис. 11.24, а. Зазор между электродами 1...2 мкм и менее. Число электродов в линейке обычно $5 \cdot 10^2 \dots 2 \cdot 10^3$, а в матрице может достигать 10^6 и более. Электроды изготавливаются из алюминия или поликремния (для улучшения прозрачности). Одна строка от другой в матричном приемнике отделяется узкими областями p^+ -стоп-канальной диффузии. Одна элементарная ячейка фотоприемника на ПЗС включает три соседних электрода одной строки, обозначенные Э1 — Э3 на рис. 11.24, а. Каждый из однотипных электродов подсоединен к своей шине тактового питания. Функцию фотоприемника выполняет одна структура, например, первая. Две другие служат для считывания, коммутации и вывода сигнала.

В течение первой фазы (диаграмма «а» на рис. 11.24) происходит прием оптического сигнала в каждой из Э1-структур. К электроду 1 прикладывается положительное напряжение смещения U_0 около 10...20 В. За это время в каждой из Э1-структур происходит накопление заряда пропорционально освещенности в данной точке.

Во время второго цикла (диаграмма «б») к электроду 2 прикладывается напряжение считывания $U_{\text{сч}} > U_0$. Происходит пере-

текание накопленного зарядового пакета в более глубокую потенциальную яму, т. е. под электрод Э2.

Во время третьей фазы (диаграмма «в») потенциал с электрода 1 снимается и заряд полностью перетекает под электрод Э2. Третий электрод Э3 играет роль буфера, обеспечивающего однонаправленное перемещение заряда. Если бы он отсутствовал, то зарядовый пакет из ячейки Э1 мог бы равновероятно перетекать как вправо, так и влево.

Итак, зарядовый пакет перенесен на один шаг вправо и подготовлен к следующему перемещению. Чтобы переместить его на полную ступень (элементарную ячейку), надо произвести три переноса из ямы в яму, для чего требуется три отдельных тактовых импульса, как показано на временных диаграммах рис. 11.24, г.

В конце каждой строки имеется элемент вывода, не показанный на рис. 11.24. Таким элементом может быть, например,

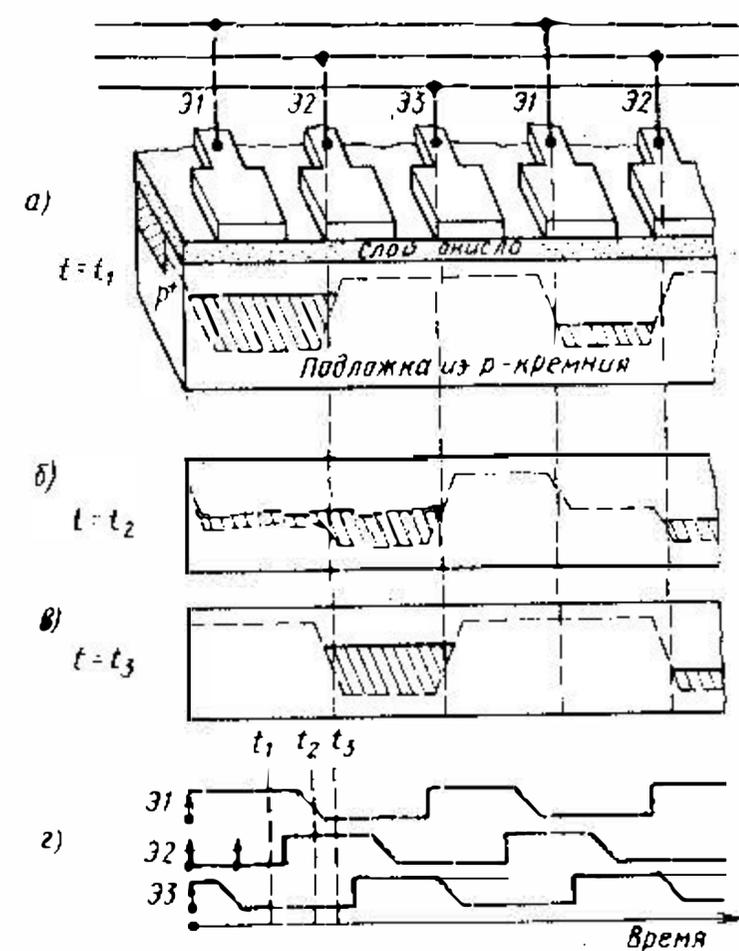


Рис. 11.24. Принцип работы твердотельного приемника изображения на ПЗС-матрице:

а) схема трехфазного n -канального ПЗС на p -Si и поперечные сечения потенциальных ям в момент времени $t=t_1$, соответствующий режиму накопления; потенциальные ямы в моменты времени t_2 и t_3 , соответствующие режимам переноса заряда (б) и его хранения (в); г) временные диаграммы тактовых импульсов

n^+ -область под последним электродом. Когда очередной зарядовый пакет достигнет этой области, он свободно пройдет через p - n^+ -переход, создавая на нагрузочном сопротивлении выходной сигнал. Таким образом все зарядовые пакеты могут детектироваться с помощью единственного выходного диода, изготовленного на той же подложке. В этом состоит одно из существенных достоинств приборов с переносом заряда.

Для удобства технической реализации перечисленных выше функций обычно их пространственно разделяют, для чего имеются секции накопления, хранения и выходной регистр. Единичный кадр возбуждается в секции накопления (Э1) в течение $1/25$ или $1/30$ с (ТВ-стандарты), затем быстро ($10^{-4} \dots 10^{-7}$ с) параллельно сдвигается в секцию хранения, из которой в течение времени накопления последующего кадра последовательно построчно переносится в выходной регистр. Обычно применяют два способа считывания из секции накопления: строчно-кадровый, при котором зарядовый пакет пробегает всю строку, и адресный или координатный, при котором зарядовый пакет от каждого элемента матрицы накопления передается в соседний с ним элемент матрицы хранения. В последнем случае обе матрицы как бы вставлены друг в друга. Такие структуры называют *фоточувствительными приборами с зарядовой инжекцией*.

Зарядовый пакет сохраняется ограниченное время (порядка $10^{-1} \dots 10^{-3}$ с). Рекомбинация и захват электронов на объемные и поверхностные центры приводит к искажению хранимой информации. При передаче зарядового пакета из ячейки в ячейку также происходит некоторая потеря информации вследствие взаимодействия электронов зарядового пакета с поверхностными ловушками, а также неполного перетекания зарядов. Для уменьшения этих нежелательных эффектов применяют ряд мер как в системе электрического питания устройства, так и при его технологическом исполнении. Кроме рассмотренной на рис. 11.24 простейшей структуры типа прибора с поверхностным каналом и одноярусным расположением электродов, существует много других разновидностей матричных фотоприемников на ПЗС, в том числе поверхностные фоточувствительные ПЗС с двух- и трехъярусным расположением электродов, фоточувствительные ПЗС с объемным (скрытым) каналом и др. Принцип их работы остается аналогичным рассмотренному выше.

Твердотельные приемники изображения делятся на две группы: *линейные (однострочные)* и *двумерные (матричные)*. Для получения двумерного изображения с помощью линейного приемника нужно применять механическое сканирование, например, вращающимся зеркалом. В матричных фотоприемниках накопление заряда происходит в течение всего времени кадра, поэтому их фоточувствительность выше, чем линейных приемников, где время накопления заряда ограничено проходом одной строки. Со-

временные матричные приемники оптических изображений на основе ПЗС характеризуются следующими основными параметрами: напряжение питания (амплитуда рабочих импульсов) — $10 \dots 30$ В; фронты управляющих трапецеидальных импульсов — $0,01 \dots 0,1$ мкс; максимальная тактовая частота — $(1 \dots 50)$ МГц; относительные потери при единичном акте передачи — $10^{-3} \dots 10^{-5}$; минимальная (пороговая) экспозиция, различимая на фоне шумов, — $(0,1 \dots 1)$ нДж/см²; динамический диапазон ~ 60 дБ; плотность темнового тока — $(5 \dots 20)$ нА/см²; чувствительность $S = 0,1 \dots 0,4$ А/Вт в спектральном диапазоне — $0,4 \dots 1,1$ мкм для кремниевых приборов; разрешающая способность — $(10 \dots 50)$ лин/мм. Основные области применения матричных фотоприемников на ПЗС — это телевизионная техника, распознавание образов, оптические измерения, фототелеграфия, ночное видение и т. д.

ГЛАВА 12

ОПТИЧЕСКИЕ МЕТОДЫ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ

Рассмотренные в предыдущих главах приборы и методы квантовой и оптической электроники настолько широко применяются в самых различных областях деятельности человека, что простое их перечисление заняло бы слишком много места. Впрочем, в этом нет нужды, поскольку каждый читатель, вероятно, сможет сам привести массу примеров применения, например, лазеров. Но кроме самостоятельного применения приборы оптической электроники в сочетании могут образовывать целые оптоэлектронные системы, позволяющие достигать качественно новых результатов. Примеры таких систем, служащих для передачи и обработки информации, рассмотрены в настоящей главе.

§ 12.1. ОСОБЕННОСТИ ОПТИЧЕСКИХ МЕТОДОВ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ

Использование оптического излучения для передачи информации дает существенные выигрыши, связанные: а) с резким возрастанием полосы частот и объема передаваемой информации (вследствие увеличения частоты несущей ν_0); б) с резким уменьшением размеров устройств (вследствие уменьшения длины волны λ_0) и в) с идеальной помехозащищенностью (вследствие нейтральности фотонов). Свойства лазерного излучения — когерентность, направленность и монохроматичность (см. § 2.6) — удовлетворяют требованиям, предъявляемым к несущей волне, чтобы с ее помощью осуществлять передачу информации, используя хорошо развитые в радиотехнике методы модуляции — амплитудную, частотную и фазовую. Оптические модуляторы позволяют кодировать информацию как в аналоговом, так и в цифровом виде до частот 10 ГГц и выше (§ 10.2), а фотоприемники (гл. 11) — принимать и декодировать эту информацию. Передающей средой может быть как естественная среда (космос, атмосфера,

вода), так и искусственные волноводы (световоды). Передача информации через атмосферу на большие расстояния возможна лишь на определенных частотах, попадающих в «окна прозрачности атмосферы» (§ 3.2), и то при хорошей погоде. Прохождение света через толщу морской или пресной воды, где имеется лишь небольшое «окно прозрачности» в сине-зеленой области спектра, еще более затруднено. Поэтому в земных условиях передачу оптических сигналов на большие расстояния осуществляют, как правило, по волоконным световодам, где достигнуты значительные результаты (см. § 12.3). В космическом пространстве лазерный луч может служить для передачи не только информации, но и энергии.

Принципы оптических методов обработки информации с помощью когерентного лазерного излучения отличаются от традиционных электронных методов, использующих, как правило, последовательную обработку сигналов. Известно, что эти методы оказались плохо приспособлены для анализа двумерных и тем более трехмерных изображений, включая буквенно-цифровые тексты и графики. Для решения таких задач создаются целые системы обработки информации, включающие специализированные процессоры, параллельные (матричные) вычислительные устройства, специальное математическое обеспечение и т. п. Все это потребовало затраты больших усилий, и напрашивается мысль, что эта техника по инерции зашла в область, ей не свойственную и не принадлежащую.

Оптические методы оказываются очень подходящими для решения целого ряда задач обработки изображений. Отметим, что оптика всегда имела дело с информацией в виде изображений и накопила здесь богатый опыт.

В качестве иллюстрации принципов обработки информации в когерентном свете рассмотрим преобразование оптического сигнала, выполняемое *тонкой положительной линзой*. Для этого представим сначала, что мы имеем плоский транспарант в виде окна с функцией пропускания $f(x, y)$, который освещается плоской монохроматической волной, распространяющейся в направлении z . Пусть комплексная амплитуда этой волны известна и равна A_0 . Тогда при прохождении транспаранта на самом окне в точке $z=0$ комплексная амплитуда будет равна $A_0 f(x, y)$. За счет явлений дифракции волновой фронт исказится. При $z>0$ он не будет плоским и его структура будет определяться функцией $f(x, y)$. Для нахождения распределения светового поля на достаточно большом расстоянии z разобьем транспарант на малые, почти точечные площадки аналогично тому, как это было сделано ранее на рис. 2.13 и 2.33. Согласно принципу Гюйгенса — Френеля, каждая такая точечная площадка излучает в пространство сферическую волну синфазно с падающей волной. Амплитуда каждой из волн с точностью A_0 равна $f(x, y)$. Распределение поля

в плоскости z (точнее — на сфере с радиусом $r=z$) будет определяться интегралом Френеля (2.34). Сравнивая (2.34) с интегралом Фурье (1.26), нетрудно заметить, что эти два преобразования имеют много общего, во всяком случае с математической точки зрения, при малых θ в (2.34). Строгий математический анализ дает следующий результат: в дальней зоне при больших z *распределение комплексных амплитуд на сфере радиусом z есть двумерный Фурье-образ функции пропускания окна (транспаранта) $f(x, y)$* . Этот результат имеет фундаментальное значение для обработки информации методами когерентной оптики. Любое смещение окна в его плоскости приводит только к появлению линейного фазового сдвига. Дальняя зона соответствует выполнению условия $z \gg r^2/\lambda$, где r — характерные поперечные размеры транспаранта (окна).

Вместо того чтобы удаляться на «бесконечность», можно взять собирающую линзу с фокусным расстоянием f , которая перенесет изображение из бесконечно удаленной плоскости в свою фокальную плоскость. Для осуществления точного преобразования Фурье необходимо транспарант $f(x, y)$ поместить в переднюю фокальную плоскость линзы, как показано на рис. 12.1. В этом случае распределение комплексных амплитуд $F(u, v)$ в задней фокальной плоскости P_2 в точности соответствует двумерному Фурье-образу распределения комплексных амплитуд $f(x, y)$ на транспаранте, т. е.

$$F(u, v) = \iint_{-\infty}^{\infty} f(x, y) e^{-\frac{2\pi i}{\lambda f}(ux + vy)} dx dy. \quad (12.1)$$

При помещении транспаранта не в фокальной плоскости в соотношении (12.1) появляется дополнительный фазовый множитель, который, впрочем, не играет существенной роли, если приемник регистрирует не амплитуду, а интенсивность волны.

Если сравнить (12.1) с (1.26), то нетрудно заметить, что переменные $u/\lambda f$ и $v/\lambda f$ в (12.1) аналогичны частоте ν в (1.26). По аналогии с терминологией, установившейся в электронике, они называются *пространственными частотами*. Их размерность — m^{-1} .

Хотя понятие пространственной частоты аналогично понятию временной частоты в электронике, между ними имеются и два важных различия.

1. **Пространственно-частотный спектр двумерен.** Это позволяет в дополнение к классическому частотному анализу выявлять другие зависимости, в частности — зависимость от ориентации в плоскости $f(x, y)$.

2. **Пространственно-частотный спектр функции $f(x, y)$ физически реален,** и для его измерения не нужно никаких спектроанализаторов. Достаточно взять линзу, осветить плоской волной транспарант с записью информации и наблюдать распределение света в задней фокальной плоскости линзы, которая будет частотной плоскостью.

Приведем два примера. Возьмем транспарант в виде окна с небольшим круглым отверстием. Это означает, что функция $f(x, y)$ имеет вид прямоугольной функции. Тогда на экране, помещенном в фокальной плоскости линзы, мы будем наблюдать дифракционную картину, аналогичную изображенной на рис. 2.33, б. Эта дифракционная картина и есть Фурье-образ прямоугольной функции.

Другой пример связан с гауссовыми пучками, рассмотренными в § 3.1. Известно, что функция Гаусса — единственная функция, инвариантная по отношению к преобразованию Фурье. Именно поэтому в линзовом волноводе, изображенном на рис. 3.3, будет выделяться и распространяться световой пучок с гауссовым распределением амплитуды.

Тот факт, что пространственно-частотный спектр функции реален и физически доступен, дает возможность выполнять над ним определенные операции, например, *пространственную фильтрацию*. Для этого достаточно поместить в фокальную плоскость P_2 на рис. 12.1 другой транспарант с амплитудным пропусканием $H(u, v)$, который будет осуществлять фильтрацию пространственных частот с передаточной функцией $H(u, v)$:

$$G(u, v) = F(u, v) H(u, v).$$

Восстановить отфильтрованное изображение очень просто с помощью дополнительной линзы L_2 , как показано на рис. 12.2. Изображенная схема называется «*схемой $4f$* » или *системой двойной дифракции*. Действительно, первая дифракция на апертуре первой линзы L_1 эквивалентна преобразованию Фурье распределения комплексных амплитуд на объекте $f(x, y)$. Затем свет

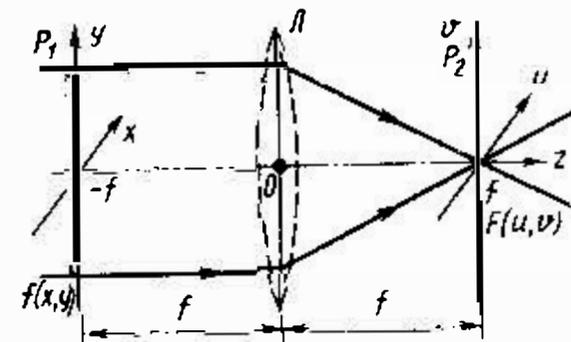


Рис. 12.1. Схема простейшей оптической системы, выполняющей преобразование Фурье

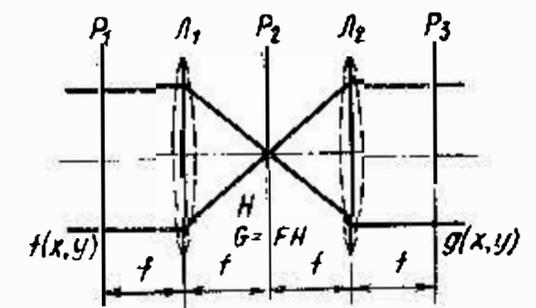


Рис. 12.2. Схема пространственной фильтрации вида $4f$ на основе двойной дифракции

проходит плоскость P_2 и, претерпев вторую дифракцию, достигает плоскости изображения P_3 . При этом вторая дифракция может быть осуществлена с помощью второй линзы или просто в результате распространения света в свободном пространстве между задней фокальной плоскостью P_2 первой линзы и плоскостью изображения. Нетрудно убедиться, что в результате этих операций выполняется двойное преобразование Фурье

$$F[F[f(x, y)]] = f(-x, -y)$$

в полном соответствии с интегральной теоремой Фурье.

В отсутствие пространственного фильтра H в плоскости P_3 будет воспроизводиться функция $f(x, y)$. При его наличии на выходе системы пространственной фильтрации формируется функция

$$g(x, y) = f * h,$$

где знак $*$ означает операцию сверки (1.154). Комплексная функция $h(x, y)$ в плоскости P_3 является Фурье-образом передаточной функции $H(u, v)$ в плоскости P_2 и называется *импульсным откликом пространственного фильтра*. Она описывает изображение, формируемое в плоскости P_3 , когда в плоскости P_1 на оптической оси помещен точечный источник, т. е. $f(x, y)$ имеет вид δ -функции. Действительно, если $f(x, y) = \delta(x, y)$, то

$$g(x, y) = \delta(x, y) * h(x, y) = h(x, y).$$

Отметим, что рассмотренная операция пространственной фильтрации *линейна и инвариантна* относительно смещения входного изображения в плоскости P_4 . Это выражается в том, что при смещении входного импульса в точку x_0, y_0 импульсный отклик системы пространственной фильтрации остается тем же самым, но смещенным, т. е. равным $h(x - x_0, y - y_0)$.

Рассмотренные простые, но важные примеры позволяют понять принципы и возможности обработки информации в когерентном свете. Из проведенного рассмотрения свойств оптической системы, изображенной на рис. 12.2, нетрудно заключить, что оптические методы позволяют:

- а) проводить параллельную (одновременную) обработку большого числа данных, например точек изображения;
- б) осуществлять двумерные свертки (в плоскости изображения P_3);
- в) осуществлять Фурье-преобразование, визуализировать Фурье-образ (в плоскости Фурье-образа P_2) и видоизменять его, осуществляя пространственную фильтрацию.

Важно, что все эти процедуры осуществляются мгновенно (со скоростью света).

Естественно, что мы рассмотрели идеализованную ситуацию и пренебрегли возможными искажениями, вносимыми реальными линзами за счет их конечных размеров, наличия аберраций и т. п. Их учет внесет некоторые погрешности, как и в любых реальных системах, но не изменит существа дела.

Другим ярким примером, иллюстрирующим возможности методов оптической обработки информации в когерентном свете, является галография, принципы которой рассматриваются в § 12.6.

Наряду с применением чисто оптических методов обработки информации в когерентном свете широко используются возможности современной *некогерентной оптоэлектроники*, примером чему могут служить оптроны и волоконно-оптические линии связи.

§ 12.2. ОПТОПАРЫ, ОПТРОНЫ И ОПТИЧЕСКИЕ ДАТЧИКИ

Оптопарой называют оптоэлектронный прибор, состоящий из излучателя (источника света) и фотоприемника, взаимодействующих между собой и объединенных в единую конструкцию. Оптопара представляет собой элементарное звено некогерентной оптоэлектроники, способное выполнять простейшие функциональные преобразования. Такое звено может использоваться либо как самостоятельный дискретный оптоэлектронный элемент электрической цепи, либо входить структурным элементом в топологию более сложных оптоэлектронных схем и устройств подобно тому, как транзистор или диод являются структурными элементами цепей полупроводниковой электроники.

Элементами оптопары служат рассмотренные в гл. 9 и 11 полупроводниковые источники и приемники оптического излучения. В качестве излучателей (И), как правило, применяют светодиоды, а в качестве фотоприемников (ФП) — фотодиоды и фоторезисторы. Связи между ними могут быть как электрическими, так и оптическими. В последнем случае важны свойства среды, посредством которой осуществляется оптическая связь. В состав оптоэлектронного прибора наряду с оптопарой могут входить дополнительные микроэлектронные или оптические элементы. Такие приборы часто называют *оптронами*. Этот термин иногда употребляется и для элементарной оптопары с внутренней оптической связью. Для расчета характеристик оптоэлектронных цепей можно пользоваться обычной теорией электрических цепей с заменой оптоэлектронных частей соответствующими электрическими эквивалентами. Оптрон как четырехполюсник характеризуется входными ($U_{вх}$; $I_{вх}$; $\Phi_{вх}$) и выходными ($U_{вых}$; $I_{вых}$; $\Phi_{вых}$) параметрами. Его передаточная функция определяется рассмотренными в §§ 9.2 и 11.1 эффективностью излучателя η и чувствительностью фотоприемника S , а также передаточными харак-

теристиками оптического ($K_{\text{опт}}$) и электрического ($K_{\text{эл}}$) каналов.

На рис. 12.3 представлены схемы двух элементарных оптронов. Оптопара с внутренней оптической и внешней электрической связью (рис. 12.3, а) обеспечивает практически полную электрическую развязку входа и выхода. Она получила очень широкое применение именно как элемент развязки, например, высоковольтных цепей с низковольтными, высокоомных с низкоомными и пр. Для достижения максимального коэффициента передачи такого оптрона необходимо не только применять излучатель и фотоприемник с высокими квантовыми выходами, но и обеспечить согласование их спектральных характеристик. Степень этого согласования оценивается корреляционным коэффициентом

$$K_{\lambda} = \frac{\int_0^{\infty} (\Phi_{\lambda}/\Phi_{\text{max}})(S_{\lambda}/S_{\text{max}}) d\lambda}{\int_0^{\infty} (\Phi_{\lambda}/\Phi_{\text{max}}) d\lambda}, \quad (12.2)$$

где $\Phi_{\lambda}/\Phi_{\text{max}}$ — относительная спектральная характеристика излучателя, а $S_{\lambda}/S_{\text{max}}$ — относительная спектральная чувствительность фотоприемника. С учетом эффективностей излучателя и приемника оптимальное сочетание представляет пара, состоящая из гетеросветодиода на основе арсенида галлия ($\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$) в качестве излучателя и кремниевого фотодиода в качестве приемника. Такая оптопара часто называется *диодной* (рис. 12.3, б). В качестве светопроводящей среды, соединяющей оптически излучатель и приемник, используют прозрачные компаунды, в том числе эпоксидные смолы, оптические клеи, вазелиноподобные полимеры и т. п. Они же осуществляют и механическое крепление элементов оптопары.

Коэффициент передачи диодной оптопары имеет низкое значение ($K \approx 1...5\%$), что определяется причинами, рассмотренными в § 9.2. Поэтому выходной сигнал, как правило, требует дополнительного усиления. Инерционность τ определяется свойствами светодиода и фотодиода и, как следует из §§ 9.2 и 11.5, может составлять 1 нс и менее. Для передачи информации в цифровом виде важным критерием является комбинированный параметр K/τ , который в диодных оптопарах достигает величины $10^7...10^8 \text{ с}^{-1}$.

Кроме диодных оптопар широкое распространение получили *резисторные оптопары*, где в качестве фотоприемника используется фоторезистор. По причинам, изложенным в § 11.3, фоторезисторы характеризуются линейностью и симметричностью вольт-амперных характеристик, отсутствием внутренних ЭДС

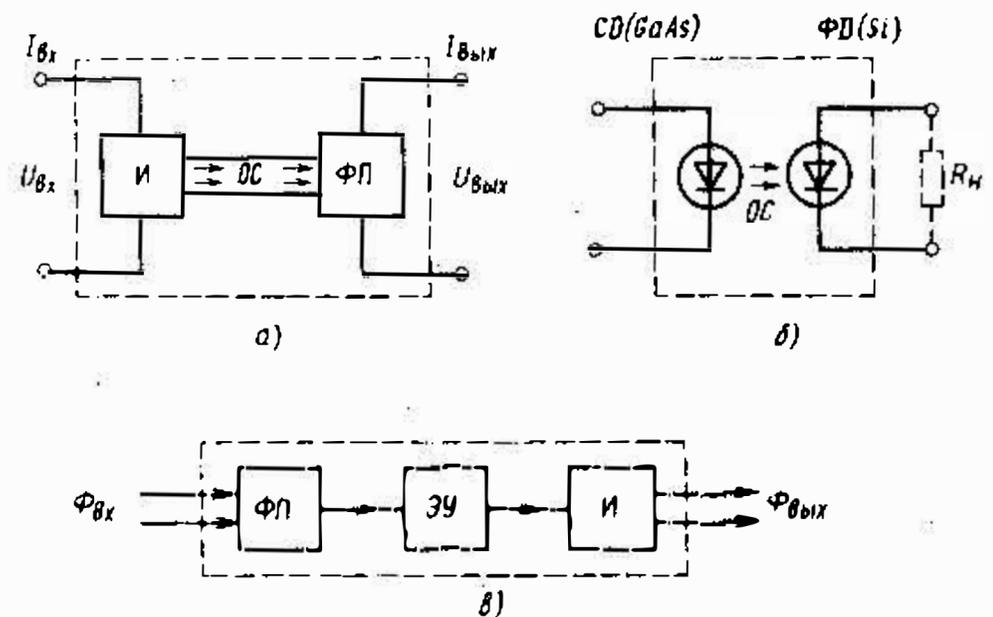


Рис. 12.3. Элементарные оптроны: а) оптопара с внутренней оптической связью и б) пример ее реализации в виде диодной оптопары; в) оптрон с оптическим входом и выходом (с внутренней электрической связью); И — излучатель; ОС — оптическая среда; ФП — фотоприемник; ЭУ — электронный усилитель, включающий схему питания И и ФП

и низким уровнем шумов. В резисторных оптопарах выходное сопротивление может изменяться в $10^7...10^8$ раз. Их недостаток — низкое быстродействие, ограниченное свойствами фоторезистора.

Существует большое число других разновидностей оптронов с внутренней оптической связью, в том числе — *транзисторные* и *тиристорные оптопары*, используемые в устройствах бесконтактного управления и коммутации высоковольтных силовых цепей. *Дифференциальные оптопары* — приборы, в которых один излучатель воздействует на два и более идентичных (или разных) фотоприемника. Такие приборы используются для передачи и корректировки аналоговых сигналов. Своеобразной разновидностью диодной оптопары является *волстон*, или *длинный оптрон*, где излучатель и фотоприемник неразрывно связаны друг с другом отрезком волоконно-оптического кабеля. Волстоны длиной от десятков сантиметров до нескольких метров незаменимы как элементы электрической развязки в высоковольтной аппаратуре.

По схеме оптрона с внутренней оптической связью работает транзистор с оптической связью, или *оптотранзистор*. Эмиттерный переход включается в прямом направлении и выполняет функцию светодиода. Сигнал через базу переносится фотонами, что существенно повышает быстродействие транзистора. Очевидно, что для достижения высокого значения K_{λ} в (12.2) такой прибор должен быть изготовлен на основе гетеропереходов и со-

держат пирозонный эмиттер и узкозонный коллектор.

Схема элементарного оптрона с внутренней электрической и внешней оптической связью приведена на рис. 12.3, в. Такой оптрон с оптическим входом и выходом представляет собой простейший преобразователь световых сигналов. Подбирая соответствующим образом спектральные характеристики фотоприемника и источника света, возможно преобразовать инфракрасное излучение в видимое. Для этого достаточно, например, взять PbS-фотосопротивление и GaP-светодиод. Если и фотоприемник, и излучатель в таком оптроне многоэлементные, то он может выполнять функцию преобразователя изображений.

Внутренняя оптическая связь в оптроне может быть не только прямой, но и обратной — как положительной, так и отрицательной. На рис. 12.4 показана схема и вольт-амперная характеристика оптрона с положительной оптической обратной связью. Характеристика такого оптрона аналогична характеристике триггера, а сам оптрон ведет себя как ключ.

Нетрудно предложить еще ряд типов оптронов, выполняющих различные операции, например, операции умножения, сложения, вычитания и т. п. Предоставляем это сделать самим читателям, выбрав при этом оптимальные для конкретной задачи источник и приемник излучения. Функциональные возможности оптронов весьма широки. По существу, они могут выполнять функции большого числа радиокомпонент или устройств. Например, переменные резисторы и потенциометры на основе оптопар не содержат подвижных механических контактов, которые являются основными источниками шумов и ненадежности.

Важным элементом оптопары является *оптический канал* между излучателем и фотоприемником. Во всех рассмотренных выше случаях через этот канал осуществлялась преломная передача энергии. Возможны также конструктивные решения, когда канал между излучателем и приемником либо делается открытым, и такая оптопара называется *оптопарой с открытым оптическим каналом*, либо он выполняется из материала, светопропускание которого изменяется при внешних воздействиях; такая система называется *оптопарой с управляемым оптическим каналом*. Оба эти прибора составляют основу для построения *оптических датчиков*.

Напомним, что *датчик*, или *сенсор**, предназначен для преобразования внешних воздействий в удобные для последующей обработки электрические или оптические сигналы. Среди подлежащих контролю величин отметим: координаты тела, линейное и угловое перемещение, скорость, ускорение, температуру, давление, влажность, химический состав газа или жидкости, скорость

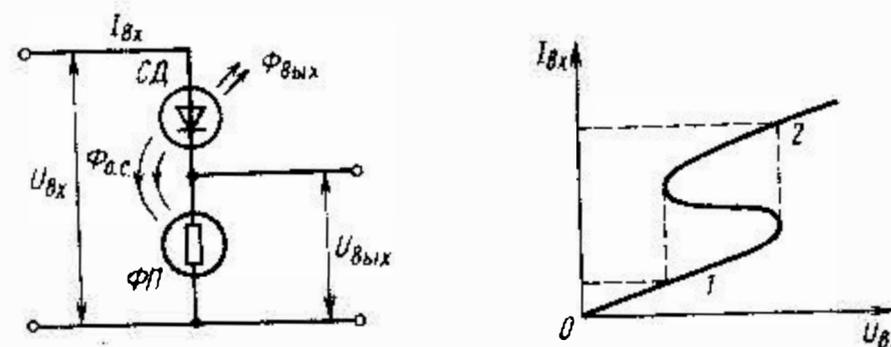


Рис. 12.4. Схема и вольт-амперная характеристика оптрона с положительной обратной связью

их потока, акустическое, электрическое и магнитное поля, световое и радиационное излучения. Эти и другие величины могут быть измерены оптическими методами с помощью рассмотренных в предыдущих главах эффектов, методов и приборов.

Простейшие оптические датчики представляют собой рассмотренные в гл. 11 полупроводниковые фотоприемники, преобразующие оптическое излучение в электрический сигнал, и светодиоды, выполняющие противоположную операцию (гл. 9). *Оптоэлектронный датчик* в общем виде представляет собой оптронную структуру, изображенную на рис. 12.3, вместе с вспомогательными элементами электронного и оптического обрамления. Чувствительным элементом датчика, как правило, является оптическая среда, свойства которой изменяются под влиянием внешних воздействий. Эти изменения могут быть непосредственными или косвенными, через предварительное изменение характеристик другой — промежуточной — среды или промежуточного элемента.

По принципу работы и конструктивным особенностям оптоэлектронные датчики можно разделить на четыре группы.

1. Оптопары с открытым оптическим каналом, где размещается или контролируемая среда, или промежуточный элемент.

2. Волоконно-оптические датчики типа волстонов, в которых чувствительным элементом является само волокно, оптические свойства которого изменяются под действием внешних факторов.

3. Интегрально-оптические датчики, использующие в качестве чувствительного элемента планарный световод (§ 3.4), изготовленный методами интегральной оптики; принцип действия таких датчиков основан на нарушении полного внутреннего отражения для лучей, распространяющихся вдоль световода, и «вытекание» их через границу раздела (§ 3.3) за счет приближения к ней промежуточной среды или изменения ее показателя преломления вследствие электро-, магнито- и пьезооптических эффектов (см. § 10.1).

*От англ. to sense — чувствовать, ощущать.

4. Датчики с волоконно-оптическими связями, в которых чувствительный элемент располагается в месте разрыва оптического волокна и воздействует на его светопередачу.

В качестве примера продемонстрируем возможности оптронов с открытым оптическим каналом как датчика. Конструктивно они могут быть двух типов: щелевого и отражательного (рис. 12.5). Эти датчики весьма удобны для контроля числа и положения объектов, считывания дискретной информации с переносителей, счетчиков ситуаций «да — нет», контроля состояния поверхности (по зеркальному и диффузному отражению), измерения скорости вращения и т. п. Их эксплуатационные характеристики можно повысить за счет создания на поверхности измеряемого или промежуточного элемента специального раstra в виде системы чередующихся определенным образом светлых и темных полос. Это позволяет преобразовывать механические перемещения в кодовые импульсы выходного сигнала.

Оптопары с открытым оптическим каналом весьма успешно используются для анализа газовых смесей и жидкостей. Принцип действия таких *оптоэлектронных газоанализаторов* основан на селективном поглощении молекулами анализируемого вещества оптического излучения в характерном для каждого вида молекул спектральном интервале. Анализируемое вещество помещается между излучателем и фотоприемником, обладающими заданными спектральными характеристиками. При известной длине оптического пути величина пропускания на заданной длине волны (или в узком спектральном диапазоне) определяется концентрацией поглощающих центров N через показатель поглощения k_{λ} (§ 1.8). Для улучшения рабочих характеристик прибора параллельно проводят измерения на опорной длине волны, соответствующей «прозрачной» области спектра. Сопоставление сигналов рабочего и опорного каналов, полученных от разных излучателей или фотоприемников, позволяет с высокой точностью определять содержание данного вещества в газе или жидкости. На этом принципе работают оптоэлектронные влагомеры, приборы экологического контроля, содержания углекислого и угарного газов,

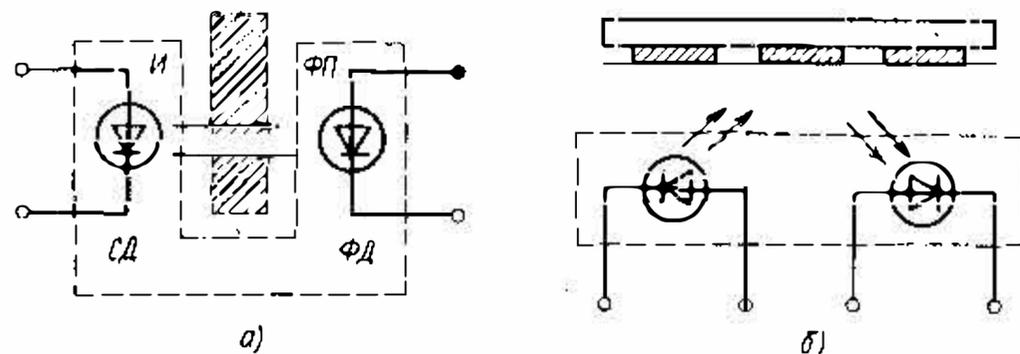


Рис. 12.5. Оптопары с открытым оптическим каналом щелевого (а) и отражательного (б) типов

сероводорода и других сернистых соединений, паров органических соединений, содержания кислорода в крови, нефти и кислот в воде, солей в нефти, дыма и пыли в атмосфере и т. п. Важно грамотно подобрать характеристики излучателя и фотоприемника.

§ 12.3. ВОЛОКОННО-ОПТИЧЕСКИЕ ЛИНИИ СВЯЗИ

Волоконно-оптические линии связи (ВОЛС) осуществляют передачу информации с помощью оптических сигналов по волоконным световодам. Этот вид связи обладает настолько большими достоинствами, а его приборная реализация достигла таких успехов, что в настоящее время ВОЛС вытесняет чисто электронные системы в таких областях, как телефония, кабельное телевидение, связь между ЭВМ, трансатлантическая и межконтинентальная связь и др.

Как и в случае электрической связи, в волоконно-оптической связи применяют две системы передачи информации и, следовательно, две системы модуляции — *аналоговую* и *цифровую* (рис. 12.6). Каждая из линий связи состоит из трех частей: *передающего*

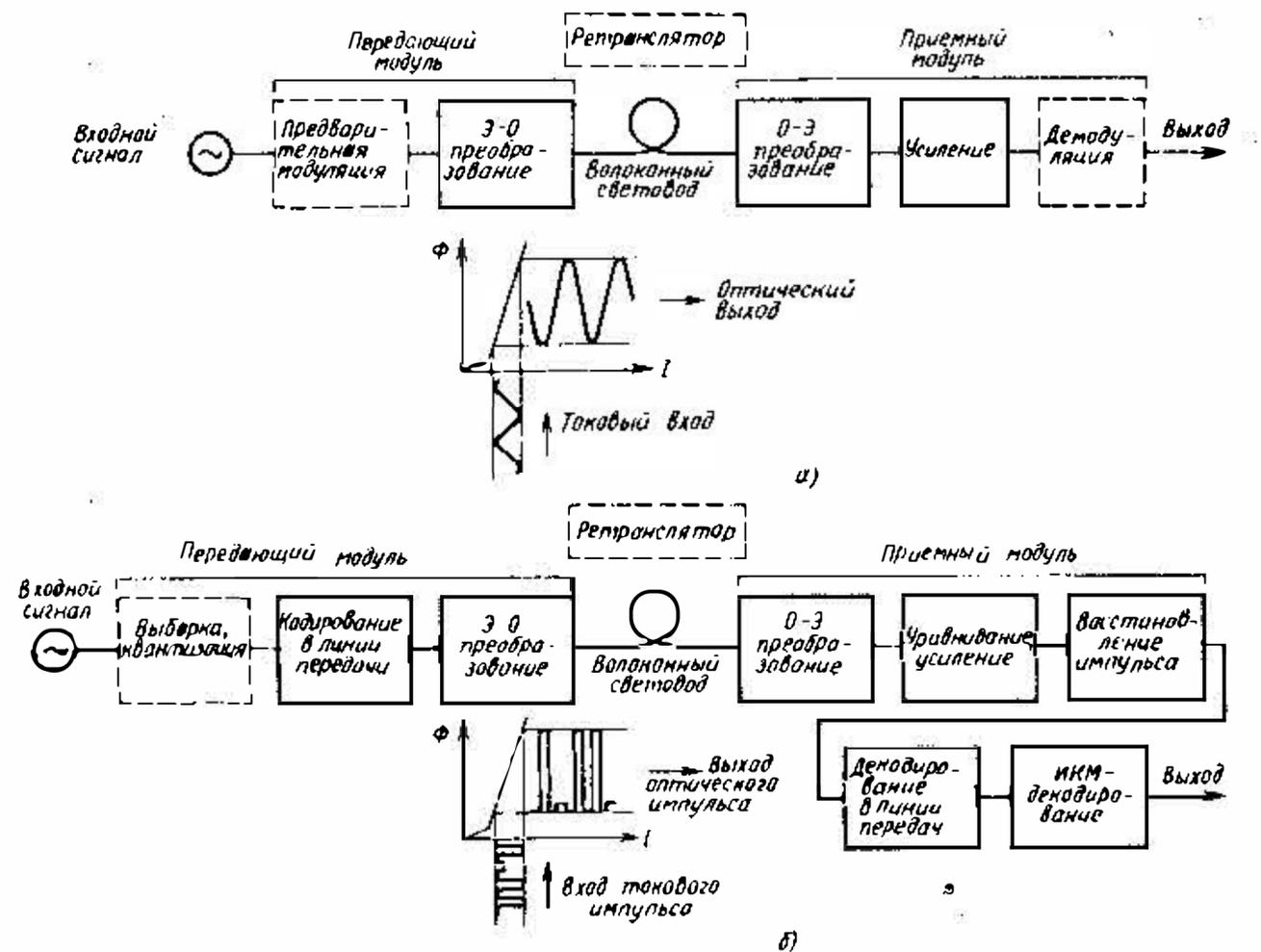


Рис. 12.6. Структурные схемы аналоговой (а) и цифровой (б) систем передачи информации по ВОЛС.

Пунктирными линиями показаны устройства, применение которых не обязательно

модуля, самой линии связи вместе с ретрансляторами (если в них есть необходимость) и приемного модуля. В передающем модуле подлежащие передаче электрические сигналы через электронную схему возбуждения и кодировки преобразуются в оптические сигналы. В качестве преобразователя электронного сигнала в оптический (Э — О преобразователя) в ВОЛС, как правило, используются или светодиоды, или инжекционные гетеролазеры. Оптическое излучение, закодированное входными сигналами, вводится в волоконный световод и распространяется по нему как по волноводу на требуемые расстояния (от нескольких метров до нескольких тысяч километров). При необходимости оно может быть усилено и подвергнуто дополнительному преобразованию в ретрансляторах. На выходе приемный модуль осуществляет преобразование оптических сигналов в электрические, их последующее усиление, декодировку и вывод на оконечное устройство.

Аналоговая система (рис. 12.6, а) заключается в модуляции интенсивности излучения пропорционально входному электрическому сигналу. Она проста, экономична, не требует использования кодирующих схем, однако на качество передачи очень сильно влияют разного рода нелинейные искажения, в частности — нелинейности вольт-амперных характеристик светодиодов и инжекционных лазеров. Поэтому для высококачественной передачи аналоговая система применяется лишь на небольшие расстояния, как правило, до 10 км.

Цифровая система (рис. 12.6, б) предполагает передачу информации в виде световых импульсов. В случае, когда на вход поступает аналоговый электрический сигнал, он предварительно кодируется в процессе импульсно-кодовой модуляции (ИКМ), мультиплексируется, затем кодируется в передающей схеме и после этого производится импульсная модуляция интенсивности излучения лазера или светодиода. В приемном модуле осуществляется обратное преобразование системы световых импульсов («цифр») в выходной сигнал, удобный для использования (например, телевизионное изображение). Такая система хотя и более сложна, но слабо чувствительна к нелинейным искажениям и возможным помехам. Она весьма удобна для реализации широкополосной передачи высокого качества, в том числе на большие расстояния.

Возможности светодиодов и полупроводниковых лазеров для передачи информации и фотоприемников для ее приема рассматривались ранее в гл. 9 и 11. Важнейшим элементом ВОЛС является волоконный световод, по которому производится направленная передача излучения. Именно создание световодов в виде тонких оптических нитей с исключительно малыми потерями обеспечило прогресс в технике оптической связи.

Волоконный световод представляет собой тонкую гибкую светопроводящую нить, выполняющую функцию диэлектрического

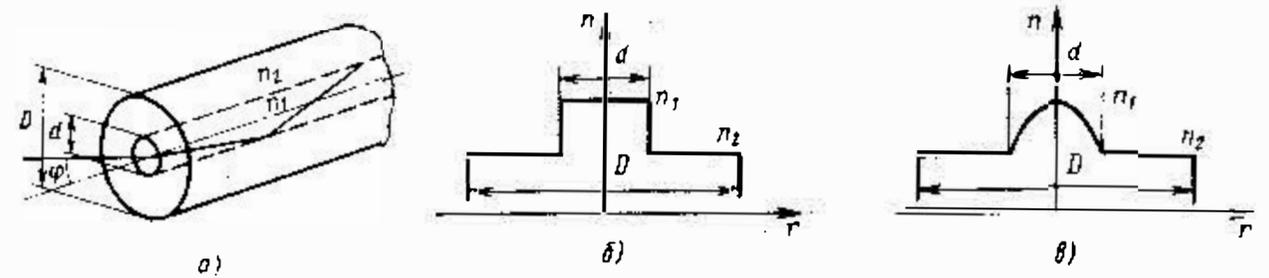


Рис. 12.7. Волоконный световод.

Общий вид (а) и радиальное распределение показателя преломления в ступенчатом (б) и градиентном (в) световодах

волновода, излучение по которому может передаваться на большие расстояния. Физика светопередачи по волоконному световоду аналогична рассмотренной в § 3.4 для планарного диэлектрического волновода. Волокно имеет цилиндрическую форму и в простейшем исполнении состоит из сердцевины диаметром d , окруженной прозрачной оболочкой с меньшим показателем преломления $n_2 < n_1$ (рис. 12.7). Удержание света, введенного в сердцевину световода, происходит за счет полного внутреннего отражения.

В зависимости от профиля радиального распределения показателя преломления различают *ступенчатые световоды* со скачкообразным изменением показателя преломления (рис. 12.7, б) и *градиентные световоды*, называемые также *градинами* или *селфоками**, показатель преломления в которых плавно увеличивается к центру, изменяясь примерно по параболическому закону.

Важнейшими параметрами, характеризующими распространение света в волоконном световоде, являются: 1) апертура; 2) потери; 3) модовая структура и 4) полоса пропускания.

1. **Апертура.** Удержание световых лучей внутри оптически более плотной сердцевины световода может быть обеспечено не для всех лучей, а лишь для тех, которые падают на его торец под углами φ меньше критического угла φ_0 (рис. 12.7, а). Этот критический угол φ_0 определяет угловую апертуру световода, а величина

$$NA = \sin \varphi_0$$

называется *числовой апертурой*. Излучение, заключенное внутри конуса с углом при вершине $\varphi = \varphi_0$, распространяется по волокну и образует *направляемые* или *канализуемые волноводные моды* световода аналогично рис. 3.8. Лучи, падающие на торец с $\varphi > \varphi_0$, после многократного повторения процесса отражения — преломления на границе сердцевина — оболочка выйдут из оболочки, и вся их энергия перейдет в оболочку, где в конце концов рассеется. Эта часть излучения образует *вытекающие* или *оболочечные моды*.

*От англ. selffocused — самофокусирующийся.

Критический угол φ_0 и числовая апертура NA легко находятся из условия полного внутреннего отражения (3.25) аналогично тому, как это было сделано в § 3.4:

$$NA = \sin \varphi_0 = \sqrt{\frac{n_1^2 - n_2^2}{n_0^2}} \approx n_1 \sqrt{2\Delta}, \quad (12.3)$$

где принято для воздуха $n_0 = 1$, а удельная разница показателей преломления Δ определяется соотношением (3.35). Для типичного световода, изготовленного из кварцевого стекла: $n_1 \approx 1,5$; $\Delta \approx 0,01$ и из (12.3) получаем $NA \approx 0,2$.

Учет изгиба световода приводит к уменьшению его эффективной апертуры. Нетрудно показать, что это будет сказываться лишь при малых радиусах изгиба, сравнимых с величиной $d/(NA)^2$.

2. Оптические потери характеризуются величиной затухания световой мощности (или интенсивности) на единичной длине световода, выраженной в децибеллах на км (дБ/км):

$$B = \frac{1}{L} 10 \lg (P_{\text{вх}}/P_{\text{вых}}). \quad (12.4)$$

Здесь $P_{\text{вх}}$ и $P_{\text{вых}}$ — мощности светового излучения на входе и выходе световода длиной L , измеряемой в км. Причинами потерь являются различные виды оптического поглощения, рассмотренные в § 4.4, и рассеяния (§ 1.7) в сердцевине световода, а также вытекание лучей из сердцевины в оболочку и потери в оболочке (§ 3.4). Если имеется несколько невзаимодействующих механизмов потерь, то полное затухание B складывается из удельных затуханий B_i , вносимых i -м механизмом потерь, т. е.

$$B = \sum B_i.$$

Спектральная зависимость затухания B определяется характерными спектрами поглощения и рассеяния для каждого из механизмов потерь. В качестве примера на рис. 12.8 представлен типичный спектр затухания в кварцевом оптическом волокне. Наибольший вклад в потери вносят:

а) Релеевское рассеяние (1.156), значение которого

убывает пропорционально четвертой степени длины волны, т. е.

$$B_{\text{рел}} = \varepsilon_{\text{рел}} \lambda^{-4}. \quad (12.5)$$

Это рассеяние происходит на различного рода нерегулярностях (флуктуациях), и потому для стекла постоянная $\varepsilon_{\text{рел}}$ тем меньше, чем ниже температура «замораживания» при его изготовлении. Для кварцевого волокна достигнуто $\varepsilon_{\text{рел}} \approx 0,7$ дБ/(км·мкм⁻⁴).

б) Собственное межзонное поглощение. Его экспоненциальный «хвост» (см. рис. 4.13 и 4.18) может простираться довольно далеко за край межзонных переходов, но поскольку длина оптического пути составляет километры, то даже ничтожное значение показателя поглощения в (1.170) сказывается на затухании. В кварцевом волокне межзонное поглощение заметно лишь в коротковолновой области при $\lambda \leq 0,6$ мкм.

в) Решеточное поглощение. Его «хвосты», вызванные многофоновыми процессами, ограничивают пропускание световода с длинноволновой стороны спектра (рис. 12.8).

г) Примесное поглощение. Полосы примесного поглощения элементов группы железа (Fe^{2+} , Cr^{3+} , Cu^{2+} , Ni^{2+}) попадают в область наивысшей прозрачности кварцевого волокна. Поэтому концентрация этих примесей должна быть менее 10^{-9} , что обеспечивают современные методы очистки.

Заметное примесное поглощение, от которого трудно избавиться в кварцевом волокне, обусловлено возбуждением локальных колебаний гидроксильной группы (ОН). Основной пик поглощения на колебаниях О — Н наблюдается при $\lambda = 2,72$ мкм. Менее интенсивные пики при $\lambda \approx 1,36$; 1,24 и 0,94 мкм вызваны обертонами этих колебаний. Рабочую длину волны ВОЛС на кварцевых волокнах необходимо выбирать между этими пиками поглощения (рис. 12.8).

Кроме указанных причин затухание света может быть связано как с качеством изготовления световода, так и с условиями его эксплуатации. Среди них: технологические разбросы параметров световода (эллиптичность сердцевины, флуктуации ее диаметра и показателя преломления); явления, связанные с дефектами эксплуатации (микронзгибы, микротрещины, механические напряжения, в том числе вследствие флуктуаций температуры и т. п.); явления, связанные с деградацией (старением) и действием ионизирующего излучения. В современных волоконных световодах, изготовленных на основе кварцевого стекла, эти потери могут быть сведены до минимума и заметных деградационных явлений в них за период эксплуатации не наблюдается. Пороговые значения плотности мощности для них составляют $10^5 \dots 10^8$ Вт·см⁻².

Из рис. 12.8 видно наличие трех окон прозрачности кварцевых световодов вблизи длин волн 0,85; 1,3 и 1,55 мкм. Именно на этих

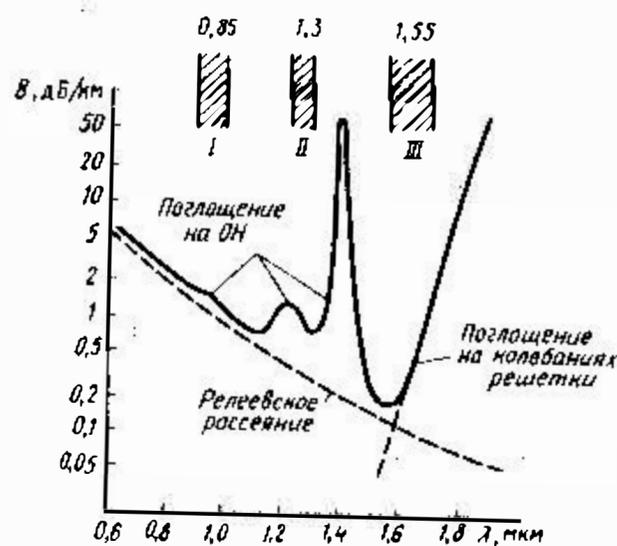


Рис. 12.8. Типичная спектральная зависимость потерь излучения в кварцевом волоконном световоде

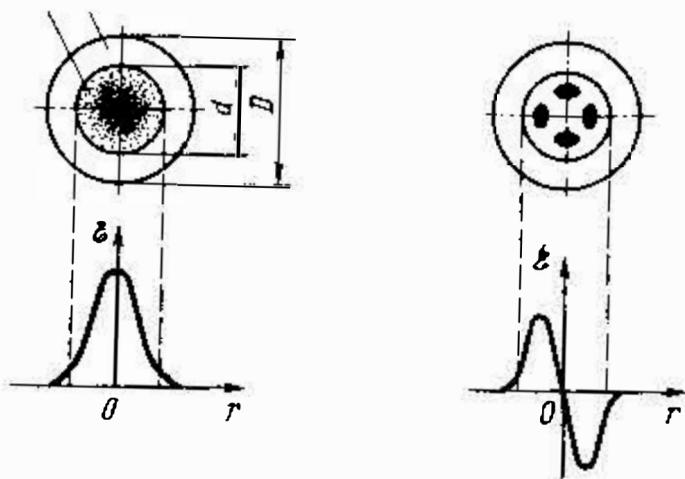


Рис. 12.9. Распределение поля электромагнитной волны для двух волноводных мод в цилиндрическом световоде

волновода (рис. 3.9). Однако цилиндрический волновод имеет другую симметрию. Он ограничен по двум координатам в отличие от планарного волновода, ограниченного по одной координате. Поэтому в оптическом волокне могут существовать более сложные моды, например, распространяющиеся по спирали и имеющие отличные от нуля проекции как вектора E , так и вектора H на ось световода. На рис. 12.9 представлены два вида распределения поля для аксиальной и первой неаксиальной моды в световоде. Сравнивая этот рисунок с рис. 2.16 и 2.17, нетрудно убедиться, что распределения поля для собственных типов колебаний в волноводе и оптическом резонаторе похожи.

Часть волноводных мод, которые могут распространяться в волноводе, определяется через параметр

$$V = \frac{\pi n_1 d}{\lambda_0} \sqrt{2\Delta}, \quad (12.6)$$

вытекающий из (3.35a). В планарном волноводе число мод N задается целым числом возможных значений m в (3.35a), т. е.

$$N \leq \left[\frac{2}{\pi} V \right] + 1. \quad (12.7)$$

В цилиндрическом волноводе $N \sim V^2$ и для больших значений V

$$N \approx V^2/2 = d^2 n^2 (2\pi/\lambda_0)^2 \Delta. \quad (12.8)$$

Для $V < 2,4$ в волокне может распространяться лишь одна TE_0 -мода. В лучевом приближении этой моде соответствует луч с $\varphi \approx 0$. Распределение интенсивности и амплитуды поля показаны в левой части рис. 12.9. Таким образом, условие $V < 2,4$ есть условие существования одномодового режима, а волокна, в которых это условие реализуется для заданной длины волны, называются

длинах волн работают ВОЛС первого, второго и третьего поколений соответственно.

3. Модовая структура. В оптическом волокне, являющемся цилиндрическим диэлектрическим волноводом, как и в планарном волноводе (§ 3.4), могут распространяться лишь вполне определенные волноводные моды, характеризуемые собственным распределением поля по сечению

ются одномодовыми волокнами. Из (12.8) видно, что одномодовый режим тем легче реализовать, чем больше λ и меньше Δ .

Проведем количественные оценки по (12.6), (12.8). Для типичного световода с диаметром сердечника $d = 50$ мкм $n_1 = 1,5$, $\Delta = 0,01$ получаем для $\lambda = 0,85$ мкм $V = 39$ и $N \approx 760$. Одномодовый режим будет реализовываться в волноводе с $d \leq 2,4\lambda_0/\pi n_1 \sqrt{2\Delta}$, т. е. в нашем случае с $d \approx 3$ мкм. Отметим, что в § 3.4 все рассмотрение было проведено для планарного волновода с резким изменением показателя преломления на границе раздела, а потому проведенные оценки справедливы для ступенчатого световода. Полезно напомнить также, что при уменьшении d , как видно из рис. 3.9, увеличивается доля мощности, просачиваемой к периферии сердечника и переносимой в оболочке световода.

4. Полоса пропускания световода определяется дисперсионными свойствами волокна и характеризует расплывание П-образного (или δ -образного) импульса при его распространении по световоду. Дисперсия* волокна определяется различием групповых скоростей распространения для различных составляющих оптического излучения. Подобно инерционным процессам в электрических цепях, дисперсия приводит к завалу частотной характеристики световода и к искажению передаваемых по нему световых импульсов. Это искажение тем существеннее, чем протяженнее световод, поскольку оно «набегает» по мере распространения света по волокну. Поэтому используют удельные временные параметры, приведенные к единице длины световода: полосу пропускания Δf_0 , МГц км; постоянную дисперсии τ_0 , нс/км, и уширение импульса σ_0 , нс/км. Величина Δf определяется по завалу частотной характеристики на 3 дБ, τ_0 — по времени нарастания (спада) импульса в e раз и σ_0 — по распылению единичного δ -импульса на уровне половины амплитуды. Между этими параметрами имеется простая взаимосвязь:

$$\Delta f_0 \tau_0 \approx 0,16; \sigma_0 \approx 2,2\tau. \quad (12.9)$$

Для оценки полосы пропускания световода длиной L величины τ_0 и σ_0 умножаются, а f_0 делится на L . По причине взаимодействия мод это может быть не всегда справедливо.

Задержка световой волны вызвана конечностью скорости ее распространения и зависит от эффективного показателя преломления $n_{эф}$. В свою очередь величина $n_{эф}$ зависит от постоянной распространения β_m в (3.36), которая различна для разных мод, и от длины волны λ вследствие дисперсии показателя преломления (§ 4.5). Поэтому выделяют модовую дисперсию и материальную дисперсию.

а) Модовая дисперсия возникает из-за разницы длины пути, проходимой разными модами $m = 0, 1, 2 \dots$ в световоде, что

*От лат. «dispergo» — раскидывать, развеивать, разгонять.

нетрудно понять из рис. 3.9, б. С увеличением номера моды увеличивается угол φ_m в (3.33), увеличивается длина пути и уменьшается эффективная скорость распространения, т. е. растет постоянная распространения β_m и эффективный показатель преломления (3.36). Это относится к *ступенчатым световодам*, где показатель преломления и скорость света постоянны по сечению сердечника. Для *градиентных световодов* (см. рис. 12.7, в) ситуация заметно улучшается. Действительно, в таких световодах световой луч вследствие рефракции (§ 3.3) распространяется по искривленной траектории, близкой к синусоиде. При этом значительную часть пути он проходит в областях с уменьшенным значением n , т. е. с большей скоростью. Это компенсирует увеличение длины пути: чем больше номер моды m , тем длиннее проходимый путь, но тем большую часть времени свет движется по периферийным областям с меньшими значениями n . Влияние модовой дисперсии на размытие светового импульса проиллюстрировано на рис. 12.10.

Оценить величину модовой дисперсии можно, пользуясь геометрическим приближением на рис. 3.9. Для этого определим разницу времен Δt распространения центрального осевого луча ($m=0$) и луча с $\varphi = \varphi_m$, после чего найдем $\Delta f = 1/\Delta t$. На единичной длине:

$$\Delta f_0 = 1/\Delta t L \approx \left(\frac{c}{n_1 \Delta} \right), \quad (12.10)$$

т. е. на единицу длины (км) с учетом (12.9) получаем межмодовую дисперсию для ступенчатого световода:

$$\sigma_0 \approx n_1 \Delta / (2\sqrt{3}c).$$

В градиентном световоде групповые скорости различных мод будут примерно одинаковы и временное расширение светового

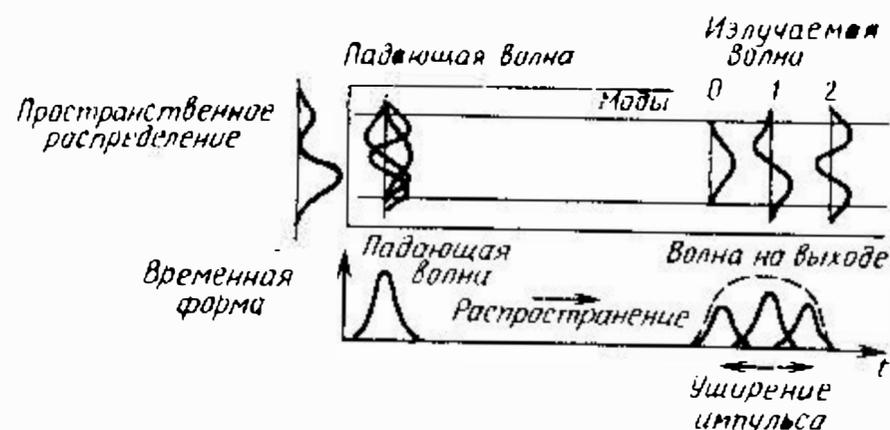


Рис. 12.10. Межмодовая дисперсия в многомодовом волоконном световоде

импульса при распространении по такому световоде будет невелико, если изменение показателя преломления по радиусу сердечника близко к параболическому, как показано на рис. 12.7, в. В оптимальном случае ширина полосы пропускания и величина дисперсии на единицу длины световода равны:

$$\Delta f_0 \approx 2 \frac{c}{n_1 \Delta^2}; \quad \sigma_0 \approx \frac{n_1 \Delta^2}{4\sqrt{3}c}. \quad (12.11)$$

Приняв $n_1 = 1,5$; $\Delta = 1\%$ для ступенчатого световода, по (12.10) получаем для $L = 1$ км $\Delta f_0 \approx 20$ МГц·км и $\sigma_0 \approx 15$ нс/км. Аналогично по (12.11) для градиентного световода находим $\Delta f_0 \approx 4$ ГГц·км и $\sigma_0 \approx 0,07$ нс/км, т. е. на два порядка лучше.

При рассмотрении модовой дисперсии мы предполагали, что взаимодействие между отдельными модами с разными индексами m отсутствует. В действительности при распространении света в волоконном световоде из-за изгибов и микронеоднородностей возникает взаимодействие между модами, происходит обмен энергией между медленной и быстрой модой. Это приводит к некоторому выравниванию времен распространения разных мод и уменьшению дисперсии. В результате уширения импульса и постоянная дисперсии не будут прямо пропорциональны длине световода L , а будут пропорциональны L^2 . Если связь между модами отсутствует, то $\gamma = 1$; при полной связи между модами $\gamma = 1/2$. Обычно параметр γ принимает промежуточные значения.

б) *Материальная дисперсия* возникает вследствие спектральной зависимости показателя преломления материала сердечника от длины волны. Как следует из § 1.8 и 4.5, в области прозрачности материала наблюдается нормальная дисперсия, т. е. $dn/d\lambda < 0$. Если на вход световода поступает световая волна со спектральным распределением шириной $\Delta\lambda$, как показано на рис. 12.11, то каждая из составляющих спектра (на рисунке выделены 3 длины волны) будет двигаться с разной скоростью. Это приведет к уширению светового импульса во времени по мере его распространения по световоде даже для одной моды $m=0$. Величина временного размытия импульса пропорциональна длине L .

Постоянная времени материальной дисперсии, приведенная к единице длины, примерно одинакова для всех волноводных мод и зависит лишь от ширины спектра излучения $\Delta\lambda$ и дисперсии показателя преломления материала:

$$\tau_0 \approx -\Delta\lambda \frac{\lambda}{c} \frac{\partial^2 n}{\partial \lambda^2}. \quad (12.12)$$

На рис. 12.12 приведена спектральная зависимость величины $\partial^2 n / \partial \lambda^2$ для чистого кварца. Вблизи $\lambda = 1,3$ мкм она проходит

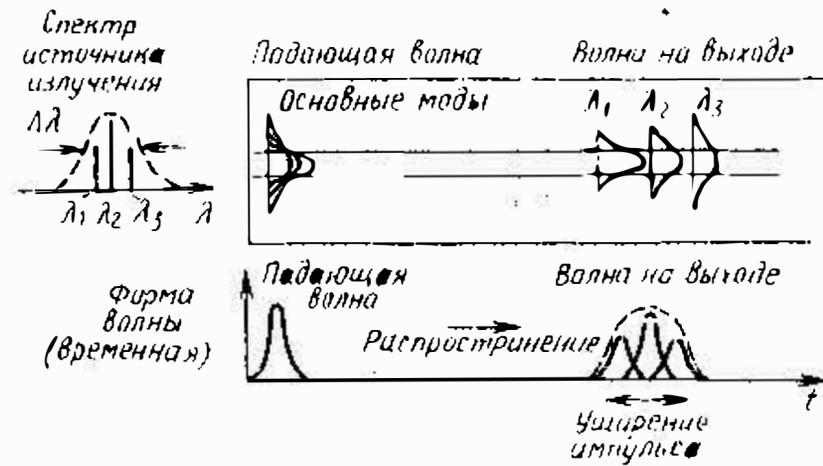


Рис. 12.11. Материальная дисперсия в одномодовом волоконном световоде

через 0. Это означает, что для кварцевого волокна материальная дисперсия при $\lambda = 1,3$ мкм минимальна. Поэтому одномодовые кварцевые волокна на этой длине волны обладают наибольшей пропускной способностью передачи информации.

Для уменьшения влияния материальной дисперсии на τ_0 , f_0 и σ_0 , как видно из (12.12), в качестве излучателя надо брать источник с узким спектром излучения. Оценим величину τ_0 . Если в качестве излучателя используется GaAs-светодиод с $\lambda_m = 0,85$ мкм и $\Delta\lambda = 50$ нм (рис. 9.10), то из рис. 12.12 находим $\frac{d^2n}{d\lambda^2} \approx 3 \cdot 10^{10} \text{ м}^{-2}$ и по (12.12) получаем $\tau_0 \approx 4$ нс/км. Аналогично, для GaAs-гетеролазера с $\Delta\lambda \approx 1$ нм (рис. 9.37) имеем $\tau_0 \leq 0,1$ нс/км. Для GaInAsP/InP-гетеролазера, работающего при $\lambda \approx 1,3$ мкм, величина τ_0 составляет сотые доли нс/км.

Спектральная зависимость скорости основной моды ($m=0$) определяется не только дисперсией материала (12.12), но и более слабыми эффектами, связанными с так называемой *дисперсией конструкции*. Последняя определяется соотношением длины волны излучения и диаметра сердцевинки световода. Как видно из (3.33), для разных длин волн при одном и том же m коэффициенты распространения различны. Это объясняется тем, что для разных λ глубина проникновения волны в оболочку различна.

Все виды дисперсии присутствуют для каждой из мод и в многомодовом световоде. Из рис. 12.10 и 12.11 можно заключить, что уширение импульса во времени носит неоднородный характер и по аналогии с рассмотренными в § 1.6 механизмами уширения спектральных линий можно заключить, что это уширение будет иметь гауссову форму с той разницей, что оно будет функцией времени, а не частоты, как в (1.151).

Таким образом, дисперсия групповой скорости D_g волны, распространяющейся по световоду, определяется материальной дисперсией $(\frac{d^2n}{d\lambda^2})_m$ и волноводной дисперсией $(\frac{d^2n}{d\lambda^2})_w$:

$$D_g(\lambda) = -\frac{\lambda}{c} \left[\left(\frac{d^2n}{d\lambda^2} \right)_m + \left(\frac{d^2n}{d\lambda^2} \right)_w \right]. \quad (12.13)$$

Последняя определяется модовой дисперсией через (12.10) и (12.11), а также дисперсией конструкции. Важно, что в некоторой узкой области длин волн знак для каждого из слагаемых в (12.13) может быть различным. Материальная дисперсия в SiO_2 проходит через 0 вблизи $\lambda = 1,3$ мкм (рис. 12.12, а) и ее положение может сдвигаться за счет легирования GeO_2 . Длина волны, при которой проходит через 0 волноводная дисперсия, зависит от диаметра сердцевинки d и относительной разницы показателей преломления n_1 и n_2 . Управляя этими факторами, возможно минимизировать дисперсию групповой скорости в одномодовом волокне и сдвинуть ее в область наивысшего пропускания к $\lambda = 1,5 \dots 1,6$ мкм, как это показано на рис. 12.12, б. Хорошие результаты получаются при $d = 4 \dots 5$ мкм и $(n_1 - n_2)/n_1 > 0,004$.

Техническая реализация световодов базируется в основном на двухслойных оптических волокнах, изготовленных из кварцевого стекла. В зависимости от диаметра сердцевинки и профиля показателя преломления различают три разновидности волокон: *многомодовые ступенчатые* (рис. 12.7, б), *многомодовые градиентные* (рис. 12.7, в) и *одномодовые*. Для первых двух $d \approx 50$ мкм, а для последних $d \approx 5$ мкм, причем для одномодовых световодов закон изменения показателя преломления внутри сердцевинки не имеет существенного значения; обычно он близок к ступенчатому. Вс-

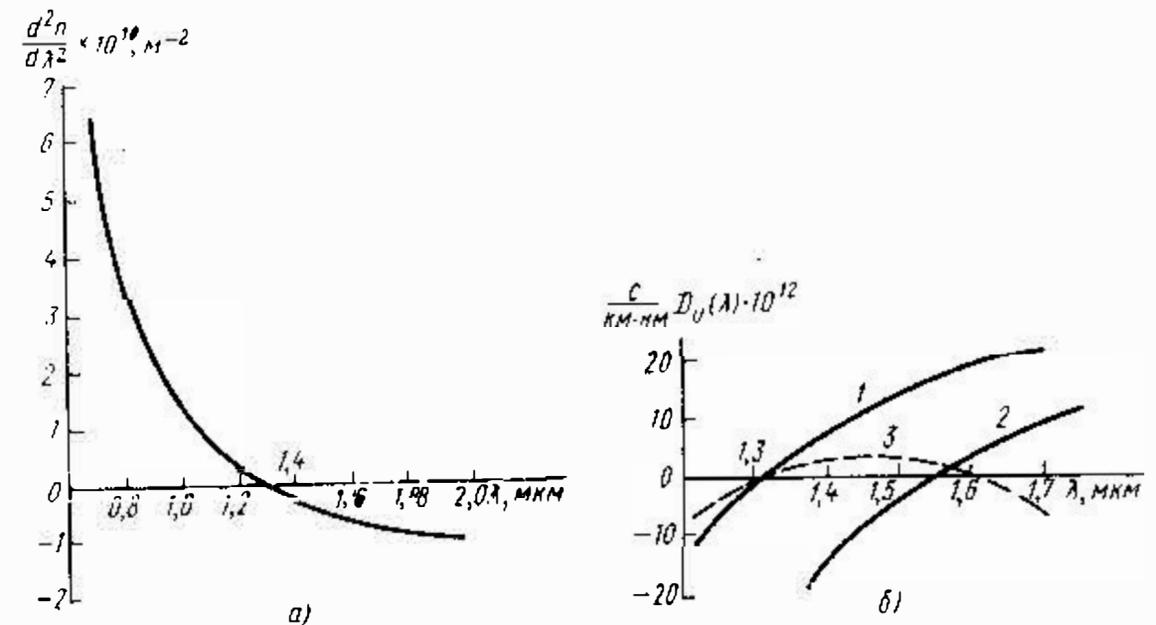


Рис. 12.12. Спектральные зависимости дисперсия показателя преломления $\frac{d^2n}{d\lambda^2}$ чистого кварца (а) и дисперсии групповой скорости D_g в кварцевом волокне (б):

1 — кварцевое одномодовое волокно, 2 — волокно со сдвинутой дисперсией, 3 — волокно с выравненной дисперсией

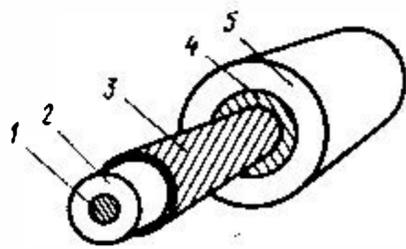


Рис. 12.13. Конструкция одножильного волоконного кабеля:

1 — сердцевина (SiO_2 ; GeO_2 ; P_2O_5); 2 — оболочка (SiO_2); 3 — тонкое защитное покрытие (кремнийорганика); 4 — амортизирующий слой; 5 — внешнее защитное покрытие

личина $\Delta \approx 0,01 \dots 0,02$ достигается за счет легирования кварцевого стекла добавками GeO_2 , P_2O_5 и др. Такие кварцевые световоды изготавливают путем вытягивания из цилиндрической кварцевой заготовки с легированным сердечником, полученным методом осаждения из газовой фазы. В качестве исходных продуктов реакции используются SiCl_4 и O_2 полупроводниковой чистоты. Одновременно с окончательным процессом вытяжки волокно покрывают тонким полимерным покрытием. Это защитное покрытие «залечивает» микротрещины в кварцевой оболочке и предохраняет волокно от влаги и смягчает механические воздействия. Такая структура из сердцевины, оболочки и тонкого защитного покрытия, получаемая в едином технологическом цикле, представляет основу для последующего оформления ее в световодный кабель. Согласно международному стандарту, предпочтителен внешний диаметр оболочки, равный 0,125 мм.

Для улучшения эксплуатационных свойств волокно покрывают дополнительными защитными покрытиями, как показано на рис. 12.13, получая одножильный *волоконный кабель*. Такой кабель может быть использован в технике оптической связи или как **самостоятельный** элемент, или на его основе могут быть изготовлены жгуты и многожильные кабели разнообразных конструкций.

Для улучшения эксплуатационных свойств волокно покрывают дополнительными защитными покрытиями, как показано на рис. 12.13, получая одножильный *волоконный кабель*. Такой кабель может быть использован в технике оптической связи или как **самостоятельный** элемент, или на его основе могут быть изготовлены жгуты и многожильные кабели разнообразных конструкций.

Естественно, что стоимость волоконно-оптического кабеля очень сильно зависит от его качества, в том числе — от величины затухания B и профиля распределения n . Для получения волокон с минимальным затуханием $< 0,2$ дБ/км в области длин волн 1,3 и 1,55 мкм требуются исключительно чистые исходные компоненты и прецизионная технология. Поэтому такие волокна дороги и применяются для связи на большие расстояния, в том числе — для сверхдальней трансатлантической связи.

Кроме двухслойных кварцевых световодов, обладающих наилучшими характеристиками, применяют: а) *кварцево-полимерные волокна*, сердцевина которых изготовлена из чистого кварца, а оболочка — из полимера (обычно кремнийорганического); б) *стеклянные волокна* из многокомпонентных оптических стекол; в) *полимерные световоды*. Последние наиболее дешевы, просты в изготовлении, удобны в эксплуатации, но обладают повышенными потерями (100 дБ/км и более).

Элементы связи и компоненты оптических схем. Как и в электрических цепях, важным моментом в создании *оптических цепей* является соединение отдельных элементов ВОЛС друг с другом, коммутация и разводка оптических сигналов. Для этих цепей

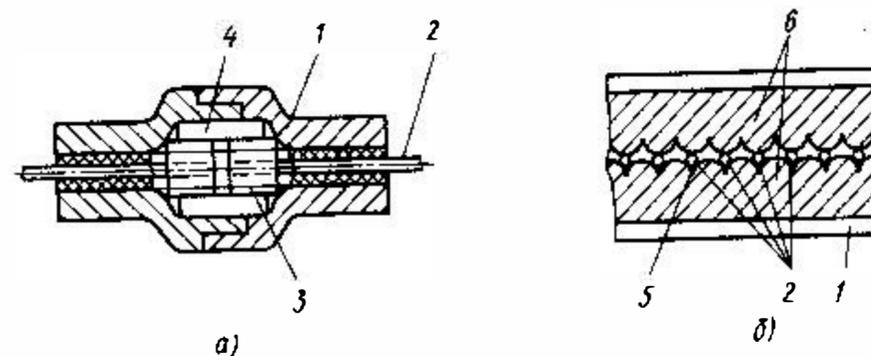


Рис. 12.14. Оптические соединители (разъемы):

а — однополосный; б — многополосный; 1 — корпус, 2 — световод (волокно), 3 — армировка, 4 — направляющая муфта, 5 — V-образные канавки, 6 — цилиндрические выравнивающие элементы

разработаны пассивные элементы оптической связи, включая оптические соединители, ответвители, смесители, переключатели, элементы ввода — вывода излучения, мультиплексоры и демultipлексоры и т. п. Все они работают на основе различных оптических эффектов, рассмотренных в гл. 3 — 5 и 10.

1. Оптический соединитель (*оптический разъем*) предназначен для разъемного сочленения друг с другом двух и более концов оптических волокон. Упрощенная конструкция одноволоконного соединителя приведена на рис. 12.14, а. По внешнему оформлению и внутреннему устройству он напоминает обычный высокочастотный электрический разъем. В многоволоконных соединителях чаще всего используются конструкции с V-образными канавками, в которые укладываются световодные жилы, как показано на рис. 12.14, б. Требования к точности изготовления и юстировке оптических разъемов значительно выше, чем электрических. Это обусловлено как малыми значениями диаметра сердцевины волокна, так и жесткими допусками на стыковку волокон друг с другом. Вносимые разъемом потери, как правило, не должны превышать 1 дБ. Это вполне реально, но требуется высокая прецизионная точность во всех компонентах сочленения, особенно для одномодовых волокон с $d \leq 5$ мкм.

2. Оптический разветвитель предназначен для распределения по заданному закону излучения, подаваемого на вход, между его выходами. В качестве примера на рис. 12.15 показана схема оптического разветвителя волноводного типа. Если использовать эту схему в обратном направлении,

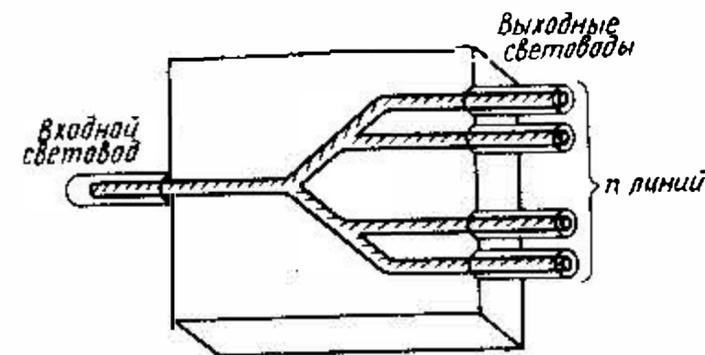


Рис. 12.15. Схема оптического разветвителя волноводного типа

то получим *оптический смеситель* световых сигналов. Передача части оптического излучения из одного световода в другой может производиться за счет эффекта «просачивания» волны из сердцевины в оболочку (рис. 12.9). Регулируя расстояние между сердцевинами волокон, можно изменять коэффициент передачи излучения. В общем случае оптический смеситель допускает смешение m световых пучков и ответвление n пучков. Он может быть изготовлен методами интегральной оптики (см. § 12.4) или путем плотного скручивания, нагрева и растягивания нескольких волокон так, что на участке сплавления световодов происходит объединение оптических каналов.

3. Оптические коммутаторы предназначены для переключения (коммутации) светового потока с m -го входа на n -й выход. В частном случае $m=n=1$ такое устройство представляет собой *оптический переключатель*, перекрывающий или устанавливающий световой контакт между излучателем и приемником. Простейшая коммутация осуществляется, например, с помощью механически управляемого зеркала или призмы. Более совершенны и обладают лучшим быстродействием коммутаторы, использующие в основе работы рассмотренные в § 10.1 электро-, магнито- и акустооптические эффекты. Весьма эффективно исполнение этих устройств методами интегральной оптики. Такие коммутаторы удобны лишь при сопряжении небольшого числа каналов (до 10). Коммутация большого числа каналов возможна с использованием принципов голографии (§ 12.5).

Основные требования к оптическим коммутаторам — надежность, эффективность (малые потери) и быстродействие.

4. Оптические мультиплексоры и демultipлексоры применяются в многоканальной оптической связи и предназначены для введения в один световод (в случае мультиплексора) или разделения (в случае демultipлексора) множества оптических сигналов различных длин волн. Кроме указанных, разработан целый ряд других элементов связи, в том числе элементы ввода — вывода оптического излучения, оптические attenuаторы, фильтры, линии задержки, светоделители и др., с помощью которых из отдельных элементов komponуются целые оптоэлектронные системы.

Таким образом, в распоряжении инженеров уже сегодня имеются все необходимые элементы для технической реализации оптической связи по волоконным световодам. Из изложенного выше можно заключить, что ВОЛС обладают следующими основными особенностями:

1. *Малые потери* и слабое искажение сигнала, что позволяет увеличить расстояние между ретрансляторами до 50...100 км, а в перспективе — свыше 1000 км. Уже сегодня это расстояние на

порядок больше, чем в применяющихся до сих пор медных линиях ВЧ-связи.

2. *Сверхвысокая пропускная способность*, достигающая 10 ГГц/км, обеспечивается широкополосностью всех элементов ВОЛС (излучатель — световод — фотоприемник) и позволяет резко увеличить объем передаваемой информации.

3. *Малые габариты* и масса дают существенный экономический выигрыш при прокладке ВОЛС. Неограниченный сырьевой ресурс для производства кварцевых волокон и их стойкость к коррозии также улучшают экономические и эксплуатационные характеристики ВОЛС.

4. *Нечувствительность* к электромагнитным воздействиям кардинально решает проблему электромагнитной совместимости, позволяя прокладывать ВОЛС вблизи силовых электрических сетей, не опасаясь удара молний и обеспечивать высокую помехозащищенность.

Все это обеспечивает быстрое и уверенное развитие ВОЛС. Во всех технически развитых странах этому направлению связи уделяется первостепенное внимание, поскольку она отвечает современным требованиям передачи больших объемов информации на большие расстояния.

Для каждого конкретного применения надо обоснованно выбирать наиболее экономичные виды излучателей (светодиод или лазерный диод), фотоприемников, световодов и определять рабочую длину волны. На рис. 12.16 схематически показаны технические данные ВОЛС, применяемых в различных областях.

По оси абсцисс на рис. 12.16 отложена дальность связи, по оси ординат — скорость передачи информации. В зависимости от соотношения этих величин различают три группы ВОЛС, в кото-

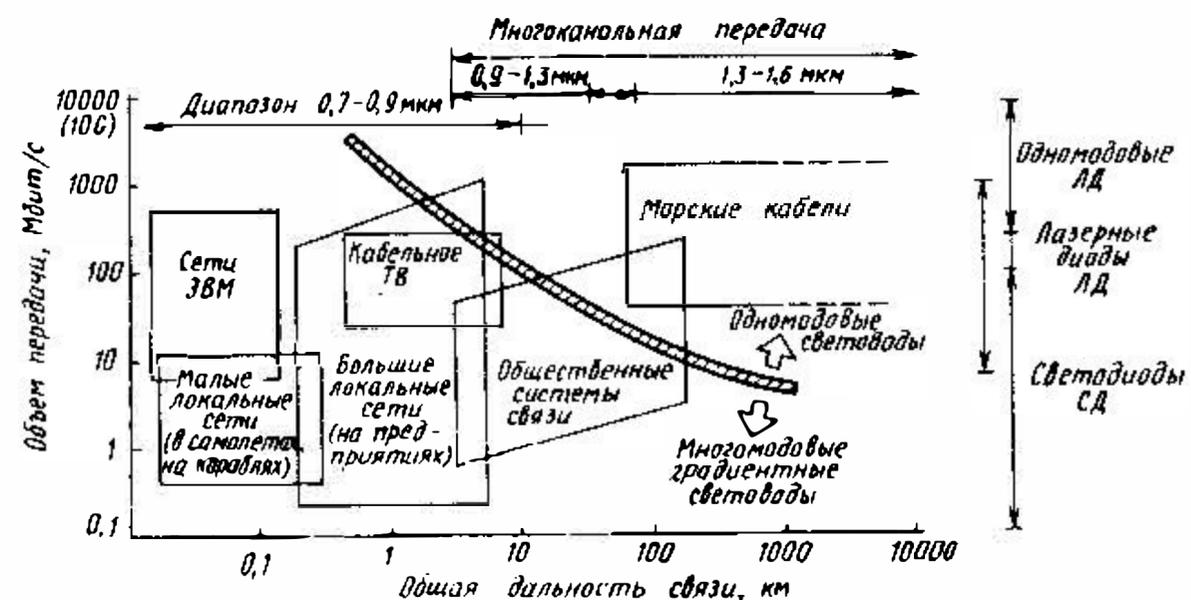


Рис. 12.16. Области применения ВОЛС

рых используются разные виды излучателей, фотоприемников и световодов. Часто они называются ВОЛС первого, второго и третьего поколений.

К первой группе относятся ВОЛС с небольшой скоростью передачи информации (десятки Мбит/с) на расстояния до десяти км. В этом случае в качестве излучателя используют светодиоды на основе арсенида галлия с рабочей длиной волны 0,85...0,9 мкм, в качестве фотоприемника — кремниевый *p-i-n*-фотодиод, а в качестве оптического кабеля — многомодовый ступенчатый световод.

Ко второй группе относятся ВОЛС с высокой скоростью передачи информации (до 100 Мбит/с и более) на средние расстояния от нескольких десятков до сотен километров. В этих системах необходимо использовать градиентные или одномодовые световоды и рабочую длину волны вблизи 1,3 мкм (где минимальна дисперсия). Поэтому излучателями здесь служат полосковые гетеролазеры в системе InGaAsP/InP, а фотоприемниками — лавинные гетерофотодиоды.

К третьей группе относятся ВОЛС на сверхдальние расстояния для морских и трансатлантических линий связи со средним и большим объемом передаваемой информации. Рабочая длина волны выбирается из условия минимума потерь (при 1,55 мкм в кварцевых волокнах $B \approx 0,2$ дБ/км, см. рис. 12.8). Используют волокна со сдвинутой или выравненной дисперсией (рис. 12.12, б). В качестве излучателя применяют одномодовый лазер, в качестве приемника — лавинный фотодиод на рабочую длину волны 1,55 мкм. Усиление сигнала в ретрансляторе осуществляется с помощью лазерного усилителя на кварцевом волокне, активированном эрбием (см. гл. 8). Накачка такого усилителя производится полупроводниковым лазером на длине волны 1,48 мкм или 0,98 мкм. Применение оптического усилителя не требует разрыва оптической цепи и особенно эффективно при импульсно-кодовой модуляции сигнала.

Развитие волоконно-оптических линий связи происходит весьма быстро. В 1988 г. завершено строительство трансатлантических и транстихоокеанских линий: подводных оптических магистралей между Европой и Америкой через Атлантический океан и между Америкой и Японией через Тихий океан. Заканчивается прокладка кабеля протяженностью 16 000 км от Австралии и Новой Зеландии через Гавайи до Северной Америки. С завершением строительства Транссибирской магистрали глобальное кольцо волоконно-оптической связи замкнется и даст возможность человечеству использовать фантастические возможности общения и обмена информацией по всей планете.

Длина регенерационных участков на дальних линиях передач составляет 100 — 150 км. При прохождении такого расстояния оптический сигнал ослабляется настолько, что его надо уси-

вать. До недавнего времени это осуществлялось специальным электронным усилителем, который встраивался в линию связи и осуществлял ретрансляцию оптического сигнала. В таком ретрансляторе (см. рис. 12.6) оптический сигнал выводился из волокна на фотоприемник, преобразовывался в электрический сигнал, усиливался электронными методами, затем с помощью полупроводникового лазера преобразовывался опять в оптический сигнал и вводился в оптическое волокно. Это требовало разрыва волокна, что наряду с необходимостью двойного $O \rightarrow \Delta$ и $\Delta \rightarrow O$ преобразования вносило дополнительные потери.

Этих недостатков лишены волоконно-оптические усилители, выполненные в виде активных оптических волокон, легированных ионами редкоземельных элементов: эрбием (Er^{3+}) для $\lambda = 1,55$ мкм (см. § 8.3 и табл. 8.1) или празеодимом (Pr^{3+}) для $\lambda = 1,3$ мкм. Они работают как твердотельные лазерные усилители на диэлектриках по схеме рис. 2.3 с той особенностью, что активный элемент выполнен в виде стеклянного волокна. Длина этого активного волокна, являющегося усилительным элементом, встроенным в световод, составляет от 0,5 до 10 см. Его накачка осуществляется полупроводниковыми лазерами или светодиодами, спектры излучения которых соответствуют спектру поглощения иона-активатора.

Для сверхдальних линий передач желательно иметь оптические волокна с потерями менее 0,01 дБ/км. Как видно из рис. 12.8, даже в идеально чистых, не содержащих примесей кварцевых волокнах эта величина не может быть достигнута из-за «хвоста поглощения» на колебаниях (Si — O). В лабораторных условиях созданы оптические волокна из тетрафторида циркония, имеющие на длине волны 2,5 мкм затухание 0,01 дБ/км и из фторида бериллия с затуханием 0,005 дБ/км на длине волны 2,1 мкм. Предполагается к началу XXI века на основе таких волокон осуществить прокладку трансатлантической линии протяженностью около 6000 км без регенерационных усилителей.

Развитие цивилизации требует постоянного увеличения объема передаваемой информации. Современные ВОЛС работают со скоростью 2,4 Гбит/с, а со спектральным уплотнением потоков на нескольких близко расположенных длинах волн вблизи 1,55 мкм — до 10 Гбит/с. Доказана возможность передачи информации по ВОЛС на скоростях 100 Гбит/с. Основной вопрос, который надо решить для реализации таких скоростей, — как компенсировать расплывание светового импульса по мере его распространения по волокну (см. рис. 12.10 и 12.11). Оказывается, что существуют структурно устойчивые волновые образования, называемые солитонами, которые ведут себя как отдельные частицы: при взаимодействии между собой или с некоторыми другими возмущениями они не разрушаются, а расходятся вновь, сохраняя свою структуру неизменной. Солитон — структурно

устойчивая уединенная волна, которая при определенных условиях может существовать в нелинейной диспергирующей среде. Его структура поддерживается неизменной за счет баланса между действием нелинейностей среды. Дисперсии как бы имеют разные знаки и компенсируют друг друга. Распространяясь в такой среде, солитон как волновой пакет не меняет свою форму. Впервые на такое образование обратил внимание в 1834 г. шотландец Дж. Рассел, который наблюдал его в виде возвышения, бегущего по поверхности воды в канале.

При скоростях передачи информации 10 Гбит/с и частотах выше 10 ГГц не могут быть использованы блоки обычной электроники. Требуется разработка новых элементов — оптических аналогов электронных приборов, которые могут быть созданы методами интегральной оптики.

§ 12.4. ЭЛЕМЕНТЫ ИНТЕГРАЛЬНОЙ ОПТИКИ

Интегральная оптика (ИО) — раздел оптоэлектроники, исследующий и применяющий оптические явления в тонкопленочных волноводах с целью создания в едином технологическом цикле пленочных оптических и оптоэлектронных элементов и оптических интегральных схем. Физическую основу интегральной оптики составляют волновые процессы в планарных оптических волноводах, рассмотренных в § 3.4, поэтому ее иногда называют «*волноводной оптикой*» и «*планарной оптикой*». В технологии ИО-элементов широко используются хорошо отработанные в микроэлектронике технологические процессы: эпитаксия, диффузия, фотолитография, ионно-лучевая обработка и др. Под термином «тонкопленочный» подразумевают диэлектрическую или полупроводниковую пленку заданной конфигурации, толщина которой сравнима с длиной волны распространяющегося в ней света.

Интегральная оптика базируется на тех же физических эффектах, математическом аппарате и представлениях волновой и геометрической оптики, что и волоконная оптика. В качестве волновода применяется световод, как правило, прямоугольного сечения, который либо располагается на поверхности подложки (рис. 12.17, а), либо погружен в нее (рис. 12.17, б). Первые формируются, например, путем нанесения однородной диэлектрической пленки на диэлектрическую подложку с меньшим показателем преломления и образуют ступенчатый световод. Вторые могут быть изготовлены, например, методами диффузии примесей в подложку, ионным легированием или эпитаксиальными методами, хорошо отработанными в полупроводниковой технологии. Здесь может быть реализован как ступенчатый, так и градиентный профиль изменения n . Типичные значения основных параметров полосковых световодов: толщина $d = 0,3 \dots 10$ мкм; ширина

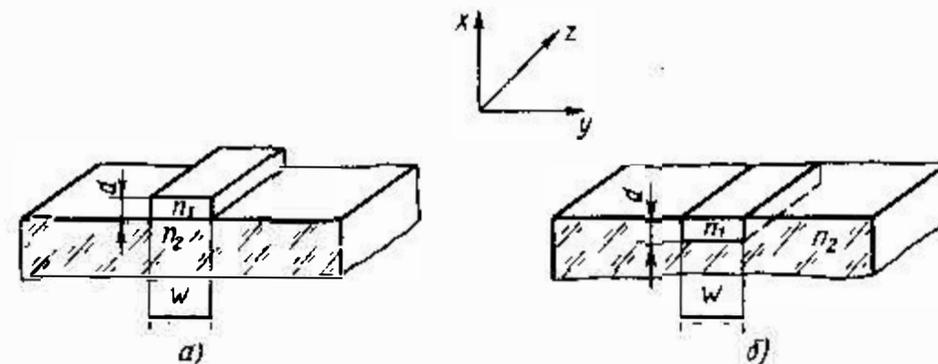


Рис. 12.17. Плаварные полосковые оптические микроволноводы

$W = 3 \dots 100$ мкм и относительная разница показателей преломления $\Delta = \Delta n/n_1 = 10^{-2} \dots 10^{-3}$. Распространение света по таким волноводам происходит согласно тем же закономерностям, что и в планарном диэлектрическом волноводе (§ 3.4) и в активном слое полоскового лазера (§ 9.6), за исключением того обстоятельства, что профиль изменения показателя преломления несимметричен по оси x .

Как и в планарном диэлектрическом волноводе, световая энергия в полосковых волноводах, изображенных на рис. 12.17, распространяется в виде волноводных мод. Их фазовая скорость распространения определяется не только показателем преломления n_1 материала волновода, но и зависит от толщины пленки и показателя преломления подложки (вследствие просачивания волны в подложку). Эффективный показатель преломления $n_{эф}$ (3.36) тонкопленочного волновода, определяющий скорость распространения волны данного типа в волноводе, будет зависеть от толщины пленки, точнее — от отношения d/λ . При $d/\lambda \gg 1$ эффект просачивания мал и $n_{эф} \approx n_1$. С уменьшением относительной толщины пленки волноводная мода все больше захватывает прилегающие области, так что в пределе при $d/\lambda \ll 1$ $n_{эф} \rightarrow n_2$. Используя пленку соответствующей толщины, как и в световоде (рис. 12.9), можно ограничить число волноводных мод полоскового волновода желаемым образом. Более того, поскольку $n_{эф} = f(d)$, то, изменяя толщину пленки в направлении y , можно управлять распространением волны в направлении x .

Эффект просачивания волны за стенки диэлектрического волновода очень широко используется в интегральной оптике. Этот эффект приводит к *оптическому туннелированию* и к обмену энергией между близко расположенными волноводами, как показано на рис. 12.18, а. Это дает возможность весьма просто методами интегральной оптики реализовать направленный разветвитель, упоминавшийся в § 12.3 и показанный на рис. 12.19. Изменяя расстояние между волноводами и длину участка взаимодействия L , можно варьировать коэффициент связи устройства. Если эти элементы выполнить из электрооптических мате-

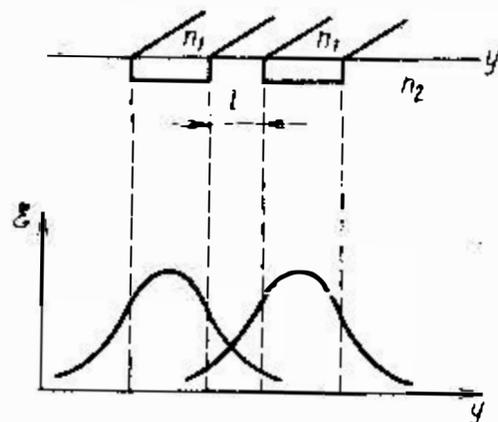


Рис. 12.18. Оптическое туннелирование между двумя связанными волноводами

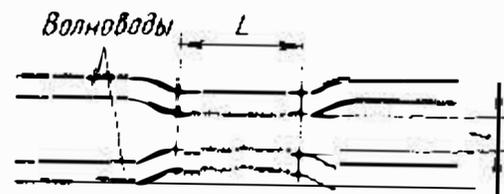


Рис. 12.19. Направленный ответвитель на основе туннельно связанных полосковых волноводов

риалов, показатель преломления которых изменяется под действием электрического поля, то величиной коэффициента связи можно управлять.

В качестве иллюстрации на рис. 12.20 приведены примеры пассивных оптоэлектронных элементов связи, выполненных методами интегральной оптики. *Фокусирующая линза* (рис. 12.20, а) может быть создана или за счет утолщения слоя в центре, или за счет изменения n_1 по координате y , создавая планарную градиентную линзу. *Планарное зеркало* (рис. 12.20, б) действует аналогично брегговским отражателям, рассмотренным в § 9.8. Работа *направленного ответвителя* (рис. 12.20, в) пояснена на рис. 12.18.

Кроме пассивных устройств, методами интегральной оптики могут быть созданы и *активные элементы*: модуляторы, излучатели, фотоприемники, дефлекторы, переключатели и т. п. Важно, что их потребляемая мощность примерно пропорциональна объему активной среды. Поэтому у планарных интегрально-оптических элементов она на два-три порядка меньше, чем у объемных. *Весьма эффективны* изготовленные методом интегральной оптики *акустические устройства* управления светом, в основу которых положена брегговская дифракция на фазовых решетках, индуцируемых электрическим полем или акустическими поверхностными волнами (рис. 10.7).

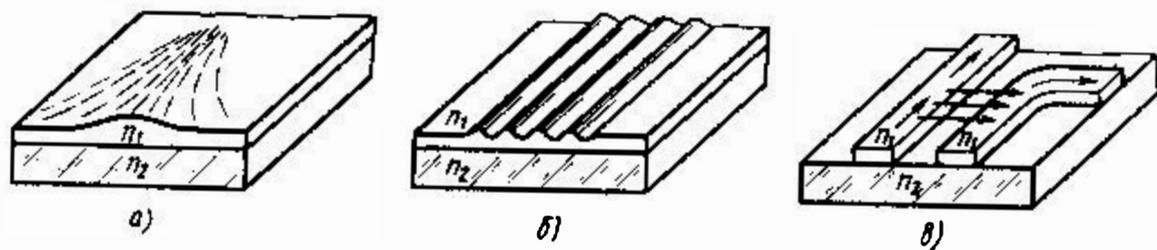


Рис. 12.20. Пассивные интегрально-оптические элементы: а — линза, б — зеркало, в — ответвитель

Нелинейные оптические явления, рассмотренные в гл. 5, также используются для создания различных интегрально-оптических преобразователей.

Важным моментом в интегрально-оптических устройствах с внешней оптической связью является ввод и вывод излучения в тонкопленочный волновод.

Эта проблема решается несколькими путями. Простейший из них, но не всегда дающий желаемые результаты — фокусировка светового пучка на торец волновода. Весьма распространен способ ввода и вывода излучения с помощью призмы полного внутреннего отражения (рис. 12.21).

Призма с большим показателем преломления n_3 располагается на малом расстоянии h (порядка λ) от поверхности пленочного волновода. Луч света входит в призму и падает на ее боковую грань под углом полного внутреннего отражения (для границы призма — воздух). Вследствие эффекта «просачивания» на границе призма — волновод возникает вытекающая волна, которая проходит слой воздуха (оптическое туннелирование) и входит в волновод. С помощью обратного процесса можно осуществить вывод энергии из тонкопленочного волновода.

Ввод и вывод излучения в планарный волновод может быть осуществлен также через нанесенные на поверхность волновода дифракционные решетки аналогично тому, как это рассматривалось в § 9.8 для гетеролазеров с распределенной обратной связью.

Исключительно важной представляется возможность интеграции оптических и электронных схем на одной подложке (например, GaAs или InP) с использованием световодных связей по разным координатам, что открывает путь для создания *трехмерных интегральных схем*.

§ 12.5. ПРИНЦИПЫ ГОЛОГРАФИИ И ГОЛОГРАФИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ

*Голография** — это метод точной записи, воспроизведения и преобразования волновых полей. Он дает возможность регистрировать и восстанавливать *полное изображение предмета* по его *нефокусированной дифракционной картине*.

*Название было предложено в 1948 г. изобретателем голографии Деннисом Габором и происходит от греческих «holos» — весь, полный и «grapho» — пишу, черчу, рисую.

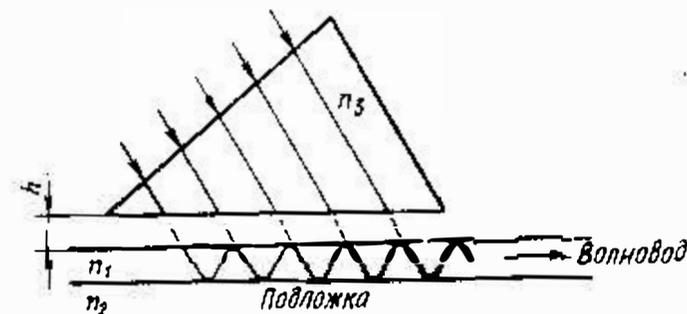


Рис. 12.21. Ввод лазерного излучения в планарный оптический волновод с помощью элемента связи — призмы

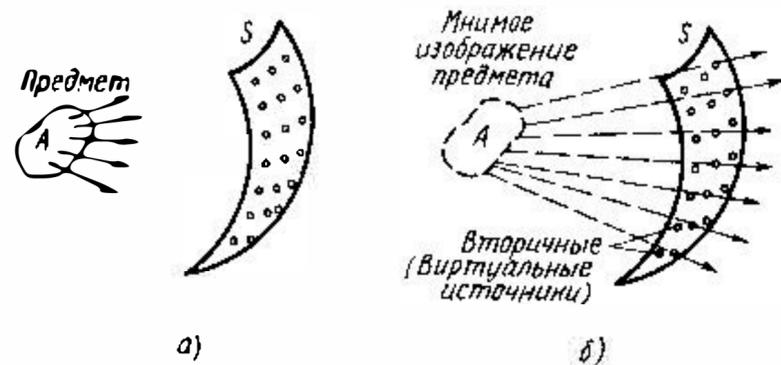


Рис. 12.22. Появление принципа Гюйгенса—Френеля

Идея голографии сводится к восстановлению «замороженного» волнового фронта и самым тесным образом связана с принципом Гюйгенса — Френеля (см. § 2.3). Если вам удастся каким-либо способом записать распределение волны (ее амплитуду и фазу) на заданной поверхности S , т. е., грубо говоря, «заморозить» эту волну на заданной поверхности, то мы получим полную информацию о предмете, которую затем, согласно принципу Гюйгенса — Френеля, сможем восстановить и получить полное (в частности, трехмерное) изображение предмета.

В соответствии с принципом Гюйгенса — Френеля действие исходной, *первичной* волны от предмета A можно заменить действием *вторичных, виртуальных* источников, расположенных на поверхности S , как показано на рис. 12.22. Эти вторичные источники должны колебаться с той амплитудой и той фазой, которые заданы дошедшей до них первичной волной от предмета A . Выбрав соответственным образом эти источники, можно «забыть» о *первичной* волне. Элементарные сферические волны, испускаемые *вторичными* источниками на поверхности S , интерферируя, восстанавливают за поверхностью копию первичного волнового поля. Глаз или любой другой приемник не сможет отличить эту копию от поля волны самого предмета.

Таким образом, задача сводится к определению способа записи амплитуды и фазы волны, рассеянной предметом на заданной поверхности S . Так как мгновенная регистрация пространственного распределения амплитуды и фазы волны за время, составляющее долю периода колебаний, в оптическом диапазоне практически невозможна, то необходимо «остановить» («заморозить») волну. Для этого было предложено использовать дополнительную, так называемую *опорную волну* (или когерентный фон) с известным распределением фазы и амплитуды, как показано на рис. 12.23, *а*. Если эта волна от опорного источника B когерентна с регистрируемой волной от точки предмета A , то на поверхности регистрации S образуется *интерференционная картина*. Контраст этой картины определяется распределением интенсивности регистрируемой волны, а частота и форма по-

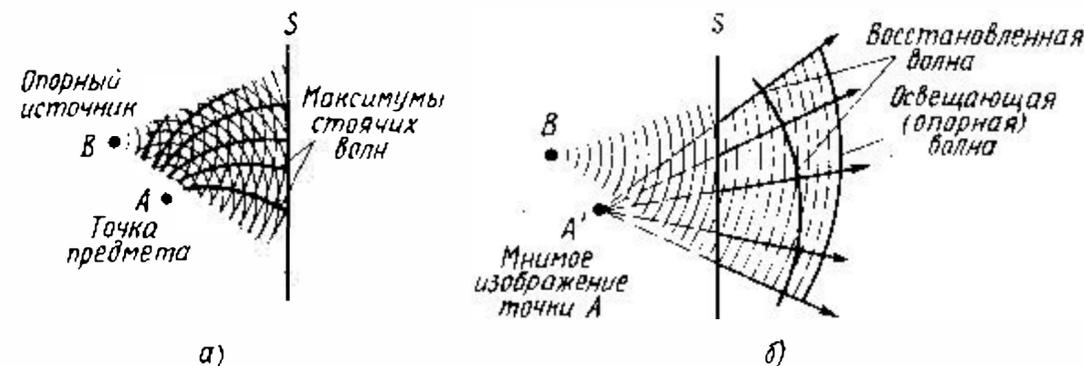


Рис. 12.23. Принцип голографирования:
а — запись, б — восстановление

лос — распределением фаз, или «фазовым рельефом». Опорный пучок «останавливает» в пространстве световую волну так, что становится возможна ее регистрация, например, на фотопластинке или каким-либо другим сигналом. Происходит *запись голограммы*, представляющей собой по существу интерференционную картину, образованную волной от предмета и опорной волной.

Если теперь на полученную голограмму направить волну, служившую при записи опорной, то в результате дифракции света за голограммой восстановится исходное волновое поле предметной волны, как показано на рис. 12.23, *б*. Таким образом происходит *восстановление*, или *считывание*, голограммы. Картина интерференции может быть записана на голограмме не только в виде вариаций коэффициента пропускания, как на фотопленке (такие голограммы называются *амплитудными* и наиболее часто применяются), но и ввиду вариаций толщины, коэффициента преломления или отражающего рельефа. Такие голограммы называются *фазовыми*.

Довольно важным обстоятельством является то, что для получения интерференционной картины опорная и предметная волны должны быть взаимно когерентны, т. е. фазы опорной и предметной волн должны строго коррелировать как в пространстве, так и во времени. Поэтому в оптическом диапазоне для освещения предмета и для получения опорной световой волны необходимо использовать лазерное излучение с высокой степенью когерентности. Именно появление лазеров привело к значительному прогрессу в голографии.

Принципиальная схема получения голограмм на фотопластинке с использованием излучения лазера показана на рис. 12.24, *а*. Когерентный свет, идущий от лазера, разделяют на два пучка, один из которых освещает объект, а другой служит опорным. Интерференционная картина, образуемая в месте перекрытия пучков, регистрируется на фотопластинке. После проявления фотопластинки получается голограмма, на которой записана интерференционная картина в виде вариаций почернений. Внешне

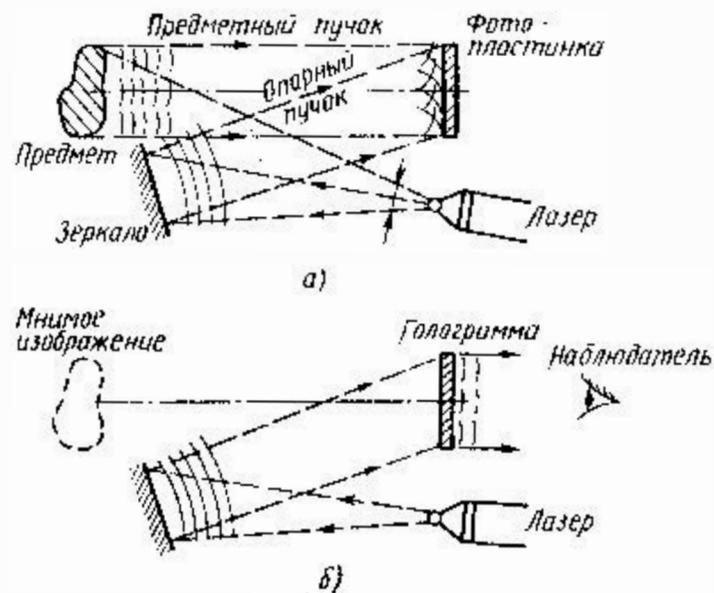


Рис. 12.24. Схемы записи (а) и восстановления (б) голограмм на фотопластинку с помощью лазера

такая проявленная голограмма на первый взгляд не отличается от равномерно засвеченной фотопластинки, но под микроскопом видна ее четкая интерференционная структура. Если теперь убрать объект и осветить проявленную голограмму тем же лазерным светом, как показано на рис. 12.24, б, то при ее просмотре мы увидим мнимое изображение предмета, не отличающееся от реального объекта.

Голограммы, зарегистрированные в тонком слое, обладают свойствами, аналогичными свойствам плоских дифракционных решеток, и называются *плоскими голограммами*. При использовании объемных сред, в частности — толстослойных эмульсий, голограмма приобретает свойства объемной дифракционной решетки. Такие голограммы были предложены Ю. Н. Денисюком и называются *объемными* или *трехмерными голограммами*.

Трехмерные голограммы представляют наиболее общий случай голографической записи и реализуются, когда запись происходит в толщине, существенно превышающей длину волны. Двумерная (плоская) голограмма может быть восстановлена только в когерентном (лазерном) свете. Особенностью трехмерной голограммы является способность воспроизводить не только фазу и амплитуду записанного на ней излучения, но и его спектральный состав. Если такую голограмму освещать источником излучения со сплошным спектром, например, солнечным излучением или излучением лампы накаливания, то она сама выберет из этого спектра те составляющие, которые участвовали в ее записи. Спектральная селективность трехмерных голограмм объясняется интерференцией волн в объемной дифракционной решетке так, что складываются синфазно и взаимно усиливают друг друга

только волны одной монохроматической составляющей — той, с помощью которой проводилась запись голограммы.

Естественно, что методы голографии могут быть применены не только к электромагнитным волнам оптического диапазона, но и к любым другим волновым процессам, в частности к акустическим, ультразвуковым, рентгеновским и другим волнам.

Области применения голографии весьма широки и не ограничиваются так называемой безлинзовой фотографией для получения трехмерного изображения предмета. Используя свойства голограмм, можно сравнительно просто реализовать разнообразные голограммные оптические элементы, осуществляющие различные преобразования световых полей: фокусировку, спектральное разложение, отражение, фильтрацию, поляризационные преобразования и др.

Голографические методы находят применение в интероскопии и в методах неразрушающих испытаний. *Голографическая интерферометрия*, как это нетрудно понять из рассмотренных принципов, идеально подходит для обнаружения очень малых деформаций (порядка длины волны). Большое применение голографические методы находят в системах накопления и обработки информации. По теоретическим оценкам, один бит информации можно накопить в кубе со стороной порядка длины волны, т. е. для видимого диапазона в объеме 1 мм^3 можно записать порядка 10^{10} бит информации. Хотя практические результаты еще довольно далеки от этого предела, уже в настоящее время получены системы с голографической памятью очень большой емкости.

Голографическая система памяти, кроме ее предельно высокой емкости, имеет также и другое важное преимущество. Двоичная память в виде картины типа шахматной доски, записанная на микрофильме, легко может быть испорчена всего лишь одной пылинкой или царапиной. Голограмма к таким дефектам почти нечувствительна, так как каждая отдельная часть голограммы содержит информацию о всем предмете, а не об отдельной его части.

Если принцип голографии хорошо усвоен, то легко убедиться, что еще одним важным ее применением может быть *распознавание образов* и символов, а также практически идеальное *кодирование* информации. Действительно, во всех предыдущих примерах на рис. 12.23 и 12.24 сложный предметный пучок интерферировал с простым плоским или сферическим опорным пучком. Предметный пучок восстанавливался при освещении голограммы таким же опорным пучком. Между тем опорный пучок может быть в принципе любым, лишь бы он коррелировал сам с собой. Например, на рис. 12.24 вместо плоского зеркала можно поставить какую-либо сложную рассеивающую поверхность. Тогда восстановить голограмму можно будет только с помощью той же самой или аналогичной ей поверхности, которая будет являть-

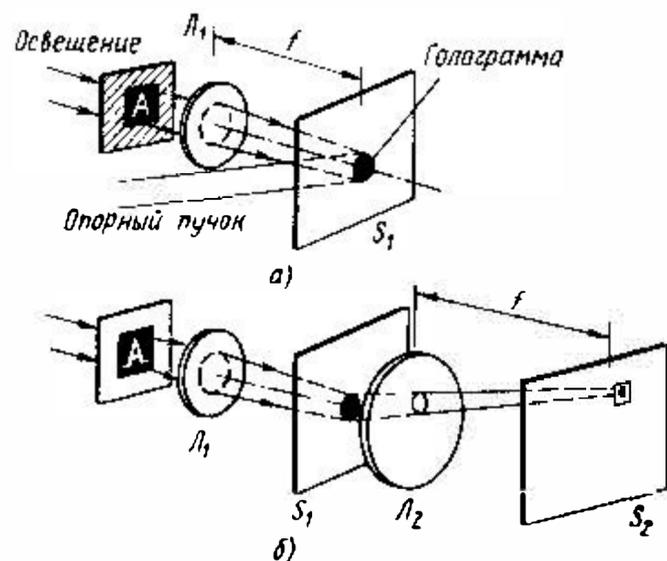


Рис. 12.25. Принцип распознавания символов с помощью голографических методов: а — получение деструктурирующей голограммы, б — считывание

ным пучком. Если эту голограмму затем осветить предметным пучком от буквы «А», как показано на рис. 12.25, б, то восстановится опорный пучок, который с помощью линзы L_2 сфокусируется во вполне определенную точку на экране S_2 и возбудит, например, вполне определенный фотозапись. Отсюда ясно, что голограмма может быть довольно универсальным преобразующим устройством (переводчиком). Все эти свойства голограмм возможно применять в вычислительных машинах. Уже в настоящее время стоит вопрос о создании интерференционно-оптических машин, работающих на голографических принципах. Элементарными операциями в таких машинах, в отличие от обычных ЭВМ, являются не арифметическое суммирование и простейшие логические сравнения, а непосредственно умножение, интегрирование, осуществление прямого и обратного преобразований Фурье и Лапласа и т. п. Возможность осуществления столь сложных действий в качестве элементарных операций вытекает из особенностей физических процессов интерференции и дифракции когерентного света и была проиллюстрирована в § 12.1 на примере осуществления преобразования Фурье с помощью обычной линзы.

ся «ключом» к дешифровке голограмм. Подобрать такой «ключ» практически невозможно.

Этот же принцип лежит в основе голографического метода распознавания образов и символов, который можно уяснить при рассмотрении рис. 12.25. С помощью плоского опорного пучка в заданном месте изготавливается, например, голограмма буквы «А», которая представляет собой зарегистрированную картину интерференции между пучком, испытавшим дифракцию на знаке, и опор-

СПИСОК ЛИТЕРАТУРЫ

1. Тамм И. Е. Основы теории электричества.— М.: Наука, 1976. 616 с.
2. Ландау Л. Д., Лифшиц Е. М. Электродинамика сплошных сред.— М.: Наука, 1982. 620 с.
3. Калпиевский Н. И. Волновая оптика.— М.: Высшая школа, 1978. 383 с.
4. Алферов Ж. И. История и будущее полупроводниковых гетероструктур. ж. «Физика и техника полупроводников». — С-Пб.: «Наука», 1998, т. 32 с. 3 — 18.
5. Карлов Н. В. Лекции по квантовой электронике.— М.: Наука, 1988. 336 с.
6. Пихтия А. Н. Физические основы квантовой электроники и оптоэлектроники.— М.: Высшая школа, 1983. 304 с.
7. Клычко Д. Н. Физические основы квантовой электроники.— М.: Наука, 1986. 296 с.
8. Зельго О. Принципы лазеров.— М.: Мир, 1984. 400 с.
9. Дмитриев В. Г., Тарасов Л. В. Прикладная нелинейная оптика.— М.: Радио и связь, 1982. 352 с.
10. Шен И. Р. Принципы нелинейной оптики.— М.: Наука, 1989. 560 с.
11. Справочник по лазерам/Под ред. А. М. Прохорова. В 2-х т.— М.: Советское радио, 1978. т. 1 — 504 с., т. 2 — 400 с.
12. Киселев Г. Л. Приборы квантовой электроники.— М.: Высшая школа, 1980. 237 с.
13. Ярив А. Введение в оптическую электронику.— М.: Высшая школа, 1983. 398 с.
14. Павлов П. В., Хохлов А. Ф. Физика твердого тела — М.: Высшая школа, 2000. 494 с.
15. Радль Б. Квантовые процессы в полупроводниках.— М.: Мир, 1986. 304 с.
16. Зи С. Физика полупроводниковых приборов: В 2-х т.— М.: Мир, 1984. Т. 2 — 456 с.
17. Ельсеев П. Г. Введение в физику инжекционных лазеров.— М.: Наука, 1983. 294 с.
18. Кейсь Х., Паншин М. Лазеры в гетероструктурах: В 2-х т.— М.: Мир, 1981. Т. 1 — 300 с., т. 2 — 364 с.
19. Носов Ю. Р. Оптоэлектроника.— М.: Радио и связь, 1989. 360 с.
20. Основы оптоэлектроники: Пер. с яп./Сузмацу Я., Катаока С., Кисино К. в др.— М.: Мир, 1988. 288 с.
21. Оптика и связь: Пер. с фр./Казание А., Флере Ж., Мэтр Г., Руссо М.— М.: Мир, 1984. 504 с.
22. Алексеев С. А., Прокопенко В. Т., Яськов А. Д. Экспериментальная оптика полупроводников.— С-Пб., Изд. «Политехник», 1994. 420 с.
23. Драгунов В. П., Неизвестный И. Г., Гродчик В. А. Основы нанополупроводников.— Новосибирск; Изд. НГТУ, 2000. 332 с.
24. Иванов В. И., Алсеев А. И., Юшин А. М. Полупроводниковые оптоэлектронные приборы: Справочник.— М.: Энергоатомиздат, 1989. 448 с.
25. Полупроводниковые фотоприемники/Под ред. В. И. Стафеева.— М.: Радио и связь, 1984. 216 с.
26. Техника оптической связи. Фотоприемники: Пер. с англ./Под ред. У. Тсанга.— М.: Мир, 1988. 526 с.
27. Ворожнов М. А., Корабин А. В., Шмальгаузен В. И. Управляемые оптические системы.— М.: Наука, 1988. 272 с.
28. Свечников Г. С. Элементы интегральной оптики.— М.: Радио и связь, 1987. 104 с.
29. Хансверджер Р. Интегральная оптика: Пер. с англ.— М.: Мир, 1985. 384 с.
30. Бугусов М. М., Верник С. М., Галкин С. Л. и др. Волоконно-оптические системы передачи.— М.: Радио и связь, 1992. 416 с.