

А.Н.Пихтин

Физические основы квантовой электроники и опто- электроники

для вузов

издательство
Высшая
Школа

А.Н.Пихтин

Физические основы квантовой электроники и опто— электроники

Допущено
Министерством высшего и среднего
специального образования СССР
в качестве учебного пособия
для студентов вузов,
обучающихся по специальностям
«Полупроводники
и диэлектрики»,
«Полупроводниковые
и микроэлектронные приборы»



МОСКВА «ВЫСШАЯ ШКОЛА» 1983

ББК 32.86
П 55
УДК 621.38

Рецензенты:

кафедра физики диэлектриков
и полупроводников Горьков-
ского государственного универ-
ситета (зав. кафедрой д-р физ.-
мат. наук, проф. *И. А. Карпович*)

д-р физ.-мат. наук, проф.
Б. С. Ринкявичюс (Московский
энергетический институт)

Пихтин А. Н.

П 55 Физические основы квантовой электроники и
оптоэлектроники: Учеб. пособие для вузов по спе-
циальностям «Полупроводники и диэлектрики»,
«Полупроводниковые и микроэлектронные прибо-
ры». — М.: Высш. шк., 1983. — 304 с., ил.

В пер.: 90 к.

В книге рассмотрены физические явления, лежащие в основе дейст-
вия лазеров, светодиодов, приборов управления оптическим излучением,
приемников оптических сигналов, приборов интегральной оптики; освеще-
ны энергетические уровни квантовых систем, взаимодействие электро-
магнитного излучения с веществом, оптические явления в твердых телах,
усиление и генерация электромагнитного излучения и другие вопросы.

П 2403000000—424
001 (01)—83 124—83

ББК 32.86
6Ф03

Александр Николаевич Пихтин

ФИЗИЧЕСКИЕ ОСНОВЫ КВАНТОВОЙ ЭЛЕКТРОНИКИ
И ОПТОЭЛЕКТРОНИКИ

Зав. редакцией Л. А. Романова. Редактор Т. И. Артемова. Мл. редактор
И. А. Исаева. Художественный редактор Т. М. Скворцова. Технический редак-
тор Л. А. Муравьева. Корректор Р. К. Косинова

ИБ № 3997

Изд. № ЭР-333. Сдано в набор 14.03.83. Подп. в печать 23.09.83. Т-18677.
Формат 84×108/32. Бум. тип. № 3. Гарнитура литературная. Печать высокая.
Объем 15,96 усл. печ. л. 15,96 усл. кр.-отт. 17,03 уч.-изд. л. Тираж
15 000 экз. Заказ № 1871 Цена 90 коп.

Издательство «Высшая школа», Москва, 101430, ГСП-4, Неглинная ул., 29/14.
Отпечатано с матриц ордена Октябрьской Революции и ордена Трудового Крас-
ного Знамени Первой Образцовой типографии имени А. А. Жданова Союзполи-
графпрома при Государственном комитете СССР по делам издательств, полигра-
фии и книжной торговли. Москва, М-54, Валовая, 28, в Московской типографии
№ 4 Союзполиграфпрома при Государственном комитете СССР по делам изда-
тельств, полиграфии и книжной торговли. Москва, И-41, Б. Переяславская 46.

© ИЗДАТЕЛЬСТВО «ВЫСШАЯ ШКОЛА», 1983

Предисловие

В Основных направлениях экономического и социального развития СССР на 1981—1985 годы и на период до 1990 года к числу важнейших направлений, на ускоренном развитии которых необходимо сосредоточить усилия, отнесены физика твердого тела, оптика, квантовая электроника и радиофизика. Иллюстрацией того, что именно эти области физики определяют прогресс в ряде отраслей современной техники, являются достижения полупроводниковой электроники и особенно оптоэлектроники. Несмотря на то что первый лазер был создан немногим более 20 лет назад, в настоящее время промышленность всех технически развитых стран освоила широкую номенклатуру этих приборов (несколько сотен различных типов) и выпускает их сотнями тысяч штук. Применение таких приборов в самых разнообразных отраслях науки и техники (физика, химия, медицина, электроника, связь, локация, строительство, приборостроение, энергетика, машиностроение и др.) позволило получить не только количественно, но и качественно новые результаты, способствуя тем самым дальнейшему развитию этих отраслей.

Успехи квантовой электроники, прежде всего оптического диапазона, привели к созданию принципиально новых оптических приборов и методов, их широкому использованию в электронике, измерительной технике, локации, связи, вычислительной технике и т. д. При этом отчетливо наблюдается тенденция к возрастанию роли оптических методов передачи, обработки и хранения информации.

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 полупроводников и диэлектриков. Основная ее цель — дать представления о фундаментальных физических процессах, лежащих в основе современной оптической и квантовой электроники, подготовить читателя к дальнейшему изучению специальной литературы по отдельным вопросам данной области и к технически грамотному применению приборов оптической электроники.

Автор намеренно ограничился рассмотрением физических основ оптической электроники, затрагивая свойства конкретных приборов и устройств лишь в общих чертах. Это обусловлено тем, что квантовая электроника, оптоэлектроника и интегральная оптика являются настолько быстро развивающимися областями современной техники, что с вероятностью выше 0,5 характеристики приборов могли устареть еще до выхода книги в свет. Более того, практика чтения курса «Квантовая электроника и оптоэлектроника» показала, что при хорошем усвоении принципов построения приборов и понимании физических процессов, лежащих в основе их работы, технические подробности приборов усваиваются довольно легко.

Существенную помощь в работе над книгой оказали своими советами, консультациями и обсуждением рукописи акад. Ж. И. Алферов, проф. О. В. Константинов, проф. В. Н. Рудаков, за что автор выражает им искреннюю благодарность. Автор особенно признателен проф. В. В. Пасынкову за постоянную помощь, поддержку и внимание к настоящей работе, а также благодарит проф. И. А. Карповича и д-ра физ.-мат. наук Б. С. Ринкявичюса за советы и замечания, сделанные ими при рецензировании рукописи.

Автор с благодарностью примет все замечания и пожелания читателей, которые просит направлять по адресу: Москва, 101430 ГСП-4, 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 генерации, усиления 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.

Оптическое излучение — электромагнитное излучение оптического диапазона.

Оптический диапазон спектра составляют электромагнитные колебания, длина волн которых лежит в пределах от 1 мм до 1 нм*. Внутри оптического диапазона выделяют *видимое* ($\lambda = 0,38 \div 0,78$ мкм), *инфракрасное* ($\lambda = 0,78 \div 1000$ мкм) и *ультрафиолетовое* ($\lambda = 0,001 \div 0,38$ мкм) излучение.

Световые волны — электромагнитные волны оптического диапазона.

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

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

Квантовый генератор — источник когерентного излу-

* Этот интервал определен Международной комиссией по освещению (МКО) и является довольно условным. Иногда длинноволновую границу оптического диапазона, примыкающую к радиодиапазону, относят к радиоволнам, а коротковолновую — к мягкому рентгеновскому излучению (рис. В.1).

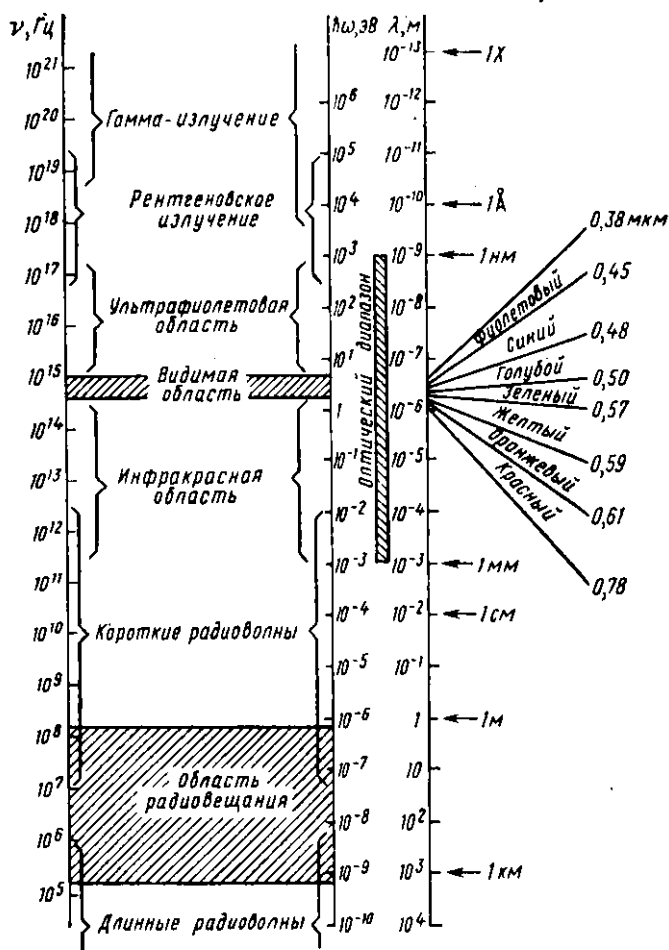


Рис. В.1. Шкала электромагнитных волн

чения, основанный на использовании вынужденного излучения.

Лазер * (оптический квантовый генератор) — кванто-

* Термины «лазер» (англ. laser) и «мазер» (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.

вый генератор оптического излучения. *Мазер* — квантовый генератор 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 ее амплитуда, частота, фаза, направление распространения и поляризация постоянны или изменяются по определенному закону (упорядоченно).

§ В.2. ОСОБЕННОСТИ ОПТИЧЕСКОЙ ЭЛЕКТРОНИКИ

Необходимость дальнейшего освоения оптического диапазона и перенесение на него хорошо развитых в настоящее время методов радиофизики, радиотехники и электроники определяются рядом принципиальных обстоятельств:

1. Частота электромагнитных колебаний (несущая частота ν_0) в оптическом диапазоне существенно выше, чем в радиодиапазоне (см. рис. В.1). Например, частота световых колебаний в наиболее освоенной видимой и ближней инфракрасной областях спектра ($\sim 10^{15}$ — 10^{18} Гц) в миллионы раз превышает частоту радиоволн в областях радио- и телевидения. Это определяет высокую информационную емкость оптического канала связи. Напомним, что для передачи обычного телевизионного изображения требуется полосу 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.

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 характерные размеры порядка длины волны. Размеры волноводов, по которым может передаваться излучение с малыми потерями, также должны быть 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 нетрудно 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 внешним 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 ее 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 электроникой, оптоэлектронику 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 информации ($\sim 10^8$ бит/см²) в оптических запоминающих устройствах.

На сегодняшний день реализована лишь небольшая часть перечисленных преимуществ оптической электроники, которая, по мнению специалистов, будет в значительной мере определять технику завтрашнего дня. Для реализации этих преимуществ необходимо прежде всего понимать фи-

зические процессы взаимодействия оптического излучения с веществом, ибо они являются фундаментом всей квантовой и оптической электроники.

§ В.3. ИСТОРИЧЕСКАЯ СПРАВКА

Оптика является одной из древнейших наук. Освоение оптического излучения можно подразделить на два больших исторических этапа.

Первый этап связан в основном с изучением наблюдаемых непосредственно глазом свойств 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 создали теорию отражения света от плоских и сферических зеркал и положили начало геометрической оптике. Теории зрительных лучей в древности противопоставлялась еще более фантастическая теория 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 собой совокупность мельчайших частиц — корпускул, движущихся вдоль определенной траектории — светового луча. Этой теорией наглядно и естественно объяснялись 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 света в неоднородных средах. Одновременно Х. Гюйгенсом была предложена интерпретация тех же явлений на основе понятий световой волны и волновой поверхности, согласно которой свет представляет собой 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 корпсускулы. Вслед за Ньютоном в XVIII в. большинство ученых стали склоняться к корпускулярной теории, и волновая теория, блестяще развитая Гюйгенсом, сохранила лишь немногих последователей.

Только на рубеже XVIII—XIX вв. англичанином Томасом Юнгом были начаты серьезные исследования интерференции и дифракции, а французом О. 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. Качественной и количественной 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 тесно связан с теми революционными преобразованиями, которые претерпела физика 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 или фотонами. Таким образом, Эйнштейну удалось объяснить открытый ранее фотоэффект. Применяя эти понятия к атому, Нильс Бор в 1913 г. объяснил простую 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), в которой он на основании рассмотрения термодинамического равновесия системы молекул ввел понятие об индуцированном излучении. На возможность использования индуцированного излучения для наблюдения отрицательного поглощения (усиления) впервые указал в 1940 г. В. А. Фабрикант.

В 1954 г. Н. Г. Басов и А. М. Прохоров разработали конкретный проект, а Ч. Таунс создал действующий мазер на пучке молекул аммиака. Это был первый прибор, работавший на квантовых принципах, в основе которого лежало явление усиления электромагнитных колебаний с помощью индуцированного излучения. За эти работы советским ученым Н. Г. Басову и А. М. Прохорову была присуждена Ленинская премия, а затем совместно с американским физиком Ч. Таунсом — Нобелевская премия. Таким образом, 1954 г. может быть назван годом рождения квантовой электроники как самостоятельной науки.

Методы, развитые первоначально в радиодиапазоне (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 на смеси гелия и неона. Наука и техника впервые получили в свое распоряжение когерентный источник световых волн. Это дало толчок развитию таких новых областей науки, как нелинейная оптика, голография и др. основополагающие работы по голографии были выполнены Д. Габором и Ю. Н. Денисюком.

Параллельно с развитием квантовой радиофизики и квантовой электроники быстрыми темпами развивалась физика полупроводников и полупроводниковая электроника. Успехи в этих областях привели к созданию эффективных полупроводниковых фотоприемников и генераторов света, т. е. приборов, составляющих основу полупроводниковой оптоэлектроники. Главными вехами в развитии данной отрасли являются следующие. В 1873 г. В. Смит обнаружил в слоях селена фотопроводимость, т. е. 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 АН СССР в Москве был рассмотрен принцип работы полупроводникового инжекционного лазера, а в 1962 г. в физико-техническом институте АН СССР 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 низкопороговые полупроводниковые лазеры, в том числе работающие в непрерывном режиме при комнатной температуре. 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). Уже в настоящее время функционируют линии ВОЛС протяженностью, составляющей сотни километров.

ГЛАВА 1. ЭНЕРГЕТИЧЕСКИЕ СОСТОЯНИЯ КВАНТОВЫХ СИСТЕМ

Квантовыми системами называют системы, состоящие из микрочастиц (электронов, ядер, атомов, молекул и т. д.) и подчиняющиеся квантовым законам, характерным для микромира.

В квантовой теории отдельный электрон, атом или молекулу (и более сложные системы) можно рассматривать только как объект, существующий в некотором разрешенном («дозволенном») квантовом состоянии. Важнейшим свойством квантовой системы, состоящей из связанных микрочастиц, является то, что ее 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.

Энергетические состояния, которыми может обладать квантовая система, для наглядности изображают графически в виде уровней энергии по аналогии с потенциальной энергией тела, поднятого на различные высоты (уровни). Самый нижний уровень, соответствующий наименьшей возможной энергии системы, называют *основным* и обычно обозначают E_0 . Все остальные уровни E_j ($j=1, 2, 3, \dots$) называют *возбужденными*, так как для перехода на них с основного уровня надо возбудить систему, т. е. сообщить ей дополнительную энергию.

Состояние квантовой системы считается полностью заданным, если заданы его волновая функция (по Шредингеру) или кет-вектор (по Дираку).

Изменение энергии системы сопровождается *квантовым переходом* — скачкообразным переходом квантовой системы с одного уровня энергии E_i на другой E_j . При $E_i > E_j$ система отдает энергию, равную $E_j - E_i$, а при

$E_i < E_j$ — поглощает ее. Если такой квантовый переход происходит при взаимодействии с электромагнитным полем, т. е. сопровождается испусканием или поглощением фотона, то он называется *оптическим переходом*, а энергия испущенного (поглощенного) фотона определяется известным *правилом частот Бора*:

$$\hbar\omega_{ij} = |E_j - E_i|. \quad (1.1)$$

§ 1.1. КВАНТОВЫЕ СОСТОЯНИЯ АТОМА ВОДОРОДА И КВАНТОВЫЕ ЧИСЛА

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 $-e$. Гамильтониан для этих частиц, взаимодействующих по закону $V(r)$, где r — расстояние между частицами, имеет вид

$$\hat{H} = -\frac{\hbar^2}{2M_p} \nabla^2(x_1 y_1 z_1) - \frac{\hbar^2}{2m} \nabla^2(x_2 y_2 z_2) + V(r). \quad (1.2)$$

Здесь ∇^2 — операторы Лапласа по координатам ядра (x_1, y_1, z_1) и электрона (x_2, y_2, z_2) : $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

Если вместо радиусов-векторов ядра \mathbf{r}_1 и электрона \mathbf{r}_2 ввести новые переменные:

$$\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1 \quad (1.3)$$

и

$$\mathbf{R} = \frac{M_p \mathbf{r}_1 + m \mathbf{r}_2}{M_p + m}, \quad (1.4)$$

то вместо (1.2) получим

$$\hat{H} = -\frac{\hbar^2}{2(M_p + m)} \nabla_{\mathbf{R}}^2 - \frac{\hbar^2}{2m_r} \nabla^2 + V(r), \quad (1.5)$$

где \mathbf{r} — вектор взаимного расстояния, не зависящий от координат каждой частицы в отдельности и определяемый лишь их взаимным расположением; \mathbf{R} — радиус-вектор центра инерции частиц, называемый также радиусом-вектором центра масс; $M_p + m$ — полная, а m_r — приведенная

• масса системы: $1/m_r = 1/M_p + 1/m$; $\nabla_{\mathbf{R}}^2$ и ∇^2 — операторы Лапласа соответственно по компонентам векторов \mathbf{R} и \mathbf{r} .

Таким образом, из выражения (1.5) видно, что приведением к центру масс, т. е. выбором переменных (1.3) и (1.4),

гамильтониан можно разбить на сумму из двух независимых частей. Поэтому волновую функцию $\psi(r_1, r_2)$, зависящую в общем от координат каждой частицы r_1 и r_2 , можно искать в виде произведения $\psi(R)\psi(r)$. Функция $\psi(R)$ описывает движение центра инерции, т. е. движение атома водорода, как свободное движение одной частицы с массой $M = M_p + m$. Функция $\psi(r)$ описывает относительное движение частиц вокруг общего центра масс, т. е. внутреннее движение в атоме водорода, как движение частицы с массой m , в центрo-симметричном поле $V(r)$.

Масса ядра (протона) существенно превышает массу электрона, поэтому в атоме водорода полная масса $M \approx M_p$, приведенная масса $m_r \approx m$, а координаты центра масс практически совпадают с координатами ядра*.

Будем интересоваться прежде всего внутренним движением и связанным с ним квантовым состоянием атома. Допустим, что электрон и протон взаимодействуют по кулоновскому закону

$$V(r) = -\frac{e^2}{4\pi\epsilon_0 r}. \quad (1.6)$$

Тогда гамильтониан (1.5) для внутреннего движения с учетом $m_r \approx m$ примет вид

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}. \quad (1.7)$$

Собственные волновые функции ψ , определяющие разрешенные внутренние состояния в атоме водорода, должны удовлетворять уравнению Шредингера

$$\hat{H}\psi = E\psi, \quad (1.8)$$

что в нашем случае приводит к уравнению для атома водорода:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(r) - \frac{e^2}{4\pi\epsilon_0 r} \psi(r) = E\psi(r). \quad (1.9)$$

Коэффициент E в уравнениях (1.8), (1.9) является параметром, называемым энергией. Его значение не зависит от r и определяет состояние системы.

Решения уравнения (1.9), известные из квантовой ме-

* Аналогичный подход применяется для описания движения взаимодействующих между собой электрона и дырки в полупроводниках. Однако там вместо m и M_p , присутствуют эффективные массы электрона m_n и дырки m_p , а координаты центра масс не совпадают с координатами ни одной из частиц (см. § 1.5).

ханики, существуют только при вполне определенных значениях параметра E , которые называются собственными значениями и определяют спектр собственных значений энергии атома водорода. Напомним, что для решения уравнения (1.9) обычно переходят от декартовых координат x, y, z к полярным (сферическим) координатам 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* . Оно 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 (1.9) не имеет решений]. Собственные значения внутренней энергии атома водорода для дискретной его части определяются соотношением

$$E_{nlm_l} = E_n = -\frac{me^4}{8(e_0\hbar)^2} \frac{1}{n^2} = -\frac{R_\infty}{n^2}, \quad (1.10)$$

$$\text{где} \quad R_\infty = \frac{me^4}{8(e_0\hbar)^2} = 2,18 \cdot 10^{-18} \text{ Дж} = 13,6 \text{ эВ} \quad (1.10a)$$

— постоянная Ридберга.

Спектр разрешенных значений внутренней энергии атома водорода изображен на рис. 1.1. Состояние $n=1$ является наименьшим, основным. При $n \rightarrow \infty$ имеем невзаимодействующие электрон и протон, т. е. ионизованное состояние. Это состояние выбирают за начало отсчета: $E_\infty = 0$. Заштрихованной области соответствует непрерывная часть энергетического спектра.

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 . Это число определяет в единицах \hbar орбитальный момент количества движения электрона, движущегося вокруг ядра. Поскольку момент количества движения является векторной величиной, обозначим его \mathbf{M} с соответствующим индексом. В частности, орбитальный механический момент количества движения обозначим \mathbf{M}_l .

Квантовомеханическое рассмотрение показывает, что значение вектора \mathbf{M}_l в единицах \hbar должно равняться не l , а $\sqrt{l(l+1)}$. Поэтому

$$|\mathbf{M}_l| = \sqrt{l(l+1)} \hbar. \quad (1.11)$$

Этому орбитальному механическому моменту соответствует орбитальный магнитный момент количества движения:

$$|\mu_l| = \sqrt{l(l+1)} \mu_B, \quad (1.12)$$

где

$$\mu_B = \frac{\hbar e}{2m} = 9,27 \cdot 10^{-24} \text{ А} \cdot \text{м}^2 \quad (1.13)$$

— магнетон Бора.

Отношение магнитного момента количества движения к механическому называется *гиромагнитным отношением*:

$$\frac{|\mu_l|}{|\mathbf{M}_l|} = \frac{e}{2m}. \quad (1.14)$$

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 \mathbf{M}_l принято обозначать бук-

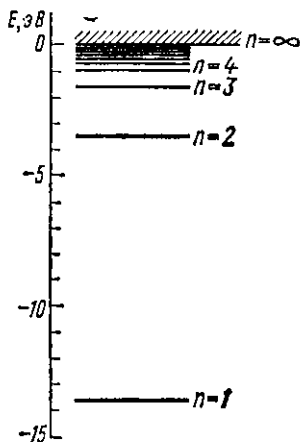


Рис. 1.1. Спектр разрешенных значений внутренней энергии атома водорода

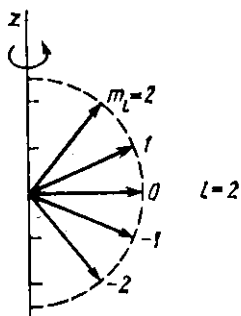


Рис. 1.2. Возможные ориентации орбитального момента количества движения электрона для d -состояния ($l=2$) относительно выделенного направления z

вами латинского алфавита со следующим соответствием:

$$\begin{array}{cccccccc} l=0 & 1 & 2 & 3 & 4 & 5 & 6 & 7, \dots \\ s & p & d & f & g & h & i & k, \dots \end{array} \quad (1.15)$$

Квантовое число l определяет величину орбитального момента, но не его направление.

3. Магнитное квантовое число m_l . Оно определяет в единицах \hbar z -компоненту орбитального момента количества движения:

$$(M_l)_z = m_l \hbar. \quad (1.16)$$

Отметим, что говорить об ориентации орбитального момента в атоме водорода не имеет смысла, поскольку там нет выделенного направления (системы отсчета): задача сферически

симметричная. Поэтому для определения ориентации орбитального момента необходимо путем внешнего воздействия выделить, по крайней мере, одно из направлений. Это можно сделать, например, поместив атом во внешнее магнитное поле \mathcal{H} . Тогда направление \mathcal{H} будет определять направление выделенной оси z и соотношение (1.16) будет определять ориентацию M_l относительно этой оси.

При заданном l квантовое число m_l может принимать значения от $+l$ до $-l$ через единицу, т. е. всего $(2l+1)$ значений: $m=0, \pm 1, \pm 2, \dots, \pm l$. Каждому разрешенному значению m_l соответствуют некоторая разрешенная ориентация вектора M_l и проекция этого вектора на ось z . В качестве примера на рис. 1.2 показаны возможные ориентации орбитального момента количества движения электрона для состояния $l=2$. Заметим, что длина вектора равна не 2, а

$$\sqrt{2(2+1)} = 2,45.$$

Надо сказать, что в уравнении (1.9) не учитывается *собственный момент количества движения электрона — спин*, наличие которого подтверждается многочисленными экспериментами. Полное теоретическое обоснование существования спина электрона может быть проведено только в рамках релятивистской квантовой теории, развитой Дираком. Будем рассматривать существование собст-

венного (спинового) момента количества движения у элементарных частиц как факт. Формализм нерелятивистской («обычной») квантовой механики допускает при рассмотрении движения произвольной частицы, в том числе электрона, приписать ей некоторый собственный момент, не связанный с движением частицы в пространстве. Это свойство элементарных частиц является специфически квантовым и, строго говоря, принципиально не допускает классической интерпретации. В частности, рассматривать спин электрона как результат его вращения вокруг «собственной оси» не совсем корректно.

Наличие спинового момента у электрона приводит к появлению дополнительных степеней свободы. 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. Как и ранее, эти величины могут принимать ограниченное число дискретных значений, которым соответствуют определенные квантовые числа. Таким образом, для полного описания состояния атома водорода наряду с квантовыми числами n , l и m_l следует ввести еще два квантовых числа, отражающих собственный момент количества движения электрона. По аналогии с l и m_l эти числа обозначают s и m_s и называют спиновыми квантовыми числами.

4. С п и н о в о е к в а н т о в о е ч и с л о s . Это число определяет в единицах \hbar спиновый момент количества движения электрона. Как и для орбитального момента, спиновый момент

$$|M_s| = \sqrt{s(s+1)} \hbar. \quad (1.17)$$

Для отдельного электрона и, следовательно, для атома водорода квантовое число s может принимать только одно значение: $s=1/2$. Значение M_s фиксировано: $|M_s| = \sqrt{0,5 \cdot 1,5} \hbar = 0,866 \hbar$, поэтому при рассмотрении атома водорода квантовое число s можно опустить. Однако для более сложных атомов этого сделать нельзя.

5. С п и н о в о е к в а н т о в о е ч и с л о m_s . Это число определяет в единицах \hbar z -компоненту спинового момента количества движения электрона:

$$(M_s)_z = m_s \hbar. \quad (1.18)$$

Как и m_l , спиновое число m_s может принимать дискретные значения от $+s$ до $-s$ через единицу. Для отдельного

электрона получаются всего два возможных значения $m_s = \pm 1/2$.

Подчеркнем, что орбитальный и спиновый моменты количества движения электрона являются различными и могут рассматриваться независимо: их операторы коммутируют.

Рассмотренные физические величины, определяемые квантовыми числами n , l , m_l , s и m_s , образуют полную систему одновременно измеряемых величин. 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 спинowego — 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. К в а н т о в о е ч и с л о j . Это число определяет в единицах \hbar полный момент количества движения как векторную сумму орбитального и спинowego моментов:

$$\mathbf{M}_j = \mathbf{M}_l + \mathbf{M}_s. \quad (1.19)$$

Как и ранее,

$$|\mathbf{M}_j| = \sqrt{j(j+1)} \hbar. \quad (1.20)$$

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 квантовомеханического сложения, величина полного момента $|\mathbf{M}_j|$, как и других моментов, квантована так, что квантовое число j в (1.20) может принимать только целые или полусуммарные 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:

$$j=l+s; l+s-1; l+s-2. \dots |l-s|. \quad (1.21)$$

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. К в а н т о в о е ч и с л о m_j . Оно определяет в единицах \hbar z -компоненту полного момента количества движения электрона

$$(M_j)_z = m_j \hbar \quad (1.22)$$

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) . Переход от одной системы волновых функций или от одной системы кет-векторов $|n, l, m_l, s, m_s\rangle$ к другой $|n, l, s, j, m_j\rangle$ может быть представлен в виде линейной комбинации собственных состояний прежней системы и наоборот, поскольку обе системы являются полными. Общее число собственных состояний в обеих системах одинаково и составляет

$$\sum_{l=|l-s|}^{l+s} (2j+1) = (2l+1)(2s+1). \quad (1.23)$$

§ 1.2. ЭЛЕКТРОННЫЕ СОСТОЯНИЯ АТОМОВ

Состояние каждого из электронов в многоэлектронных атомах 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* *.

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

Самосогласованное поле обладает центральной симметрией, поэтому каждое состояние электрона можно характеризовать определенным 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 атома водорода. Например, если в атоме водорода энергия не зависит от l , то в сложных атомах она возрастает по мере увеличения l .

Энергию электрона можно представить как функцию n и l в виде, аналогичном (1.10):

$$E_{nl} = -\frac{(Z^*)^2}{n^2} R_\infty, \quad (1.24)$$

где вместо заряда ядра Z введен эффективный заряд Z^* , значение которого зависит как от n , так и от l . Обычно $Z^* < Z$, и можно записать

$$Z^* = Z - \sigma_{nl}, \quad (1.25)$$

где σ_{nl} — постоянная экранирования, являющаяся функцией n и l .

Таким образом, в одноэлектронном приближении состояние электрона в сложном атоме характеризуется совокупностью квантовых чисел n , l , m_l , s и m_s по аналогии с атомом водорода. 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

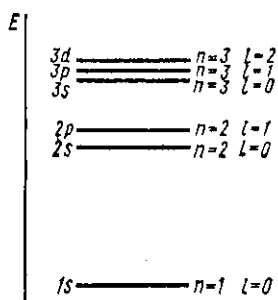


Рис. 1.3. Уровни энергии многоэлектронного атома (одноэлектронное приближение)

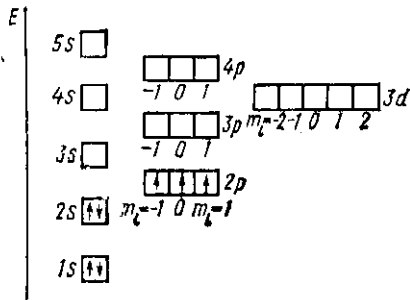


Рис. 1.4. Схематическое изображение собственных состояний электрона в многоэлектронном атоме в приближении самосогласованного центрального поля

символикой (1.15). 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*. В одноэлектронном приближении уровни энергии остаются вырожденными по m_l и m_s , а вырождение по l снимается. При этом постоянная экранирования σ_{nl} в (1.25) возрастает с увеличением l :

$$\sigma_{ns} < \sigma_{np} < \sigma_{nd} < \sigma_{nf}, \quad (1.26)$$

а эффективный заряд Z^* в (1.24) уменьшается. Соответствующая энергетическая диаграмма изображена на рис. 1.3. В атоме может быть не более $2(2l+1)$ электронов с заданными n и l . В s -оболочках может быть максимально два, в p -оболочках шесть, в d -оболочках 10 и в f -оболочках 14 электронов.

Совокупность электронов с заданным значением главного квантового числа n образует *слой*, содержащий при полном его заполнении $2n^2$ электронов, а именно: 2, 8, 18, 32, ... для $n=1, 2, 3, 4$ соответственно.

Таким образом, в первом приближении каждый электрон в многоэлектронном атоме считают находящимся в одном из собственных состояний и характеризуют некоторой определенной системой квантовых чисел n, l, m_l, m_s . Для того чтобы определить заполнение электронных состояний в атоме, пользуются условной схемой, изображенной на рис. 1.4. Эта схема вытекает из рис. 1.3, но масштаб по оси энергий здесь произволен. Каждому квадрату на рис. 1.4

сопоставляется состояние электрона с заданными n , l и m_l . Так как m_s может принимать два значения $\pm 1/2$, то согласно принципу Паули в каждом квадрате может располагаться по два электрона, которые условно изображают в виде стрелок \uparrow для $m_s = +1/2$ и \downarrow для $m_s = -1/2$. Тогда электронной конфигурацией, указывающей заполнение электронных состояний, например, для атома азота, будет $1s^2 2s^2 2p^3$, как изображено на рис. 1.4. Действительно, в атоме азота ($Z=7$) необходимо разместить семь электронов в самых нижних квадратах. Уровни $1s$ и $2s$ заняты полностью, а оставшиеся три электрона занимают $2p$ -состояния. Аналогично электронной конфигурацией для атома натрия ($Z=11$) будет $1s^2 2s^2 2p^6 3s^1$.

Для упрощения записи заполненные оболочки обычно не указывают, так как способ их заполнения является единственно возможным. Кроме того, если верхний индекс равен единице, то его можно опустить. В частности, электронная конфигурация для атома натрия обозначается $3s^1$ или просто $3s$.

Как видно из рис. 1.4, при больших n возможны случаи, когда собственные состояния с различными n и l имеют примерно одинаковую энергию. Возникает вопрос, в какой последовательности будет происходить заполнение электронных состояний в этом случае. Более того, даже в простом примере с атомом азота на рис. 1.4 остается неясным, по каким из шести $2p$ -состояний необходимо располагать три электрона. Ответ на эти вопросы дают полуэмпирические правила, известные как *правила Хунда*, которые определяют порядок размещения электронов по квадратам в частично заполненных оболочках. Эти правила гласят: внутри оболочки электроны распределяются таким образом, чтобы: 1) сумма чисел m_s по всем электронам данной оболочки была максимальной; 2) сумма чисел m_l по всем электронам данной оболочки также была максимальной. В первую очередь должно выполняться первое правило, затем — второе.

Согласно этим правилам прежде всего заполняется состояние $m_s = 1/2$, $m_l = l$, затем состояние $m_s = 1/2$, $m_l = l-1$ и т. д. Последним заполняется состояние $m_s = -1/2$, $m_l = -l$. Поэтому для атома азота наименьшему (основному) состоянию будет соответствовать заполнение, изображенное на рис. 1.4. Каждый элемент Периодической системы Д. И. Менделеева имеет свою характерную электронную конфигурацию, которая определяет его физико-химические

свойства. Периодичность в ϵ в заполнении электронных состояний определяет периодичность свойств элементов.

Из рассмотренной схемы собственных состояний электрона в многоэлектронном атоме видно, что имеются случаи, когда внешняя (например, $4s$) и внутренняя (например, $3d$) оболочки имеют примерно одинаковую энергию. Это

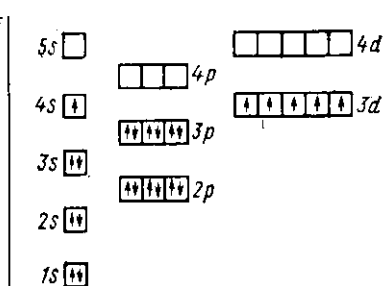


Рис. 1.5. Заполнение электронных состояний в атоме хрома

приводит к возникновению переходных групп элементов с частично заполненными внутренними оболочками. Примером может служить атом хрома ($Z=24$), основному состоянию которого отвечает электронная конфигурация $1s^2 2s^2 2p^6 3s^2 3p^6 3d^5 4s^1$, как изображено на рис. 1.5. Переходным группам обычно присваивают название, соответствующее наиболее важному ее представителю. В квантовой электронике важную роль играют элементы группы железа с незастроенной $3d$ -оболочкой и элементы группы редкоземельных металлов, или лантаноидов, с незастроенной $4f$ -оболочкой.

Квантовыми состояниями и квантовыми числами, строго говоря, можно характеризовать только полные и независимые системы, а приписывать отдельному электрону в атоме свои квантовые числа можно лишь в первом, довольно грубом приближении. Поэтому рассмотренные электронные конфигурации весьма приближенно отражают энергетическое состояние атома в целом. Так, для атома азота запись $2s^2 2p^3$ не отражает полностью распределение электронов в p -оболочке в соответствии с рис. 1.4. Сам факт существования правил Хунда свидетельствует о том, что энергия атома зависит от состояния всех электронов в его валентной оболочке.

В отличие от электрона свободный атом представляет собой полную и независимую систему. Поэтому говорить о квантовых состояниях отдельного атома вполне 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, и 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

$$L = [l_1 + l_2], \quad (1.27)$$

а для i электронов

$$L = \sum_i l_i. \quad (1.27a)$$

Аналогично для суммарного спинового квантового числа S можно записать

$$S = [s_1 + s_2], \quad (1.28)$$

$$S = \sum_i s_i. \quad (1.28a)$$

Квадратные скобки указывают, что в соотношениях (1.27), (1.28) проводится квантовое суммирование. Это означает, что, например, L в (1.27) может принимать значения от $l_1 + l_2$ до $|l_1 - l_2|$ через единицу. Это же справедливо для S : в (1.28) $S = s_1 + s_2$; $s_1 + s_2 - 1$; $s_1 + s_2 - 2$, ..., $|s_1 - s_2|$.

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.

** Иногда ее называют рассел-саундерсовской связью по фамилиям спектроскопистов Рассела и Саундерса, предложивших ввести квантовые числа L и S для характеристики состояния атома как целого.

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 наличием 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 (1.21) квантовое число J , определяющее в единицах \hbar полный момент количества движения атома J , может принимать значения от $(L+S)$ до $(L-S)$ через единицу:

$$J=[L+S]. \quad (1.29)$$

В большинстве случаев релятивистские эффекты относительно малы, и их можно учитывать как малое возмущение. Под действием этого возмущения происходит частичное снятие вырождения и 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 (1.29) 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:

$$\boxed{{}^{2S+1}L_J} \quad (1.30)$$

Состояния с различными значениями полного орбитального момента L обозначают прописными буквами латинского алфавита с соответствием, аналогичным (1.15):

$$\begin{array}{cccccccccccc} L=0 & 1 & 2 & 3 & 4 & 5 & 6 & 7, & \dots \\ & S & P & D & F & G & H & I & K, & \dots \end{array} \quad (1.31)$$

Слева сверху у этого символа указывают число $2S+1$, называемое *мультиплетностью термина*, справа внизу — квантовое число J . Например, символы 1S_0 , ${}^3P_{1/2}$ 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 (1.30) 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 (1.29) $J=0$. Поэтому при нахождении L и S всего электронного облака необходимо учитывать только неполностью 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 . Наименшие возбужденные состояния отвечают случаю, когда один электрон из $1s$ -оболочки перейдет в $2s$ -оболочку (см. рис. 1.4), т. е. электронной конфигурации $1s^1 2s^1$. Для определения термов возбужденных состояний удобно рассматривать отдельно состояния 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

* Иногда в символе (1.30) справа вверху указывают четность состояния: нечетные состояния помечают индексом «0», а четные состояния не отмечают. Четность многоэлектронного состояния совпадает с четностью суммарной волновой функции и в атоме определяется сум-

мой орбитальных квантовых чисел для всех электронов, т. е. $\sum_{i=1}^Z l_i$.

приписать только значения $s=1/2$ и $l=0$. Отсюда для конфигурации $1s^1 2s^1$ согласно (1.28) возможны два значения спинового числа S атома гелия в возбужденном состоянии: $S=0$ и $S=1$ и единственное значение $L=0$ [см. (1.27)]. Каждому из этих состояний соответствуют в силу (1.29) единственные значения $J=0$ и $J=1$. Им отвечают два возможных термина: 1S_0 и 3S_1 . Чтобы указать,

что эти термы принадлежат возбужденному состоянию с электронной конфигурацией $1s^1 2s^1$, перед символом термина (1.30) 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 $2s^3 S_1$. Состояние $2s^3 S_1$ по правилам, аналогичным правилам Хунда, должно быть расположено ниже состояния $2s^1 S_0$, как показано на рис. 1.6. 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 . Наинизшим 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 (1.27) и (1.28) возможны следующие комбинации квантовых чисел для атома в целом:

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

Им соответствуют четыре уровня энергии, как показано на рис. 1.6. Очевидно, каждой из электронных configura-

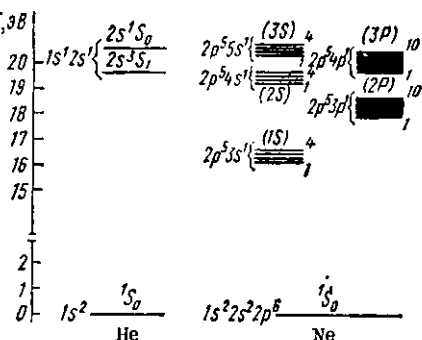


Рис. 1.6. Энергетические состояния атомов гелия и неона (за начало отсчета энергии принята энергия основного состояния)

ний $2p^4s^1$, $2p^5s^1$, $2p^6s^1$ и т. д. также будут соответствовать по четыре разрешенных состояния.

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

Рассмотренная систематика энергетических уровней нейтральных атомов сохраняется в своей основе и для ионов. Электронная конфигурация и собственные состояния атома с атомным номером Z , у которого удалены N электронов, будут аналогичны и для нейтрального атома с атомным номером $Z' = (Z - N)$. Это не означает, что энергетические уровни иона, содержащего $(Z - N)$ электронов, будут в точности совпадать с уровнями нейтрального атома с $Z' = (Z - N)$, но общий характер спектра, систематика термов и общие закономерности в структуре энергетических уровней будут одинаковыми.

§ 1.3. ЭНЕРГЕТИЧЕСКИЕ СОСТОЯНИЯ МОЛЕКУЛ

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:

$$E = E_{эл} + E_{кол} + E_{вр}. \quad (1.32)$$

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 электронвольт. 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 электронвольта.

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 в принципе 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. 1.7 для основного 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 снижении R 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 . Обычно среднее межатомное расстояние для молекулы в возбужденном состоянии больше, чем в основном. Энергия $D = E_{R \rightarrow \infty} - E_{R=R_0} = E_{\infty} - E_{\min}$ называется энергией диссоциации и равна энергии разрыва молекулы на атомы.

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

$$U(R) = \frac{1}{2} \kappa (R - R_0)^2 + \text{const}, \quad (1.33)$$

как для гармонического осциллятора. Здесь κ — упругая постоянная, а константа определяет начало отсчета. Соб-

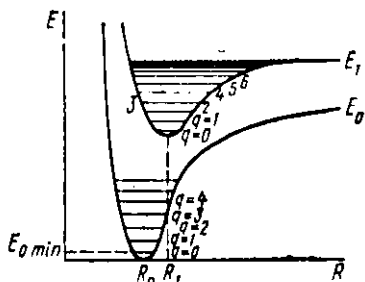


Рис. 1.7. Уровни разрешенных значений энергии для основного и первого возбужденного состояний двухатомной молекулы (R — расстояние между ядрами атомов; q — колебательное квантовое число)

ственные значения энергии гармонического осциллятора, как известно, квантованы и могут принимать значения

$$E_{\text{кол}} = \hbar \omega_0 (q + 1/2), \quad (1.34)$$

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

$$\omega_0 = \sqrt{\kappa/M} \quad (1.35)$$

— частота собственных колебаний осциллятора.

Величина $E_0 = \frac{1}{2} \hbar \omega_0$ называется *энергией нулевых колебаний*.

Приближение гармонического осциллятора дает серию эквидистантно расположенных уровней энергии (1.34) с расстоянием между ними $\hbar \omega_0$. При больших отклонениях от положения равновесия, как видно из рис. 1.7, зависимость $U(R)$ отклоняется от закона (1.33), и колебательные уровни начинают сближаться, сходясь к границе диссоциации. В этом случае энергетический колебательный спектр двухатомной молекулы можно представить в виде

$$E_{\text{кол}} = \hbar \omega_0 (q + 1/2) - \hbar \omega_0 x (q + 1/2)^2, \quad (1.36)$$

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 в предположении, что расстояния между ядрами атомов остаются неизменными. Квантование вращательной энергии дается выражением

$$E_{\text{вр}} = \frac{\hbar^2}{2I} K(K+1) = BK(K+1), \quad (1.37)$$

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:

$$|M_{\text{вр}}| = \sqrt{K(K+1)} \hbar. \quad (1.38)$$

Расстояние между соседними вращательными уровнями энергии

$$\Delta E_{\text{вр}} = E_{K+1} - E_K = 2B(K+1) \quad (1.39)$$

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 (1.16) определяет в единицах \hbar возможные значения проекции вращательного момента на выделенное направление z :

$$(M_{\text{вр}})_z = M_K \hbar, \quad (1.40)$$

где $M_K = K, K-1, \dots, -K$.

Электронные термы молекул по своей природе не отличаются от электронных термов 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. 1.7 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 $|A|$, а соответствующее квантовое число A называют *орбитальным квантовым числом молекулы*. Оно может принимать значения 0, 1, 2, ... Термы с различными значениями A обозначают прописными буквами греческого алфавита со следующим соответствием:

$$\begin{array}{c} A = 0 \quad 1 \quad 2, \dots \\ \Sigma \quad \Pi \quad \Delta, \dots \end{array} \quad (1.41)$$

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 термина.

По аналогии с (1.20) и (1.29) вводят *внутреннее квантовое число молекулы*:

$$\Omega = [L + S]. \quad (1.42)$$

Систематика электронных термов молекулы по числам Λ , S и Ω является обобщением систематики электронных термов атома по числам L , S и J . Аналогично (1.30) молекулярный терм обозначается символом

$$\boxed{^{2S+1}\Lambda_{\Omega}} \quad (1.43)$$

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

Энергетический спектр многоатомной молекулы также может быть представлен в виде (1.32). 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 (1.34) или (1.36) со своими частотами нормальных колебаний ν_{oi} .

Схема вращательных уровней энергии для линейных многоатомных молекул аналогична рассмотренной ранее для двухатомной молекулы [см. соотношение (1.37)]. Для нелинейных многоатомных молекул по мере снижения их симметрии схема вращательных уровней энергии усложняется.

§ 1.4. 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* *.

* Впервые эти эффекты наблюдались как расщепление спектральных линий под действием магнитного (П. Зееман, 1896) и электрического (И. Штарк, 1913) полей. В настоящее время под эффектами Зее-

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, строго говоря, полный момент количества движения атома J уже не будет сохраняться, а будет сохраняться лишь проекция M_J полного момента на выделенное внешним полем направление. Поэтому состояния с разными значениями M_J 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).

Рассмотрим сначала эффект Зеемана, а затем эффект Штарка. Как следует из § 1.2, атом обладает полным моментом количества движения J , а следовательно, магнитным моментом, который обозначим μ_J . Если магнитный диполь, обладающий магнитным моментом μ_J , поместить в постоянное магнитное поле, то энергия диполя будет зависеть от его ориентации относительно поля. Эта ориентационная энергия диполя в поле с индукцией B будет равна

$$\Delta E = -\mu_J B = -|\mu_J| |B| \cos(\widehat{\mu_J B}). \quad (1.44)$$

Поскольку вектор J , а значит, и вектор μ_J могут иметь только вполне определенные ориентации, определяемые квантовым числом M_J , ориентационная энергия атома может принимать соответствующие дискретные значения.

Найдем величину μ_J . В приближении LS -связи атому можно приписать орбитальный L и спиновый S моменты количества движения, которые, подобно (1.11) и (1.17), равны:

$$|L| = \sqrt{L(L+1)} \hbar, \quad (1.45)$$

$$|S| = \sqrt{S(S+1)} \hbar. \quad (1.46)$$

Аналогично (1.12) им соответствуют суммарный орбитальный магнитный момент атома

$$|\mu_L| = \sqrt{L(L+1)} \mu_B \quad (1.47)$$

мана и Штарка понимают не только расщепление в магнитном или электрическом поле спектральных линий, но также смещение и расщепление на несколько подуровней самих уровней энергии.

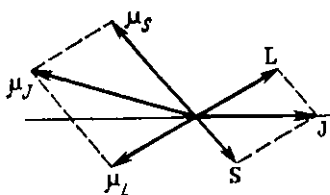


Рис. 1.8. Векторная диаграмма, поясняющая сложение механических и магнитных моментов атома

и суммарный спиновый магнитный момент атома

$$|\mu_S| = \sqrt{S(S+1)} 2\mu_B, \quad (1.48)$$

где μ_B — магнетон Бора, представляющий собой элементарную единицу магнитного момента и определяемый соотношением (1.13).

Полный магнитный момент атома равен векторной сумме μ_L и μ_S :

$$\mu_J = -(\mathbf{L} + 2\mathbf{S}) \mu_B. \quad (1.49)$$

Знак минус здесь является следствием отрицательного заряда электрона. Заметим, что спинового моменту соответствует в два раза больший магнитный момент, чем такому же орбитальному моменту [множитель 2 в соотношениях (1.48) и (1.49) является следствием *гиромагнитной аномалии* спина]. По этой причине полный магнитный момент атома μ_J в общем случае не будет антипараллелен его полному механическому моменту \mathbf{J} , как показано на рис. 1.8.

В предыдущих параграфах указывалось, что векторы \mathbf{L} и \mathbf{S} не являются полностью независимыми, поскольку в сумме они должны давать полный момент \mathbf{J} , который квантован. Аналогично векторы μ_L и μ_S также должны быть связаны друг с другом, чтобы в сумме получился полный магнитный момент атома μ_J . Магнитный момент μ_J связан с полным механическим моментом \mathbf{J} соотношением

$$\mu_J = -g_J \mu_B \mathbf{J}, \quad (1.50)$$

где безразмерный коэффициент g_J называют *g-фактором* Ланде. Он может принимать различные значения от нуля до двух в зависимости от того, каким образом комбинируются векторы \mathbf{L} и \mathbf{S} при образовании вектора \mathbf{J} .

Расчет *g-фактора* может быть выполнен с помощью векторной модели атома, изображенной на рис. 1.8. Из соотношения (1.50) следует, что полный магнитный момент μ_J должен быть коллинеарен \mathbf{J} . В то же время сумма $\mu_L + \mu_S$ не удовлетворяет этому требованию. Согласно векторной модели атома векторы \mathbf{L} и \mathbf{S} , а следовательно, μ_L и μ_S быстро прецессируют вокруг вектора \mathbf{J} . При этом перпендикулярные \mathbf{J} составляющие магнитных моментов в среднем за период равны нулю и полный магнитный момент атома μ_J определяется только параллельными \mathbf{J}

составляющими, т. е.

$$|\mu_J| = \mu_B |L| \cos(\hat{LJ}) + 2\mu_B |S| \cos(\hat{SJ}). \quad (1.51)$$

Воспользовавшись тем, что косинус угла между сторонами a и b треугольника ABC

$$\cos(a, b) = \frac{a^2 + b^2 - c^2}{2ab},$$

находим

$$|\mu_J| = \left[1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \right] \sqrt{J(J+1)} \mu_B = g_J \sqrt{J(J+1)} \mu_B, \quad (1.52)$$

откуда

$$g_J = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \quad (1.53)$$

В случае $S=0$, $J=L$ (полный момент обусловлен чисто орбитальным моментом) $g_J=1$. Для чисто спинного момента $L=0$, $J=S$, $g_J=2$. В общем случае g_J может принимать различные значения от нуля до двух (например, для термов с $L=2S-1$, $J=S-1$, $g_J=0$).

Разрешенные проекции магнитного момента на направление поля определяются магнитным квантовым числом M_J . Тогда из (1.44) окончательно получаем величину энергетического расщепления:

$$\Delta E = -(\mu_J)_z B = g_J \mu_B M_J B. \quad (1.54)$$

Разрешенные проекции вектора μ_J на ось z , совпадающую с направлением внешнего поля, удовлетворяют соотношению (1.50).

Таким образом, магнитное поле полностью снимает вырождение уровня (если только $g_J \neq 0$), приводя к его расщеплению на $(2J+1)$ компонент, как показано на рис. 1.9. Такое расщепление называется *зеemanовским расщеплением*, а соответствующие ему уровни — *зеemanовскими подуровнями*. Для магнитных полей с индукцией порядка нескольких десятых долей тесла зеemanовское расщепление имеет порядок нескольких см^{-1} ($1 \text{ эВ} = 8066 \text{ см}^{-1}$), а соответствующие частоты $\omega = \Delta E/\hbar$ лежат в диапазоне нескольких тысяч мегагерц.

Из выражения (1.54) и рис. 1.9 видно, что зеemanовское расщепление пропорционально внешнему магнитному полю. Линейная зависимость сохраняется лишь в слабых магнитных полях. При этом слабым считается магнитное поле, вызывающее расщепление ΔE , существенно меньшее

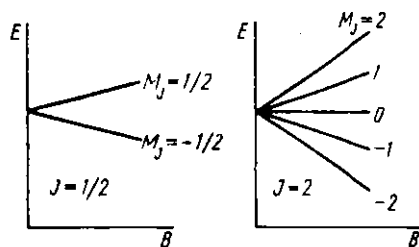


Рис. 1.9. Зеемановское расщепление энергетических уровней атома для случая $J=1/2$ и $J=2$

расстояния между соседними уровнями энергии E_i и E_j в отсутствие поля. В сильных магнитных полях, когда произведение μ_B сравнимо с интервалами между уровнями энергии невозмущенного атома, их расщепление отклоняется от предсказываемого формулой (1.54).

Аналогично эффект Зеемана проявляется в молекулах. При этом магнитный момент, возникающий вследствие движения ядер, в молекуле очень мал по сравнению с магнитным моментом электронов.

Рассмотрим теперь поведение атомов и молекул во внешнем электрическом поле (эффект Штарка). Квантовая система во внешнем электрическом поле приобретает дополнительную энергию ΔE . Здесь, как и с приложением магнитного поля, появляется выделенное направление, что приводит к снятию вырождения уровней энергии. Но в однородном электрическом поле это снятие неполное: дополнительная энергия ΔE зависит от абсолютного значения $|M_J|$, поэтому состояния, отличающиеся лишь знаком M_J , остаются вырожденными. Наличие такого двукратного вырождения можно объяснить тем, что электрическое поле (в отличие от магнитного) одинаково изменяет траектории движения заряженных частиц независимо от направления их вращения вокруг выделенной оси z , т. е. от знака M_J .

Расщепление и смещение энергетических уровней под действием электрического поля называют *штарковским*. Величина штарковского расщепления может по-разному зависеть от напряженности внешнего поля \mathcal{E} . Различают *линейный* и *квадратичный эффекты Штарка*, когда $\Delta E \sim \mathcal{E}$ и $\Delta E \sim \mathcal{E}^2$ соответственно.

Дополнительная энергия системы, обладающей дипольным электрическим моментом d и находящейся в электрическом поле напряженностью \mathcal{E} :

$$\Delta E = -d\mathcal{E}. \quad (1.55)$$

Свободный атом обладает центром симметрии и потому не имеет постоянного дипольного момента. Но под дей-

ствием внешнего электрического поля в атоме индуцируется электрический дипольный момент, обусловленный поляризацией атома и пропорциональный напряженности поля:

$$d_{\text{инд}} = \alpha \mathcal{E}, \quad (1.56)$$

где α — поляризуемость атома.

Тогда дополнительная энергия

$$\begin{aligned} \Delta E &= - \int_0^{\mathcal{E}} \mathcal{E} d(d_{\text{инд}}) = \\ &= - \alpha \int_0^{\mathcal{E}} \mathcal{E} d\mathcal{E} = - \frac{1}{2} \alpha \mathcal{E}^2, \quad (1.57) \end{aligned}$$

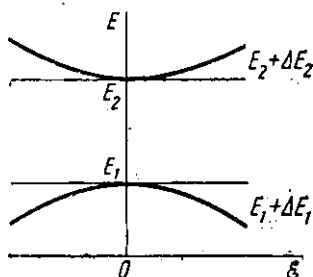


Рис. 1.10. Смещение двух «взаимодействующих» уровней энергии под действием электрического поля

т. е. расщепление пропорционально квадрату напряженности поля и для атомов характерен квадратичный эффект Штарка. Исключение составляет атом водорода и водородоподобные атомы, где наряду с квадратичным проявляется гораздо более сильный линейный эффект Штарка. Он обусловлен наличием у атома водорода случайного вырождения энергетических уровней (термов) по квантовому числу l .

Для вращательных уровней энергии молекул, обладающих постоянным дипольным моментом, согласно соотношению (1.55) наблюдается линейный эффект Штарка.

При квадратичном эффекте Штарка значение дополнительной энергии ΔE в поле \mathcal{E} зависит не только от свойств рассматриваемого энергетического состояния, но также от расположения и свойств ближайших уровней энергии. Каждый уровень E_i как бы «отталкивает» соседний уровень E_j , притом тем больше, чем ближе эти уровни расположены друг к другу. Для простейшего случая двух невырожденных уровней E_1 и E_2 их взаимное «отталкивание» приводит к смещению уровней в разные стороны, как показано на рис. 1.10. Величины смещений равны:

$$\Delta E_1 = - \frac{\beta^2 \mathcal{E}^2}{E_2 - E_1}, \quad \Delta E_2 = \frac{\beta^2 \mathcal{E}^2}{E_2 - E_1}, \quad (1.58)$$

где коэффициент β характеризует «взаимодействие» уровней.

Такой характер смещения уровней дает возможность сортировать молекулы или атомы по энергиям с помощью

неоднородных электрических полей, что, в частности, применяется в аммиачном лазере. Действительно, в неоднородном электрическом поле молекулы будут двигаться в направлении, соответствующем уменьшению их потенциальной энергии. Это означает, что молекулы, находящиеся в энергетическом состоянии E_1 , будут выталкиваться в область больших полей, а молекулы в состоянии E_2 — стремиться в область $\mathcal{E}=0$, где их энергия минимальна.

§ 1.5. ЭНЕРГЕТИЧЕСКИЕ СОСТОЯНИЯ КРИСТАЛЛОВ

Кристалл можно в принципе рассматривать как гигантскую молекулу, в которой разрешено движение электронов в пределах остова, образующего кристаллическую решетку. Этому движению соответствуют электронные уровни энергии. Возможно также колебательное движение решетки, которому соответствуют колебательные уровни энергии.

Определение энергетического спектра кристалла как многоядерной и многоэлектронной системы исходя лишь из «первых принципов» не представляется реальным, так как задача исключительно сложна (напомним, что даже для простейших атомов, содержащих более одного электрона, уравнение Шредингера нельзя решить в аналитическом виде). Поэтому прибегают к различным упрощениям. Подавляющее большинство задач физики твердого тела решают с использованием адиабатического и одноэлектронного приближений, сущность которых аналогична адиабатическому и одноэлектронному приближениям для многоэлектронных атомов (см. § 1.2 и 1.3). Поскольку эти приближения являются фундаментальными в физике твердого тела, напомним их основное содержание.

В *адиабатическом приближении* рассматривается отдельно и независимо движение электронов и ядер вследствие сильного различия в их массах и скоростях. Адиабатическое приближение позволяет разделить волновые функции электронов и ядер и рассматривать электроны как движущиеся в поле неподвижных ядер. Отклонения от адиабатического приближения, например электронно-колебательное (электрон-фононное) взаимодействие, считают малыми возмущениями, которые могут быть при необходимости учтены с помощью хорошо разработанных методов теории возмущений.

В *одноэлектронном приближении* каждый из электронов рассматривается движущимся независимо под ус-

редненным влиянием не только ядер, но и всех других электронов. Корреляция в движении электронов не учитывается, а взаимодействие данного электрона со всеми другими электронами и ядрами заменяется воздействием на него стационарного электрического поля, называемого *самосогласованным*. Как и в случае многоэлектронных атомов, такое движение описывается одноэлектронной волновой функцией, удовлетворяющей одноэлектронному уравнению Шредингера:

$$\hat{H}\psi = -\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{r})\psi = E\psi. \quad (1.59)$$

Здесь $V(\mathbf{r})$ — потенциал, создаваемый усредненным в пространстве зарядом всех электронов и ядер и называемый *самосогласованным потенциалом*.

В отличие от атомов и молекул этот потенциал, а следовательно, и гамильтониан (1.59) не обладают ни сферической, ни аксиальной симметрией, но имеют новый тип симметрии, обусловленный пространственной периодичностью кристаллической решетки:

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}), \quad (1.60)$$

где вектор \mathbf{R} определяется через базисные векторы \mathbf{a}_1 , \mathbf{a}_2 и \mathbf{a}_3 , образующие элементарную ячейку кристалла:

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (1.61)$$

(n_1 , n_2 и n_3 — любые целые числа). Такой тип *трансляционной симметрии* отражает основное свойство кристаллической решетки и определяет фундаментальные характеристики кристалла.

Так как гамильтониан (1.59) не обладает сферической симметрией, то квантовые числа n , l , m_l , m_s , L , S , J и т. д. в кристалле теряют физический смысл. В то же время условие (1.60) 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 (1.59) могут быть выбраны в виде произведения экспоненциальной функции $e^{i\mathbf{k}\mathbf{r}}$ (плоской волны) на функцию, периодическую с периодом решетки:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}), \quad (1.62)$$

где

$$u_{nk}(\mathbf{r} : \mathbf{R}) = u_{nk}(\mathbf{r}).$$

Это означает, что волновая функция электрона, движущегося в периодическом поле, представляет собой модулированную плоскую волну и состояние электрона определено по всему кристаллу. Функции вида (1.62) называют *блоховскими функциями*, а описываемые ими электроны — *блоховскими электронами*. Вектор \mathbf{k} в (1.62) называется *волновым вектором*.

Из (1.62) видно, что с каждым из собственных состояний оператора \hat{H} связан некоторый волновой вектор \mathbf{k} , который определяет состояние системы и, следовательно, является квантовым числом для рассматриваемой задачи. Собственные значения энергии E в (1.59) для состояния $\psi_{\mathbf{k}}$ будут зависеть от \mathbf{k} , аналогично тому как они зависели от квантовых чисел n, l, m_l, \dots в атоме. Но в отличие от атома квантовое число \mathbf{k} может принимать непрерывный, точнее, квазинепрерывный ряд значений. Действительно, при образовании кристалла из N атомов каждый из «атомарных» уровней энергии будет расщепляться и смещаться вследствие эффекта Штарка, что приведет к образованию соответствующих полос или зон энергии. Количество уровней в каждой из таких энергетических зон имеет порядок числа атомов в кристалле, т. е. 10^{23} см^{-3} . Так как каждый из этих уровней имеет конечную ширину δE , определяемую соотношением неопределенности $\delta E \delta t \approx \hbar$, они сливаются, образуя непрерывную полосу разрешенных значений 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 (стремится к нулю при $N \rightarrow \infty$), и для кристаллов обычных размеров по тем же причинам, что и ранее, 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 \mathbf{k} уравнение Шредингера (1.59) может иметь не одно, а бесконечное семейство решений с собственными дискретными значениями, которые можно классифицировать с помощью второго квантового числа n , присутствующего в индексах волновых функций

(1.62). Это квантовое число не имеет глубокого физического смысла и по существу означает номер зоны n .

Из теоремы Блоха (1.62) видно, что состояния, описываемые волновыми векторами \mathbf{k} и \mathbf{k}' , отличающимися друг от друга на вектор обратной решетки \mathbf{q} :

$$\mathbf{k}' = \mathbf{k} + \mathbf{q}, \quad (1.63)$$

полностью эквивалентны. Действительно, для любого вектора обратной решетки $e^{i\mathbf{q}\cdot\mathbf{R}} = 1$, поэтому, если соотношение (1.62) справедливо для \mathbf{k}' , то оно должно быть справедливым также для \mathbf{k} . При заданном n собственные значения энергии электрона $E_{n\mathbf{k}}$ и его собственные волновые функции $\psi_{n\mathbf{k}}(\mathbf{r})$ представляют собой периодические функции \mathbf{k} в пространстве обратной решетки:

$$\begin{aligned} \psi_{n, \mathbf{k}+\mathbf{q}}(\mathbf{r}) &= \psi_{n, \mathbf{k}}(\mathbf{r}), \\ E_{n, \mathbf{k}+\mathbf{q}} &= E_{n, \mathbf{k}}. \end{aligned} \quad (1.64)$$

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.

Из сказанного следует, что для полного описания всей совокупности энергетических состояний кристалла область значений \mathbf{k} достаточно ограничить одной элементарной ячейкой обратной решетки, т. е. ограничиться рассмотрением одной зоны Бриллюэна. Так как волновой вектор \mathbf{k} может принимать квазинепрерывный ряд значений, то энергетические состояния электрона в поле, обладающем трансляционной симметрией (1.60), необходимо характеризовать семейством непрерывных функций $E_{n, \mathbf{k}}$ [или $E_n(\mathbf{k})$]. Эти функции определяют *зонную структуру* кристалла. Совокупность всех электронных уровней, описываемых функцией $E_n(\mathbf{k})$ при заданном n , называют *энергетической зоной*.

Волновое число k является основным квантовым числом, характеризующим состояние электрона в кристалле. Поэтому зависимость $E(\mathbf{k})$ имеет фундаментальное значение для физики твердого тела. В частности, скорость движения электрона в кристалле можно отождествить с групповой скоростью волнового пакета

$$v = \frac{d\omega}{dk} = \frac{dE}{\hbar dk}. \quad (1.65)$$

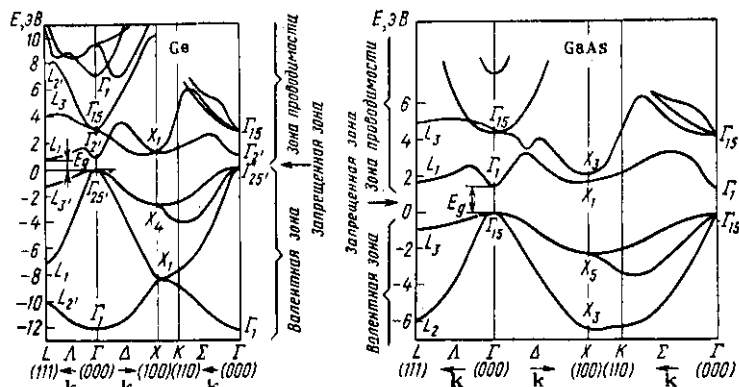


Рис. 1.11. Зонная структура германия и арсенида галлия

Пусть на электрон, описываемый блоховской функцией (1.62), действует сила F . Тогда изменение энергии за время dt будет равно $Fvdt$. С другой стороны, $dE = \left(\frac{dE}{dk}\right) dk = v d(\hbar k)$, откуда $F = d(\hbar k)/dt$. Это равенство можно рассматривать как аналог закона Ньютона, если величину $\hbar k$ определить 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

$$m^* = \frac{\hbar^2}{d^2E/dk^2} \quad (1.66)$$

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 квантовой электронике и оптоэлектронике наряду с элементарными полупроводниками — германием и кремнием — наиболее широкое применение получили полупроводниковые соединения элементов III и V групп Периодической системы Д. И. Менделеева (полупроводники группы $A^{III}B^V$) и их твердые растворы.

Рассмотрим особенности энергетического спектра этих материалов.

В качестве примера на рис. 1.11 приведены части энергетических диаграмм $E(k)$ германия и арсенида галлия. Наибольшее влияние на электрофизические и оптические свойства материала оказывают наивысшие заполненные состояния (валентная зона) и наинизшие свободные состояния (зона проводимости). Именно эти части зонной структуры изображены на рис. 1.11.

Обычно зонную структуру изображают в виде зависимостей $E(k)$ 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 со следующим соответствием: точки (100) — X , (111) — L и (110) — K на границе зоны при $k=2\pi/a$ и $[100]$ — Δ , $[111]$ — Λ и $[110]$ — Σ 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 точке зоны Бриллюэна. Чтобы отличить состояния, принадлежащие валентной зоне и зоне проводимости, их помечают верхним индексом v или c соответственно.

Как видно из рис. 1.11, 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 простирается от нуля до 0,7 эВ, а для арсенида галлия — от нуля 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

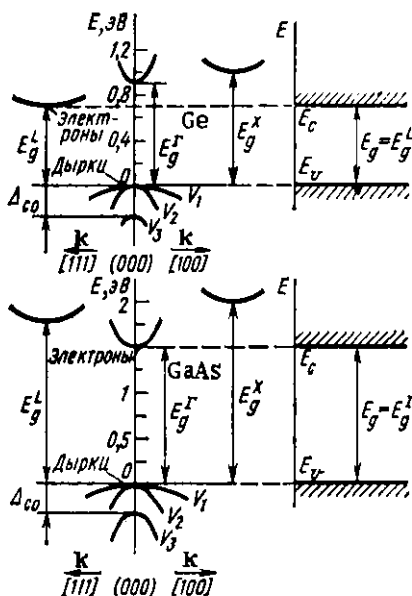


Рис. 1.12. Упрощенные энергетические диаграммы для электронных состояний в кристаллах германия и арсенида галлия

точке Γ . В арсениде галлия абсолютный минимум зоны проводимости и максимум валентной зоны в k -пространстве совпадают: оба они 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^Γ) и непрямыми (например, E_g^L или E_g^X).

В полупроводниках группы $A^{III}B^V$ наблюдается изменение их основных свойств по мере изменения суммы атомных номеров элементов, составляющих данное соеди-

вие (рис. 1.12). 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. 1.12. Ширина запрещенной зоны соответствует минимальному энергетическому зазору. Для Ge $E_g = E_g^L = L_1^\Gamma - \Gamma_{25}^c$, а для арсенида галлия $E_g = E_g^\Gamma = \Gamma_1^\Gamma - \Gamma_{25}^c$.

Как видно из рис. 1.11 и 1.12, наинизший (абсолютный) минимум зоны проводимости германия (L_1^Γ -минимум) расположен на границе зоны Бриллюэна, в то время как максимум валентной зоны — в центре зоны, в

нение. С увеличением $\Sigma Z = Z_{III} + Z_V$ происходит монотонное уменьшение соответствующих энергетических зазоров, как показано на рис. 1.13. Скорость изменения «прямого» энергетического зазора E_g^r 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 $\Sigma Z \leq 56$ — непрямоzonной структурой, как схематически показано на рис. 1.13. 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 изменяются сравнительно слабо. Например, для тяжелых дырок отношение m_p^V/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 соотношение

$$\frac{m_n^r}{m} \approx 0,05 E_g^r,$$

где энергия E_g^r выражена в электронвольтах.

Закон дисперсии, т. е. зависимость $E(k)$, вблизи Γ -минимума изотропен. Отличие от простой квадратичной зависимости

$$E_n^r = \frac{\hbar^2 k^2}{2m_n^r}, \quad (1.67)$$

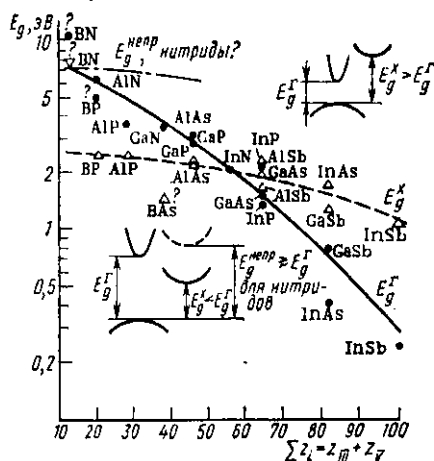


Рис. 1.13. Общие закономерности в изменении «прямого» (E_g^r , сплошная кривая) и «непрямого» (E_g^x , пунктирная кривая) энергетических зазоров в полупроводниках группы $A^{III}B^V$

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

$$\frac{3}{m_n} = \frac{1}{m_{n\parallel}} + \frac{2}{m_{n\perp}}. \quad (1.68)$$

Кроме бинарных соединений элементов III и V групп Периодической системы, большое значение для оптоэлектроники и лазерной техники имеют твердые растворы на их основе. Принципиально важной особенностью твердых растворов является возможность плавно и в широких пределах изменять параметры материала путем изменения его состава. Это позволяет получать материал с точно заданным набором основных характеристик, что особенно ценно при создании приборов интегральной оптики, оптоэлектронных устройств, гетеролазеров, фотоприемников, полупроводниковых модуляторов и т. д.

Твердые растворы довольно широко распространены в классе алмазоподобных полупроводников. Многие соединения группы $A^{III}B^V$ образуют непрерывный ряд твердых растворов замещения. Их можно рассматривать как соединения переменного состава с узкой областью гомогенности. Идеальным твердым раствором замещения считается кристалл с идеальной решеткой, по соответствующим узлам которой статистически независимо распределены различные атомы составляющих компонентов. В таких твердых растворах сохраняется строгая трансляционная симметрия кристаллической решетки, поскольку каждый из атомов располагается в узле решетки. Но периодичность в потенциальной энергии кристалла здесь нарушена. Это иллюстрирует рис. 1.14, где схематически показано изменение потенциала в неупорядоченном твердом растворе замещения A_xB_{1-x} .

Нарушение трансляционной симметрии потенциальной энергии кристалла приводит, строго говоря, к несправедливости теоремы Блоха и невозможности рассматривать волновой вектор \mathbf{k} как квантовое число для электронов. Однако во многих твердых растворах (в том числе в твер-

дых растворах полупроводников группы $A^{III}B^V$) это нарушение сравнительно невелико. Средняя длина пробега L носителя заряда существенно больше межатомных расстояний a , так что $kL \gg 1$. Поэтому нарушение трансляционной симметрии гамильтониана, вызванное статистическим распределением атомов в узлах кристаллической решетки твердого раствора, можно рассматривать как малое возмущение, т. е. в (1.59) представить $V(r)$ в виде

$$V(r) = V_0(r) + V_1(r), \quad (1.69)$$

где $V_0(r)$ — периодический потенциал, обладающий трансляционной симметрией; $V_1(r)$ — неперіодическая часть потенциала, обусловленная статистически беспорядочным распределением атомов твердого раствора типа $A_xB_{1-x}C$ по узлам подрешетки R_{AB} .

Обычно $V_1(r) \ll V_0(r)$, влияние неперіодической части потенциала можно рассматривать как малое возмущение и соответственно пользоваться методами теории возмущений, как это делается, например, для кристаллов с дефектами. Поэтому при рассмотрении свойств твердых растворов правомерно пользоваться теми же понятиями зонной теории, что и для идеальных кристаллов, в частности, такими понятиями, как «волновой вектор», «зонная структура», «правила отбора» и т. д.

Образование твердых растворов в полупроводниках группы $A^{III}B^V$ возможно при замещении как анионной, так и катионной подрешеток, например, трехкомпонентные растворы в системах $GaAs - GaP (GaAs_{1-x}P_x)$ и $GaAs - AlAs (Ga_xAl_{1-x}As)$. Возможны и четырехкомпонентные твердые растворы типа $Ga_xIn_{1-x}As_yP_{1-y}$.

Параметры зонной структуры плавно изменяются с изменением состава твердого раствора. Поскольку структуры валентных зон в полупроводниках группы $A^{III}B^V$ сходны, основные изменения в энергетическом спектре их твердых растворов определяются различием в структурах зон проводимости составляющих компонентов. Наиболее важным здесь является характер изменения соответствующих энергетических зазоров в зависимости от состава твердого раствора. Эти зависимости отклоняются от ли-

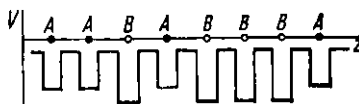


Рис. 1.14. Схематическое изображение изменения потенциальной энергии с координатой z в твердом растворе замещения $A_xB_{1-x}C$

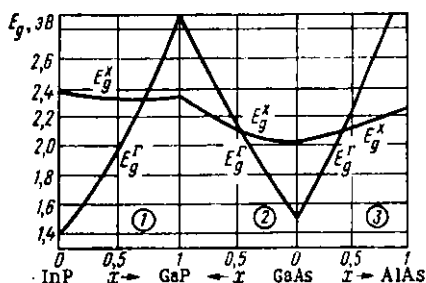


Рис. 1.15. Изменение «прямого» (E_g^r) и «непрямого» (E_g^x) энергетических зазоров в квазибинарных твердых растворах $\text{In}_{1-x}\text{Ga}_x\text{P}$ (1), $\text{GaAs}_{1-x}\text{P}_x$ (2) и $\text{Al}_x\text{Ga}_{1-x}\text{As}$ (3)

Эти зависимости, естественно, тесно связаны с общими закономерностями в изменении E_g (см. рис. 1.13). Для кристаллов $\text{GaAs}_{1-x}\text{P}_x$ с $x < 0,46$ Г-минимум зоны проводимости расположен ниже X-минимума, и эти твердые растворы обладают «прямозонной» структурой. Твердые растворы $\text{GaAs}_{1-x}\text{P}_x$ при $x > 0,46$ обладают «непрямозонной» структурой. Аналогично кристаллы $\text{Ga}_x\text{Al}_{1-x}\text{As}$ при $x < 0,36$ обладают «непрямозонной», а при $x > 0,36$ — «прямозонной» структурой. Состав твердого раствора, при котором происходит переход от «прямозонной» к «непрямозонной» структуре, называется *переходным*.

С другой стороны, нарушение периодичности в потенциальной энергии твердого раствора может вызывать появление новых эффектов, например: кажущееся «нарушение» закона сохранения квазимпульса, расщепление и сдвиг энергетических уровней примесных состояний, возможное «размазывание» энергетического спектра.

Рассмотрим энергетические состояния, связанные с колебательным движением 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 колебательных движений ядер в трехмерной решетке приводит к ограниченному числу колебательных ветвей в противоположность бесконечному числу электронных энергетических зон. Об-

нейных и могут быть представлены квадратичной функцией состава x в виде

$$E_g(x) = E_g(0) + [E_g(1) - E_g(0)]x - cx(1-x). \quad (1.70)$$

На рис. 1.15 показан характер изменения энергетических зазоров в твердых растворах, часто применяемых в качестве основных рабочих материалов оптоэлектронных устройств.

шее число различных колебательных ветвей, называемых также *типами колебаний* или *модами*, равно утроенному числу атомов в примитивной ячейке кристалла. Состояния в каждой из ветвей будут определяться тем же волновым вектором k , что и электронные состояния.

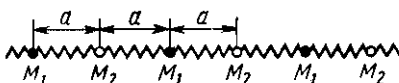


Рис. 1.16. Колебания в одномерном кристалле с двумя атомами в элементарной ячейке

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 квантованным 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 (1.34):

$$E_{\text{кол}} = \hbar\Omega_k (q + 1/2). \quad (1.71)$$

Здесь $\hbar\Omega_k$ — энергия фонона с волновым вектором k ; q — число фононов.

Полная колебательная энергия кристалла определяется суммированием выражения (1.71) по всем значениям k и всем колебательным ветвям.

Представим себе одномерную модель решетки в виде цепочки, состоящей из двух атомов с массами M_1 и M_2 , которые связаны между собой упругой силой, как показано на рис. 1.16. Собственные частоты колебаний в такой бесконечной цепочке определяются соотношением

$$v_{\pm} = \left[a, \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \pm a, \sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2} \right)^2 - \frac{4 \sin^2(ka)}{M_1 M_2}} \right]^{1/2}, \quad (1.72)$$

где a_j — константа связи; k — волновое число; a — равновесное расстояние между атомами в цепочке.

Частоты ν_- пропорциональны k вблизи $k \approx 0$. Эта ветвь называется акустической, ибо вблизи $k \approx 0$ частотам ν_- соответствуют частоты звуковых колебаний, распространяющихся в среде.

Частота ν_+ стремится к $\left[2a_j \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \right]^{1/2}$ при $k \rightarrow 0$.

Эта ветвь существенно отличается от акустической. Она называется оптической ветвью, так как вблизи $k=0$ частота ν_+ лежит в оптическом (инфракрасном) диапазоне.

Для акустической ветви две подрешетки атомов с различными массами движутся в одном направлении (фазы совпадают), для оптической — в противоположных направлениях (фазы противоположны).

В трехмерной решетке общее число типов колебаний будет в три раза больше, чем в одномерной. Довольно часто ветви колебаний решетки (моды) можно классифицировать как продольные и поперечные. Это означает, что поляризация ветви соответственно параллельна и перпендикулярна k или близка к этому. Такая классификация справедлива для всех k вдоль осей симметрии. Для продольных колебаний направление движения ядер решетки приблизительно совпадает с направлением движения волны, т. е. с направлением вектора k , а для поперечных колебаний — приблизительно перпендикулярно k . Эти ветви колебаний, или типы фононов, обозначают ТО — поперечные оптические, ЛО — продольные оптические, ТА — поперечные акустические и ЛА — продольные акустические.

В качестве примера на рис. 1.17 приведены дисперсии фоновых ветвей для кристаллов Ge и GaAs. Вместо шести ветвей имеются лишь четыре, поскольку каждые из ТА- и ТО-колебаний в направлениях $[100]$ и $[111]$ дважды вырождены. В направлении $[110]$ для акустических колебаний это вырождение снимается.

До сих пор рассматривались энергетические состояния идеальных кристаллов. Примеси, как известно, определяют ряд важнейших свойств кристаллов, особенно полупроводников. Примесные состояния в кристаллах условно можно считать слабо связанными («мелкими») состояниями, для которых энергия связи существенно меньше ширины запрещенной зоны, и сильно связанными («глу-

бокими») состояниями, для которых энергия связи 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 (1.6), а действие кристалла учитывается введением эффективной диэлектрической проницаемости ϵ_r :

$$V(r) = -\frac{e^2}{4\pi\epsilon_0\epsilon_r r}. \quad (1.73)$$

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 (1.9) с той разницей, что вместо массы свободного электрона m будет эффективная масса m_n :

$$-\frac{\hbar^2}{2m_n} \nabla^2 \psi(r) - \frac{e^2}{4\pi\epsilon_0\epsilon_r r} \psi(r) = E \psi(r). \quad (1.74)$$

Решения уравнения (1.74) аналогичны решениям урав-

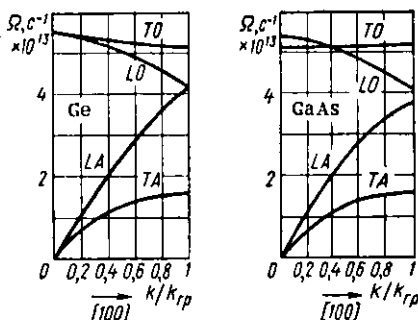


Рис. 1.17. Дисперсионные кривые для фононов в направлении [100] для германия и арсенида галлия

нения для атома водорода. В случае изотропной эффективной массы спектр разрешенных значений энергии, например для водородоподобного донора, будет аналогичен (1.10), где вместо массы свободного электрона m будет фигурировать масса m_n , а вместо диэлектрической проницаемости вакуума — диэлектрическая проницаемость среды:

$$E_n = -\frac{m_n e^4}{8(\epsilon_0 \epsilon_r \hbar)^2} \frac{1}{n^2} = -\frac{R_\infty}{\epsilon_r^2} \left(\frac{m_n}{m}\right) \frac{1}{n^2} = -\frac{E_d}{n^2}. \quad (1.75)$$

Здесь R_∞ — постоянная Ридберга, а n — главное квантовое число.

За начало отсчета энергии в (1.75) аналогично (1.10) 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*

$$E_d = \frac{R_\infty}{\epsilon_r^2} \left(\frac{m_n}{m}\right). \quad (1.76)$$

Найдем энергию ионизации водородоподобного донора, например, для арсенида галлия. При $\epsilon_r=13$, $m_n^{\Gamma}=0,07m$ из (1.76) имеем $E_d=5,6$ мэВ. Для акцепторных состояний энергия отсчитывается от потолка-валентной зоны. Аналогично (1.76) для энергии ионизации водородоподобного акцептора в GaAs при $m_p \approx 0,5m$ получаем $E_a \approx 40$ мэВ.

По аналогии с атомом водорода вводят понятие *эффективного боровского радиуса* для доноров:

$$a_d = \frac{\hbar^2 \epsilon_0 \epsilon_r}{\pi m_n e^2} = \epsilon_r \left(\frac{m}{m_n}\right) a_B, \quad (1.77)$$

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 (1.77) 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 основание 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:

$$\psi_d(\mathbf{r}) = \sum_{\mathbf{k}} C(\mathbf{k}) \psi_{\mathbf{k}}(\mathbf{r}). \quad (1.78)$$

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 ожидать, что в (1.78) наибольший вклад в формирование функции $\psi_d(\mathbf{r})$ будут давать ближайшие состояния, и со значительной амплитудой под знак суммы войдут лишь члены с \mathbf{k} вблизи экстремума зоны. Действительно, если (1.78) подставить в (1.74) и провести преобразование Фурье, то с точностью до постоянного множителя получим

$$C(\mathbf{k}) \sim \frac{1}{[(ka_d)^2 + 1]^2}. \quad (1.79)$$

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. 1.18. Для сравнения там же изображен спектр энергетических уровней донора на диаграмме $E(x)$. Соотношение

$$\Delta k \approx 1/a_d \quad (1.80)$$

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 (1.80).

Из (1.79), (1.80) следует, что чем сильнее электрон (дырка) локализован вблизи примесного центра в обычном пространстве, тем в большей степени он «размазан» в \mathbf{k} -пространстве.

В многодолинных полупроводниках типа германия и кремния волновая функция для донорного состояния должна строиться из комбинаций компонент, соответствующих каждому эквивалентному минимуму зоны прово-

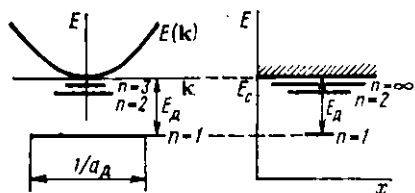


Рис. 1.18. Изображение состояний водородоподобного донора на диаграмме $E(k)$, т. е. в k -пространстве и в геометрическом g -пространстве

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 (1.73) 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.

ГЛАВА 2. ВЗАИМОДЕЙСТВИЕ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ С АТОМНЫМИ СИСТЕМАМИ

В основе работы всех приборов квантовой и оптической электроники лежат процессы взаимодействия электромагнитного излучения с веществом. Если в радиодиапазоне в большинстве случаев проявляются волновые свойства электромагнитного излучения и его можно рассматривать в виде определенного набора электромагнитных волн, то 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.

Элементарными процессами, определяющими взаимодействие света с веществом, являются испускание, погло-

димости. 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 энер-

щение 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. Однако нередко можно ограничиться полуженменологической теорией, которой и воспользуемся далее.

§ 2.1. СПОСОБЫ ОПИСАНИЯ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

Строго говоря, оптическое излучение необходимо рассматривать как электромагнитное поле, подчиняющееся квантовым закономерностям. Однако при обсуждении многих практически важных вопросов можно ограничиться приближенными представлениями и рассматривать электромагнитное излучение в виде: 1) световых лучей (или пучков); 2) световых (электромагнитных) волн; 3) 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 наикратчайшего 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

$$\int_A^B n \, dl. \quad (2.1)$$

Для однородной среды $n = \text{const}$ и, очевидно, $s = nl$.

Так как $dl = \frac{c}{n} dt$, то

$$\int_A^B n dl = c \int_A^B dt. \quad (2.2)$$

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.

Закон преломления (закон Снеллиуса) гласит, что при прохождении границы раздела двух сред с показателями преломления n_1 и n_2 отношение синуса угла преломления θ_2 к синусу угла падения θ_1 равно отношению обратных показателей преломления:

$$\sin \theta_2 / \sin \theta_1 = n_1 / n_2 = n_{12} = 1 / n_{21}. \quad (2.3)$$

Углы θ_1 и θ_2 образованы падающим и преломленным лучами с нормалью к поверхности раздела и лежат с ней в одной плоскости. Показатели преломления n_1 и n_2 равны отношению скорости света в вакууме c к фазовой скорости света в среде v :

$$n = c/v$$

и называются *абсолютными показателями преломления*. Величина n_{12} в (2.3) называется *относительным показателем преломления* одной среды относительно другой.

Закон отражения гласит, что угол отражения от границы раздела двух сред равен по абсолютной величине углу падения и лежит с ним в одной плоскости.

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 $c = 299\,792 \pm 0,5$ км/с. Волновая оптика базируется на уравнениях классической электродинамики, основу которой составляют уравнения Макс-

велла. 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:

$$\begin{aligned} \operatorname{rot} \mathcal{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \quad \operatorname{div} \mathbf{D} = \rho, \\ \operatorname{rot} \mathcal{H} &= \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \quad \operatorname{div} \mathbf{B} = 0. \end{aligned} \quad (2.4)$$

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

$$\mathbf{D} = \epsilon_r \epsilon_0 \mathcal{E}, \quad \mathbf{B} = \mu_r \mu_0 \mathcal{H}, \quad \mathbf{j} = \sigma \mathcal{E}, \quad (2.5)$$

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 по существу отражают реакцию среды на внешнее электромагнитное возмущение.

На границе раздела двух сред выполняются граничные условия для тангенциальных τ и нормальных n составляющих:

$$\begin{aligned} \mathcal{E}_{\tau 1} &= \mathcal{E}_{\tau 2}, \quad D_{n1} - D_{n2} = \sigma_{\text{пов}}, \\ \mathcal{H}_{\tau 1} - \mathcal{H}_{\tau 2} &= j_{\text{пов}}, \quad B_{n1} = B_{n2}. \end{aligned} \quad (2.6)$$

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 (2.4), уравнения (2.5) и граничные условия (2.6), называется системой уравнений Максвелла и играет в классической 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 (2.4) 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 $j=0$, $\rho=0$, $\mu_r=1$, $\epsilon_r=1$ и система (2.4) имеет вид:

$$\begin{aligned}\operatorname{rot} \mathcal{E} &= -\mu_0 \frac{\partial \mathcal{H}}{\partial t}, \quad \operatorname{div} \mathcal{E} = 0, \\ \operatorname{rot} \mathcal{H} &= \epsilon_0 \frac{\partial \mathcal{E}}{\partial t}, \quad \operatorname{div} \mathcal{H} = 0.\end{aligned}\tag{2.7}$$

Отсюда находим

$$\operatorname{rot} \operatorname{rot} \mathcal{E} = -\mu_0 \frac{\partial}{\partial t} \operatorname{rot} \mathcal{H} = -\epsilon_0 \mu_0 \frac{\partial^2 \mathcal{E}}{\partial t^2}.$$

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

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

откуда получаем волновое уравнение для вектора напряженности электрического поля:

$$\Delta \mathcal{E} - \epsilon_0 \mu_0 \frac{\partial^2 \mathcal{E}}{\partial t^2} = 0\tag{2.8}$$

(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:

$$\Delta \mathcal{H} - \epsilon_0 \mu_0 \frac{\partial^2 \mathcal{H}}{\partial t^2} = 0.\tag{2.8a}$$

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

$$\mathbf{B} = \operatorname{rot} \mathbf{A}, \quad \mathcal{E} = -\operatorname{grad} \varphi - \frac{\partial \mathbf{A}}{\partial t}.\tag{2.9}$$

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

$$\mu_0 \mathcal{H} = \operatorname{rot} A, \quad \mathcal{E} = -\frac{\partial}{\partial t} A. \quad (2.10)$$

Подставляя (2.10) в (2.7), находим

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

откуда аналогично (2.8) получаем, что векторный потенциал A удовлетворяет, как и векторы \mathcal{E} и \mathcal{H} , волновому уравнению

$$\Delta A - \epsilon_0 \mu_0 \frac{\partial^2 A}{\partial t^2} = 0. \quad (2.11)$$

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, где постоянная c называется электродинамической постоянной или скоростью света.

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

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

$$\rho = \frac{\epsilon_0 \mathcal{E}^2}{2} + \frac{\mu_0 \mathcal{H}^2}{2}. \quad (2.12)$$

Распространение электромагнитной энергии в пространстве характеризуется *вектором Пойнтинга* S :

$$S = [\mathcal{E} \times \mathcal{H}]. \quad (2.13)$$

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

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

В качестве энергетической характеристики часто пользуются понятием интенсивности излучения I , под которым понимают величину, пропорциональную квадрату амплитуды электромагнитного колебания. Очевидно, $I \sim |S|$.

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 :

$$p = \frac{[E \times H]}{c^2} = \frac{S}{c^2}. \quad (2.14)$$

В общем случае решения волновых уравнений (2.8) имеют вид:

$$E = E_1 \left(t - \frac{rn}{c} \right) + E_2 \left(t + \frac{rn}{c} \right), \quad (2.15)$$

$$H = H_1 \left(t - \frac{rn}{c} \right) + H_2 \left(t + \frac{rn}{c} \right), \quad (2.15a)$$

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). Следовательно, если в некоторый момент времени $t = 0$ в некоторой точке r пространства поле имело определенное значение, то через время t такое же значение поле будет иметь на расстоянии ct в направлении r .

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 волн, распространяющихся только в сторону положительных значений r , т. е. учесть лишь первые слагаемые в (2.15). 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 \mathcal{E} и \mathcal{H} зависят только от времени и одной декартовой координаты. Например, для плоской волны, распространяющейся в направлении z ($\mathcal{E}_z = \mathcal{H}_z = 0$), волновые уравнения (2.8) принимают вид:

$$\frac{\partial^2}{\partial z^2} \mathcal{E}(z, t) - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \mathcal{E}(z, t) = 0, \quad (2.16)$$

$$\frac{\partial^2}{\partial z^2} \mathcal{H}(z, t) - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \mathcal{H}(z, t) = 0. \quad (2.16a)$$

В плоской волне все лучи параллельны друг другу. Энергия электрического поля световой волны равна энергии магнитного поля: $\epsilon_0 \mathcal{E}^2/2 = \mu_0 \mathcal{H}^2/2$, так что

$$\mathcal{H} = \sqrt{\frac{\epsilon_0}{\mu_0}} [\mathbf{n} \times \mathcal{E}], \quad (2.17)$$

где \mathbf{n} — единичный вектор в направлении распространения волны.

Подставляя (2.17) в (2.13), с учетом (2.12) находим

$$\mathbf{S} = \mathbf{n}rc. \quad (2.18)$$

Электромагнитная волна называется *монохроматической*, если компоненты векторов \mathcal{E} и \mathcal{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 \mathcal{E} $\partial^2 \mathcal{E} / \partial t^2 = -\omega^2 \mathcal{E}$, и распределение поля по пространству в монохроматической волне соответствует уравнению

$$\Delta \mathcal{E} + \frac{\omega^2}{c^2} \mathcal{E} = 0, \quad (2.19)$$

которое вытекает из (2.8). Аналогичные соотношения получаются из уравнений (2.8), (2.11) для векторов \mathcal{H} и \mathbf{A} .

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

$$\mathcal{E} = \mathcal{E}_0 \cos \omega(t - z/c), \quad (2.20)$$

или

$$\mathcal{E} = \operatorname{Re} \{ \mathcal{E}_0 e^{-i\omega(t - z/c)} \} = \operatorname{Re} \{ \mathcal{E}_0 e^{i\omega(t - z/c)} \}, \quad (2.21)$$

где Re — вещественная часть комплексного выражения, стоящего в фигурных скобках, а \mathcal{E}_0 — некоторый постоян-

ный комплексный вектор, называемый *амплитудным вектором*. Укажем, что выражения (2.20) и (2.21) полностью эквивалентны. Часто знак Re в записи вида (2.21) опускают, подразумевая при этом, что физический смысл имеет только вещественная часть.

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

$$\mathbf{k} = \frac{\omega}{c} \mathbf{n} = \frac{2\pi}{\lambda} \mathbf{n} \quad (2.22)$$

называется *волновым вектором*. С его помощью выражение (2.21) удобно представить в виде, не зависящем от выбора осей координат:

$$\mathcal{E} = \mathcal{E}_0 e^{i(\mathbf{k}\mathbf{r} - \omega t)}. \quad (2.23)$$

Аналогично для вектора напряженности магнитного поля можно написать

$$\mathcal{H} = \mathcal{H}_0 e^{i(\mathbf{k}\mathbf{r} - \omega t)} \quad (2.23a)$$

и для векторного потенциала

$$\mathbf{A} = \mathbf{A}_0 e^{i(\mathbf{k}\mathbf{r} - \omega t)}. \quad (2.23b)$$

Величину, стоящую с множителем i в показателе экспоненты, называют *фазой волны*. Скорость c в выражениях (2.20), (2.21) есть фазовая скорость света.

Рассмотрим подробнее вопрос о направлении поля монохроматической волны. Известно, что векторы \mathcal{E} и \mathcal{H} и соответственно амплитудные векторы \mathcal{E}_0 и \mathcal{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 \mathcal{E} и \mathcal{H} в любой точке

пространства остаются неизменными во времени. *Плоскостью поляризации* при этом является плоскость, проходящая через вектор \mathbf{E} , и направление распространения электромагнитной волны. Для линейно поляризованного света, распространяющегося в направлении z :

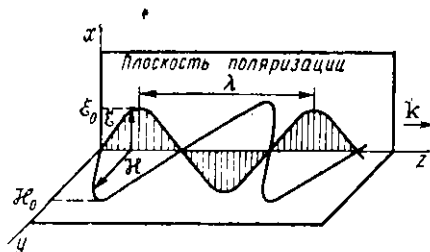


Рис. 2.1. «Мгновенная фотография» плоской линейно поляризованной монохроматической волны

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}r - \omega t)}, \quad (2.24)$$

где \mathbf{E}_0 — амплитуда волны.

«Мгновенная фотография» плоской линейно поляризованной монохроматической волны представлена на рис. 2.1.

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 поляризации (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 (2.13), (2.18):

$$I = \langle S \rangle = \frac{1}{T} \int_0^T S dt = \rho c, \quad (2.25)$$

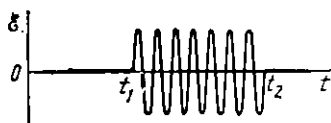


Рис. 2.2. Пример немонохроматической волны

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

Для плоской линейно поляризованной мономонхроматической волны (2.24) $I \sim \mathcal{E}_0^2$.

Понятие плоской мономонхроматической волны, описываемой соотношениями (2.23), (2.24), 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. 2.2, то она не является мономонхроматической волной [соотношения (2.23), (2.24) должны быть справедливы при любых значениях t] и ее можно представить состоящей из совокупности плоских мономонхроматических волн, т. е. в виде разложения (или интеграла) Фурье. Совокупность мономонхроматических волн, образующих реальную волну, называется *группой волн* или *волновым пакетом*. Важность понятия плоской поляризованной мономонхроматической волны и заключается в том, что любую реальную волну можно представить набором (суперпозицией) идеализированных мономонхроматических волн с разными частотами ω . Действительно, любую функцию $f(t)$ можно представить в виде *разложения Фурье по мономонхроматическим (гармоническим) волнам*:

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{-i\omega t} d\omega. \quad (2.26)$$

Здесь под $f(t)$ понимают любую из компонент векторов \mathcal{E} , \mathcal{H} или \mathbf{A} . Интеграл Фурье (2.26) 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

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \quad (2.27)$$

и имеет смысл амплитуды мономонхроматической 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 частоте, т. е. спектральное распределение, вводят функцию $\rho(\omega)$, называемую *спектральной плотностью излучения* и связанную с интегральной плотностью излучения ρ [см. (2.12)] соотношением *

$$\rho = \int_0^{\infty} \rho(\omega) d\omega. \quad (2.28)$$

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.

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

$$E_{\text{фот}} = \hbar\omega, \quad (2.29)$$

$$p_{\text{фот}} = \hbar k, \quad (2.30)$$

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

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

** От греч. *phos* (род. падеж *photos*) — свет; термин «фотон» был введен в физику Г. Н. Льюисом (1929).

2. Масса покоя фотона равна нулю (экспериментально доказано, что у фотона, как элементарной частицы, масса покоя, по крайней мере, меньше $10^{-24} m$, где m — 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.

Необходимо подчеркнуть, что представления света в виде волн или корпускул не исключают одно другое. Это приближенное описание одних и тех же процессов. Оказывается, что квантово-волновой дуализм присущ природе в целом. Другое дело, что 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. Оптический диапазон расположен как раз посередине (см. рис. В.1) и его особенностью является то, что здесь в одинаковой мере проявляются и волновые, и корпускулярные свойства излучения.

Наиболее последовательное и полное рассмотрение свойств электромагнитного излучения проводится в рамках квантовой электродинамики с единых позиций — как квантованное электромагнитное поле.

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, казалось бы, такие непонятные с точки зрения «здравого смысла» (точнее, с точки зрения привычных житейских представлений) 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.

Рассмотрим в качестве примера вопрос, связанный с поляризацией света. Пусть имеется линейно поляризованный свет, пропускаемый через поляроид. Поляроид, в качестве которого могут применяться, например, некоторые анизотропные кристаллы, обладает тем свойством, что пропускает только свет, поляризованный в направлении, перпендикулярном оптической оси кристалла. Тогда пучок, поляризованный перпендикулярно оптической оси, пройдет через кристалл полностью, а параллельно оси — не пройдет совсем. Если пучок поляризован под углом α к оптической оси кристалла, то через кристалл пройдет часть пучка, пропорциональная $\sin^2 \alpha$. С точки зрения классической электродинамики эти результаты ясны и понятны.

Рассмотрим эти же результаты, представив свет как пучок линейно поляризованных фотонов. Если каждый фотон поляризован перпендикулярно или параллельно оси кристалла, то все фотоны либо пройдут, либо не пройдут через кристалл. Это тоже понятно. Основная сложность возникает, когда падающий пучок поляризован наклонно, т. е. под углом α к оси кристалла. Каждый из падающих фотонов поляризован наклонно и вопрос о том, что произойдет с отдельным фотоном, не совсем ясен. Во всяком случае можно с уверенностью сказать, что фотон не делится и не может пройти через кристалл лишь частично: он либо пройдет, либо не пройдет.

Современная квантовая теория дает следующее описание этого явления. Чтобы можно было говорить о поляризации фотона в том или ином направлении, необходимо иметь начало отсчета, т. е. какое-то выделенное направление, которое в нашем случае создается с помощью анизотропного кристалла — поляроида (напомним, что при рассмотрении эффектов Зеемана и Штарка в § 1.4 аналогичное выделенное направ-

ление z создавалось полем). Фотон, поляризованный под углом α к оптической оси кристалла (к выделенному направлению) можно считать находящимся частично в состоянии поляризации, параллельной оси, и частично в состоянии поляризации, перпендикулярной оси. Следовательно, состояние поляризации под углом к оси можно рассматривать как результат наложения (суперпозиции) двух взаимно перпендикулярных состояний поляризации. Пропуская фотон через поляризатор, его подвергают наблюдению, в результате которого фотон переводится полностью в одно из двух возможных состояний поляризации: либо параллельной, либо перпендикулярной относительно оптической оси. После прохождения фотона через поляризатор на выходе всегда будет обнаружен фотон с энергией, равной энергии падающего фотона, или не обнаружено ничего. При этом целый фотон будет поляризован перпендикулярно оптической оси кристалла. При многократном повторении таких опытов с отдельными фотонами окажется, что число случаев, в которых фотон будет обнаружен на выходе, пропорционально $\sin^2 \alpha$. Это означает, что фотон, поляризованный под углом α к оси поляризатора, пройдет через него с вероятностью $\sin^2 \alpha$ (и при этом будет поляризован перпендикулярно оптической оси кристалла) или не пройдет (поглотится) с вероятностью $\cos^2 \alpha$.

Аналогично, используя принцип суперпозиции состояний, можно объяснить явления, связанные с интерференцией отдельных фотонов. Подчеркнем, что каждый фотон интерферирует лишь с самим собой. При наличии большого числа фотонов интерференция принимает классический характер, что отчасти маскирует ее квантовомеханическую природу.

Таким образом, с точки зрения квантовых представлений свет 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 запишутся в сле-

дующем виде:

$$\hbar\omega + E = \hbar\omega' + E', \quad (2.31)$$

$$\hbar\mathbf{k} + \mathbf{p} = \hbar\mathbf{k}' + \mathbf{p}'. \quad (2.32)$$

В этих соотношениях, как и в (2.29), (2.30), 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:

$$\rho = \sum_{\alpha=1}^2 \sum_{\mathbf{k}} \hbar\omega N_{k\alpha}. \quad (2.33)$$

Аналогично полный импульс

$$\mathbf{p} = \sum_{\alpha=1}^2 \sum_{\mathbf{k}} \hbar\mathbf{k} N_{k\alpha}. \quad (2.34)$$

Сравнение этих соотношений с аналогичными классическими выражениями (2.12) и (2.14) позволяет перейти от числа фотонов в $k\alpha$ -состоянии к амплитуде электромагнитной волны с поляризацией α и волновым вектором \mathbf{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 полуцелым 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. (2.12) и (2.14)], а при квантовомеханическом — скачками [см. (2.33) и (2.34)]. Квантованием поля, т. е. его дискретностью, можно пренебречь, если полная энергия ρ существенно меньше энергии кванта $\hbar\omega$, т. е. если

$$N_{k\alpha} \gg 1. \quad (2.35)$$

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.

§ 2.2. 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 в силу внутренних причин. В большинстве случаев возмущение, под воздействием которого совершается квантовый переход, можно считать малым по сравнению с внутренней энергией системы и решать задачу с помощью хорошо разработанных методов теории возмущений.

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

$$i\hbar \frac{\partial \varphi^{(0)}}{\partial t} = \hat{H}^{(0)} \varphi^{(0)}, \quad (2.36)$$

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 (2.36) являются функции вида

$$\varphi^{(n)}(\mathbf{r}, t) = e^{-\frac{i}{\hbar} E_n t} \psi_n^{(0)}(\mathbf{r}), \quad (2.37)$$

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

$$\hat{H}^{(0)} \psi_n^{(0)} = E_n \psi_n^{(0)}. \quad (2.38)$$

Индекс n здесь указывает квантовое число или набор квантовых чисел.

Состояние системы, описываемое волновой функцией $\varphi^{(n)}(\mathbf{r}, t)$ (2.37), 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:

$$|\varphi^{(n)}(\mathbf{r}, t)|^2 = \varphi^{(n)} \varphi^{(n)*} = |\psi_n^{(0)}(\mathbf{r})|^2. \quad (2.39)$$

Пусть теперь на систему действует какое-то 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

$$i\hbar \frac{\partial \varphi(\mathbf{r}, t)}{\partial t} = (\hat{H}^{(0)} + V) \varphi(\mathbf{r}, t). \quad (2.40)$$

Функции $\varphi^{(n)}$ вида (2.37) уже не будут удовлетворять этому уравнению, решение которого удобно искать в форме

$$\varphi(\mathbf{r}, t) = \sum_n C_n(t) \varphi_n^{(0)}(\mathbf{r}, t) = \sum_n C_n(t) e^{-\frac{i}{\hbar} E_n t} \psi_n^{(0)}(\mathbf{r}), \quad (2.41)$$

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

$$\sum_n |C_n|^2 = 1, \quad (2.42)$$

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 (2.41), 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 :

$$W_{ba}(t) = |C_b(t)|^2 \quad (2.43)$$

при условиях

$$\begin{aligned} |C_a(0)|^2 &= 1, \\ |C_{n \neq a}(0)| &= 0. \end{aligned} \quad (2.44)$$

Таким образом, определение вероятности перехода W_{ba} из состояния a в состояние b сводится к вычислению амплитуды $C_b(t)$ с начальными условиями (2.44).

Для нахождения C_b обычно пользуются теорией возмущений. Подставляя (2.41) в (2.40), с учетом (2.37) получаем

$$\sum_n \psi_n^{(0)} \frac{\partial C_n}{\partial t} = -\frac{i}{\hbar} \sum_n C_n \hat{V} \psi_n^{(0)}. \quad (2.45)$$

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

$$\frac{\partial C_m}{\partial t} = -\frac{i}{\hbar} \sum_n V_{mn} e^{i\omega_{mn}t} C_n, \quad (2.46)$$

где

$$V_{mn} = \int \psi_m^{(0)*} \hat{V} \psi_n^{(0)} d\tau \quad (2.47)$$

— матричный элемент возмущения между состояниями m и n . Частота $\omega_{mn} = (E_m - E_n)/\hbar$ в (2.46) есть боровская частота перехода $m \rightarrow n$. Физический смысл матричных элементов поясним в § 2.4, сопоставляя классический и квантовый осцилляторы.

Систему уравнений (2.46) удобно решать методом последовательных приближений. Сначала подставляем в правую часть исходные значения коэффициентов C_n (2.44)

и находим уравнение для первого приближения:

$$\frac{\partial C_m^{(1)}}{\partial t} = -\frac{i}{\hbar} \sum_n V_{mn} e^{i\omega_{mn}t} C_k^{(0)}, \quad (2.48)$$

откуда

$$C_m^{(1)}(t) = -\frac{i}{\hbar} \int_0^t V_{mn}(\tau) e^{i\omega_{mn}\tau} d\tau. \quad (2.49)$$

Подставляя это первое приближение $C_m^{(1)}(t)$ в правую часть уравнения (2.46), находим второе приближение и т. д. В результате при достаточно больших t (по сравнению с периодом колебаний) получаем следующее выражение для коэффициента $C_b(t)$:

$$C_b(t) = \frac{M_{ba}}{\hbar\omega_{ba}} (1 - e^{i\omega_{ba}t}), \quad (2.50)$$

где

$$M_{ba} = V_{ba} + \sum_{n_1} \frac{V_{bn_1} V_{n_1a}}{(E_a - E_{n_1})} + \sum_{n_1, n_2} \frac{V_{bn_2} V_{n_2n_1} V_{n_1a}}{(E_a - E_{n_1})(E_a - E_{n_2})} + \dots \quad (2.51)$$

— матричный элемент перехода. Первому слагаемому в (2.51) 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 (2.43) и (2.50) вероятность перехода

$$W_{ba} = \frac{|M_{ba}|^2}{\hbar^2} \frac{\sin^2(\omega_{ba}t/2)}{(\omega_{ba}/2)^2}. \quad (2.52)$$

Рассмотрим функцию $f(x) = \sin^2 xt/x^2$, где $x = \omega_{ba}/2$. График этой функции изображен на рис. 2.3. Амплитуда в максимуме при $x=0$ пропорциональна t^2 , а ширина пика — $1/t$. Поэтому при больших t функцию $f(x)$ можно заменить дельта-функцией $\delta(x)$:

$$f(x) = A\delta(E_b - E_a). \quad (2.53)$$

Постоянный множитель A определяется из условия нормировки

$$\int_0^\infty f\left(\frac{\omega_{ba}}{2}\right) dE_b = \int_0^\infty A\delta(E_b - E_a) dE_b. \quad (2.54)$$

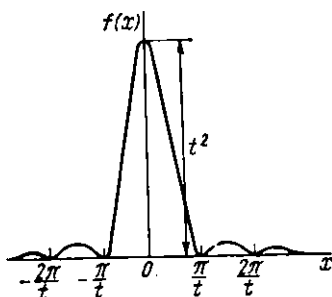


Рис. 2.3. График функции $f(x) = \sin^2 xt/x^2$

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

$$A = 2\pi\hbar t \quad (2.55)$$

и вероятность перехода

$$W_{ba} = \frac{2\pi}{\hbar} |M_{ba}|^2 \delta(E_b - E_a) t. \quad (2.56)$$

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

$$w_{ba} = \frac{2\pi}{\hbar} |M_{ba}|^2 \delta(E_b - E_a). \quad (2.57)$$

Это важное соотношение связывает вероятность квантового перехода из состояния a в состояние b под воздействием возмущения V с соответствующим матричным элементом (2.51). 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 (2.51). 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 (2.51) сопоставим график

$$\begin{array}{c} a \qquad \qquad b \\ \longrightarrow \bullet \longrightarrow \end{array} = V_{ba}. \quad (2.58)$$

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

$$\begin{array}{c} a \qquad n_i \qquad b \\ \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \end{array} = \sum \frac{V_{bn_i} V_{n_i a}}{n_i (E_a - E_{n_i})}. \quad (2.59)$$

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

$$\begin{array}{c} a \quad n_1 \quad n_2 \quad b \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \hline \sum_{n_1, n_2} \frac{V_{bn_2} V_{n_2 n_1} V_{n_1 a}}{(E_a - E_{n_1})(E_a - E_{n_2})} \end{array} \quad (2.60)$$

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 (2.58) — (2.60), этими правилами являются следующие:

1. Каждой точке соответствует матричный элемент вида (2.47), например, переходу $\overline{n_1 \bullet n_2}$ — матричный элемент возмущения между состояниями n_1 и n_2 , т. е.

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

2. Каждому участку между точками, например $\overline{\bullet \quad n_1 \quad \bullet}$ 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. 2.4 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

$$\begin{array}{c} a \quad n_1 \quad b \\ \bullet \quad \bullet \quad \bullet \\ \hline \sum_{n_1} V_{n_1 a} \frac{1}{E_{n_1} - E_a} V_{b n_1} \end{array}$$

Рис. 2.4. Диаграмма взаимодействия, изображающая квантовый переход из состояния a в состояние b через виртуальное состояние n_1 и соответствующие ей аналитические выражения

b , т. е. переход $a \rightarrow b$, возможен, если матричный элемент возмущения V_{ba} между этими состояниями не равен нулю и не равна нулю δ -функция в (2.57), т. е. если выполняется закон сохранения энергии. В этом случае, т. е. при $V_{ba} \neq 0$, остальные слагаемые в (2.51) 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 (2.51), т. е. переходы через промежуточные, виртуальные состояния. Как видно из (2.57), для осуществления перехода из состояния a в состояние b через виртуальное

состояние n_1 , т. е. для перехода $a \rightarrow n_1 \rightarrow b$, закон сохра-

нения 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, называемого виртуальным переходом, требуется только отличие от нуля соответствующего матричного элемента. В виртуальных состояниях система может существовать лишь очень короткое время δt , определяемое соотношением неопределенности

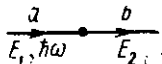
$$\delta t \delta E \approx \hbar, \quad (2.61)$$

где δE — неопределенность энергии.

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 состояний не выполняются обычные соотношения между энергией, импульсом и массой. Возможность существования таких состояний является чисто квантовомеханическим 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 пол-

ная 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

$$\begin{array}{c}
 \text{~~~~~}\hbar\omega \\
 \diagdown \\
 \text{-----} \\
 \text{-----} \quad E_2
 \end{array}
 \quad (2.62)$$

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 ($V_{E_2, E_1}^{e-\text{фот}}$), отличен от нуля.

Аналогичные результаты получаются, если электромагнитное поле представить в виде электромагнитной волны. Как указывалось, любую электромагнитную волну можно представить в виде совокупности линейно поляризованных монохроматических волн, чему соответствует разложение в интеграл Фурье (2.26). 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 собственными частотами, равными боровским частотам ω_{mn} . При воздействии внешнего переменного

возмущения возбуждаются только те осцилляторы, частоты которых совпадают с частотами, присутствующими во внешнем воздействии.

§ 2.3. 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 практически важных случаев процессы испускания и поглощения фотонов можно рассматривать менее строго, опираясь на полуклассическую теорию излучения, развитую Эйнштейном еще в 1917 г., и на теорию квантовых переходов. Такое представление основывается на законах сохранения энергии (2.31) и импульса (2.32) при взаимодействии квантовых систем с электромагнитным излучением.

Из § 2.2 следует, что непосредственный (разрешенный) в первом приближении 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 энер-

гии (2.31) следуют три возможных процесса взаимодействия:

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

Фотоны с энергией $\hbar\omega \neq E_m - E_n$ во взаимодействии не участвуют, поэтому можно положить $\hbar\omega = 0$. 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*.

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

$\hbar\omega' = 0$, т. е. 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 (2.31) следует, что в этом случае фотон $\hbar\omega$ не исчезает, а появляется дополнительный фотон $\hbar\omega' = \hbar\omega$, т. е. в правой части соотношения (2.31) 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*.

На перечисленных трех элементарных процессах взаимодействия электромагнитного излучения с квантовыми системами (схематически они изображены на рис. 2.5, 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 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 воз-

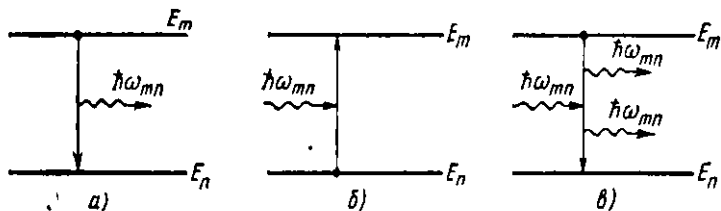


Рис. 2.5. Квантовые переходы при взаимодействии с фотоном:
 а — спонтанный переход с испусканием фотона; б — вынужденный переход с поглощением фотона; в — вынужденный (индуцированный) переход с испусканием фотона

бужденного 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 (1.8), 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 величиной E_0 в соотношении (1.34)]. Управлять указанным процессом невозможно, и в этом смысле спонтанный переход происходит самопроизвольно, случайно во времени, аналогично радиоактивному распаду. 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

$$dW_{mn}^{\text{спонт}} = A_{mn} dt, \quad (2.63)$$

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

Величина, обратная A_{mn} , называется *спонтанным временем жизни* для переходов $m \rightarrow n$, т. е. $\tau_{\text{спонт}} = 1/A_{mn}$. Значение коэффициента Эйнштейна A_{mn} для переходов в оптическом диапазоне может изменяться от $\sim 10^8$ с $^{-1}$ для

разрешенных) наиболее вероятных (переходов до $\sim 1 \text{ с}^{-1}$ 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 направленно, 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 $\omega_{mn} = (E_m - E_n)/\hbar$ (рис. 2.5, б), обусловленная переходом $n \rightarrow m$, за интервал времени dt

$$dW_{nm}^{\text{погл}} = B_{nm}\rho(\omega) dt, \quad (2.64)$$

где B_{nm} — коэффициент Эйнштейна для вынужденных (индуцированных) переходов с поглощением; $\rho(\omega)$ — спектральная плотность излучения [см. (2.28)].

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 нижнее состояние (рис. 2.5, 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. $\rho(\omega)$:

$$dW_{mn}^{\text{исп}} = B_{mn}\rho(\omega) dt. \quad (2.65)$$

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:

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

а число квантов света, испущенных в результате индуцированного испускания:

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

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

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

или

$$N_m (A_{mn} + B_{mn} \rho(\omega)) = N_n B_{nm} \rho(\omega). \quad (2.66)$$

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

$$\begin{aligned} N_m &= \frac{N}{\sum} g_m \exp(-E_m/kT), \\ N_n &= \frac{N}{\sum} g_n \exp(-E_n/kT), \end{aligned} \quad (2.67)$$

где N — полное число частиц в системе; g_m и g_n — статистический вес уровней E_m и E_n .

Для невырожденных уровней $g_m = g_n = 1$. Знаком Σ обозначена статистическая сумма:

$$\Sigma = \sum_i g_i \exp(-E_i/kT),$$

где суммирование проводится по всем энергетическим состояниям i . Из (2.67) следует, что

$$\frac{N_m/g_m}{N_n/g_n} = e^{-\left(\frac{E_m - E_n}{kT}\right)}. \quad (2.68)$$

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 (2.67) в (2.66), получаем

$$A_{mn} + B_{mn}\rho(\omega) = \frac{g_n}{g_m} B_{mn}\rho(\omega) e^{\frac{E_m - E_n}{kT}} \quad (2.69)$$

или

$$\frac{A_{mn}}{\rho(\omega)} + B_{mn} = \frac{g_n}{g_m} B_{mn} e^{\frac{E_m - E_n}{kT}}. \quad (2.70)$$

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 Рэлея — Джинса:

$$\rho(\omega) = \frac{\omega^2}{\pi^2 c^3} kT. \quad (2.71)$$

При $T \rightarrow \infty$ $\rho(\omega) \rightarrow \infty$, $\exp\left(\frac{E_m - E_n}{kT}\right) \rightarrow 1$, и из (2.70)

получаем первое соотношение между коэффициентами Эйнштейна для вынужденных переходов:

$$g_m B_{mn} = g_n B_{nm}. \quad (2.72)$$

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 (2.70) получаем

$$\rho(\omega) = \frac{A_{mn}}{B_{mn}} \frac{1}{e^{\frac{E_m - E_n}{kT}} - 1}. \quad (2.73)$$

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

$$\rho(\omega) = \frac{A_{mn}}{B_{mn}} \frac{kT}{(E_m - E_n)}. \quad (2.74)$$

В условиях термодинамического равновесия при $kT \gg \hbar\omega$ спектральное распределение $\rho(\omega)$ должно определяться формулой (2.71). Сравнивая (2.71) и (2.74), видим, что эти выражения будут тождественно равны при выполнении условий

$$E_m - E_n = \hbar\omega \quad (2.75)$$

и

$$\frac{A_{mn}}{B_{mn}} = \frac{\hbar\omega^3}{\pi^2 c^3}. \quad (2.76)$$

Таким образом, условие частот Бора (2.75) вытекает из проведенного рассмотрения. Второе соотношение между коэффициентами Эйнштейна (2.76) позволяет связать коэффициент спонтанного испускания с коэффициентом поглощения, который может быть найден непосредственно из измерений. Соотношения (2.72), (2.76) являются достаточно общими и не зависят от внешних условий (в частности, от температуры) и типа вещества.

Подставляя (2.75), (2.76) в (2.73), получаем известную формулу Планка для спектрального распределения излучения, находящегося в термодинамическом равновесии со стенками сосуда, называемого также спектром излучения абсолютно черного тела *:

$$\rho(\omega) = \frac{\omega^3}{\pi^2 c^3} \frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}} - 1}. \quad (2.77)$$

Обратим внимание, что коэффициент $\omega^3/\pi^2 c^3$ есть число типов колебаний в единичном объеме и в единичном интервале частот для свободного пространства. Величина $1/[\exp(\frac{\hbar\omega}{kT}) - 1]$ представляет собой среднее число фотонов в заданном типе колебаний (напомним, что фотоны подчиняются статистике Бозе — Эйнштейна), а $\hbar\omega/[\exp(\frac{\hbar\omega}{kT}) - 1]$ — среднюю энергию, приходящую на один тип колебаний. В «классической» области при $\hbar\omega \ll kT$ (например, в

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

радиодиапазоне при комнатной температуре) эта средняя энергия сводится к kT .

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

$$d\omega_{mn}^{\text{спонт}} = a_{mn\alpha}(\omega, \Omega) d\omega d\Omega. \quad (2.78)$$

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

$$A_{mn} = \sum_{\alpha=1}^2 \iint_{\omega\Omega} a_{mn\alpha}(\omega, \Omega) d\omega d\Omega. \quad (2.79)$$

Аналогично вводят дифференциальные коэффициенты Эйнштейна для поглощения $b_{nm\alpha}(\omega, \Omega)$ и вынужденного испускания $b_{mn\alpha}(\omega, \Omega)$. Между дифференциальными коэффициентами имеется такая же связь, как и между интегральными коэффициентами Эйнштейна:

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

и

$$g_m b_{mn\alpha}(\omega, \Omega) = g_n b_{nm\alpha}(\omega, \Omega).$$

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 выразится формулой

$$-dN_m(t) = A_{mn} N_m(t) dt, \quad (2.80)$$

поскольку A_{mn} есть вероятность того, что один атом покинет это состояние за единицу времени. Так как $A_{mn} =$

$= \text{const}$, то решение уравнения (2.80) имеет вид

$$N_m(t) = N_m^0 e^{-A_{mn}t} = N_m^0 e^{-t/\tau_{mn}}, \quad (2.81)$$

где $\tau_{mn} = 1/A_{mn}$.

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

$$\langle t \rangle = \frac{1}{N_m^0} \int_0^\infty t dN(t) = \frac{1}{N_m^0} \int_0^\infty t N_m^0 A_{mn} e^{-A_{mn}t} dt = 1/A_{mn} = \tau_{mn}, \quad (2.82)$$

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

Величина $\tau_{mn} = 1/A_{mn}$ выражает время жизни атома в возбужденном состоянии, обусловленное спонтанными переходами $m \rightarrow n$. Она равна среднему времени пребывания атома в возбужденном состоянии.

Из (2.80), (2.81) получаем закон затухания мощности спонтанного излучения:

$$P(t) = P_0 e^{-t/\tau_{mn}}, \quad (2.83)$$

где

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

Кроме оптических излучательных переходов, обусловленных взаимодействием с электромагнитным излучением, возможны неоптические квантовые переходы из одного состояния в другое, которые называются *безызлучательными квантовыми переходами*. Такие переходы, в частности, могут происходить при столкновениях атомов и молекул в газе 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 $\omega_1 = 1/\tau_1, \omega_2 = 1/\tau_2, \dots, \omega_i = 1/\tau_i, \dots$), согласно теореме о сложении вероятностей среднее время жизни атома в состоянии m

$$\frac{1}{\tau_m} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \dots + \frac{1}{\tau_i} + \dots \quad (2.84)$$

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.

§ 2.4. ДИПОЛЬНОЕ ИЗЛУЧЕНИЕ

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 один из коэффициентов, например A_{mn} или B_{nm} , и воспользовавшись полученными в § 2.3 соотношениями между ними, можно без особых затруднений определить интенсивности спонтанного и вынужденного излучения или поглощения. Как следует из теории квантовых переходов (см. § 2.2), для этого необходимо прежде всего определить матричные элементы возмущения (2.47). 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 уравнений Максвелла (2.4) следует, что энергия E , излучаемая за-

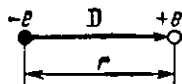


Рис. 2.6. Электрический диполь

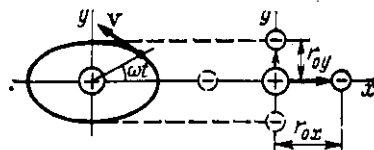


Рис. 2.7. Представление атома в виде двух осцилляторов

рядом e в единицу времени, пропорциональна квадрату ускорения a :

$$\frac{dE}{dt} = \frac{e^2}{6\pi\epsilon_0 c^3} a^2. \quad (2.85)$$

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

$$r(t) = r_m \cos \omega t, \quad (2.86)$$

то из (2.85) получаем мгновенную мощность излучения

$$P = -\frac{dE}{dt} = \frac{e^2 r_m^2 \omega_0^4}{6\pi\epsilon_0 c^3} \cos^2 \omega_0 t. \quad (2.87)$$

Усредненная по времени за период колебаний мощность излучения

$$\langle P \rangle = \frac{e^2 r_m^2 \omega_0^4}{12\pi\epsilon_0 c^3}, \quad (2.88)$$

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 величине 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. 2.6). Если дипольный момент D гармонически изменяется с частотой ω , то такой диполь называют *осциллирующим диполем* или *осциллятором*. Средняя мощность излучения такого осциллятора получается из выражения (2.88):

$$\langle P \rangle = \frac{\omega_0^4}{12\pi\epsilon_0 c^3} D_0^2, \quad (2.89)$$

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. Для иллюстрации этого рассмотрим электрон, движущийся по эллиптической орбите вокруг положительно заряженного ядра. Как видно из рис. 2.7, такое движение можно заменить гармоническими колебаниями двух диполей (линейных осцилляторов), собственные частоты которых равны угловой скорости обращения электрона по орбите, а фазы сдвинуты на угол $\pi/2$. Дипольные моменты таких излучателей равны:

$$\begin{aligned} D_x &= -er_{0x} \cos \omega_0 t, \\ D_y &= -er_{0y} \sin \omega_0 t. \end{aligned}$$

Естественно, можно ограничиться рассмотрением свойств одного из осцилляторов.

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

$$E_0 = \frac{m\omega_0^2 r_0^2}{2}. \quad (2.90)$$

С течением времени такой классический осциллятор передает энергию полю по закону (2.87). Этому соответствует уменьшение амплитуды колебаний r_m . При $t > 0$

$$\langle E(t) \rangle = \frac{m\omega_0^2 r_m^2(t)}{2}. \quad (2.91)$$

Сопоставляя это выражение с (2.88), получаем

$$\langle P \rangle = -\frac{d\langle E \rangle}{dt} = \frac{e^2 \omega_0^2}{6\pi\epsilon_0 c^3 m} \langle E \rangle. \quad (2.92)$$

Вводя обозначения

$$\gamma = \frac{e^2 \omega_0^2}{6\pi\epsilon_0 c^3 m}, \quad (2.93)$$

получаем уравнение для энергии излучателя

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

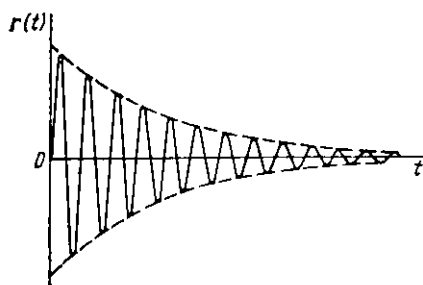


Рис. 2.8. График функции (2.96)

Отсюда

$$\langle E(t) \rangle = E_0 e^{-\gamma t}, \quad (2.94)$$

где $E_0 = E_{t=0}$ определяется соотношением (2.90).

Сопоставляя выражения (2.94) и (2.91), находим изменение амплитуды колебаний классического осциллятора во времени:

$$r_m = r_0 e^{-\frac{\gamma}{2} t}, \quad (2.95)$$

а с учетом (2.86) получаем

$$r(t) = r_0 e^{-\frac{\gamma}{2} t} \cos \omega_0 t. \quad (2.96)$$

График этой функции представлен на рис. 2.8.

Из (2.94) находим, что средняя за период мощность излучения электрического диполя изменяется во времени по экспоненциальному закону:

$$\langle P \rangle = -\frac{d\langle E \rangle}{dt} = \frac{e^2 \omega_0^4}{12\pi e_0 c^3} r_0^2 e^{-\gamma t} = \frac{\omega_0^4}{12e_0 \pi c^3} D_0^2 e^{-\gamma t} = P_0 e^{-\gamma t}, \quad (2.97)$$

где

$$P_0 = \frac{\omega_0^4}{12\pi e_0 c^3} D_0^2.$$

Это выражение полностью совпадает с выражением для мощности спонтанного излучения (2.83), если положить $\gamma = A_{mn}$. 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. 2.8). Для определения спектра излучения $r(t)$ следует представить в виде интеграла Фурье, т. е. функцию (2.96), изображенную на рис. 2.8, представить в виде суммы (интеграла) гармонических состав-

ляющих. Аналогично (2.26) и (2.27) имеем

$$r(t) = \int_{-\infty}^{\infty} r(\omega) e^{i\omega t} dt, \quad (2.98)$$

где амплитуды гармонических составляющих

$$r(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(t) e^{-i\omega t} dt. \quad (2.99)$$

Зависимость (2.96) можно записать в экспоненциальной форме:

$$r(t) = \frac{r_0}{2} e^{-\frac{\gamma}{2}t} (e^{i\omega_0 t} + e^{-i\omega_0 t}).$$

Подставляя это выражение в интеграл Фурье (2.99), 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 $|r(\omega)|^2 = r^*(\omega)r(\omega)$ равна

$$|r(\omega)|^2 = \frac{r_0^2}{4\pi} \frac{1}{(\omega_0 - \omega)^2 + (\gamma/2)^2}. \quad (2.100)$$

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

$$g_L(\omega) = \frac{A}{(\omega_0 - \omega)^2 + (\gamma/2)^2}, \quad (2.101)$$

где A — нормировочный множитель, называется *кривой Лоренца*. Она определяет распределение энергии по частотному спектру. Из условия нормировки $\int_{-\infty}^{\infty} g_L(\omega) d\omega = 1$ определяем нормировочный множитель $A = \gamma/2\pi$.

Функция Лоренца представлена на рис. 2.9. 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

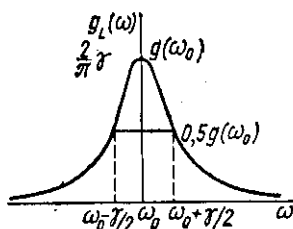


Рис. 2.9. Функция Лоренца

было показано, $\gamma = 1/\tau$, где τ имеет смысл времени релаксации. Отсюда вытекает фундаментальное соотношение спектрального анализа

$$\Delta\omega\tau \approx 1. \quad (2.102)$$

Сравнив его с (2.61), 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. 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

$$D_{nn} = e \int \varphi_n^* \mathbf{r} \varphi_n dV, \quad (2.103)$$

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 \mathbf{r} — радиус-вектор, проведенный из начала координат, в котором помещено ядро атома.

Поскольку зависимость волновой функции состояния от времени имеет вид (2.37), произведение $\varphi_n^*\varphi_n$ не зависит от времени и, следовательно, распределение заряда в атоме, находящемся в стационарном состоянии, не изменяется по времени. 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:

$$\rho_{mn} = e\psi_n^*\psi_m.$$

С учетом (2.37) произведение $\psi_n^*\psi_m$ будет зависеть от времени по закону

$$\psi_n^*\psi_m = \psi_n^*\psi_m e^{i\left(\frac{E_m - E_n}{\hbar}\right)t}. \quad (2.104)$$

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 осцилляции дипольного момента

$$D_{mn}(t) = D_{mn}^0 \exp\left(i\frac{E_m - E_n}{\hbar}t\right). \quad (2.105)$$

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

$$D_{mn}^0 = e \int_V \psi_m^* r \psi_n dV. \quad (2.106)$$

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

$$\begin{vmatrix} D_{11} & D_{12} & \dots & D_{1n} & \dots \\ D_{21} & D_{22} & \dots & D_{2n} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ D_{n1} & D_{n2} & \dots & D_{nn} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix} \quad (2.107)$$

Числа, составляющие матрицу, называются *матричными элементами*. Нondiagonalные 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

$$\begin{vmatrix} 0 & \omega_{12} & \dots & \omega_{1n} & \dots \\ \omega_{21} & 0 & \dots & \omega_{2n} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \omega_{m1} & \omega_{m2} & \dots & \omega_{mn} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix} \quad (2.108)$$

Частоты ω_{ii} , расположенные по диагонали, тождественно равны нулю. Им соответствуют диагональные элементы D_{ii} матрицы (2.107), 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 (2.89) или (2.97), 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

$$\langle P \rangle = \frac{\omega_{mn}^4 N_m}{3\pi\epsilon_0 c^3} (D_{mn})^2. \quad (2.109)$$

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

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

имеем

$$A_{mn} = \frac{\omega_{mn}^3}{3\pi\epsilon_0 \hbar c^3} (D_{mn})^2. \quad (2.110)$$

Это выражение соответствует излучению осциллятора с дипольным моментом

$$D_{mn} = 2er_{mn} \cos \omega_{mn} t,$$

где

$$r_{mn} = \int \psi_m^* r \psi_n dV.$$

Воспользовавшись соотношениями (2.72) и (2.76) между коэффициентами Эйнштейна, получим

$$B_{mn} = \frac{\pi}{12\hbar^2 \epsilon_0} (D_{mn})^2. \quad (2.111)$$

Соотношения (2.110) и (2.111) 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. § 2.2). При этом необходимо учитывать, что вынужденный переход (например, определяющий поглощение) вызывается взаимодействием между электрическим полем электромагнитной волны и электрическим дипольным моментом атома. Тогда возмущение

$$V = e\mathcal{E}(r, t)r.$$

Вспомогательное, поскольку длина электромагнитной волны существенно больше размеров атома, можно считать, что $\mathcal{E}(r, t) \approx \mathcal{E}(0, t)$ и

$$\begin{aligned} V_{mn} &= e \int \psi_m^* \mathcal{E} r \psi_n dV \approx \\ &\approx e\mathcal{E}(0, t) \int \psi_m^* r \psi_n dV = \mathcal{E}(0, t) D_{mn}. \end{aligned} \quad (2.112)$$

Подставляя это выражение в (2.49), с учетом (2.57), (2.63) получаем формулу (2.110).

Проведем оценку, например, коэффициента Эйнштейна A_{mn} . Дипольный момент D_{mn} имеет порядок величины ea_B . Положив $a_B \approx 1 \text{ \AA} = 10^{-10} \text{ м}$, по формуле (2.110) находим для видимой области спектра ($\lambda \approx 0,5 \text{ мкм}$, $\omega = 2\pi\nu = \frac{2\pi}{\lambda}c$) $A_{mn} \approx 10^8 \text{ с}^{-1}$.

Таким образом, коэффициенты Эйнштейна, вероятности спонтанных и вынужденных переходов определяются через соответствующие дипольные матричные элементы, которые

являются недиагональными элементами матрицы (2.107). 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 $\psi(\mathbf{r}) = \psi(-\mathbf{r})$, и нечетные, для которых $\psi(\mathbf{r}) = -\psi(-\mathbf{r})$. 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 и т. д.) — нечетными. Дипольный матричный элемент (2.106) отличен от нуля только для уровней с различной четностью [для уровней с одинаковой четностью вклад, получаемый при вычислении подынтегральной части выражения (2.106) в точках \mathbf{r} и $-\mathbf{r}$, одинаков по величине, но противоположен по знаку]. Поэтому дипольные переходы возможны лишь между четными и нечетными уровнями энергии или наоборот.

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:

$$\Delta n = 0; 1; 2; \dots$$

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.

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, т. е. $D_{mn}=0$, 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 приближения (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, называется *метастабильным уровнем*. Время жизни атомов в этом состоянии порядка 10^{-3} с и больше.

§ 2.5. УШИРЕНИЕ СПЕКТРАЛЬНЫХ ЛИНИЙ

При рассмотрении энергетических состояний квантовых систем в гл. I 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 фиксиро-

ванной частоте, т. е. идеальная монохроматическая волна. В действительности подобная ситуация невозможна хотя бы по той причине, что время жизни τ в возбужденном состоянии конечно, а в силу соотношения неопределенностей $\delta E \tau \approx \hbar$ это приводит к неопределенности в энергии состояния, т. е. к «размытию» энергетического уровня на величину порядка δE . Для изолированных атомов и молекул это уширение очень мало ($\delta E \ll E$), но принципиально 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. Исследования энергетического положения, формы и интенсивности спектральных линий в оптическом диапазоне дают ценную информацию о свойствах излучаемых систем. Этими вопросами занимается *оптическая спектроскопия*, являющаяся целым разделом физики.

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:

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

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.

1. 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 спектраль-

ной линии, определяемая конечным временем жизни в возбужденном состоянии, была получена в § 2.4, где было показано, что спектр излучения затухающего осциллятора описывается функцией Лоренца (2.101), или, другими словами, спектральная линия имеет лоренцеву форму. Выражение (2.101) перепишем в виде

$$g_L(\omega) = \frac{\Delta\omega}{2\pi} \frac{1}{(\omega_0 - \omega)^2 + (\Delta\omega/2)^2}, \quad (2.113)$$

где $\Delta\omega = \gamma = A_{mn}$ — ширина линии на уровне 0,5 от максимума, а ω_0 — резонансная частота. Точный квантово-механический расчет показывает, что естественная форма спектральной линии в точности описывается выражением (2.101), 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 (2.37) можно записать в виде

$$\varphi_m(\mathbf{r}, t) = \psi_m(\mathbf{r}) e^{-\frac{i}{\hbar} E_m t} e^{-t/2\tau}, \quad (2.114)$$

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. 2.8. Поэтому разложение по монохроматическим волнам (разложение Фурье) даст спектр вида (2.101), т. е. функцию Лоренца (2.113). 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_{mn} , тем больше естественная ширина спектральной линии. Типичное значение естественной ширины линии для разрешенных в дипольном приближении переходов в видимой области спектра $\Delta\omega = A_{mn} \approx 10^8 \text{ с}^{-1}$, 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_{mn} пропорционален кубу частоты [см. выражение (2.110)], то в радиодиапазоне естественная ширина линии существенно меньше, чем в

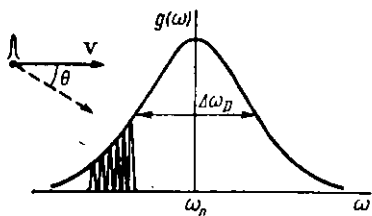


Рис. 2.10. Доплеровское уширение спектральной линии

оптическом диапазоне. Например, для газообразного аммиака NH_3 спектральная линия перехода с частотой $\nu_0 = 24\,870\text{ МГц}$ ($\lambda_0 = 1,25\text{ см}$) 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 (2.93), получим

$$\Delta\lambda = \frac{e^2}{3\varepsilon_0 c^2 m} \approx 1,2 \cdot 10^{-14}\text{ м},$$

что приблизительно равно двум классическим диаметрам электрона d_e .

2. Доплеровское уширение. Эффект Доплера есть изменение частоты (длины волны), наблюдаемое при движении источника волн 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. 2.10), воспринимаемая частота излучения

$$\omega = \omega_0 \frac{\sqrt{1-v^2/c^2}}{1 - \frac{v}{c} \cos \theta}. \quad (2.115)$$

При $v \ll c$ (c — скорость света)

$$\omega = \omega_0 \left(1 + \frac{v}{c} \cos \theta \right). \quad (2.115a)$$

Частота излучения увеличивается, если источник приближается к наблюдателю, и уменьшается, если он уда-

ляется 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. 2.10. 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):

$$f(v) dv = \frac{1}{V \pi v_0} \exp \left[- \left(\frac{v}{v_0} \right)^2 \right] dv, \quad (2.116)$$

где $v_0 = \sqrt{2kT/M}$ — наиболее вероятная скорость движения частиц в газе, и учитывая выражение (2.115а), можно получить форму спектральной линии, обусловленную доплеровским уширением:

$$g_G(\omega) = \frac{2 \sqrt{\ln 2}}{V \pi \Delta\omega_D} \exp \left\{ - \left[\frac{2 \sqrt{\ln 2} (\omega - \omega_0)}{\Delta\omega_D} \right]^2 \right\}. \quad (2.117)$$

Здесь

$$\Delta\omega_D = 2 \sqrt{\ln 2} \frac{\omega_0}{c} \sqrt{\frac{2kT}{M}}$$

— ширина доплеровской линии на уровне 0,5 от максимального значения при $\omega = \omega_0$.

Функция (2.117) есть функция Гаусса, поэтому говорят, что линия, уширенная за счет эффекта Доплера, имеет гауссову форму. Множитель перед экспонентой определяется условием нормировки.

На рис. 2.11 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 $\Delta\nu_D = \Delta\omega_D/2\pi \approx 3$ ГГц. В радио-

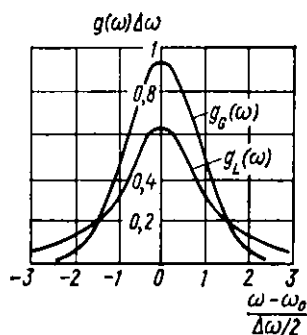


Рис. 2.11. Сравнение лоренцевой $g_L(\omega)$ и гауссовой $g_G(\omega)$ форм спектральных линий при одинаковой их ширине на уровне 0,5

диапазоне, например, для линии аммиака $\nu_0 = 24\,870$ МГц, при комнатной температуре доплеровское уширение $\Delta\nu_D \approx 70$ кГц, т. е. в миллионы раз превышает естественную ширину линии.

3. 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 $\Delta\omega = 1/\tau_{mn} = A_{mn}$ в (2.113) должно стоять $\Delta\omega = 1/\tau_p$, где τ_p — время релаксации, определяемое процессами столкновений.

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

$$\tau_p \approx \tau_{ст} \approx \frac{V \sqrt{MkT}}{16 \sqrt{\pi} a_B^2 p}, \quad (2.118)$$

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

Отметим, что величина $\tau_{ст}$ 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_p может составлять 10^{-9} — 10^{-11} с, приводя к существенным уширениям соответствующих спектральных линий.

4. Уширение за счет влияния внутренних (внутрикристаллических) и внешних электрического и магнитного полей. Этот механизм уширения определяется рассмотренными в § 1.4 эффектами Зеемана и Штарка. Если величины зеемановского или штарковского расщепления 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, относится 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 $\mathcal{E}_{кр}$ 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 $g_L(\omega)$ (2.113), а форма неоднородно уширенных линий — функцией Гаусса $g_G(\omega)$ (2.117). 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:

$$g(\omega) = g_1(\omega) * g_2(\omega) = \int_{-\infty}^{\infty} g_1(\omega') g_2(\omega - \omega') d\omega' \quad (2.119)$$

(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 (2.84).

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 имеют один порядок величины, то задача определения формы линии сводится к нахождению свертки лоренцевой и гауссовой функций (соответствующие интегралы табулированы и в математике известны как интегралы Воята).

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.

$$a_{mn}(\omega) = A_{mn} g(\omega). \quad (2.120)$$

Аналогично для остальных коэффициентов Эйнштейна имеем:

$$b_{mn}(\omega) = B_{mn} g(\omega), \quad (2.120a)$$

$$b_{nm}(\omega) = B_{nm} g(\omega), \quad (2.120b)$$

где A_{mn} , B_{mn} и B_{nm} — интегральные коэффициенты Эйнштейна, определяемые через дипольные матричные элементы соотношениями (2.110) и (2.111).

Для однородного уширения функция $g(\omega)$ имеет лоренцеву форму (2.113), а для неоднородного — гауссову форму (2.117). В общем случае $g(\omega)$ дается сверткой (2.119).

§ 2.6. РАССЕЯНИЕ СВЕТА И ДВУХФОТОННОЕ ПОГЛОЩЕНИЕ

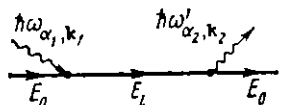
Ранее всюду принималось, что взаимодействие света с квантовой системой возможно только в случае соблюдения правила частот Бора (1.1). В частности, предполагалось, что поглощение света имеет место лишь тогда, когда энергия квантов падающего излучения $\hbar\omega$ равна разнице энергий основного E_0 и одного из возбужденных E_m состояний системы в пределах точности, допускаемой соотношением неопределенности.

Рассмотрим теперь взаимодействие атомов и молекул с фотонами, когда правило частот Бора не выполняется даже приближенно. Как следует из теории квантовых переходов (см. § 2.2), 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 может все же быть передан от поля излучения атому, который перейдет в виртуальное состояние (см. § 2.2). 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 поляризации 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:



$$(2.121)$$

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_I . 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 (2.31) и импульса (2.32). Но они должны соблюдаться для процесса в целом. При переходе в виртуальные состояния, как отмечалось в § 2.2, закон сохранения энергии не соблюдается. Для осуществления такого перехода необходимо лишь, чтобы матричный элемент электрон-фотонного взаимодействия $V_{E_I, E_0}^{e-фот}$ 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 (2.31) и (2.32) для процесса (2.121) 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 (2.121) $\hbar\omega \approx \hbar\omega'$, 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 *комптоновским 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

$$(2.122)$$

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. 2.12, а — в.

При комбинационном рассеянии могут быть два случая: 1) энергия начального состояния (обычно основного состояния E_0) меньше энергии конечного (возбужденного) состояния (рис. 2.12, б). При этом частота рассеянного

* Эффект комбинационного рассеяния света наблюдался одновременно советскими 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*.

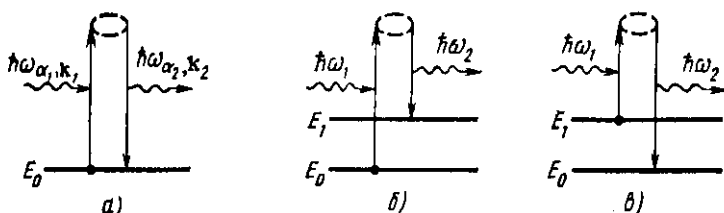


Рис. 2.12. Энергетические диаграммы, поясняющие эффекты рассеяния: а — рэлеевское рассеяние; б, в — комбинационное рассеяние (б — стоксова компонента; в — антистоксова компонента)

фотона смещена в область меньших частот на величину $\Delta\omega = (E_1 - E_0)/\hbar$ от частоты падающего фотона. Такое смещение в область меньших частот называется *стоксовым смещением*; 2) энергия начального состояния больше энергии конечного состояния (рис. 2.12, в). Этому процессу соответствует диаграмма взаимодействия:

$$(2.123)$$

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 температуры. Стоксовы и антистоксовы компоненты в спектре рассеяния будут располагаться симметрично относительно возбуждающей линии, как показано на рис. 2.13. 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

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

произведению вероятностей переходов из начального состояния в промежуточное и из промежуточного состояния в конечное, как следует из диаграмм взаимодействия (2.121) — (2.123). 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. 2.12, б и на диаграмме взаимодействия (2.122), не будет зависеть от матричного элемента перехода $V_{E_1, E_0}^{e\text{-фот}}$ между состояниями E_1 и E_0 . Согласно диаграмме (2.122) она будет определяться лишь произведением матричных элементов $V_{E_1, E_0}^{e\text{-фот}} V_{E_0, E_1}^{e\text{-фот}}$. Поэтому комбинационное рассеяние света является удобным методом изучения состояний, непосредственные оптические переходы в которые запрещены правилами отбора.

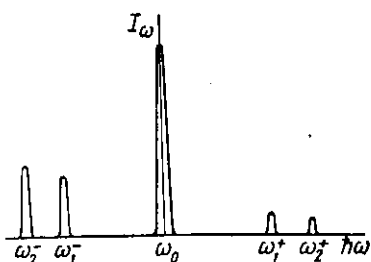


Рис. 2.13. Спектр рассеяния: ω_0 — частота возбуждающего излучения и резонансного рассеяния; ω_1^- , ω_2^- — стоксовы, ω_1^+ , ω_2^+ — антистоксовы компоненты в спектре комбинационного рассеяния

Кроме рассмотренных процессов рассеяния во втором приближении теории возмущений возможны процессы так называемого двухфотонного поглощения света. Сущность этих процессов сводится к следующему. Поглощая фотон $\hbar\omega_1$, атом переходит в виртуальное состояние E_1 , в котором он может успеть поглотить второй фотон $\hbar\omega_2$. Если для суммы энергий этих двух фотонов выполняется правило частот Бора ($\hbar\omega_1 + \hbar\omega_2 = E_m - E_n$), то атом перейдет из одного стационарного состояния E_n в другое E_m . Отвечающие этому процес-



Рис. 2.14. Двухфотонное поглощение света:

а — диаграмма взаимодействия; б — энергетическая диаграмма

су диаграмма взаимодействия и энергетическая диаграмма показаны на рис. 2.14, а, б. В частном случае возможно равенство $\hbar\omega_1 = \hbar\omega_2 = (E_m - E_n)/2$. Вероятность двухфотонных процессов пропорциональна квадрату ин-

тенсивности света и обычно очень малá.

Подчеркнем, что виртуальные энергетические состояния, изображенные на рис. 2.12 и 2.14 пунктирными линиями, на самом деле не существуют и показаны лишь для наглядности. Рассмотренные процессы рассеяния и двухфотонного поглощения 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.

ГЛАВА 3. УСИЛЕНИЕ И ГЕНЕРАЦИЯ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

Из § 2.3 непосредственно следует, что при вынужденном испускании появляется фотон, неразличимый с фотоном, индуцировавшим соответствующий квантовый переход. Этот процесс может служить основой для усиления и генерации оптического излучения, т. е. для создания лазеров.

§ 3.1. 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. 3.1.

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, определяемый соотношениями (2.64) и (2.120б). 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 соотношениями (2.65) и (2.120 а). Сечение слоя в плоскости, перпендикулярной оси z , для простоты примем равным единице.

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

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

Здесь учтена связь (2.120) 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 (2.72), получаем окончательно

$$dI_\omega = \frac{\hbar\omega n B_{12} g(\omega)}{c} \left(\frac{g_1}{g_2} N_2 - N_1 \right) I_\omega dz. \quad (3.1)$$

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 процес-

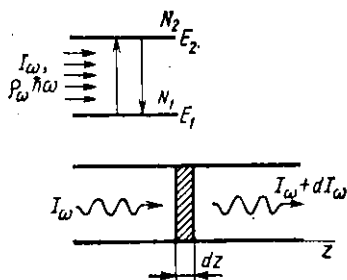


Рис. 3.1. Прохождение плоской электромагнитной волны через слой вещества толщиной dz

сов вынужденного испускания, над всеми возможными потерями. Если частота перехода ω_{21} лежит в оптическом диапазоне, то соответствующий усилитель называется *лазерным*, если в СВЧ-диапазоне — *лазерным* (см. терминологию на с. 6).

Обозначим коэффициент пропорциональности в (3.1), стоящий перед $I_\omega dz$, через $-k_\omega$:

$$k_\omega = \frac{\hbar \omega n B_{12}}{c} g(\omega) \left(N_1 - \frac{g_1}{g_2} N_2 \right). \quad (3.2)$$

Если форма спектральной линии $g(\omega)$ не зависит от интенсивности электромагнитной волны I_ω , то для однородной среды, где населенности уровней 1 и 2 не зависят от координаты, величину k_ω можно считать постоянной. Тогда выражение (3.1) приобретает вид

$$-dI_\omega/I_\omega = k_\omega dz. \quad (3.3)$$

Элементарное интегрирование этого уравнения дает экспоненциальный закон изменения интенсивности:

$$I_\omega(z) = I_\omega(0) e^{-k_\omega z}. \quad (3.4)$$

Это есть закон Бугера — Ламберта. Коэффициент k_ω называется *натуральным показателем поглощения на частоте ω* или просто *показателем поглощения* *. Обычно он измеряется в см^{-1} и численно равен обратной величине расстояния, на котором интенсивность волны изменяется в e раз. При $N_1/g_1 > N_2/g_2$ показатель поглощения положителен и интенсивность волны уменьшается с расстоянием по экспоненциальному закону, как показано на рис. 3.2. При инверсии населенностей, т. е. при $N_1/g_1 < N_2/g_2$, показатель поглощения отрицателен и интенсивность волны возрастает по экспоненциальному закону (кривая 3 на рис. 3.2). Поэтому 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 функ-

* Часто коэффициент k_ω называют коэффициентом поглощения, а показателем поглощения называют безразмерный коэффициент κ (см. гл. 4), который связан с k_ω соотношением (4.9). В настоящей книге мы будем придерживаться терминологии, установленной ГОСТ 7601—78 «Физическая оптика. Термины и определения», и коэффициент k_ω в (3.4) называть показателем поглощения, а κ — главным показателем поглощения.

ция $g(\omega)$ нормирована

$$\left(\int g(\omega) d\omega = 1\right),$$

$$K = \int k_{\omega} d\omega = \frac{\hbar \omega_0 \pi B_{12}}{c} \left(N_1 - \frac{g_1}{g_2} N_2 \right). \quad (3.5)$$

Здесь учтено, что ширина спектральной линии мала и в пределах спектральной линии ω изменяется слабо по сравнению с $g(\omega)$, поэтому при интегрировании выражения (3.2) множитель $\hbar \omega$ можно вынести из-

под интеграла. Величина $K = \int k_{\omega} d\omega$ называется *интегральным показателем поглощения* для данного типа переходов.

Очевидно, максимальное усиление будет наблюдаться на частоте $\omega = \omega_0$ вблизи максимума спектральной линии. Тогда для лоренцевой формы спектральной линии (2.113) имеем следующий показатель усиления в максимуме спектральной линии при $\omega = \omega_{\text{ли}} = \omega_0$:

$$\alpha_{\text{max}}(\omega_0) = -k_{\text{max}}(\omega_0) = -\frac{2}{\pi \Delta \omega} K = \frac{2 \hbar \omega_0 \pi B_{12}}{\pi c \Delta \omega} \left(\frac{g_1}{g_2} N_2 - N_1 \right), \quad (3.6)$$

где $\Delta \omega$ — ширина линии на уровне 0,5 от максимума, определяемая процессами уширения (см. § 2.5).

Пусть система находится в условиях термодинамического равновесия. Тогда населенности энергетических уровней определяются распределением Больцмана (2.67), 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

$$k_{\omega} = \frac{\hbar \omega \pi B_{12}}{c} g(\omega) N \quad (3.7)$$

пропорционален числу поглощающих центров в единице объема. Поэтому величина $\sigma(\omega) = k_{\omega}/N$ характеризует способность каждой из частиц поглощать электромагнитную

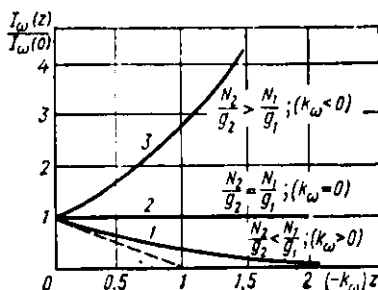


Рис. 3.2. Изменение интенсивности волны при прохождении через среду в случаях поглощения (1), просветления (2), усиления (3)

энергию на частоте ω . Она имеет размерность площади и называется *поперечным сечением поглощения* или просто сечением поглощения. Для данного типа переходов $n \rightarrow m$ величина σ связана с коэффициентами Эйнштейна:

$$\sigma_{nm}(\omega) = \frac{\hbar \omega n}{c} b_{nm}(\omega) = \frac{\hbar \omega n}{c} B_{nm} g(\omega). \quad (3.8)$$

По аналогии с интегральным показателем поглощения (3.5) вводят *интегральное сечение поглощения*:

$$\kappa = \int \sigma_{nm}(\omega) d\omega = \frac{\hbar \omega n}{c} B_{nm}. \quad (3.9)$$

Таким образом, система, находящаяся в термодинамическом равновесии, не может усиливать электромагнитное излучение. Для получения усиления необходимо создать в среде инверсию населенностей, для чего принципиально важно 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 среды (активного вещества) с целью получения инверсии населенностей называют *накачкой*, а источник этого возбуждения — *источником накачки*.

Общая схема квантового усилителя представлена на рис. 3.3. Источник 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 вы-

ходе 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 $\alpha_\omega = -k_\omega$), экспонен-

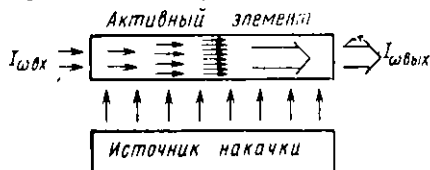


Рис. 3.3. Общая схема квантового усилителя

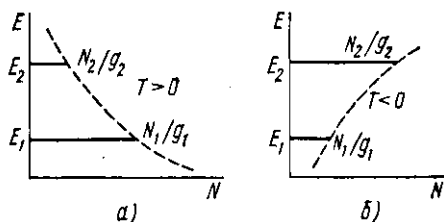


Рис. 3.4. Населенности энергетических уровней:

а — в состоянии термодинамического равновесия; б — в состоянии инверсии населенностей (с отрицательной температурой)

циально возрастает согласно (3.4) в соответствии с длиной активного элемента и степенью инверсии ($N_2/g_2 - N_1/g_1$), которая определяется числом активных частиц в единице объема вещества и интенсивностью накачки.

Состояние вещества с инверсией населенностей иногда называют

состоянием с отрицательной температурой. Смысл этого понятия поясняет рис. 3.4. Распределение атомов по энергетическим уровням в равновесном состоянии (рис. 3.4, а) определяется статистикой Больцмана. Чисто формально состояние с инверсией населенностей (рис. 3.4, б) можно описать, используя соотношение (2.68), 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.

Для превращения усилителя в генератор необходимо, как обычно, ввести 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. 3.5. Плоская электромагнитная волна, распространяющаяся в на-

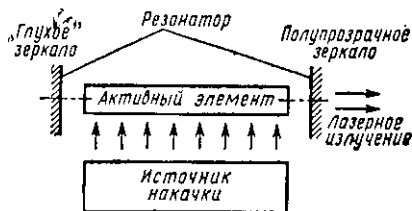


Рис. 3.5. Общая схема лазера

правлении, 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 (3.2) зависит от степени перенаселенности рабочих уровней, а она в свою очередь определяется интенсивностью накачки, то существует некоторая критическая (пороговая) мощность накачки, при которой начинает развиваться процесс генерации. Пороговой мощности накачки соответствует критическая инверсия $N_2/g_2 - N_1/g_1$. По достижении критической инверсии генерация развивается 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 в оптическом резонаторе 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, совпадающих с собственными частотами резонатора.

§ 3.2. ВОЗБУЖДЕНИЕ АКТИВНОГО ВЕЩЕСТВА (НАКАЧКА)

Как следует из § 3.1, усиление электромагнитных колебаний с помощью вынужденного излучения возможно лишь в сре-

де с инверсией населенностей энергетических уровней. 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 электромагнитным излучением, называемым вспомогательным излучением или излучением накачки. Это излучение выбирают таким образом, 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, источники солнечного излучения 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 мазерах. 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. Невозбужденные 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 позво-

ляет непосредственно, без промежуточных процессов, преобразовывать электрическую энергию источника в когерентное электромагнитное излучение лазера. Для данного метода характерны процессы диффузии и дрейфа носителей заряда в полупроводниковых гомо- и гетеро-*p-n*-переходах, смещенных в прямом направлении.

5. Возбуждение частицами высоких энергий, например ускоренными электронами. Такой метод используется в полупроводниковых лазерах с электронной накачкой. 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; в ней 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 равновесное 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

$$\frac{dN_m}{dt} = \sum_{n \neq m} N_n w_{nm} - N_m \sum_{n \neq m} w_{mn}. \quad (3.10)$$

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 опре-

деляться различными излучательными и безызлучательными элементарными процессами. Обозначив скорость безызлучательных переходов S_{mn} , получим

$$\omega_{mn} = A_{mn} + B_{mn}\rho_{mn} + S_{mn}, \quad (3.11)$$

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

Величина $B_{mn}\rho_{mn}$ определяет 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 (3.11), как и во многих последующих формулах, вероятность $B_{mn}\rho_{mn}$ представляет собой сокращенную запись интеграла

$$B_{mn}\rho_{mn} = \int b_{mn}(\omega) \rho_{mn}(\omega) d\omega, \quad (3.12)$$

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 величину $\rho_{mn}(\omega)$ можно вынести из-под знака интеграла. Тогда получится обычное выражение для вероятности индуцированных переходов:

$$\int b_{mn}(\omega) \rho_{mn}(\omega) d\omega = \rho_{mn}(\omega_{mn}) \int b_{mn}(\omega) d\omega = \rho_{mn}(\omega_{mn}) B_{mn}, \quad (3.13)$$

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

$$\begin{aligned} \int b_{mn}(\omega) \rho_{mn}(\omega) d\omega &= b_{mn}(\omega_r) \int \rho_{mn}(\omega) d\omega \approx \\ &\approx b_{mn}(\omega_{ген}) \rho_{mn}(\omega_r) \Delta\omega_{ген}^{ген}, \end{aligned} \quad (3.14)$$

где $\omega_{ген}$ — 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 $\Delta\omega_{ген}^{ген}$.

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

$$\sum_{m=1}^K N_m = N \quad (3.15)$$

(суммирование проводится по всем K энергетическим уровням).

Для каждого уровня может быть записано свое уравнение баланса, так что получится система из K уравнений (3.10), из которых $K-1$ уравнений линейно независимы. Опуская одно из уравнений (3.10) и учитывая (3.15), получим систему из K уравнений, содержащую K неизвестных N_m .

По достижению стационарного состояния, когда число частиц на каждом из уровней остается неизменным во времени, $dN_m/dt=0$, и получается система линейных однородных уравнений:

$$N_m \sum_{n \neq m} \omega_{mn} - \sum_{n \neq m} N_n \omega_{nm} = 0. \quad (3.16)$$

Эти уравнения совместно с (3.15) 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. 1.6) 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 уров-

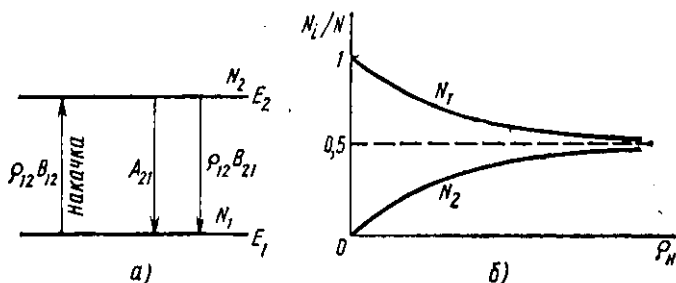


Рис. 3.6. Двухуровневая схема (а) и зависимость относительной населенности уровней от интенсивности (плотности) накачки (б)

ни — возбужденными. Рассмотрим особенности указанных схем.

1. 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. 3.6, 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

$$\left. \begin{aligned} \frac{dN_1}{dt} &= (A_{21} + \rho_{12}B_{21})N_2 - \rho_{12}B_{12}N_1 = 0, \\ \frac{dN_2}{dt} &= \rho_{12}B_{12}N_1 - (A_{21} + \rho_{12}B_{21})N_2 = 0, \\ N_1 + N_2 &= N. \end{aligned} \right\} \quad (3.17)$$

Учитывая, что $B_{12} = B_{21}$, находим населенности уровней:

$$N_2 = \frac{\rho_{12}B_{21}}{A_{21} + 2\rho_{12}B_{21}} N, \quad (3.18)$$

$$N_1 = \frac{A_{21} + \rho_{12}B_{21}}{A_{21} + 2\rho_{12}B_{21}} N. \quad (3.18a)$$

Изменение населенности уровней в зависимости от плотности излучения накачки $\rho_n = \rho_{12}$ в двухуровневой схеме представлено на рис. 3.6, б. В отсутствие накачки все частицы находятся в основном состоянии 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 пре-

дельном случае $\rho_n \rightarrow \infty$ населенности уровней выравниваются. Это явление, называемое *насыщением*, приводит к «просветлению» среды, т. е. при $\rho_n \rightarrow \infty$ показатель поглощения $k \rightarrow 0$ (см. кривую 2 на рис. 3.2). Но ни при каких условиях, используя оптическую накачку, в двухуровневой схеме нельзя 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 указывают, что в двухуровневой схеме квантовые усилители и генераторы с оптической накачкой работать не могут. Аналогичные результаты получатся, если в системе (3.17) 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 — лазера на пучке молекул аммиака.

2. Трехуровневые схемы. Предложение использовать для создания инверсии населенностей более сложные трех- и четырехуровневые схемы накачки вызвало прогресс в квантовой электронике.

Механизм создания инверсии населенностей в трехуровневых схемах поясняет рис. 3.7. В зависимости от того, между какими уровнями достигается инверсия, различают трехуровневые схемы первого и второго типов. В схемах первого типа рабочий переход заканчивается в основном состоянии (рис. 3.7, а), а в схемах второго типа — в возбужденном (рис. 3.7, б). Накачка осуществляется по возможности селективно на уровень E_3 . По трехуровневой схеме первого типа работает рубиновый лазер, а по схеме второго типа — гелий-неоновый газовый лазер.

В трехуровневых схемах канал накачки и канал усиления частично разделены. Это позволяет использовать для достижения инверсии наиболее универсальный метод оптиче-

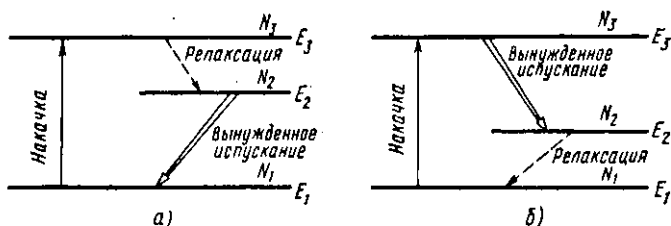


Рис. 3.7. Трехуровневые схемы работы квантовых усилителей и генераторов первого (а) и второго (б) типов

ской накачки, а также накачку с помощью газового разряда. Возможность получения инверсии населенностей с помощью оптической накачки в трехуровневой схеме довольно очевидна. Например, если в схеме второго типа осуществить селективный переход $E_1 \rightarrow E_3$, то уровень E_3 окажется инверсно заселенным относительно уровня E_2 (при $kT \ll E_2 - E_1$). Из рис. 3.7 можно заключить, что накопление частиц на верхнем лазерном уровне (E_2 в схеме первого типа и E_3 — в схеме второго типа) будет в том случае, если релаксационные процессы $E_3 \rightarrow E_2$ в схеме а и $E_2 \rightarrow E_1$ в схеме б идут достаточно быстро, а верхний рабочий уровень является метастабильным.

Определим зависимость населенностей уровней от плотности (интенсивности) возбуждающего излучения накачки $\rho_n = \rho_{13}$. В качестве примера рассмотрим трехуровневую схему первого типа. Предположим, что возбуждение системы (накачка) осуществляется чисто оптическим путем в канале $1 \rightarrow 3$, а внешнее возбуждение в каналах $2 \rightarrow 3$ и $1 \rightarrow 2$

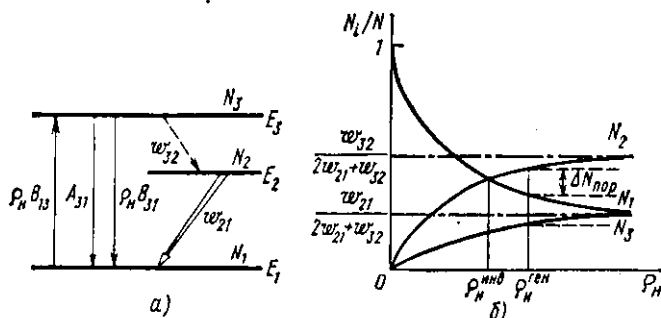


Рис. 3.8. Трехуровневая схема (а) и зависимость относительной населенности уровней от интенсивности накачки (б) в отсутствие (сплошные линии) и при наличии (пунктирные линии) резонатора

отсутствует (или пренебрежимо мало). Скорость релаксации $3 \rightarrow 2$ обозначим ω_{32} . Она может осуществляться за счет излучательных и безызлучательных переходов, так что $\omega_{32} = A_{32} + S_{32}$. Рассмотрим сначала режим усиления, когда активное вещество не находится в резонаторе. Соответствующие переходы изображены на рис. 3.8, а. Кинетические уравнения (3.10) в этом случае для стационарного режима будут иметь вид

$$\left. \begin{aligned} \frac{dN_3}{dt} &= \rho_n B_{13} N_1 - [\rho_n B_{31} + (\omega_{32} + A_{31})] N_3 = 0, \\ \frac{dN_2}{dt} &= \omega_{32} N_3 - \omega_{21} N_2 = 0, \\ N_1 + N_2 + N_3 &= N. \end{aligned} \right\} \quad (3.19)$$

Приняв для простоты кратности вырождения уровней $g_1 = g_2 = g_3 = 1$ и решая систему уравнений (3.19), находим населенности уровней:

$$N_1 = \left[\frac{\omega_{21} (A_{31} + \omega_{32} + \rho_n B_{31})}{\omega_{21} (\omega_{32} + A_{31}) + (2\omega_{21} + \omega_{32}) \rho_n B_{31}} \right] N, \quad (3.20)$$

$$N_2 = \left[\frac{\omega_{32} \rho_n B_{31}}{\omega_{21} (\omega_{32} + A_{31}) + (2\omega_{21} + \omega_{32}) \rho_n B_{31}} \right] N, \quad (3.20a)$$

$$N_3 = \left[\frac{\omega_{21} \rho_n B_{31}}{\omega_{21} (\omega_{32} + A_{31}) + (2\omega_{21} + \omega_{32}) \rho_n B_{31}} \right] N. \quad (3.20б)$$

Зависимость относительной населенности уровней N_i/N ($i=1, 2, 3$) от плотности накачки согласно выражениям (3.20) представлена на рис. 3.8, б. При больших плотностях накачки населенности основного и верхнего состояний в пределе $\rho_n \rightarrow \infty$ стремятся к

$$\lim_{\rho_n \rightarrow \infty} N_1 = \lim_{\rho_n \rightarrow \infty} N_3 = \frac{\omega_{21}}{2\omega_{21} + \omega_{32}} N,$$

а населенность уровня E_2 при $\rho_n \rightarrow \infty$ стремится к

$$\lim_{\rho_n \rightarrow \infty} N_2 = \frac{\omega_{32}}{2\omega_{21} + \omega_{32}} N.$$

При $\omega_{32} > \omega_{21}$, как видно из рис. 3.8, б и соотношений (3.20), начиная с некоторого значения плотности накачки ρ_n между уровнями E_2 и E_1 будет наблюдаться инверсия населенностей ($N_2 > N_1$). Величина $\rho_n^{\text{нв}}$ называется *пороговой плотностью накачки по инверсии*. С увеличением $\rho_n > \rho_n^{\text{нв}}$ инверсия увеличивается.

Отметим, что пороговая накачка для генерации будет превышать пороговую накачку по инверсии, поскольку

для возникновения генерации необходимо выполнить еще условия самовозбуждения (см. § 3.4). Приравнявая выражения (3.20) и (3.20а), найдем $\rho_n^{\text{инв}}$:

$$\rho_n^{\text{инв}} = \frac{\omega_{21} (\omega_{32} + A_{31})}{B_{31} (\omega_{32} - \omega_{21})}. \quad (3.21)$$

Из проведенного рассмотрения вытекает, что для накопления частиц на уровне E_2 и создания максимальной инверсии населенностей наиболее выгодны системы с большим значением ω'_{32} (переход $3 \rightarrow 2$ должен быть быстрым), малым значением ω_{21} (уровень E_2 должен быть метастабильным) и большим коэффициентом Эйнштейна B_{13} (оптический переход $1 \rightarrow 3$ должен быть разрешен)

В соотношениях (3.19), а следовательно, и (3.20) вероятность перехода ω_{21} считалась постоянной, не зависящей от скорости накачки. Это справедливо в отсутствие генерации, когда опустошением уровня E_2 за счет вынужденных переходов можно пренебречь. Если же активное вещество помещено в резонатор, то после превышения инверсии над некоторым пороговым значением начинает развиваться процесс генерации. Наличие интенсивного излучения на частоте ω_{21} и связанных с ним вынужденных переходов $2 \rightarrow 1$ и $1 \rightarrow 2$ вызывает изменение населенности уровней. Возрастание интенсивности накачки ρ_n приводит к увеличению инверсии $\Delta N = N_2 - N_1$ и согласно (3.2) — к увеличению усиления. Это увеличивает ρ_{21} , что в свою очередь обуславливает выравнивание населенностей N_2 и N_1 , т. е. происходит процесс насыщения инверсии населенностей и коэффициента усиления. Увеличение поступления частиц на уровень E_2 , вызываемое ростом накачки, компенсируется возрастанием числа активных переходов $2 \rightarrow 1$. Поэтому в режиме генерации инверсия $\Delta N = N_2 - N_1$ остается приблизительно постоянной, как изображено пунктирными линиями на рис. 3.8, б. Ее значение примерно равно пороговой перенаселенности, при которой усиление превышает потери в генераторе и которая достигается при пороговой накачке, равной $\rho_n^{\text{ген}}$.

3. 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 лазерны-

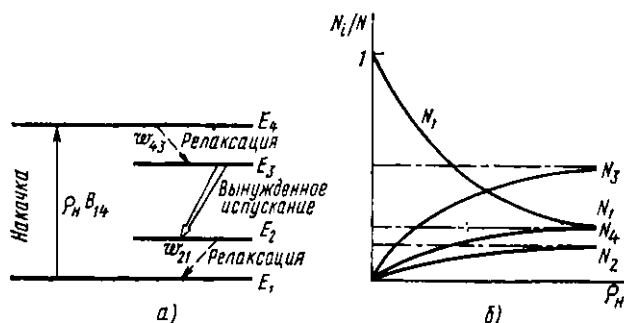


Рис. 3.9. Четырехуровневая схема работы квантовых усилителей и генераторов (а) и зависимость относительной населенности уровней от интенсивности накачки (б)

ми уровнями E_3 и E_2 в четырехуровневой схеме и переходы, определяющие сущность процесса, изображены на рис. 3.9, а. Так же, как и в трехуровневой системе, накачка переводит атомы из основного состояния E_1 в верхнее состояние E_4 . Инверсия достигается между уровнями E_3 и E_2 . Для этого необходимо, чтобы уровень E_3 быстро заполнялся, а уровень E_2 опустошался за счет процессов $E_4 \rightarrow E_3$ и $E_4 \rightarrow E_2$ соответственно. Наличие реально существующих переходов $4 \rightarrow 1$, $4 \rightarrow 2$, $3 \rightarrow 1$, $1 \rightarrow 2$ и $3 \rightarrow 4$ 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 в трехуровневой схеме. В данном случае эти уравнения будут лишь более громоздкими. Зависимость населенности уровней от интенсивности накачки в четырехуровневой схеме представлена на рис. 3.9, б. Инверсия населенностей между уровнями E_3 и E_2 будет достигаться при соблюдении неравенства

$$w_{21}w_{43} > [w_{41}(w_{31} + w_{32}) + w_{31}w_{42}] \frac{g_3}{g_2}, \quad (3.22)$$

где w_{mn} — вероятности переходов $m \rightarrow n$.

Из рис. 3.9, б видно, что в четырехуровневой схеме (при $kT \ll E_2 - E_1$) ничтожное накопление частиц на третьем уровне приводит к инверсии населенностей, т. е. пороговая плотность излучения накачки по инверсии $\rho_n^{\text{порог}}$ в четырехуровневой схеме равна нулю. Поскольку здесь учитываются потери в активном веществе, реальная поро-

говая накачка в активных материалах, работающих по четырехуровневой схеме, несколько больше нуля, но может быть очень малá.

Так же, как и в трехуровневой схеме, зависимость населенности уровней в режиме генерации будет отличаться от изображенной на рис. 3.9, б. За счет интенсивных вынужденных переходов с испусканием в канале генерации $3 \rightarrow 2$ значение N_3 будет уменьшаться, а N_2 — возрастать, приводя к насыщению коэффициента усиления и населенностей уровней.

В заключение отметим следующие моменты, относящиеся как к трехуровневым, так и к четырехуровневым схемам.

а. Для исключения термического заселения необходимо, чтобы энергетические расстояния между уровнями $E_4 - E_3$ и $E_3 - E_1$ 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 к. п. д. [в пределе $\eta < (E_3 - E_1)/(E_4 - E_1)$] и разогреву активного вещества, если 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_4 (или система уровней) был достаточно широким. Это нужно для более полного использования излучения накачки [интеграл (3.12) должен быть максимальным].

в. Для исключения самопоглощения, приводящего к переходам $E_1 \rightarrow E_2$ и $E_3 \rightarrow E_4$, желательно, чтобы релаксационные процессы осуществлялись за счет неоптических (безызлучательных) переходов.

г. Время жизни на верхнем лазерном уровне E_3 должно определяться излучательными процессами, а вероятность безызлучательных переходов с этого уровня должна быть минимальной.

§ 3.3. ОПТИЧЕСКИЕ РЕЗОНАТОРЫ

Основные понятия. Важнейшим элементом квантовых генераторов радио- и оптического диапазонов является резонатор (см. общую схему квантовых генераторов на рис. 3.5). За счет многократного отражения в резонаторе достигается наиболее полное взаимодействие электромагнитного излучения с активной средой. Обеспечивая положительную об-

ратную 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 — 10^{11} Гц ($\lambda \approx 0,3 \div 30$ 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

$$\mathcal{E}(\mathbf{r}, t) = \mathcal{E}_0 U(\mathbf{r}) e^{i\omega t}, \quad (3.23)$$

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_{mnq} находят из соотношения

$$\frac{1}{\lambda_{mnq}} = \sqrt{\left(\frac{m}{2L_1}\right)^2 + \left(\frac{n}{2L_2}\right)^2 + \left(\frac{q}{2L_3}\right)^2}. \quad (3.24)$$

Здесь m , n и q — целые числа, определяющие число полу-волн, укладываемых вдоль сторон L_1 , L_2 и L_3 .

Соотношение (3.24) выражает по существу условие образования стоячей волны при ее отражениях от стенок резо-

натора. Например, для $m=n=0$ имеем $\frac{1}{\lambda_{00q}} = \frac{q}{2L_z}$, или $L_z = \frac{\lambda}{2} q$. На длине L_z укладывается целое число полуволн, 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 \mathcal{E} , \mathcal{H} , k взаимно перпендикулярны. Электромагнитные колебания в объемном резонаторе не являются поперечными; один из векторов \mathcal{E} или \mathcal{H} имеет проекцию на направление распространения волны — продольную составляющую. Если продольную составляющую имеет вектор \mathcal{E} , то электромагнитное колебание называют электрическим и обозначают Е или ТМ. Если продольную составляющую имеет вектор \mathcal{H} , то колебание называют магнитным и обозначают Н или ТЕ. Каждое колебание, например, в прямоугольном резонаторе характеризуют тремя индексами m , n и q (3.24), соответствующими числу полуволн в направлениях x , y и z . Для цилиндрического объемного резонатора, часто применяемого в СВЧ-диапазоне, индексы m , n и q указывают число полуволн 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 характеризуют добротностью. *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 величины 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 (3.24):

$$k_x = \frac{\pi m}{L}, \quad k_y = \frac{\pi n}{L}, \quad k_z = \frac{\pi q}{L},$$

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:

$$\frac{1}{V} \frac{dN_k}{d\omega} = \frac{\omega^2}{\pi^2 c^3}. \quad (3.25)$$

Заметим, что именно этот множитель стоит в функции Планка (2.77) и в формуле Рэлея — Джинса (2.71). Оценим число типов колебаний ΔN_k , приходящихся на спектральный диапазон $\Delta\nu = \Delta\omega/2\pi = 3$ ГГц. Величина $\Delta\nu$, как следует из § 2.5, соответствует доплеровскому уширению спектральной линии при комнатной температуре. Приняв $\lambda = 1$ мкм, получим в объеме 1 см^3 $\Delta N_k \approx 2 \cdot 10^8$. Из (3.25) определим, что частотный интервал между ближайшими типами колебаний при тех же условиях имеет величину порядка 10 Гц. Ширина каждой резонансной линии определяется ее добротностью: $\Delta\omega = \omega/Q$. Даже если бы удалось создать резонатор с добротностью 10^8 , 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.

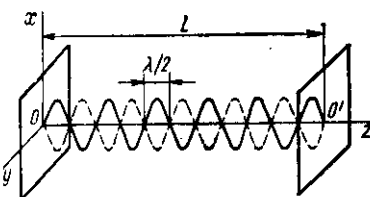


Рис. 3.10. Схема оптического резонатора с плоскими зеркалами

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. 3.10). 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. 3.10. 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,

* Резонансные свойства системы, состоящей из двух параллельных плоских зеркал, были известны давно в оптике, где подобные системы называются интерферометрами. Плоский резонатор иногда называют резонатором Фабри—Перо по имени французских физиков, впервые изучавших плоский интерферометр.

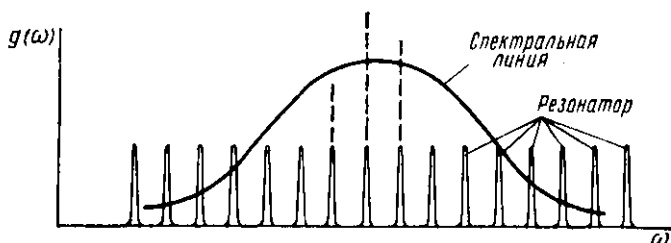


Рис. 3.11. Сравнение формы спектральной линии со спектром собственных частот оптического резонатора (пунктирными линиями обозначены частоты, наиболее благоприятные для развития генерации)

собственно, 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, например, 1 м. На этой длине будет укладываться $2 \cdot 10^8$ полувольт ($q=2 \cdot 10^8$). Расстояние между ближайшими типами колебаний

$$\Delta\nu_q = \nu_q - \nu_{q-1} = \frac{c}{\lambda_q} - \frac{c}{\lambda_{q-1}} = \frac{qc}{2L} - \frac{(q-1)c}{2L} = \frac{c}{2L}. \quad (3.26)$$

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).

Несмотря на столь существенное разрежение спектра собственных частот, частотный интервал между соседними типами колебаний в оптических резонаторах с размерами $L \approx 0,1 \div 2$ 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, как показано на рис. 3.11. 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. 3.11, будет определяться собственными частотами резонатора, лежащими вблизи максимума спектральной линии.

В радиодиапазоне наблюдается обратная картина. Ширина спектральной линии мала (см. оценки, приведенные в § 2.5), 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 сот отражений) установившееся распределение электромагнитного поля не изменяется. Такое стационарное распределение поля и определяет собственные колебания резонатора. Эти колебания, имеющие пренебрежимо малые проекции \mathcal{E} и \mathcal{H} на ось z , называются *трансверсальными электромагнитными колебаниями* и обозначаются $ТЕМ_{mnp}$. Смысл индексов m , n и q тот же, что в объемных резонаторах.

Величина q , указывающая число полуволн, укладываемых на длине L резонатора, в оптическом диапазоне очень велика, и при обозначении типов колебаний ее обычно опускают. Величины m и n , наоборот, малы. Для аксиальных колебаний $m=n=0$ и их обозначают $ТЕМ_{00}$. Кроме $ТЕМ_{00}$ -колебаний возможны $ТЕМ_{10}$ - $ТЕМ_{20}$ - $ТЕМ_{01}$ - $ТЕМ_{11}$ -колебания и т. д., которые называются *неаксиальными колебаниями*. Казалось бы, неаксиальные моды должны быстро выходить из резонатора, поскольку они распространяются не строго параллельно оптической оси резонатора. На самом деле из-за явлений дифракции на краях зеркал неаксиальные колебания, дойдя до края зеркала, будут отражаться внутри резонатора. Это явление аналогично отражению электромагнитной волны от границы волновода,

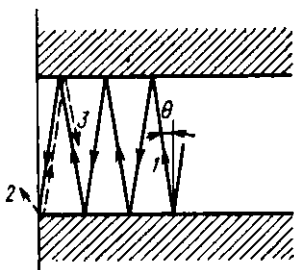


Рис. 3.12. К объяснению явления дифракции на открытом конце волновода

как показано на рис. 3.12. 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 к единице. Понятно, что рис. 3.12 и проведенные рассуждения отражают процесс условно, поскольку такой «геометрический» подход не объясняет сущность дифракционных явлений. Важно, что волны, распространяющиеся под малыми углами θ к оси резонатора, т. е. 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

$$Q = 2\pi \frac{E_{\text{полн}}}{E_{\text{потерь}}(T)} = \omega \frac{E_{\text{полн}}}{E_{\text{потерь}}(1\text{с})}, \quad (3.27)$$

где $E_{\text{полн}}$ — полная энергия, запасенная в резонаторе; $E_{\text{потерь}}(T)$ — энергия, теряемая за период; $E_{\text{потерь}}(1\text{с})$ — энергия, теряемая за 1 с.

Добротность определяется потерями в резонаторе. Из (3.27) 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

$$E_{\omega}(t) = E_{\omega}(0) e^{-\frac{\omega}{Q} t}, \quad (3.28)$$

т. е. аналогично изменению во времени мощности излучения классического осциллятора (2.97). Поэтому каждая спектральная линия резонатора будет описываться функцией (2.101) с $\gamma = \omega/Q$, т. е. иметь лоренцеву форму с полушириной $\Delta\omega = \gamma = \omega/Q$. Из (3.28) 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 $E_{\omega} (1-R) \frac{c}{L}$. Из (3.27) получаем добротность резонатора, определяемую потерями на излучение:

$$Q = \frac{\omega L}{(1-R)c} = \frac{kL}{(1-R)}, \quad (3.29)$$

где $k=2\pi/\lambda$. Взяв типичные значения $\lambda=1$ мкм, $L=1$ м и $R=0,9$, получаем $Q \approx 6 \cdot 10^7$, 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

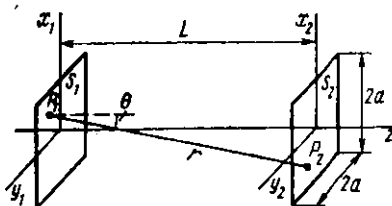


Рис. 3.13. К определению свойств плоского резонатора

точно настраивают (юстируют). Угол между плоскостями зеркал, называемый углом разъюстировки, обычно 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. 3.13. Собственные частоты для аксиальных колебаний в первом приближении определяются выражением (3.26). Определим теперь собственные частоты для неаксиальных колебаний. Их можно найти из (3.24), положив $L_1 = L_2 = 2a$ и $L_3 = L$. Поскольку $(m, n) \ll q$, разлагая (3.24) в степенной ряд, имеем

$$\nu_{mnq} = \frac{c}{\lambda_{mnq}} = \frac{c}{2} \left[\frac{q}{L} + \frac{1}{2} \frac{(m^2 + n^2)}{q} \frac{L}{4a^2} \right]. \quad (3.30)$$

Отсюда, как и следовало ожидать, получаем разность частот между соседними аксиальными колебаниями в виде (3.26). Разность частот между двумя неаксиальными колебаниями, отличающимися друг от друга на единицу по индексу m , получается из (3.30):

$$\Delta \nu_m = \nu_{m+1} - \nu_m = \frac{c}{4} \frac{L[(m+1)^2 - m^2]}{4a^2 q} = \frac{cL}{8a^2 q} \left(m + \frac{1}{2} \right). \quad (3.31)$$

Эту формулу с учетом (3.26) можно переписать в виде

$$\Delta \nu_m = \Delta \nu_q \frac{L^2}{4qa^2} \left(m + \frac{1}{2} \right). \quad (3.32)$$

Аналогично находят расстояние по частоте для двух типов колебаний (мод), отличающихся друг от друга на единицу по индексу n :

$$\Delta \nu_n = \Delta \nu_q \frac{L^2}{4qa^2} \left(n + \frac{1}{2} \right). \quad (3.33)$$

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 (3.32) и (3.33) можно переписать в виде:

$$\Delta v_m = \Delta v_q \frac{m + 1/2}{8N}, \quad (3.34)$$

$$\Delta v_n = \Delta v_q \frac{n + 1/2}{8N}, \quad (3.34a)$$

где

$$N = a^2 / (\lambda L) \quad (3.35)$$

— 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 (Δv_m , Δv_n) порядка мегагерц, т. е. существенно меньше расстояния между аксиальными колебаниями ($\Delta v_q \approx 150$ МГц) [см. (3.26)]. 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. 3.14. Расстояние по частоте между аксиальными модами Δv_q увеличивается при уменьшении длины резонатора L , а между неаксиальными модами Δv_m и Δv_n — при уменьшении числа Френеля N . Из (3.30) 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 частично вырожденными.

Более строгое рассмотрение свойств оптических резонаторов, как отмечалось, 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 характери-

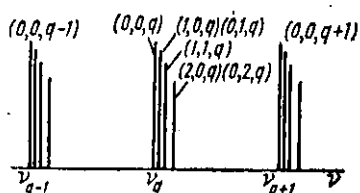


Рис. 3.14. Спектр собственных частот плоского резонатора

стик оптических резонаторов основывается на принципе Гюйгенса — Френеля *. 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 в общем случае произвольное распределение электромагнитного поля u_1 на зеркале S_1 (см. рис. 3.13), 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:

$$u_2(P_2) = \frac{ik}{4\pi} \int_{S_1} u_1(P_1) \frac{\exp(-ikr)}{r} (1 + \cos \theta) dS_1. \quad (3.36)$$

Здесь $k=2\pi/\lambda$, а все обозначения приведены на рис. 3.13.

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 (3.36). 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 поверхности 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

$$u_p = \left(\frac{1}{\gamma}\right)^p u, \quad (3.37)$$

где u — функция, не изменяющаяся от отражения к отра-

* Напомним, что 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 $t+\Delta t$ дается поверхностью, огибающей фронты всех вторичных волн.

жению: γ — комплексная постоянная, не зависящая от пространственных координат.

Подставляя (3.37) в (3.36), получаем интегральное уравнение

$$\left(\frac{1}{\gamma}\right)^{p+1} v = \frac{ik}{4\pi} \int_S \left(\frac{1}{\gamma}\right)^p \frac{v}{r} e^{ikr} (1 + \cos \theta) dS,$$

или

$$v = \gamma \frac{ik}{4\pi} \int_S \frac{v}{r} e^{ikr} (1 + \cos \theta) dr. \quad (3.38)$$

Функция распределения v , удовлетворяющая интегральному уравнению (3.38), 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 (3.38) 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 вперед и назад 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 (3.38) упрощается:

$$v(x, y) = \gamma \frac{ie^{-ikL}}{\lambda L} \int_S v(x, y) e^{-ik(R-L)} dS. \quad (3.39)$$

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:

$$\ln \gamma_{mn} = \beta_{mn} + i(\alpha_{mn} + kL). \quad (3.40)$$

Для каждого собственного колебания резонатора величина

на 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. 3.13. Если в качестве начального, «затравочно-го» распределения поля u_1 вы-

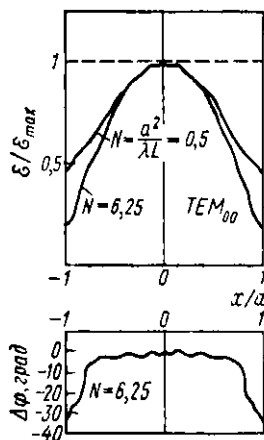


Рис. 3.15. Распределение относительной амплитуды $\varepsilon/\varepsilon_{\max}$ и фазы $\Delta\varphi$ в направлении x для TEM_{00} -колебаний в плоском резонаторе (пунктирной линией указано начальное распределение поля)

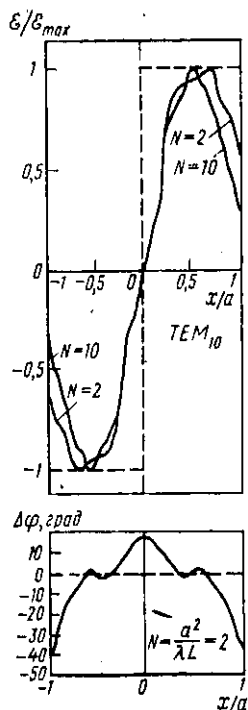


Рис. 3.16. Распределение относительной амплитуды $\varepsilon/\varepsilon_{\max}$ и фазы $\Delta\varphi$ в направлении x для TEM_{10} -колебаний в плоском резонаторе (пунктирной линией указано начальное распределение поля)

брать 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. 3.15. 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 наименший четно-симметричный тип аксиальных TEM_{00} -колебаний.

Если в качестве начального распределения задана нечетно-симметричная функция u_1 с одинарной сменой знака

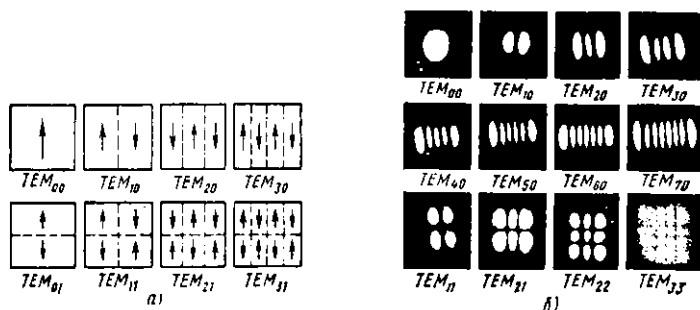


Рис. 3.17. Структура электрического поля для некоторых типов колебаний в плоском резонаторе с квадратными зеркалами (а), фотографии структуры поля на зеркалах резонатора гелий-неонового лазера (б)

(например, $u_1(x) = -u_1(-x)$, причем $|u(x)| = \text{const}$, как показано на рис. 3.16), то после нескольких сот отражений получится стационарное распределение, изображенное на рис. 3.16. Это распределение 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. 3.17, а показана структура электрического поля для простейших типов колебаний, а на рис. 3.17, б — фотографии структуры светового поля на зеркалах резонатора гелий-неонового лазера.

С увеличением индексов m и n амплитуда поля на краях зеркала возрастает (ср. рис. 3.15 и 3.16). 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 значе-

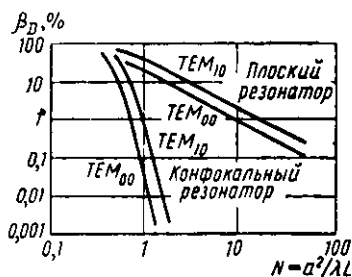


Рис. 3.18. Зависимость дифракционных потерь β_D за один проход для простейших типов колебаний от числа Френеля N для круглых зеркал плоского и конфокального резонаторов

ния 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. 3.18.

Распределение поля внутри резонатора с плоскими зеркалами в разных сечениях, перпендикулярных оптической оси, изменяется сравнительно слабо. Это позволяет достаточно эффективно использовать активный объем рабочего вещества. Отличительной чертой плоского резонатора является то, что фронт волны на его зеркале почти плоский. Искривление фронта волны хотя и имеется (см. рис. 3.15), но оно мало. Это дает возможность получить малую расходимость луча, выходящего из такого резонатора.

Различные типы резонаторов. Кроме открытого резонатора с плоскими зеркалами в квантовой электронике используются системы других конфигураций. Широкое распространение получили оптические резонаторы, образованные двумя сферическими зеркалами или одним сферическим и одним плоским зеркалами. Различные типы резонаторов со сферическими зеркалами представлены на рис. 3.19, а—г. Для специальных целей применяются более сложные системы, например составные и кольцевые резонаторы.

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 (3.36) — (3.38) справедливы для общего случая.

Рассмотрим свойства различных типов оптических резонаторов, вытекающие из проведенных расчетов.

1. **Конфокальный резонатор** (рис. 3.19, а). Конфокальным называют открытый резонатор, образованный одинаковыми сферическими зеркалами, оси и фокусы которых совпадают. Так как фокус сферического зеркала радиусом R расположен на расстоянии $R/2$, то это означает, что радиусы кривизны зеркал равны длине резонатора. Конфокальный резонатор имеет следующие особенности.

а. Распределение фазы по поверхности зеркал однородно, т. е. отражающие поверхности являются поверхностями равных фаз.

б. По сравнению с плоским резонатором поле в конфокальном резонаторе более плотно сконцентрировано у

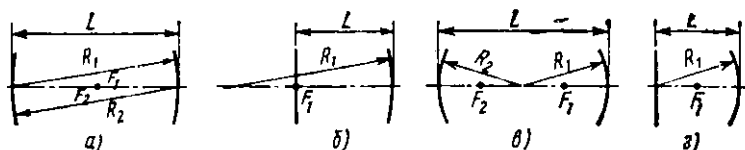


Рис. 3.19. Схемы различных типов резонаторов со сферическими зеркалами:

а — конфокальный ($R_1=R_2=L$); б — полуконфокальный ($R_1=2L$, $R_2=\infty$); в — concentric ($R_1=R_2=L/2$); г — полуконцентрический ($R_1=L$, $R_2=\infty$).

его оси и спадает на краях значительно быстрее. Это приводит к тому, что дифракционные потери в конфокальном резонаторе значительно меньше, чем для аналогичных типов колебаний в плоском резонаторе. Распределение поля для разных N приведено на рис. 3.20, а на рис. 3.18 сравниваются дифракционные потери в конфокальном и плоском резонаторах.

в. В распределении поля не наблюдается характерных для плоского резонатора перегибов (ср. рис. 3.20 и 3.15). Основное TEM_{00} -колебание имеет гауссово распределение как в направлении x , так и в направлении y . Собственные функции записываются в виде $u_{00}(x, y) = \exp[-\pi(x^2 + y^2)/\lambda L]$. Радиус светового пятна $r_{1/e}$, соответствующий уменьшению поля в TEM_{00} -моде в e раз,

$$r_{1/e} = \sqrt{\lambda L / \pi}. \quad (3.41)$$

Пучок, распределение поля в поперечном направлении которого характеризуется функцией Гаусса, называется *гауссовым пучком*.

г. Собственные частоты определяются по формуле

$$\nu_{mnq} = \frac{\omega_{mnq}}{2\pi} = \frac{c}{4L} (2q + 1 + m + n). \quad (3.42)$$

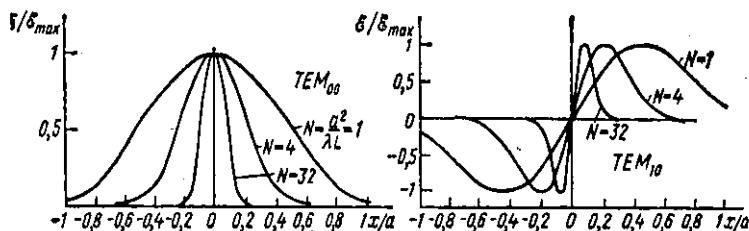


Рис. 3.20. Распределение относительной амплитуды E/E_{\max} для TEM_{00} - и TEM_{10} -колебаний в конфокальном резонаторе

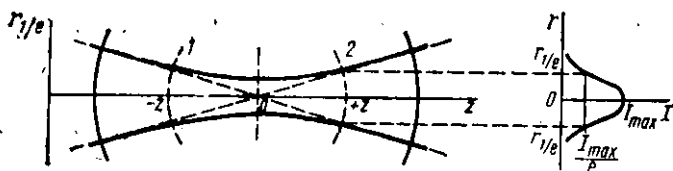


Рис. 3.21. Радиус светового пятна $r_{1/e}$, синфазные поверхности и распределение интенсивности для аксиальной TEM_{00} -моды в конфокальном резонаторе

Отсюда видно, что спектр собственных частот в конфокальном резонаторе сильно вырожден.

д. Поле внутри резонатора в наибольшей степени сконцентрировано вблизи общего фокуса зеркал, т. е. в центре резонатора. Поверхности равных фаз представляют собой сферические поверхности. Радиус кривизны синфазной поверхности, пересекающей оптическую ось резонатора в точке z (начало отсчета $z=0$ совпадает с общим фокусом зеркал), определяется соотношением

$$R(z) = z(1 + L^2/4z^2). \quad (3.43)$$

След этих поверхностей изображен на рис. 3.21 пунктирными линиями. С приближением к центру резонатора радиусы кривизны синфазных поверхностей увеличиваются. При $z=0$ $R \rightarrow \infty$, т. е. синфазная поверхность является плоской.

е. Распределение поля внутри резонатора в поперечном направлении для TEM_{00} -моды приближенно описывается функцией Гаусса, причем радиус светового пятна $r_{1/e}$ изменяется с координатой z по закону

$$r_{1/e} = \left(\frac{\lambda L}{2\pi}\right)^{1/2} \sqrt{1 + \left(\frac{2z}{L}\right)^2}. \quad (3.44)$$

Эта зависимость представлена на рис. 3.21 сплошной линией. Радиус пятна минимален в центре резонатора. Он в $\sqrt{2}$ раз меньше, чем на поверхности зеркал [см. (3.41)].

Перечисленные особенности распределения поля в конфокальном резонаторе имеют как преимущества, так и недостатки. То, что благодаря фокусирующему действию зеркал поле стремится сконцентрироваться вдоль оси резонатора, с одной стороны, обуславливает минимальные дифракционные потери. По этой же причине конфокальный резонатор менее чувствителен к разъюстировке зеркал. С дру-

гой стороны, это иногда мешает полностью использовать объем активного вещества в резонаторе.

2. Резонаторы с произвольными сферическими зеркалами. Это резонаторы, состоящие 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 (3.43) 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 (3.44), где вместо L будет 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 (3.43) получаем, что длина эквивалентного конфокального резонатора

$$L_{\text{эк}} = \sqrt{L(2R - L)}. \quad (3.45)$$

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_{mnq} в общем случае будут определяться не соотношением (3.42), 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 по существу теряет свои резонансные свойства). Например, если резонатор образован двумя одинаковыми сферическими зеркалами радиусами R , причем $L > 2R$, то, как следует из (3.45), 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 кон-

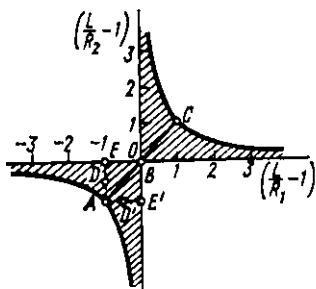


Рис. 3.22. Диаграмма устойчивости оптических резонаторов со сферическими зеркалами (незаштрихованные области соответствуют неустойчивым резонаторам с большими потерями)

фокальный резонатор, если выполняется неравенство

$$0 \leq \left(\frac{L}{R_1} - 1 \right) \left(\frac{L}{R_2} - 1 \right) \leq 1. \quad (3.46)$$

На рис. 3.22 приведена диаграмма, иллюстрирующая условие устойчивости оптических резонаторов. Устойчивые области соответствуют 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. 3.19, 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 применяют для селекции (отбора) неаксиальных колебаний: за счет большой разницы в дифракционных потерях между аксиальными TEM_{00} и неаксиальными модами последние в концентрическом резонаторе обладают малой добротностью и, следовательно, гасятся.

Точкам D и D' с координатами $(-1; -0,5)$ и $(-0,5; -1)$ соответствует *полуконфокальный резонатор*, т. е. резонатор, образованный одним плоским и одним сферическим зеркалом, радиус кривизны которого равен удвоенной длине резонатора (см. рис. 3.19, б). Он находится в середине устойчивой области и его свойства аналогичны конфокальному резонатору с удвоенной длиной.

Точкам E и E' с координатами $(-1; 0)$ и $(0; -1)$ соответствует *полуконцентрический резонатор* (см. рис. 3.19, г). Он обладает особенностями, характерными для концентрического резонатора, а также применяется для селекции неаксиальных колебаний.

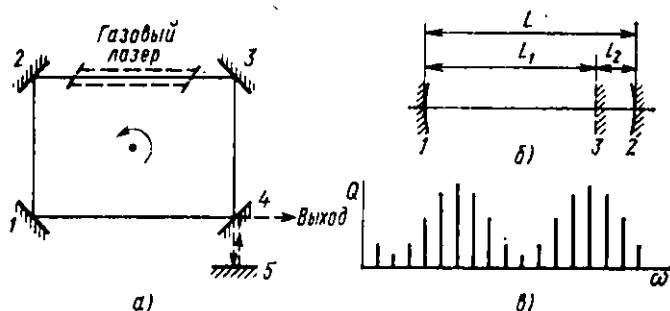


Рис. 3.23. Схемы кольцевого (а) и составного (б) резонаторов; зависимость добротности аксиальных колебаний от частоты в составном резонаторе (в)

3. Кольцевой резонатор. Кольцевым называют открытый резонатор, зеркала которого обеспечивают распространение электромагнитных волн по замкнутому контуру (рис. 3.23, а). 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 оптических 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. 3.23, а, то в таком резонаторе можно осуществить режим бегущей волны. Иногда режим стоящей волны является энергетически менее выгодным, поскольку наиболее полно взаимодействие излучения с веществом здесь происходит лишь в пучностях светового поля. В бегущей волне поле в среднем более однородно, что позволяет полнее 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.

4. Составной резонатор. Такой резонатор представляет собой два (или более) связанных между собой резонатора. Схема составного резонатора приведена на рис. 3.23, б. Зеркало 3 является полупрозрачным. Свойства такого резонатора аналогичны свойствам двух связанных контуров. Как видно из рис. 3.23, в, составной резонатор

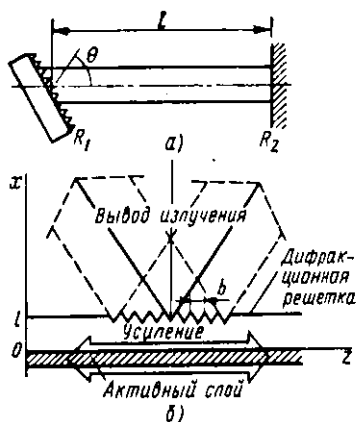


Рис. 3.24. Схемы резонаторов с брэгговским зеркалом (а) и с распределенной обратной связью (б)

можно использовать для селекции аксиальных колебаний (ср. с рис. 3.11). Дополнительная спектральная селекция колебаний в составном резонаторе определяется тем, что наибольшей добротностью будут обладать те типы колебаний, для которых условие резонанса наилучшим образом удовлетворяется в каждой из составных частей резонатора.

5. Резонатор с брэгговским зеркалом. Иногда возникает необходимость плавно перестраивать собственную частоту резонатора и тем самым плавно изменять частоту генерации лазера в пределах контура

спектральной линии активного вещества. В принципе это может быть достигнуто путем изменения оптической длины L резонатора [см. выражения (3.30) и (3.42),] но лишь в очень малом спектральном диапазоне, соответствующем частотному расстоянию $\Delta\nu$ между двумя ближайшими аксиальными модами. В более широком диапазоне перестройка частоты может быть осуществлена, если одно из зеркал сделать селективным, т. е. обладающим большим коэффициентом отражения в узком спектральном диапазоне. При этом необходимо, чтобы была возможность каким-либо образом изменять спектральный диапазон. В качестве селективного элемента может быть использована, например, дифракционная решетка, которую устанавливают вместо одного из зеркал резонатора. Подобный резонатор называется *резонатором с брэгговским зеркалом* (рис. 3.24, а). Дифракционная решетка располагается под углом к оптической оси резонатора. Если на такое «зеркало» падает плоская электромагнитная волна, распространяющаяся вдоль оптической оси резонатора, то за счет дифракции на решетке она будет отражаться назад точно в противоположном направлении при выполнении условия Вульфа — Брэгга:

$$b \sin \theta = m\lambda/2,$$

где b — период дифракционной решетки; θ — угол между нормалью к плоскости решетки и оптической осью резонатора; $m=1; 2; 3; \dots$ — порядок дифракции.

Изменяя угол θ , можно изменять длину волны λ , соответствующую максимуму отражательной способности зеркала в направлении оптической оси. Такой резонатор применяется, в частности, для перестройки частоты генерации жидкостных лазеров на органических красителях.

6. Резонатор с распределенной обратной связью. Это особый вид резонатора, в котором торцевые зеркала вообще могут отсутствовать, а положительная обратная связь обеспечивается рассеянием на периодических неоднородностях, образующих строго периодичную решетку. Важно, чтобы эта решетка образовывалась пространственными периодическими изменениями одного из параметров, от которого зависят условия распространения света. Это может быть коэффициент преломления, коэффициент усиления, толщина пленки или кристалла. Расстояние b между неоднородностями должно удовлетворять условию Вульфа — Брэгга. 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. 3.24, б.

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. 3.24, б). 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 ра-

спространения на обратное, приводя тем самым к возникновению положительной обратной связи. Очевидно, 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. Например, если период b решетки равен длине волны λ (в данном случае λ есть длина световой волны в активном слое), то обратная связь будет осуществляться во втором порядке ($m=2$, $\theta=90^\circ$). В первом порядке дифракции ($m=1$) можно осуществлять вывод энергии излучения под углом $\theta=30^\circ$ к плоскости волновода. Из симметрии волновода следует, что в нем будут распространяться как минимум две волны во встречных направлениях $(+z)$ и $(-z)$. По этой же причине излучение будет выводиться в двух симметричных направлениях, как схематически показано на рис. 3.24, б.

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. Величины обратной связи и коэффициента потерь на излучение в такой системе можно изменять, например, изменяя расстояние l .

При наличии нескольких типов колебаний в резонаторе, т. е. нескольких волноводных мод, вследствие их возможного взаимодействия общая картина излучения может значительно усложниться.

Очевидно, с помощью системы, аналогичной изображенной на рис. 3.24, б, можно в обратном порядке осуществлять не вывод, а ввод излучения в тонкопленочный волновод через его боковую поверхность. Это часто используют в устройствах интегральной оптики.

§ 3.4. УСЛОВИЕ САМОВОЗБУЖДЕНИЯ И НАСЫЩЕНИЕ УСИЛЕНИЯ

Как было показано в § 3.1, 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 такой величины 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 (3.2), т. е.

$$\alpha_\omega = \frac{\hbar \omega n B_{nm}}{c} g(\omega) \left(\frac{g_n}{g_m} N_m - N_n \right). \quad (3.47)$$

В оптическом резонаторе, как видно из § 3.3, существуют конфигурации поля с малыми потерями. Очевидно, именно для таких конфигураций, определяющих слабозатухающие собственные моды резонатора, условие самовозбуждения будет выполняться в первую очередь. Потери в резонаторе для данного типа колебаний определяются его добротностью Q . На основании выражения (3.27) уменьшение относительной интенсивности излучения $\delta I/I$ за время δt , вызванное потерями, определяется соотношением

$$-\frac{\delta I}{I} = \frac{\omega}{Q} \delta t = \frac{\omega}{Q} \frac{n}{c} \delta z. \quad (3.48)$$

Здесь учтено, что время прохождения слоя активного вещества толщиной δz есть $\delta t = \frac{n}{c} \delta z$.

С другой стороны, увеличение относительной интенсивности излучения в активном веществе, обусловленное процессами вынужденного испускания, согласно (3.3) описывается формулой

$$\frac{\delta I_\omega}{I_\omega} = -k_\omega \delta z = \alpha_\omega \delta z. \quad (3.49)$$

Тогда условие превышения усиления над потерями за-

пишется в виде

$$\alpha_{\omega} > \frac{\omega}{Q} \frac{n}{c}.$$

С учетом (3.47) получаем условие самовозбуждения квантового генератора:

$$\hbar \omega B_{mn} g(\omega) \left(\frac{g_n}{g_m} N_m - N_n \right) \geq \frac{\omega}{Q}. \quad (3.50)$$

С учетом соотношений (2.72) и (2.76) между коэффициентами Эйнштейна A_{mn} , B_{mn} и B_{nm} условие самовозбуждения может быть переписано таким образом:

$$\frac{\pi^2 c^3 A_{mn}}{\omega^2} \left(N_m - \frac{g_m}{g_n} N_n \right) g(\omega) \geq \frac{\omega}{Q}. \quad (3.50a)$$

Здесь N_m — число частиц на верхнем лазерном уровне E_m со статистическим весом g_m , а N_n — то же, на нижнем уровне E_n со статистическим весом g_n . Функция $g(\omega)$, как и ранее, есть нормированный контур спектральной линии и определяется лоренцевым контуром (2.113) при однородном уширении и гауссовым контуром (2.117) при неоднородном уширении.

Знак равенства в (3.50) определяет пороговое условие самовозбуждения квантового генератора и пороговую инверсию населенностей для генерации, т. е. минимальную инверсию населенностей, при которой усиление в генераторе полностью компенсирует потери. Энергия (мощность) накачки, при которой достигается пороговая инверсия для генерации, называется *пороговой энергией (мощностью) накачки для генерации*.

Из (3.50) вытекает, что пороговая энергия накачки для генерации минимальна, если добротность Q максимальна. В свою очередь добротность будет максимальной, если минимальны потери, в том числе потери на частичное пропускание полупрозрачного зеркала [см. выражение (3.29)]. Это не означает, что коэффициент отражения R зеркал резонатора в лазерах должен быть всегда близок к единице. Очевидно, при $R \rightarrow 1$ коэффициент пропускания зеркала $T \rightarrow 0$ и выходная мощность $P_{\text{вых}} \rightarrow 0$. Поэтому оптимальной скорости накачки соответствует оптимальный коэффициент пропусканий полупрозрачного зеркала, при котором достигается максимальная выходная мощность.

Условие самовозбуждения будет выполняться в первую очередь для колебаний, обладающих наивысшей доброт-

ностью, резонансные частоты которых расположены наиболее близко к максимуму спектральной линии $g(\omega_0)$. Это означает, что лазер начнет генерировать на аксиальной моде, собственная частота которой наиболее близка к максимуму спектральной линии активного вещества. Энергия этой моды будет возрастать по экспоненциальному закону (3.4) до тех пор, пока ее не начнут ограничивать эффекты насыщения.

Вероятность вынужденного испускания пропорциональна интенсивности вынуждающего излучения. Поэтому после достижения условия самовозбуждения энергия накачки, подводимая к активному веществу, будет «перекачиваться» в генерируемую моду. Подчеркнем, что система, состоящая из активного вещества, помещенного в резонатор, будет не просто отфильтровывать определенные частоты, выступая как пассивный фильтр, а именно «перекачивать» энергию в моды резонатора с малыми потерями.

Таким образом, по достижении в активном веществе определенной степени инверсии населенностей, при которой выполняется условие самовозбуждения (3.50), процессы вынужденного испускания преобладают над процессами поглощения, осуществляя передачу энергии от активной среды в моды резонатора. Этот процесс поставляет в резонатор энергию электромагнитного излучения, которое по амплитуде, фазе и поляризации скоррелировано с полем излучения, вызывающим вынужденное испускание.

Рассмотрим теперь, что произойдет при увеличении интенсивности накачки выше порога генерации $\rho_n^{\text{ген}}$. В отсутствие генерации увеличение ρ_n привело бы к увеличению инверсии населенностей согласно рис. 3.7 и 3.8 и соответственно к возрастанию коэффициента усиления. При наличии генерации, т. е. мощного излучения на частоте лазерного перехода, происходит выравнивание населенностей рабочих лазерных уровней, т. е. ограничение (насыщение) коэффициента усиления, о чем говорилось в § 3.2. В результате эффекта насыщения инверсия населенностей в лазере автоматически поддерживается на уровне $\Delta N_{\text{пор}}$, как показано пунктирными линиями на рис. 3.8, б. Отсутствие эффекта насыщения привело бы к безграничному возрастанию интенсивности излучения.

Когда поле излучения в лазере достигает стационарного значения, происходит насыщение усиления. Насыщение усиления в лазере, как и в генераторах любого другого типа, осуществляется за счет эффектов, нелинейных по ин-

тенсивности излучения. В приближении слабого сигнала, т. е. при малых интенсивностях I_ω , показатель усиления α_ω (3.47), так же как и показатель поглощения (3.3), можно считать не зависящими от интенсивности. При больших интенсивностях такое приближение оказывается несправедливым, поскольку инверсия, а следовательно, и показатель усиления α_ω , становятся зависящими от I_ω .

Характер насыщения усиления в лазере очень сильно зависит от типа уширения спектральной линии. Поясним это на примере изменения спектра поглощения (отрицательного усиления) под влиянием мощного монохроматического излучения I_ω в случае однородно и неоднородно уширенных спектральных линий.

При однородном уширении форма спектра поглощения для каждой из активных частиц совпадает с общим контуром спектральной линии. Поэтому включение мощного монохроматического излучения на частоте ω приводит к равномерному перераспределению всех частиц по энергетическим уровням и к равномерному по спектру уменьшению коэффициента поглощения, как показано на рис. 3.25, а. 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 после включения излучения I_ω уменьшение показателя поглощения (просветление) будет происходить не равномерно по всему спектру, а преимущественно вблизи частоты ω , т. е. в общем спектре образуется характерный провал (рис. 3.25, б). Очевидно, такая же картина будет и для показателя усиления (отрицательного коэффициента поглощения).

Для количественного описания эффекта насыщения усиления необходимо записать кинетические уравнения, аналогичные рассмотренным в § 3.2, и определить инверсию между рабочими лазерными уровнями как функцию интенсивности, а затем подставить полученные выражения в формулу (3.47) для показателя усиления. Соответствующие алгебраические преобразования приводят к тому, что в первом приближении зависимость показателя усиления, определяемого из соотношения (3.49), т. е. $\alpha = \frac{1}{I} \frac{dI}{dz}$, можно представить в виде:

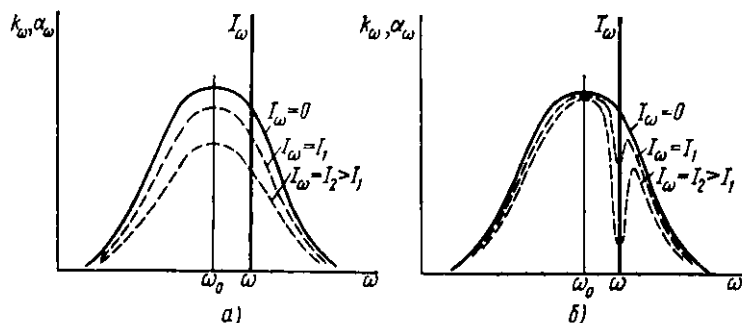


Рис. 3.25. Насыщение показателя поглощения k_ω и усиления α_ω в случае однородно (а) и неоднородно (б) уширенной спектральной линии

для однородно уширенной линии

$$\alpha_{\text{одн}} = \frac{\alpha_0 \text{ одн}}{1 + I/I_S}, \quad (3.51)$$

для неоднородно уширенной линии

$$\alpha_{\text{неодн}} = \frac{\alpha_0 \text{ неодн}}{(1 + I/I_S)^{1/2}}. \quad (3.52)$$

В начальный момент интенсивность I поля излучения лазера равна нулю и усиление характеризуется величиной $\alpha_0 = \alpha_0(\omega)$, называемой *ненасыщенным показателем усиления* или *показателем усиления слабого сигнала*. Она определяется соотношением (3.47) при условии, что населенности N_m и N_n вычислены без учета вынужденных лазерных переходов. Величина I_S называется *параметром насыщения*. Она является мерой интенсивности, требуемой для достижения данной степени насыщения, и определяется через вероятности всех переходов, участвующих в создании инверсии населенностей, т. е. характеризуется свойствами активного вещества.

Когда под воздействием накачки инверсия населенностей достигает определенной величины, начинается генерация на одной моде, обладающей наивысшей добротностью и расположенной ближе всех к максимуму спектральной линии. При дальнейшем увеличении интенсивности накачки условие самовозбуждения начинает выполняться и для других мод, обладающих меньшими добротностями или расположенных дальше от максимума спектральной линии. Этот процесс зависит как от параметров резонатора, так и от параметров активного вещества.

Режим работы, при котором квантовый генератор генерирует преимущественно на одной из мод резонатора, называется *режимом одномодовой генерации*. Режим работы, при котором генерация осуществляется одновременно на нескольких собственных частотах, называется *режимом многомодовой генерации*. При многомодовой генерации возможно возникновение эффектов взаимодействия мод между собой, так что в этом случае их нельзя рассматривать как полностью независимые.

Для получения одномодовой генерации лазера принимают меры, обеспечивающие селекцию типов колебаний, например используют резонаторы специальной формы.

§ 3.5. НЕСТАЦИОНАРНАЯ ГЕНЕРАЦИЯ, МОДУЛЯЦИЯ ДОБРОТНОСТИ И СИНХРОНИЗАЦИЯ МОД

Для количественного описания процессов, происходящих при нестационарном режиме работы трех- и четырехуровневых лазеров, необходимо решить соответствующие кинетические уравнения. 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. 3.26). В момент времени t_1 накачка включена, в момент времени t_1 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. 3.26). Дальнейшее возрастание функции $N_2(t)$ 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 на 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 пичков), амплитуда которых уменьшается приблизительно по экспоненциальному закону (диаграммы 2 и 3 на рис. 3.26). Такую генерацию называют *режимом регулярных пичков*. При больших t в отсутствие допол-

нительных 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 возникнуть 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 заметно превышает время жизни фотона в резонаторе и время обхода резонатора. Обычно это время, приблизительно равное $2(t_2 - t_3)$, составляет 0,1—10 мкс.

Режим регулярных затухающих осцилляций может возникнуть лишь при одномодовой генерации и неизменности параметров лазера во времени. В реальных условиях, особенно для твердотельных лазеров на диэлектриках и для полупроводниковых лазеров, наблюдается *режим нерегу-*

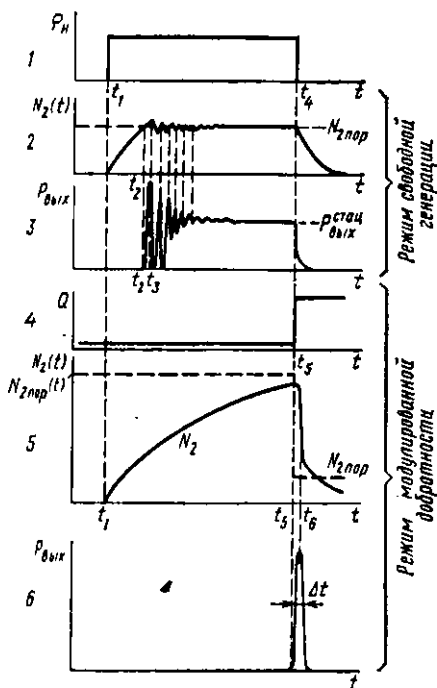


Рис. 3.26. Временные диаграммы развития лазерного импульса

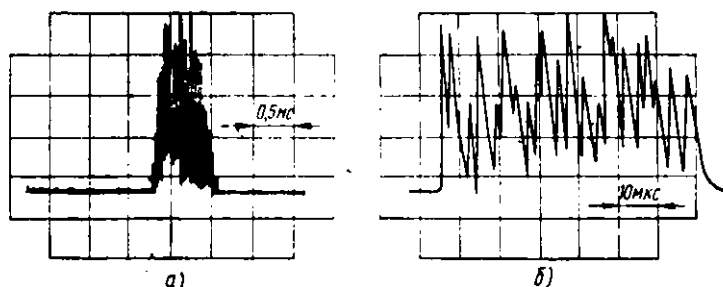


Рис. 3.27. Осциллограммы «пиковой» структуры лазерного импульса: а — развертка 500 мкс/см; б — развертка 10 мкс/см

лярных пиков, интенсивность которых хаотически изменяется и не затухает во времени. Типичные временные зависимости $P_{\text{вых}}(t)$ для этого режима, называемого иногда пиковым режимом генерации, приведены на осциллограммах рис. 3.27, 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 неустойчивостью параметров лазера в реальных условиях, а также взаимодействием различных мод при многомодовой генерации.

Временные диаграммы 1—3 на рис. 3.26 соответствуют так называемому режиму свободной генерации лазера: добротность Q резонатора и соответственно пороговая населенность $N_{2\text{пор}}$ остаются неизменными во времени. Генерация в таком режиме работы начинается самопроизвольно в момент времени t_s .

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. 3.26 и состоит в следующем. Поместим внутри лазер-

ного 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 (диаграммы 4 и 5 на рис. 3.26). Запасенная в активном веществе энергия выделится в виде короткого (длительностью порядка нескольких наносекунд) и интенсивного импульса света (диаграмма 6 на рис. 3.26).

Предположим, что затвор открывается за время, меньшее 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 $\Delta N = \left(N_m - \frac{g_m}{g_n} N_n \right)$ 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 $\Delta N(0) = \Delta N_{\text{нач}}$ и $\rho_{\omega}(0) = \rho_{\omega}^n$. 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, вызванное накачкой и релаксационными процессами, можно пренебречь. Изменение во времени величины $\Delta N(t)$ будет определяться лишь индуцированными переходами, т. е.

$$\frac{d\Delta N(t)}{dt} = \frac{d \left(N_m - \frac{g_m}{g_n} N_n \right)}{dt} = -2B_{\omega}\rho_{\omega}(t)\Delta N(t). \quad (3.53)$$

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

$$\frac{d\Phi_{\omega}}{dt} = \frac{d\rho_{\omega}}{\hbar\omega dt} = \left[B_{\omega} \Delta N(t) \hbar\omega - \frac{1}{t_{\phi}} \right] \Phi_{\omega}. \quad (3.54)$$

Результаты совместного решения уравнений (3.53), (3.54) с начальными условиями $\Delta N(0) = \Delta N_{\text{нач}}$ и $\Phi(0) = \Phi_{\text{сп}}$ представлены на рис. 3.28. Здесь по существу изображена в более крупном масштабе часть рис. 3.26. Подставляя в уравнение (3.54) $d\Phi_{\omega}/dt = 0$, нетрудно найти инверсию населенностей, соответствующую максимуму лазерного импульса:

$$\Delta N = \frac{1}{B_{\omega} t_{\phi} \hbar\omega} = \frac{\omega}{Q B_{\omega} \hbar\omega}. \quad (3.55)$$

Сравнение с (3.50) показывает, что эта величина равна критической инверсии [знак равенства в (3.50)] для непрерывного режима работы лазера. Этот результат вполне объясним: до момента времени $t_{\text{и}}$, когда $\Delta N > \Delta N_{\text{пор}}$, интенсивность лазерного импульса нарастает, после чего, достигнув пика, импульс ослабляется. Форма лазерного импульса при этом асимметрична: время нарастания обычно несколько короче времени спада.

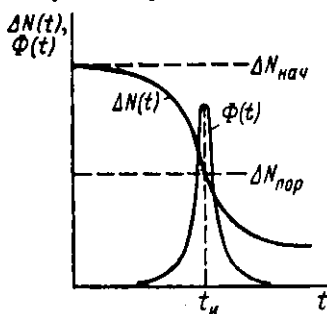


Рис. 3.28. Изменение инверсии $\Delta N(t)$ и числа фотонов $\Phi(t)$ в резонаторе в процессе генерации гигантского импульса

Максимальное число фотонов, запасенных внутри резонатора, равно $\Delta N_{\text{нач}}/2$. Если начальная инверсия намного превышает пороговую, то мощность лазерного излучения в максимуме

$$P_{\text{max}} \approx \frac{\hbar\omega}{2t_{\phi}} V \Delta N_{\text{нач}}, \quad (3.56)$$

где V — объем, занимаемый генерируемой модой в активном веществе.

Развитие генерации в рассмотренном виде будет проис-

ходить 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. Каждый генерируемый импульс приводит к тому, что усиление становится ниже мгновенного порогового значения, подавляя генерацию до тех пор, пока действие затвора не уменьшит потери в резонаторе и, следовательно, порог генерации лазера.

Укажем три наиболее эффективных метода модуляции добротности лазера.

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.

2. Внутри резонатора имеется специальный элемент — оптический модулятор, оптические свойства которого можно изменять с помощью внешних воздействий. Наиболее часто для этих целей используют электрооптические модуляторы, работающие на основе электрооптических эффектов в кристаллах (см. § 4.2).

3. Внутри резонатора имеется 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 (3.2) и рис. 3.2].

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 (3.26): $\Delta\omega = \omega_{q+1} - \omega_q = \pi c/L$. Обозначим собственную частоту центральной моды ω_0 . Тогда полное электрическое поле $\mathcal{E}(t)$ электромагнитной волны в некоторой произвольной точке, например на зеркале резонатора, можно запи-

сать в комплексной форме:

$$\mathcal{E}(t) = \sum_{l=-N}^N \mathcal{E}_l e^{i[(\omega_0 + l\Delta\omega)t + \varphi_l]}, \quad (3.57)$$

где \mathcal{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 $\mathcal{E}_l = \mathcal{E}_0$ и все фазы равны нулю. Тогда выражение (3.57) переписывается в виде

$$\mathcal{E}(t) = \sum_{l=-N}^N \mathcal{E}_0 e^{i(\omega_0 + l\Delta\omega)t}, \quad (3.58)$$

что при суммировании дает

$$\mathcal{E}(t) = A(t) e^{i\omega_0 t}, \quad (3.59)$$

где

$$A(t) = \mathcal{E}_0 \frac{\sin[(2N+1)\Delta\omega t/2]}{\sin(\Delta\omega t/2)}. \quad (3.60)$$

Выражение (3.59) отображает гармоническую волну с частотой несущей ω_0 и амплитудой $A(t)$, промодулированной по закону (3.60). Выходная мощность лазера пропорциональна $|A(t)|^2$, т. е.

$$P_{\text{вых}}(t) \sim \frac{\sin^2[(2N+1)\Delta\omega t/2]}{\sin^2(\Delta\omega t/2)}. \quad (3.61)$$

На рис. 3.29 приведена временная зависимость выходной мощности, рассчитанная по соотношению (3.61), в случае генерации $2N+1=7$ мод с синхронизованными фазами и одинаковыми амплитудами. Отметим следующие важные свойства функции (3.61), являющиеся следствием интерференции мод.

1. Энергия излучается в виде последовательности коротких световых импульсов. Максимумам соответствуют моменты времени, когда знаменатель в (3.61) 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 Δt между вершиной импульса и ближайшим минимумом:

$$\tau_n = \Delta t = \frac{\tau}{2N+1}. \quad (3.62)$$

Число генерируемых мод $(2N+1)$ можно оценить как отношение спектральной ширины линии усиления $\Delta\nu$ к межмодовому интервалу. Подставляя это отношение в (3.62), получаем выражение для длительности импульса

$$\tau_n \approx 1/\Delta\nu, \quad (3.63)$$

которое хорошо иллюстрирует соотношение неопределенностей $\delta E \delta t = \hbar \Delta\nu \Delta t \approx \hbar$.

Из (3.63) 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, как было показано в § 2.5, ширина спектральной линии $\Delta\nu \approx 10^9$ Гц. Поэтому в таких лазерах возможно получить импульсы длительностью порядка 1 нс. В твердотельных лазерах линий уширены за счет влияния сильного поля кристаллической решетки, 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.

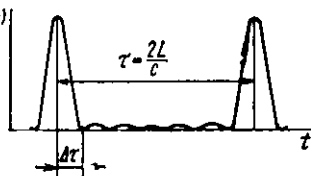


Рис. 3.29. Временная зависимость выходной мощности в случае генерации семи мод в режиме их синхронизации

СВОЙСТВА ЛАЗЕРНОГО ИЗЛУЧЕНИЯ

...отмечалось ранее и подтверждается материалом настоящей главы, свойства лазерного излучения принципиально отличаются от свойств оптического излучения, испускаемого «классическими» источниками. Основными свойствами лазерного излучения являются: 1) монохроматичность; 2) когерентность; 3) направленность; 4) мощность и яркость. Рассмотрим коротко каждое из этих свойств.

Монохроматичность*. Монохроматичность характеризует степень «размазанности» излучения по спектру. Количественной характеристикой степени монохроматичности является ширина спектральной линии на уровне 0,5 от ее максимума 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:

$$\Delta\omega_c = \omega_0/Q = 1/t_\Phi, \quad (3.64)$$

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 (3.64) для пассивного резонатора. Если внутри резонатора находится активная среда, компенсирующая потери на частоте генерации, то для такой моды в режиме генерации $Q \rightarrow \infty$ 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 форму (2.113) с полу-

* От греч. monos — единственный, chroma — цвет.

шириной

$$\Delta\omega = \frac{4\hbar\omega_0 (\Delta\omega_c)^2}{P}, \quad (3.65)$$

где P — выходная мощность лазерного излучения (остальные обозначения те же).

В формуле (3.65) предполагается, что лазер генерирует в стационарном непрерывном режиме на одной моде и его параметры не изменяются во времени. Величину $\Delta\omega$ (3.65) 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 § 3.3 ($Q\approx 10^6$), из (3.65) 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 (3.30) 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 составляют величину порядка 0,1 нм $=1\text{А}$, 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} — 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. § 2.5). На первый взгляд это может показаться парадокс-

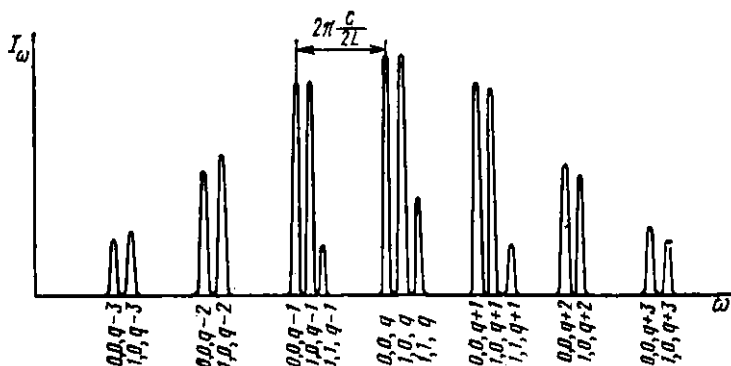


Рис. 3.30. Спектр излучения лазера в многомодовом режиме

сальным, так как $\Delta\nu_{\text{ест}}$ определяется соотношением неопределенностей $\hbar\Delta\nu_{\text{ест}}\tau_{\text{сп}} \approx \hbar$, где $\tau_{\text{сп}}$ — 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 одномодовой 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. 3.30 представлен спектр генерации на семи аксиальных TEM_{00} -модах и семи неаксиальных TEM_{10} -модах.

При генерации в импульсном режиме спектр излучения для каждой моды будет уширен до $\Delta\nu_{\text{ген}} \approx 1/\tau_{\text{и}}$, где $\tau_{\text{и}}$ — длительность импульса. Поэтому в режиме генерации гигантских импульсов при $\tau_{\text{и}} \approx 1$ нс $\Delta\nu_{\text{ген}} \approx 1000$ МГц даже при одномодовой генерации. 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.

Когерентность*. Когерентность характеризует согла-

* От англ. coherency — согласованность, связность, связь.

сованность (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 идеальное монохроматическая линейно поляризованная волна, которая, как отмечалось в § 2.1, 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», нелазерных 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 спектрального 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.

Рассмотрим линейно поляризованную электромагнитную волну. Во многих практически важных случаях диапазон частот $\Delta\omega$, занимаемых волной, существенно меньше среднего значения частоты спектра ω_0 . Такую волну можно записать в общем виде как функцию координаты и времени:

$$V(r, t) = A(r, t)e^{i(\omega_0 t + \varphi(r, t))}. \quad (3.66)$$

Для плоской волны $V(r, t)$ — скалярная величина, которая может

означать проекцию напряженности электрического $\mathcal{E}(r, t)$ или магнитного $\mathcal{H}(r, t)$ поля, $A(r, t)$ — амплитуда волны. В отличие от плоской монохроматической волны она зависит от r и t . Величина $\varphi(r, t)$ есть фаза волны, набегаяемая дополнительно к фазе $\omega_0 t$.

Зафиксируем какую-либо точку P_1 в пространстве с координатами r_1 и определим для нее корреляционную функцию первого порядка $\Gamma^{(1)}$:

$$\Gamma^{(1)}(r_1, r_1, \tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T V(r_1, t+\tau) V^*(r_1, t) dt. \quad (3.67)$$

Она представляет собой автокорреляционную функцию $V(r, t)$ и определяет среднее значение произведения $V(r_1, t+\tau) V^*(r_1, t)$. Используя, как и ранее, усреднение по времени, перепишем соотношение (3.67) в виде

$$\Gamma^{(1)}(r_1, r_1, \tau) = \langle V(r_1, t+\tau) V(r_1, t) \rangle. \quad (3.68)$$

Определим также нормированную функцию $\gamma^{(1)}(r_1, r_1, t)$:

$$\gamma^{(1)}(\tau) = \frac{\Gamma^{(1)}}{\langle V(r, t) V(r, t) \rangle} = \frac{\Gamma^{(1)}}{\langle I(r, t) \rangle}, \quad (3.69)$$

где $I(r, t) = V(r, t) V^*(r, t)$ — интенсивность пучка.

Функции $\Gamma^{(1)}$ и $\gamma^{(1)}(\tau)$ характеризуют меру корреляции между сигналами в одной и той же точке пространства для двух моментов времени, разделенных интервалом τ . Модуль функции $\gamma^{(1)}(r_1, r_1, \tau)$ определяет *степень временной когерентности*, которая равна $0 \leq |\gamma^{(1)}| \leq 1$. В случае когда временная когерентность полностью отсутствует, из (3.69) имеем $\gamma^{(1)}(\tau) = 0$ при $\tau > 0$. Для идеальной монохроматической волны: $V(r, t) = A(r) \exp(-i\omega t)$ из (3.69) получаем $|\gamma^{(1)}(\tau)| = 1$ при любых значениях τ . Это соответствует полной временной когерентности.

В общем случае функция $|\gamma^{(1)}(\tau)|$ представляет собой плавную зависимость от τ с максимумом при $\tau = 0$ [из выражений (3.67) и (3.69) следует, что для любых $V(r, t)$ $\gamma^{(1)}(\tau) = 1$ при $\tau = 0$]. Функция $|\gamma^{(1)}(\tau)|$ является симметричной функцией τ : $|\gamma^{(1)}(-\tau)| = |\gamma^{(1)}(\tau)|$.

Из рис. 3.31 можно определить характерное время τ_c , называемое *временем когерентности*. Это есть время, за которое величина $|\gamma^{(1)}(\tau)|$ уменьшается вдвое. Очевидно, для полностью когерентной волны $\tau_c = \infty$. Расстояние, которое проходит волна за время, равное

времени когерентности, называется *длиной когерентности*, т. е. $\Delta_c = c\tau_c$.

Временная когерентность связана со степенью монохроматичности излучения. Предположим, что $\gamma^{(1)}(\tau)$ описывается функцией Гаусса. Найдем спектр этой функции, для чего проведем преобразование Фурье (2.27). Фурье-образ от функции Гаусса также является функцией Гаусса с полушириной $\Delta\omega = 1/2\tau_c$. Это означает, что временная когерентность определяется шириной спектральной линии и для гауссова контура линии времени когерентности связано с полушириной спек-

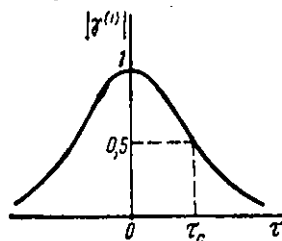


Рис. 3.31. Типичная зависимость степени временной когерентности от τ

ра соотношением

$$\tau_c = \frac{1}{2\Delta\omega}. \quad (3.70)$$

Чем меньше ширина спектральной линии, т. е. чем выше чистота спектра и степень монохроматичности, тем больше время когерентности. Для лазерной линии с $\Delta\omega/2\pi = 100$ Гц получаем $\tau_c \approx 10^{-8}$ с. Этому соответствует длина когерентности $\Delta\phi = c\tau_c \approx 300$ км. Отметим, что для обычных, нелазерных источников длина когерентности обычно не превышает нескольких миллиметров, т. е. $\tau_c < 10^{-10}$ с.

Аналогично определяется корреляционная функция первого порядка между двумя различными точками P_1 и P_2 с координатами r_1 и r_2 в один и тот же момент времени:

$$\Gamma^{(1)}(r_1, r_2, 0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T V(r_1, t) V^*(r_2, t) dt = \langle V(r_1, t) V(r_2, t) \rangle \quad (3.71)$$

и нормированная функция

$$\begin{aligned} \gamma^{(1)}(r_1, r_2, 0) &= \frac{\Gamma^{(1)}(r_1, r_2, 0)}{[\Gamma^{(1)}(r_1, r_1, 0) \Gamma^{(1)}(r_2, r_2, 0)]^{1/2}} = \\ &= \frac{\Gamma^{(1)}(r_1, r_2, 0)}{V \langle I(r_1) \rangle \langle I(r_2) \rangle}. \end{aligned} \quad (3.72)$$

Модуль функции $\gamma^{(1)}(r_1, r_2, 0)$ определяет *степень пространственной когерентности*. По аналогии с τ_c можно определить длину пространственной когерентности: если координата r_1 фиксирована, то на некотором характерном расстоянии $r = r_2 - r_1$ величина $\gamma^{(1)}$ как функция r_2 уменьшается вдвое.

Лазер, генерирующий на одной моде, дает излучение с полной пространственной когерентностью, т. е. для любых точек r_1 и r_2 $\gamma^{(1)}(r_1, r_2, 0) = 1$. Для протяженных нелазерных источников степень пространственной когерентности зависит от размеров источника и расстояния от источника до точки наблюдения, точнее от угла, под которым виден источник излучения из точки наблюдения. Если источник точечный, то его излучение всегда пространственно когерентно. Чем выше направленность пучка (меньше угол α), тем больше степень его пространственной когерентности.

Понятия временной и пространственной когерентности можно объединить, введя функцию взаимной когерентности

$$\Gamma^{(1)}(r_1, r_2, \tau) = \langle V(r_1, t + \tau) V(r_2, t) \rangle \quad (3.73)$$

и соответствующую ей нормированную функцию

$$\gamma^{(1)}(r_1, r_2, \tau) = \frac{\Gamma^{(1)}(r_1, r_2, \tau)}{[\Gamma^{(1)}(r_1, r_1, 0) \Gamma^{(1)}(r_2, r_2, 0)]^{1/2}}. \quad (3.74)$$

Модуль функции (3.74) является мерой степени когерентности между двумя различными точками волны в разные моменты времени.

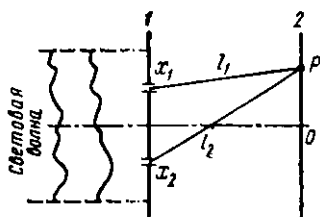


Рис. 3.32. Измерение степени когерентности электромагнитной волны с помощью интерферометра Юнга

Поясним понятие когерентности световых волн и смысл функции взаимной когерентности на следующем примере. Поместим на пути световой волны экран 1 с малыми отверстиями в точках x_1 и x_2 (рис. 3.32). За экраном 1 расположим экран 2, на котором свет, прошедший через оба отверстия, создает интерференционную картину (такое устройство называется интерферометром Юнга). Определим взаимную

когерентность световых колебаний в точках x_1 и x_2 .

Пусть в точке x_1 вектор напряженности электрического поля равен \mathcal{E}_1 , а в точке x_2 — \mathcal{E}_2 . Для простоты рассмотрим линейно поляризованный свет. Это позволяет использовать одну компоненту вектора \mathcal{E} . Тогда поле в точке P в момент времени t с точностью до постоянных множителей, зависящих от геометрии задачи, будет определяться суммой

$$\mathcal{E}(t) = \mathcal{E}_1(t-t_1) + \mathcal{E}_2(t-t_2),$$

где времена запаздывания $t_1 = l_1/c$ и $t_2 = l_2/c$. Интенсивность световой волны I в точке P пропорциональна $|\mathcal{E}(t)|^2$:

$$I = \langle [K_1 \mathcal{E}_1(x_1, t-t_1) + K_2 \mathcal{E}_2(x_2, t-t_2)] \times \\ \times [K_1 \mathcal{E}_1^*(x_1, t-t_1) + K_2 \mathcal{E}_2^*(x_2, t-t_2)]^* \rangle. \quad (3.75)$$

Множители K_1, K_2 зависят от размеров отверстий и обратно пропорциональны расстояниям l_1, l_2 . Звездочка *, как обычно, означает комплексное сопряжение.

Интенсивность света в точке P при закрытом отверстии x_2

$$I = \langle K_1 \mathcal{E}_1(t-t_1) K_1 \mathcal{E}_1^*(t-t_1) \rangle = K_1^2 \langle \mathcal{E}_1(t) \mathcal{E}_1^*(t) \rangle. \quad (3.76)$$

Аналогично при закрытом отверстии x_1

$$I_2 = \langle K_2 \mathcal{E}_2(t-t_2) K_2 \mathcal{E}_2^*(t-t_2) \rangle = \\ = K_2^2 \langle \mathcal{E}_2(t) \mathcal{E}_2^*(t) \rangle. \quad (3.77)$$

Обозначая $\tau = t_2 - t_1$ и учитывая, что смещение начала отсчета времени не влияет на результат усреднения, получаем

$$I = I_1 + I_2 + 2K_1 K_2 \operatorname{Re} \langle \mathcal{E}_1(x_1, t+\tau) \mathcal{E}_2^*(x_2, t) \rangle. \quad (3.78)$$

Здесь учтено, что

$$\begin{aligned} & \langle K_1 \mathcal{E}_1(t-t_1) K_2 \mathcal{E}_2^*(t-t_2) + K_1 \mathcal{E}_1^*(t-t_1) K_2 \mathcal{E}_2(t-t_2) \rangle = \\ & = K_1 K_2 \langle \mathcal{E}_1(t+\tau) \mathcal{E}_2^*(t) + \mathcal{E}_1^*(t+\tau) \mathcal{E}_2(t) \rangle = \\ & = 2K_1 K_2 \operatorname{Re} \{ \langle \mathcal{E}_1(x_1, t+\tau) \mathcal{E}_2^*(x_2, t) \rangle \}. \end{aligned}$$

Эффект взаимной корреляции рассматриваемых волн определяется последним слагаемым в (3.78) и нормированная функция взаимной когерентности будет иметь вид

$$\begin{aligned} \gamma^{(1)}(x_1, x_2, \tau) &= \frac{K_1 K_2 \langle \mathcal{E}_1(x_1, t+\tau) \mathcal{E}_2^*(x_2, t) \rangle}{\sqrt{I_1 I_2}} = \\ &= \frac{\langle \mathcal{E}_1(x_1, t+\tau) \mathcal{E}_2^*(x_2, t) \rangle}{\langle \mathcal{E}_1(x_1, t+\tau) \rangle \langle \mathcal{E}_2^*(x_2, t+\tau) \rangle}. \end{aligned} \quad (3.79)$$

Такой же результат можно получить с помощью соотношений (3.73) и (3.74), если заменить $V(r, t) \rightarrow \mathcal{E}(r, t)$.

Таким образом, интенсивность света в точке P

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \operatorname{Re} \{ \gamma^{(1)}(x_1, x_2, \tau) \}. \quad (3.80)$$

При полном отсутствии когерентности фазы $\varphi(r, t)$ в (3.66) для волн \mathcal{E}_1 и \mathcal{E}_2 меняются хаотически и усреднение за промежуток времени, существенно больший периода колебаний, даст нуль, ибо при беспорядочном изменении

фазы $\frac{1}{T} \int_0^T \cos(\varphi_2 - \varphi_1) dt = 0$. В этом случае $\gamma^{(1)} = 0$ и результирующая интенсивность есть просто сумма I_1 и I_2 , т. е. интерференционные эффекты отсутствуют.

В другом крайнем случае, когда разность фаз $\varphi_2 - \varphi_1$ остается неизменной, $\cos(\varphi_2 - \varphi_1)$ может изменяться от -1 до $+1$, что соответствует изменению интенсивности света от $I_{\min} = (\sqrt{I_1} - \sqrt{I_2})^2$ до $I_{\max} = (\sqrt{I_1} + \sqrt{I_2})^2$. При $I_1 = I_2$ ин-

тенсивность изменяется от $I_{\min} = 0$ до $I_{\max} = 4I_1$, т. е. $|\operatorname{Re} \gamma^{(1)}| = 1$ и наблюдается полная интерференция.

Таким образом, степень взаимной когерентности можно экспериментально оценить по контрасту интерференционной картины κ :

$$\kappa = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (3.81)$$

где I_{\min} и I_{\max} — интенсивности в минимуме и максимуме интерференционных полос.

Измерив вблизи выбранной точки экрана интенсивности

I_{\max} и I_{\min} , можно определить функцию $\gamma^{(1)}$ по формуле

$$\gamma^{(1)} = \frac{\kappa}{2} \left(\frac{I_1 + I_2}{\sqrt{I_1 I_2}} \right). \quad (3.82)$$

Для наблюдения только пространственной когерентности волны в точках x_1 и x_2 на рис. 3.32 необходимо положить $\tau=0$, т. е. $l_1=l_2$, 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, в 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. Направленность определяет расходимость светового пучка в пространстве и характеризуется плоским или телесным углом, в котором распространяется большая часть излучения. Расходимость пучка служит мерой его отклонения от параллельности.

«Классические» источники света обладают очень низкой направленностью излучения и обычно излучают в пределах телесного угла от 2π до 4π . 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 направленность 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. § 3.3). 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. 3.33). 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. 3.33, 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 concentрическими кольцами. Поскольку 84 % энергии излучения, проходящего через отверстие, заключено в центральном пятне, мерой дифракции и мерой расходимости может служить угол α_D , под которым из отверстия виден диаметр первого темного кольца. Вычисления дают,

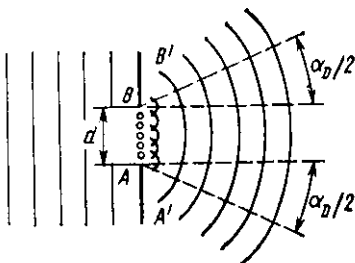


Рис. 3.33. Дифракционная расходимость плоской электромагнитной волны

что при $d \gg \lambda$ угол

$$\alpha_D = 2,44 \frac{\lambda}{d}, \quad (3.83)$$

где λ — длина волны, а d — диаметр отверстия.

Очевидно, дифракционная расходимость лазерного излучения будет определяться соотношением (3.83) для пучка диаметром d , имеющего «жесткие» края. Для гауссова пучка, который получается в резонаторе со сферическими зеркалами, распределение интенсивности в поперечном сечении представляется гладкой экспоненциальной кривой: с увеличением расстояния r от оси пучка интенсивность убывает по закону $\exp(-r/r_{1/e})^2$, где $r_{1/e}$ определяется формулой (3.44) (см. рис. 3.21). Как показывают расчеты, для такого пучка на больших расстояниях радиус пятна $r_{1/e}$ увеличивается линейно с координатой z : $r_{1/e} \approx \lambda z / \pi r_0$, где r_0 — радиус в шейке пучка внутри резонатора при $z=0$, т. е. $r_0 = (\lambda L / 2\pi)^{1/2}$. Поэтому угол дифракционной расходимости гауссова пучка определяется выражением

$$\alpha_D = \frac{2\lambda}{\pi r_0} \approx 1,27 \frac{\lambda}{d_0}, \quad (3.84)$$

где d_0 — диаметр пучка в наиболее узкой его части.

Из сравнения выражений (3.83) и (3.84) видно, что при одинаковых диаметрах расходимость гауссова пучка приблизительно вдвое меньше расходимости плоского пучка. Можно показать, что при прочих равных условиях гауссов пучок имеет минимально возможную дифракционную расходимость.

Если электромагнитная волна не обладает полной пространственной когерентностью, то ее расходимость будет больше определяемой соотношениями (3.83) и (3.84) для пространственно-когерентной волны. Это можно понять из рис. 3.33. Если волна не обладает пространственной когерентностью, то вторичные волны, испускаемые с поперечного сечения AB , не будут находиться в фазе и волновой фронт, образуемый в результате дифракции, будет иметь большую расходимость, чем α_D в формуле (3.83).

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 этих «нитей» или «горячих точек» a существенно меньше общего диаметра светового пуч-

ка d . Если отдельные пучки взаимно некоррелированы, то расходимость пучка в целом $\sim 2,44 \lambda/a$. Если бы такие пучки были коррелированы, то общий угол расходимости $\sim 2,44 \lambda/d$. Последнее эквивалентно множеству синхронно излучающих антенн.

Яркость. Яркость какого-либо источника электромагнитных волн характеризует мощность излучения, испускаемого с единицы поверхности в единичном телесном угле в направлении, перпендикулярном излучательной поверхности. Эта величина имеет размерность $(\text{Вт}/(\text{м}^2 \cdot \text{ср}))$ и называется также *энергетической яркостью*, или излучательной способностью. Энергетическая яркость является объективной физической характеристикой электромагнитного излучения. Кроме энергетической яркости вводится понятие *фотометрической яркости*, или просто 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) в качестве стандартной кривой видности. Она приведена на рис. 3.34 и выражает чувствительность глаза к свету с различными длинами волн.

Таким образом, переход от энергетических величин к фотометрическим осуществляется 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 B_v связана с энергетической яркостью B_E соотношением

$$B_{v\lambda} = K_\lambda B_{E\lambda}. \quad (3.85)$$

Для немонохроматического света связь между B_v и B_E ,

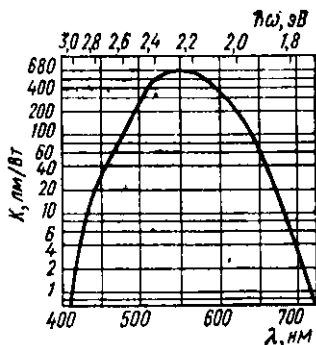


Рис. 3.34. Кривая видности для дневного зрения, утвержденная Международной комиссией по освещению

очевидно, будет определяться интегралом

$$B_v = \int B_{v\lambda} d\lambda = \int K_\lambda B_{E\lambda} d\lambda. \quad (3.86)$$

Аналогично полный световой (фотометрический) поток Φ_v источника света связан со спектральной плотностью потока излучения соотношением

$$\Phi_v = \int K_\lambda \Phi_\lambda d\lambda. \quad (3.87)$$

Для нормального зрения максимум функции видности приходится на длину волны $\lambda = 555$ нм (зеленый свет). На этой длине волны, как видно из рис. 3.34, мощность излучения 1 Вт эквивалентна световому потоку 680 лм, т. е. $K_{\lambda=555 \text{ нм}} = K_{\max} = 680 \text{ лм/Вт}$. При $\lambda_{\min} = 380$ нм и $\lambda_{\max} = 780$ нм функция видности уменьшается почти до нуля. Эти длины волн приняты за фиолетовую и красную границы видимой области спектра.

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. Если, например, уменьшить диаметр лазерного пучка от D_1 до D_2 с помощью двух линз, фокусы которых совпадают (при этом $D_2/D_1 = f_2/f_1$), то интенсивность пучка увеличится, но расходимость выходного пучка ($\sim \lambda/D_2$) будет больше расходимости входного ($\sim \lambda/D_1$), так что яркость останется неизменной (если пренебречь потерями). 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.

Для оценки яркости лазерного излучения рассмотрим малоомощный гелий-неоновый газовый лазер, генерирующий в красной области спектра на длине волны $0,63 \text{ мкм} = 630 \text{ нм}$ мощность 10^{-3} Вт . Диаметр пучка на выходном зеркале равен 3 мм . Угол расходимости определяется дифракционными эффектами, и из (3.83) получаем $\alpha_D = 2,44 \frac{0,63 \cdot 10^{-6}}{10^{-3}} \approx \approx 10^{-3} \text{ рад}$. Соответствующий телесный угол пучка равен 10^{-6} ср . Выходной мощности 1 мВт на длине волны 630 нм соответствует $K_\lambda = 180 \text{ лм/Вт}$, т. е. световой поток $0,18 \text{ лм}$. Разделив эту величину на площадь источника и телесный угол, получим значение яркости $B_v \approx 2,5 \cdot 10^{10} \text{ кд/м}^2$. Для того чтобы оценить эту величину, укажем, что средняя яркость солнечного диска за пределами земной атмосферы составляет $2 \cdot 10^9 \text{ кд/м}^2$. Это означает, что яркость излучения, генерируемого даже малоомощным 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 B_λ будет на 10 — 12 порядков превышать спектральную яркость Солнца.

ГЛАВА 4. ОПТИЧЕСКИЕ ЯВЛЕНИЯ В ТВЕРДЫХ ТЕЛАХ

Современная оптоэлектроника, как и «обычная» электроника, является твердотельной. Это означает, что работа как отдельных элементов, так и оптоэлектронных систем в целом построена на основе различных оптических явлений, протекающих в твердых телах. Анализ этих явлений базируется на рассмотрении процессов взаимодействия электромагнитного излучения с атомными системами, проведенном в гл. 2.

§ 4.1. ОПТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВЕЩЕСТВА И ЭЛЕМЕНТЫ КРИСТАЛЛООПТИКИ

Распространение электромагнитных волн в веществе полностью описывается в рамках макроскопической электромагнитной теории с помощью уравнений Максвелла (2.4) и материальных уравнений (2.5). Уравнения (2.5) по существу отражают реакцию среды на электромагнитное возмущение.

Относительную магнитную проницаемость μ_r в оптическом диапазоне частот для большинства применяемых в квантовой и оптической электронике материалов можно положить равной единице. Кроме того, обычно объемная плотность зарядов $\rho=0$, что хорошо выполняется для однородных материалов.

Рассмотрим сначала однородную изотропную среду, оптические свойства которой не зависят от направления распространения и поляризации электромагнитной волны. Для такой среды при $\mu_r=1$ и $\rho=0$ систему уравнений (2.5) можно переписать в виде

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}, \quad \mathbf{B} = \mu_0 \mathcal{H}, \quad \mathbf{j} = \sigma \mathbf{E}, \quad (4.1)$$

где относительная диэлектрическая проницаемость ϵ_r и удельная электрическая проводимость σ являются скалярными величинами. Отметим, что здесь σ есть проводимость на оптической частоте, которая не равна проводимости на низкой или нулевой частоте. При тех же условиях уравнения Максвелла (2.4) переписутся в виде

$$\begin{aligned} \operatorname{rot} \mathbf{E} &= -\mu_0 \frac{\partial \mathcal{H}}{\partial t}, \quad \operatorname{div} \mathbf{E} = 0, \\ \operatorname{rot} \mathcal{H} &= \sigma \mathbf{E} + \epsilon_0 \epsilon_r \frac{\partial \mathbf{E}}{\partial t}, \quad \operatorname{div} \mathcal{H} = 0. \end{aligned} \quad (4.2)$$

Выполняя преобразования, аналогичные проведенным в § 2.1, получаем волновое уравнение, например, для вектора \mathbf{E} :

$$\Delta \mathbf{E} - \sigma \mu_0 \frac{\partial \mathbf{E}}{\partial t} - \epsilon_0 \epsilon_r \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0. \quad (4.3)$$

Оно отличается от волнового уравнения (2.8) для электромагнитной волны в вакууме дополнительным вторым слагаемым и множителем ϵ_r 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.

Решением уравнения (4.3) для одной из компонент вектора \mathbf{E} будет

$$E_x = E_0 e^{i\omega \left(t - \frac{\hat{n}z}{c} \right)}. \quad (4.4)$$

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 (4.4) удовлетворяет уравнению (4.3) при условии

$$(\hat{n})^2 = c^2 \left(\epsilon_0 \epsilon_r \mu_0 - i \frac{\sigma \mu_0}{\omega} \right). \quad (4.5)$$

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

$$\hat{n} = n - i\kappa. \quad (4.6)$$

Подставляя (4.6) в (4.4), получаем

$$\mathcal{E}_x = \mathcal{E}_0 e^{i \left(\omega t - \frac{\omega n z}{c} \right)} e^{-\frac{\omega \kappa z}{c}}. \quad (4.7)$$

Это выражение описывает волну с угловой частотой ω , распространяющуюся с фазовой скоростью c/n и испытывающую ослабление. Первый множитель определяет амплитуду волны в точке $z=0$, второй — фазу волны в точке z в момент времени t , а третий — затухание волны с расстоянием z .

На практике прибор измеряет всегда не амплитуду и фазу волны, а ее интенсивность или мощность, которые пропорциональны $|\mathcal{E}|^2$. Из (4.7) имеем

$$I \sim |\mathcal{E}|^2 = \mathcal{E}_0^2 e^{-\frac{2\omega\kappa}{c} z}. \quad (4.8)$$

Таким образом, коэффициент κ характеризует свойства среды ослаблять электромагнитное излучение и называется *главным показателем поглощения* *. Сравнение выражения (4.8) с (3.4) показывает, что величина, стоящая в показателе экспоненты перед z , есть не что иное, как введенный ранее показатель поглощения k_ω .

Отсюда

$$k_\omega = \frac{2\omega\kappa}{c} = \frac{4\pi\nu}{c} \kappa = \frac{4\pi}{\lambda} \kappa. \quad (4.9)$$

Естественно, что и главный показатель поглощения κ и показатель поглощения k_ω зависят от частоты (длины волны) электромагнитного излучения. Эта зависимость определяет спектр поглощения вещества.

Наряду с комплексным показателем преломления \hat{n} оптические свойства твердых тел можно характеризовать комплексной относительной диэлектрической проницаемостью

$$\hat{\epsilon}_r = (\hat{n})^2 = (n - i\kappa)^2 = n^2 - \kappa^2 - 2in\kappa. \quad (4.10)$$

* См. сноску на с. 114.

С учетом соотношения $\epsilon_0\mu_0=1/c^2$ из (4.5) получаем

$$\hat{\epsilon}_r = \epsilon - i \frac{\sigma}{\epsilon_0\omega}. \quad (4.11)$$

Разобьем $\hat{\epsilon}_r$ по аналогии с (4.6) на вещественную и мнимую части:

$$\hat{\epsilon}_r = \epsilon_1 - i\epsilon_2. \quad (4.12)$$

Сравнение соотношений (4.10), (4.11) и (4.12) дает простую связь между n и κ , с одной стороны, вещественной ϵ_1 и мнимой ϵ_2 частями комплексной относительной диэлектрической проницаемости — с другой:

$$\epsilon_1 = \epsilon_r = n^2 - \kappa^2, \quad (4.12a)$$

$$\epsilon_2 = \frac{\sigma}{\epsilon_0\omega} = 2n\kappa. \quad (4.12b)$$

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

Таким образом, мнимая часть κ комплексного показателя преломления характеризует способность среды поглощать (или усиливать при создании инверсии населенностей) электромагнитное излучение. На микроскопическом уровне процессы поглощения определяются квантовыми переходами из более низкого в более высокое энергетическое состояние и были рассмотрены ранее.

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 соотношений (4.12). 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 показа-

теля 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 сдвиг фазы (запаздывание) первоначальной волны. Вследствие того что суммарный сдвиг фазы 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 κ не являются полностью независимыми. Связь между ними имеет довольно глубокую основу и может быть выведена из принципа причинности. Эта связь устанавливается соотношением Крамерса — Кронига:

$$n(\omega) - 1 = \frac{2}{\pi} \int_0^{\infty} \frac{\omega'}{(\omega')^2 - \omega^2} \kappa(\omega') d\omega', \quad (4.13)$$

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

Рассмотрим отражение и преломление света. Задача о прохождении электромагнитной волны через границу раздела двух сред с показателями преломления n_1 и n_2 решается в курсах электродинамики. Напомним основные результаты.

При падении пучка света с интенсивностью I_0 на границу раздела двух сред часть его I_R отражается, а часть I_T проходит. При этом:

1) угол отражения θ_R равен углу падения θ_1 :

$$\theta_R = \theta_1; \quad (4.14)$$

2) произведение $n \sin \theta$ одинаково как для падающего (индекс 1), так и для преломленного (индекс 2) лучей:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2; \quad (4.15)$$

3) интенсивность отраженного света I_R зависит как от угла падения θ_1 , так и от направления поляризации. Эти зависимости устанавливаются соотношениями Френеля. Для электромагнитной волны I_{\perp} , вектор \mathbf{E} которой перпендикулярен плоскости падения, коэффициент отражения по интенсивности

$$R_{\perp} = \frac{I_R}{I_{\perp}} = \frac{\sin^2 (\theta_1 - \theta_2)}{\sin^2 (\theta_1 + \theta_2)}. \quad (4.16)$$

Для волны I_{\parallel} , вектор \mathbf{E} которой параллелен плоскости падения, коэффициент отражения

$$R_{\parallel} = \frac{I_R}{I_{\parallel}} = \frac{\operatorname{tg}^2 (\theta_1 - \theta_2)}{\operatorname{tg}^2 (\theta_1 + \theta_2)}; \quad (4.17)$$

4) при нормальном падении луча ($\theta_1 = 0$) для любой поляризации

$$R = \frac{I_R}{I_1} = \frac{(n_2 - n_1)^2 + \kappa^2}{(n_2 + n_1)^2 + \kappa^2}. \quad (4.18)$$

Из (4.17) видно, что при $\theta_1 + \theta_2 = \pi/2$ $R_{\parallel} = 0$. При этом $\operatorname{tg} \theta_1 = n_2/n_1$ и этот угол называется углом Брюстера. Зависимости коэффициента отражения для полупроводникового кристалла арсенида галлия ($n_2 = 3,5$, $n_1 = 1$) от угла падения, рассчитанные по формулам (4.16), (4.17), приведены на рис. 4.1.

Отметим, что когда угол $\theta_1 + \theta_2$ проходит через $\pi/2$, фаза отраженной волны I_R резко изменяется на противоположную, проходя через нуль.

Зеркала, работающие под углом Брюстера и изготовленные из пластин германия, кремния, арсенида галлия и других материалов с большими показателями преломления, могут служить хорошими инфракрасными поляризаторами.

Ранее рассматривалось распространение света в оптически изотропных средах, когда диэлектрическую проницаемость ϵ можно было считать скалярной величиной. Для анизотропных сред связь между компонентами векторов электрической индукции \mathbf{D} и напряженности электрического поля \mathbf{E} задается в общем случае более сложными соот-

ношениями:

$$\begin{aligned} D_1 &= \epsilon_{11}\epsilon_0\mathcal{E}_1 + \epsilon_{12}\epsilon_0\mathcal{E}_2 + \epsilon_{13}\epsilon_0\mathcal{E}_3, \\ D_2 &= \epsilon_{21}\epsilon_0\mathcal{E}_1 + \epsilon_{22}\epsilon_0\mathcal{E}_2 + \epsilon_{23}\epsilon_0\mathcal{E}_3, \\ D_3 &= \epsilon_{31}\epsilon_0\mathcal{E}_1 + \epsilon_{32}\epsilon_0\mathcal{E}_2 + \epsilon_{33}\epsilon_0\mathcal{E}_3. \end{aligned} \quad (4.19)$$

Девять величин ϵ_{ij} являются постоянными среды (на самом деле они зависят от частоты ω и температуры T) и составляют *тензор диэлектрической проницаемости*. Вектор D равен произведению этого тензора на вектор \mathcal{E} . Из (4.19) следует, что в общем случае векторы D и \mathcal{E} неколлинеарны.

Соотношения (4.19) можно записать в более компактной форме:

$$D_i = \sum_j \epsilon_{ij}\epsilon_0\mathcal{E}_j, \quad (4.20)$$

или

$$D_i = \epsilon_{ij}\epsilon_0\mathcal{E}_j. \quad (4.20a)$$

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.

Изучением распространения света в анизотропных средах занимается *кристаллооптика*. Знание ее законов необходимо в оптоэлектронике для решения задач электрооптической, магнитооптической и акустооптической модуляции света, при рассмотрении условий фазового синхронизма, нелинейных оптических эффектов и т. д.

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 , z , для которых справедливы соотношения

$$D_x = \epsilon_x\epsilon_0\mathcal{E}_x, \quad D_y = \epsilon_y\epsilon_0\mathcal{E}_y, \quad D_z = \epsilon_z\epsilon_0\mathcal{E}_z. \quad (4.21)$$

Другими словами, выбором системы координат матрица

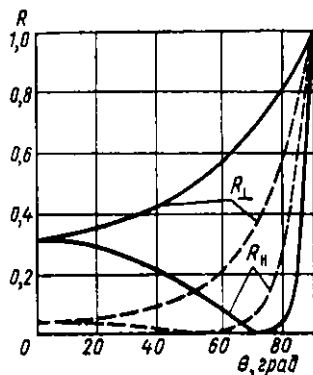


Рис. 4.1. Зависимости коэффициента отражения R от угла падения θ для арсенида галлия ($n=3,5$, сплошные кривые) и кварца ($n=1,5$, пунктирные кривые) для двух поляризаций падающего излучения

ϵ_r может быть диагонализирована и приведена к виду

$$\epsilon_r = \begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix}. \quad (4.22)$$

Система координат, в которой тензор ϵ_r диагонализирован, называется *главной системой координат*, а ее 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 $\epsilon_x = \epsilon_y = \epsilon_z$, т. е. они являются оптически изотропными. В кристаллах с гексагональной, тригональной и тетрагональной сингониями $\epsilon_x = \epsilon_y \neq \epsilon_z$. В кристаллах, имеющих более низкую симметрию, $\epsilon_x \neq \epsilon_y \neq \epsilon_z$.

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 $\hat{\epsilon}$ вещественна).

В общем случае плотность электрической энергии в электромагнитной волне

$$\rho_g = \mathcal{E} D / 2. \quad (4.23)$$

В главной системе координат

$$\begin{aligned} 2\rho_g &= \epsilon_x \epsilon_0 \mathcal{E}_x^2 + \epsilon_y \epsilon_0 \mathcal{E}_y^2 + \epsilon_z \epsilon_0 \mathcal{E}_z^2 = \\ &= \frac{D_x^2}{\epsilon_0 \epsilon_x} + \frac{D_y^2}{\epsilon_0 \epsilon_y} + \frac{D_z^2}{\epsilon_0 \epsilon_z}, \end{aligned} \quad (4.24)$$

откуда следует, что поверхность постоянной плотности энергии ρ_g в пространстве D_x, D_y, D_z имеет форму эллипсоида.

Действительно, если обозначить вектор $\frac{1}{\sqrt{2\rho_g \epsilon_0}} D$ через a , то уравнение для поверхности постоянной плотности энергии (4.24) переписывается в виде

$$\frac{a_x^2}{n_x^2} + \frac{a_y^2}{n_y^2} + \frac{a_z^2}{n_z^2} = 1. \quad (4.25)$$

Это есть уравнение эллипсоида с главными осями вдоль направлений x, y, z . Длины полуосей эллипсоида равны со-

ответственно 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 \mathbf{k} , и групповая (лучевая) скорость, определяемая вектором Пойнтинга \mathbf{S} , совпадают. В анизотропных кристаллах этого нет. Чтобы показать это, предположим, что в кристалле распространяется плоская монохроматическая волна, характеризуемая фазовым множителем $e^{i\omega(t - \frac{\mathbf{n}}{c} \cdot \mathbf{r})}$, где \mathbf{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 $\mathbf{k} = \frac{\omega n}{c} \mathbf{n}$, а ее фазовая скорость

$$\mathbf{v}_\phi = \frac{c}{n} \mathbf{n}.$$

Найдем для такой волны вектор Пойнтинга, характеризующий направление распространения потока энергии. При этом учтем, что для плоской монохроматической волны дифференцирование по времени $\partial/\partial t$ сводится к умножению на $i\omega$, а действие оператора Δ (дифференцирование по координатам) формально можно заменить векторным умножением на $-\frac{i\omega n}{c} \mathbf{n}$. Благодаря этому уравнения Максвелла $\text{rot} \mathcal{H} = \partial \mathbf{D} / \partial t$ и $\text{rot} \mathcal{E} = -\partial \mathbf{B} / \partial t$ преобразуются к виду

$$\mathbf{D} = -\frac{n}{c} [\mathbf{n} \times \mathcal{H}], \quad \mathcal{H} = \frac{n}{\mu_0 c} [\mathbf{n} \times \mathcal{E}].$$

Эти уравнения показывают, что вектор \mathbf{D} перпендикулярен \mathcal{H} и оба вектора перпендикулярны \mathbf{n} . Вместе с тем вектор Пойнтинга $\mathbf{S} = [\mathcal{E} \times \mathcal{H}]$ перпендикулярен векторам \mathcal{E} и \mathcal{H} . Фазовая скорость света \mathbf{v}_ϕ направлена по \mathbf{k} , а групповая скорость \mathbf{v}_{gr} по \mathbf{S} . Связь между величинами $|\mathbf{v}_{gr}|$ и $|\mathbf{v}_\phi|$ определяется углом β между векторами \mathbf{D} и \mathcal{E} :

$$|\mathbf{v}_{gr}| = |\mathbf{v}_\phi| / \cos \beta.$$

В оптически изотропных средах (например, в кубических кристаллах, для которых $\epsilon_x = \epsilon_y = \epsilon_z = \epsilon$) векторы \mathbf{D} и \mathcal{E} коллинеарны и фазовая скорость совпадает с групповой, т. е. $\mathbf{v}_\phi = \mathbf{v}_{gr}$. В анизотропных средах в общем случае направле-

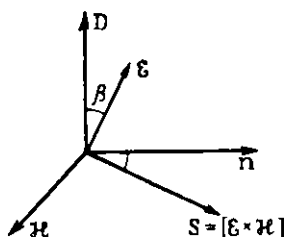


Рис. 4.2. Относительная ориентация векторов и направление распространения электромагнитной энергии в анизотропном кристалле (векторы D , E , n и S лежат в одной плоскости)

ние потока энергии неколлинеарно направлению распространения фазы, т. е. $v_{\text{ф}} \neq v_{\text{гр}}$. Этот случай поясняет рис. 4.2.

Подчеркнем, что в анизотропных кристаллах понятие «показатель преломления» не имеет такого простого отношения к закону преломления (2.3), 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 к волновому фронту, а не направление потока энергии. В кристаллах кубической симметрии ($n_x = n_y = n_z = n$) оптическая индикатриса представляет собой сферу. Любое ее 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.

В одноосных кристаллах уравнение эллипсоида пока-

телей преломления (4.25) упрощается:

$$\frac{a_x^2}{n_0^2} + \frac{a_y^2}{n_0^2} + \frac{a_z^2}{n_e^2} = 1. \quad (4.26)$$

Здесь оптическая ось совпадает с осью симметрии кристалла и направлена вдоль z . 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, при $n_e > n_0$ — оптически положительным.

Названия «обыкновенный» и «необыкновенный» соответствуют различному поведению лучей в кристалле. Как указывалось, в анизотропной среде в произвольном направлении распространяются две линейно поляризованные волны. Это означает, что имеется две поверхности волновых векторов. В одноосном 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. Это обуславливает особенности преломления света: при прохождении границы раздела в одноосном кристалле падающий луч раздваивается. Первый луч — обыкновенный — ведет себя аналогично обычным преломленным волнам в изотропной среде (для него поверхность волновых векторов есть сфера). Второй луч — необыкновенный — аномален: в общем случае он лежит не в плоскости падения. Двойное лучепреломление не испытывают только волны, распространяющиеся вдоль оптической оси кристалла.

§ 4.2. ЭЛЕКТРООПТИЧЕСКИЕ, МАГНИТООПТИЧЕСКИЕ И УПРУГООПТИЧЕСКИЕ ЭФФЕКТЫ

Под электрооптическими, магнитооптическими и упругооптическими 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 для данного кристалла являются постоянными лишь при определенных условиях. Воздействие внешних механических напряжений, электрических и магнитных полей может приводить 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 диэлектрической проницаемости ϵ тензор диэлектрической непроницаемости a , компоненты a_{ij} которого обратны компонентам тензора ϵ_{ij} :

$$a_{ij} = 1/\epsilon_{ij} = 1/n^2. \quad (4.27)$$

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 $\mathcal{E} = 0$

$$a_{11} = 1/\epsilon_{11} = 1/\epsilon_x = 1/n_x^2, \quad a_{22} = 1/\epsilon_{22} = 1/\epsilon_y = 1/n_y^2, \quad a_{33} = 1/\epsilon_{33} = 1/\epsilon_z = 1/n_z^2.$$

При наложении электрического поля эллипсоид оптической индикатрисы (4.25) поворачивается и деформируется. Главные оси этого эллипсоида в общем случае не будут совпадать с главными осями исходного эллипсоида, в системе координат xyz которого уравнение оптической индикатрисы принимает вид

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{31}zx = 1. \quad (4.28)$$

Изменение коэффициентов оптической индикатрисы под воздействием внешнего электрического поля в общем случае будет описываться соотношением

$$\Delta a_{ij} = r_{ijk} \mathcal{E}_k + R_{ijkl} \mathcal{E}_k \mathcal{E}_l + \dots, \quad (4.29)$$

где \mathcal{E}_k и \mathcal{E}_l — компоненты вектора напряженности внешнего поля \mathcal{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 трехиндексовой системы обозначений r_{ijk} используют двухиндексовую (матричную) систему, заменяя два индекса, по которым тензор симметричен, одним со следующим соответствием: 11 \rightarrow 1; 22 \rightarrow 2; 33 \rightarrow 3; 23 \rightarrow 4; 31 \rightarrow 5; 12 \rightarrow 6. Тогда линейное изменение поляризационных констант под воздействием внешнего электрического поля \mathcal{E} будет определяться выражением

$$\Delta a_i = \sum_{j=1}^6 r_{ij} \mathcal{E}_j, \quad (4.30)$$

где индекс суммирования j означает 1=x; 2=y и 3=z. Уравнение (4.30) можно записать в матричной форме:

$$\begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta a_3 \\ \Delta a_4 \\ \Delta a_5 \\ \Delta a_6 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{bmatrix} \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \mathcal{E}_3 \end{bmatrix}. \quad (4.31)$$

Матрица 6×3 с элементами r_{ij} есть матрица электрооптических коэффициентов. Зная правила перемножения матриц (4.31) в виде (4.30), находим, например, $\Delta a_4 = r_{41} \mathcal{E}_1 + r_{42} \mathcal{E}_2 + r_{43} \mathcal{E}_3$. Вид матрицы r_{ij} , но не численные 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 $r_{41}=r_{52}$. Для кристаллов типа KDP (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 матрицу, элементами которой являются константы эллипсоида (4.28).

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 \mathcal{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 — линей-

ный (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:

$$\varphi = -VIB, \quad (4.32)$$

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.

Упругооптические эффекты. В твердых телах под действием упругих механических напряжений изменяется показатель преломления вещества. Это явление называется *фотоупругостью*, а связанные с ним эффекты — 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

$$\Delta a_i = \sum_j \pi_{ij} \sigma_j. \quad (4.33)$$

Здесь π_{ij} — пьезооптические коэффициенты, которые являются компонентами соответствующего тензора четвертого ранга и хорошо известны из кристаллографии.

Упругооптические 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 упругооптический эффект наблюдается во всех кристаллах и аморфных средах. В частности, он заметно проявляется в стеклах, плавленом 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:

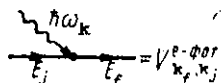
$$\Delta n = n_o - n_e = K\sigma, \quad (4.34)$$

где коэффициент пропорциональности K зависит от свойств вещества.

§ 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 (1.62). 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



$$E_j \xrightarrow{\hbar\omega_k} E_f = V_{k_f, k_j}^{e-фот} \quad (4.35)$$

Здесь $V_{k_f, k_j}^{e-фот}$ — матричный элемент электрон-фотонного

взаимодействия для состояний с волновыми векторами k_j и k_f .

Рассмотрим условия, при которых переход (4.35) возможен, т. е. возможен непосредственный оптический переход $k_j \rightarrow k_f$ при взаимодействии только с фотоном. Как и ранее (см. § 2.2), матричный элемент (4.35)

$$V_{k_f, k_j}^{e\text{-фот}} = \int \psi_{k_f} V \psi_{k_j} d\mathbf{r}. \quad (4.36)$$

Возмущение V в нашем случае можно записать в виде плоской монохроматической волны:

$$V = A e^{i k_{\text{фот}} \mathbf{r}} = A e^{i \frac{2\pi}{\lambda} (\mathbf{n} \mathbf{r})}, \quad (4.37)$$

где $k_{\text{фот}}$ — 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 (1.62), т. е.

$$\psi_{k_j} = e^{i k_j \mathbf{r}} u_{k_j}(\mathbf{r}), \quad (4.38)$$

$$\psi_{k_f} = e^{i k_f \mathbf{r}} u_{k_f}(\mathbf{r}), \quad (4.38a)$$

где $u_{k_j}(\mathbf{r})$ и $u_{k_f}(\mathbf{r})$ — периодические функции с периодом решетки.

Подставляя (4.37) и (4.38) в (4.36), нетрудно убедиться, что подынтегральная функция в матричном элементе будет содержать множитель $\exp[i(-k_f + k_j + k_{\text{фот}})\mathbf{r}]$, который быстро осциллирует в пространстве, за исключением случая $k_j + k_{\text{фот}} - k_f = 0$. 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 $j \rightarrow f$:

$$k_j + k_{\text{фот}} = k_f, \quad (4.39)$$

где $k_{\text{фот}}$ — волновой вектор фотона ($k_{\text{фот}} = \frac{2\pi}{\lambda} \mathbf{n}$).

Соотношение (4.39) аналогично закону сохранения импульса в классической физике. Действительно, если обе части выражения (4.39) умножить на \hbar , то получим правило отбора по квазиимпульсу:

$$p_j + p_{\text{фот}} = p_f. \quad (4.40)$$

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 (4.39), (4.40) выражают условия, при которых матричный элемент $V_{k_f k_j}^{\text{фот}}$ будет отличен от нуля. Для осуществления реального перехода, как отмечалось в § 2.2 [см. соотношение (2.57)], необходимо выполнение закона сохранения энергии. В нашем случае для непосредственного перехода $j \rightarrow f$ с поглощением фотона $\hbar\omega$ это приводит к условию*

$$E_f + \hbar\omega = E_j.$$

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:

$$E_f \pm \hbar\omega = E_j, \quad (4.41)$$

$$\mathbf{k}_f \pm \mathbf{k}_{\text{фот}} = \mathbf{k}_j. \quad (4.42)$$

Знак плюс соответствует поглощению, а знак минус — испусканию. Поскольку волновой вектор \mathbf{k} в кристаллах определен с точностью до вектора обратной решетки \mathbf{a} , соотношение (4.42) справедливо с точностью до $2\pi/|\mathbf{a}|$.

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 $\mathbf{k}_{\text{фот}}$ и $\mathbf{p}_{\text{фот}}$ в соотношениях (4.39), (4.40) можно пренебречь. Тогда в соответствии с законом сохранения (4.42) при взаимодействии только с фотоном могут осуществляться только переходы без изменений волнового вектора:

$$\mathbf{k}_j = \mathbf{k}_f. \quad (4.43)$$

* Это условие легко получается из (4.36), если вместо волновых функций $\psi_{\mathbf{k}}$ записать волновые функции $\varphi_{\mathbf{k}}$, зависящие от времени:

$$\varphi_{\mathbf{k}_j} = e^{i \frac{E_j}{\hbar} t} \psi_{\mathbf{k}_j}(\mathbf{r}), \text{ и учесть временной множитель в (4.37).}$$

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. 4.3, а). Диаграмма взаимодействия для подобного перехода может быть аналогичной диаграмме (4.35), 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 (4.35). При таком описании трудно проследить 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. 4.3, а). Тогда волновая функция для состояния с волновым вектором k в валентной зоне может быть записана аналогично (2.37):

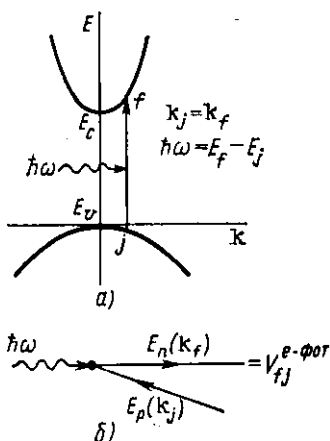


Рис. 4.3. Прямой оптический переход из j -состояния валентной зоны в f -состояние зоны проводимости с поглощением фотона $\hbar\omega$ (а) и соответствующая диаграмма взаимодействия (б)

$$\begin{aligned}\varphi_{vk} &= \psi_{vk}(\mathbf{r}) e^{-\frac{i}{\hbar} [E_v(0) - E_v(k)] t} = \\ &= \psi_{vk}(\mathbf{r}) e^{-\frac{i}{\hbar} [E_v(k) - E_v(0)] (-t)} = \psi_{vk}(\mathbf{r}) e^{\frac{i}{\hbar} E_p(k) t}, \quad (4.44)\end{aligned}$$

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

$$\varphi_{ck} = \psi_{ck} e^{-\frac{i}{\hbar} E_n(k)t} \quad (4.45)$$

Функции φ_{ok} и ψ_{ck} являются блоховскими функциями (4.38).

Из сравнения временных множителей в (4.44) и (4.45) 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. 4.3, б. До того как фотон поглотился, валентная зона была полностью заполнена, а зона проводимости — пустая. Следовательно, до акта взаимодействия имелаась 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. В соответствии с этим дырочная линия направлена в сторону уменьшения времени, и на графике в явном виде присутствуют электронная и дырочная линии, т. е. можно наглядно проследить за более сложными процессами взаимодействия.

Из рассмотрения следует, что 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, то непосредственный оптический переход $j \leftrightarrow f$ между этими состояниями будет невозможен, поскольку для них $V_{jf} = 0$.

Как указывалось в § 2.2, равенство $V_{jf} = 0$ 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 про-

межуточное виртуальное состояние. В этом случае необходимо рассмотреть второе слагаемое в выражении (2.51) для матричного элемента перехода, т. е. процесс, описываемый диаграммой (2.59). 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 (4.42) вытекает из условия трансляционной симметрии кристалла, поскольку волновые функции электронов представлены в виде блоховских функций (4.38). Различные нарушения трансляционной симметрии, определяемые дефектами, будут выполнять роль рассеивающих центров и приводить к возможности «нарушения» закона сохранения (4.42).

Таким образом, для реализации оптического перехода из j -состояния 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 $\hbar\Omega_k$. Этот процесс разрешен во втором приближении теории возмущений и ему соответствует следующая диаграмма взаимодействия:

$$E_j, \mathbf{k}_j \quad E_l, \mathbf{k}_l \quad E_f, \mathbf{k}_f \quad \frac{V_{lj}^{e-\text{фот}} V_{fl}^{e-\text{фон}}}{(E_l - E_j - \hbar\omega)} \quad (4.46)$$

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 $V^{e-\text{фон}}$ учитывает электрон-фононное взаимодействие, 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,

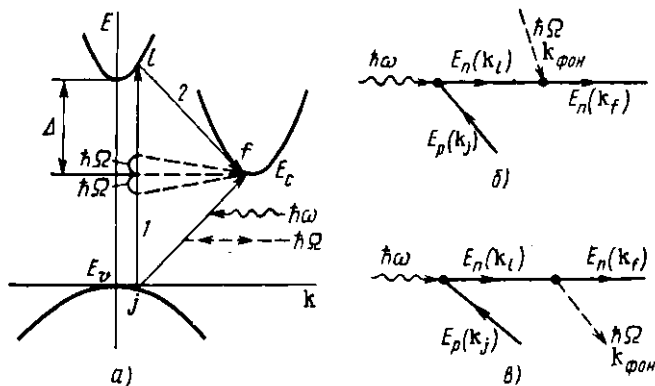


Рис. 4.4. Непрямой оптический переход из j -состояния валентной зоны в f -состояние зоны проводимости при поглощении фотона $\hbar\omega$ и рассеянии на фононе:

a — изображение на диаграмме $E(k)$; $б, в$ — диаграммы взаимодействия для процессов поглощения фотона соответственно при одновременном поглощении и испускании фонона с энергией $\hbar\Omega$ и волновым вектором $k_{\text{фон}} = k_f - k_j$

то процесс поглощения фотона $\hbar\omega$ будет изображаться, как на рис. 4.4. 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. 4.4, a). Время δt пребывания электрона в этом промежуточном состоянии чрезвычайно мало, так что из соотношения неопределенностей $\delta E \delta t \geq \hbar$ 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. 4.4, 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 $k_{\text{фон}}$, соответствующие законы сохранения при оптических переходах с участием фононов, как видно из рис. 4.4, переписываются в виде

$$E_j + \hbar\omega \pm \hbar\Omega_{\text{фон}} = E_f, \quad (4.47)$$

$$k_j \pm k_{\text{фон}} = k_f. \quad (4.48)$$

Знаки в этих выражениях соответствуют поглощению и

испусканию фонона. Им отвечают диаграммы взаимодействия, изображенные на рис. 4.4, б, в.

Оптические переходы между состояниями с различными значениями волнового вектора называются *непрямыми* (рис. 4.4). Они разрешены лишь во втором приближении теории возмущений, поэтому вероятность таких переходов существенно меньше, чем прямых.

Заметим, что в матричных элементах для не прямых 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 переходом 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 собственное межзонное фундаментальное поглощение и собственное решеточное поглощение. В качестве примера на рис. 4.5 представлены спектры собственного поглощения нелегированных кристаллов кремния (Si), германия (Ge) и арсенида галлия (GaAs) в широком диапазоне энергий фотонов (длины 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. 4.6. Рассмотрим особенности перечисленных механизмов поглощения.

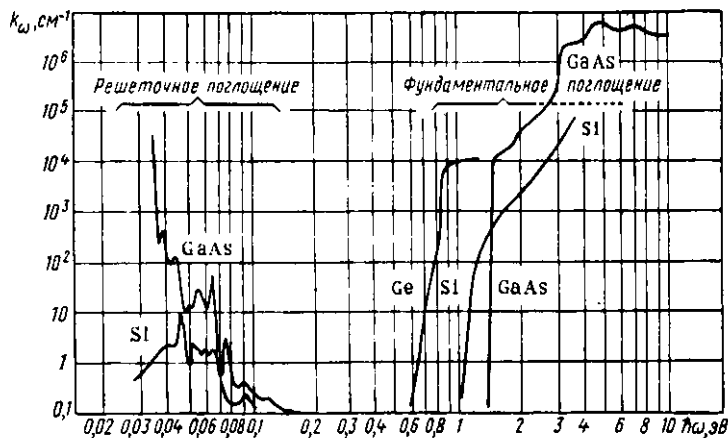


Рис. 4.5. Спектры собственного поглощения нелегированных кристаллов кремния, германия и арсенида галлия при комнатной температуре

Фундаментальное (собственное межзонное) поглощение. Фундаментальное поглощение определяется межзональными 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.

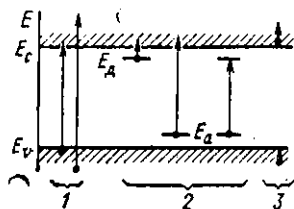


Рис. 4.6. Схематическое изображение электронных переходов, ответственных за процессы оптического поглощения в полупроводниках:

1 — фундаментальное поглощение; 2 — примесное поглощение; 3 — поглощение свободными носителями заряда

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 минимума 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. § 1.5) край собственного поглощения определяется прямыми оптическими переходами из j -состояния валентной зоны в f -состояние зоны проводимости, как изображено на рис. 4.3. В кристаллах с непрямой структурой зон край собственного поглощения определяется непрямыми оптическими переходами $j \rightarrow f$ (см. рис. 4.4).

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 (4.41), (4.42). Число таких пар состояний, приходящихся на единичный интервал энергий и единицу объема кристалла, называется комбинированной плотностью состояний и обозначается $N(\hbar\omega)$. Тогда

$$k_\omega \sim w_{ff} N(\hbar\omega). \quad (4.49)$$

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

$$E_f = \frac{\hbar^2 k^2}{2m_n}, \quad E_g = \frac{\hbar^2 k^2}{2m_p},$$

откуда

$$\hbar\omega - E_g = \frac{\hbar^2 k^2}{2m_r}. \quad (4.50)$$

Здесь m_r — приведенная эффективная масса:

$$\frac{1}{m_r} = \frac{1}{m_n} + \frac{1}{m_p}.$$

Так как для прямых межзонных переходов должно вы-

полняться условие (4.43), 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 (4.50) число состояний в интервале энергий от $\hbar\omega$ до $\hbar\omega + d(\hbar\omega)$:

$$N(\hbar\omega) d(\hbar\omega) = \frac{8\pi k^2 dk}{(2\pi)^3} = \frac{(2m_r)^{3/2}}{2\pi^2 \hbar^3} (\hbar\omega - E_g)^{1/2} d(\hbar\omega). \quad (4.51)$$

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_{k_f} и u_{k_i} в (4.38).

Для большинства полупроводниковых материалов симметрия волновых функций u_{k_f} и u_{k_i} 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 (4.49), (4.51) получаем спектральную зависимость показателя поглощения для прямых разрешенных переходов:

$$k_\omega = A (\hbar\omega - E_g)^{1/2}. \quad (4.52)$$

В коэффициент пропорциональности A входят квадрат матричного элемента $|V_{cv}|^2$, приведенная эффективная масса $(m_r)^{1/2}$ и другие константы. Его оценка может быть проведена с помощью соотношения

$$A \approx \frac{e^2 (2m_r)^{3/2}}{nch^2 m \epsilon_0} f_{cv}. \quad (4.53)$$

Коэффициент f_{cv} , называемый силой осциллятора, имеет величину порядка единицы. Для полупроводников группы A^{III} B^V

$$f_{cv} = \frac{2}{3} \frac{|V_{cv}(0)|^2}{m\hbar\omega} \approx \frac{1}{3} \left(\frac{m}{m_n} - 1 \right). \quad (4.54)$$

Для оценки коэффициента A примем $m_r = m$, $m_n = 0,25m$, $n=4$, что дает $A \approx 10^4 \text{ см}^{-1} \cdot \text{эВ}^{-1/2}$. Из (4.52) 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 пе-

реходов очень быстро возрастает с увеличением энергии фотонов при $\hbar\omega \geq E_g$.

На рис. 4.7 показан край собственного поглощения арсенида галлия, измеренный при различных температурах. Точки обозначают эксперимент, сплошные линии — расчет по соотношениям (4.53), (4.54), пунктирные линии — положение и относительную интенсивность линий экситонного

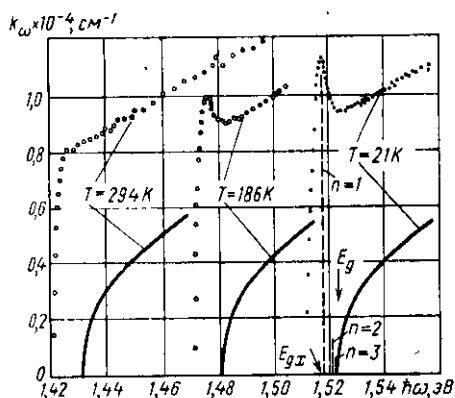


Рис. 4.7. Край собственного поглощения арсенида галлия, измеренный при различных температурах

поглощения. Сдвиг спектров по шкале энергий определяется температурным изменением ширины запрещенной зоны. Видно, что экспериментальные данные существенно отличаются от теоретических не только по значению показателя поглощения, но, что более важно, по форме спектра. Это отличие объясняется тем, что при выводе формул (4.49) — (4.54) не учитывалось взаимодействие между созданными в результате поглощения электроном и дыркой. На самом деле электрон и дырка, являясь противоположно заряженными частицами (точнее квазичастицами), взаимодействуют между собой в основном за счет сил кулоновского притяжения, образуя связанную электронно-дырочную пару. Эту пару взаимодействующих электрона и дырки можно рассматривать как некую квазичастицу, называемую *экситоном*. Картина здесь по существу получается аналогичной рассмотренной в § 1.1 для атома водорода. Действительно, взаимодействующие между собой электрон и протон образуют атом водорода, который существует как единое целое и которому можно приписать свою кинетическую энергию, связанную с движением центра масс, и внутреннюю энергию, обусловленную взаимодействием электрона с ядром. Аналогично взаимодействующие между собой электрон и дырка образуют качественно новое состояние — экситон, которому можно приписать свою кинетическую энергию, связанную с движением центра масс, и внутреннюю энер-

гию, обусловленную кулоновским взаимодействием электрона с дыркой. Но в отличие от атома водорода экситон является возбужденной, неравновесной квазичастицей *. Он может исчезнуть либо в результате теплового «довозбуждения», т. е. термической диссоциации, сопровождающейся возникновением свободных электрона и дырки, либо за счет отдачи своей энергии решетке, либо вследствие аннигиляции с испусканием кванта света.

Учитывая сказанное, уравнение движения экситона

$$\hat{H}\Phi(\mathbf{r}_e, \mathbf{r}_p) = E\Phi(\mathbf{r}_e, \mathbf{r}_p)$$

можно записать в виде

$$\left[-\frac{\hbar^2}{2m_n} \nabla_e^2 + E_g - \frac{\hbar^2}{2m_p} \nabla_p^2 - \frac{e^2}{4\pi\epsilon_r\epsilon_0 |\mathbf{r}_e - \mathbf{r}_p|} \right] \Phi = E\Phi. \quad (4.55)$$

Здесь гамильтониан \hat{H} аналогичен (1.2), но вместо массы ядра и свободного электрона стоят эффективные массы дырки m_p и электрона m_n , а ослабление кулоновского взаимодействия в кристалле учтено введением относительной диэлектрической проницаемости ϵ_r . Приведением к центру масс аналогично (1.3), (1.4) гамильтониан задачи разбивается на две независимые части. Тогда решение уравнения (4.55) можно представить в виде

$$\Phi_{\mathbf{K}, n} = \frac{1}{\sqrt{V}} e^{i\mathbf{K}\mathbf{R}} \psi_n(\mathbf{r}), \quad (4.56)$$

где $\mathbf{K} = \mathbf{k}_e + \mathbf{k}_p$ — экситонный волновой вектор; \mathbf{R} — координаты центра масс; V — объем кристалла, $\mathbf{r} = \mathbf{r}_e - \mathbf{r}_p$.

Плоская волна $e^{i\mathbf{K}\mathbf{R}}$, как и в случае атома водорода, отображает движение центра масс, а волновая функция $\psi_n(\mathbf{r})$ описывает состояние внутреннего движения. Она удовлетворяет уравнению

$$\left(-\frac{\hbar^2}{2m_r} \nabla_r^2 - \frac{e^2}{4\pi\epsilon_r\epsilon_0 r} \right) \psi_n(\mathbf{r}) = E_n \psi_n(\mathbf{r}), \quad (4.57)$$

аналогичному уравнению (1.9) для атома водорода**.

* Экситон от англ. to excite — возбуждать.

** Уравнения (4.55) и (4.57) и приведенные рассуждения справедливы до тех пор, пока действует приближение эффективной массы, т. е. для экситонов, эффективный боровский радиус которых существенно превышает размеры элементарной ячейки кристалла. Такие экситоны обычно существуют в полупроводниках и называются экситонами большого радиуса или экситонами Ванье. Экситоны малого радиуса называются экситонами Френкеля. Уравнение (4.57) записано для изотропных параболических зон.

Поэтому внутренняя энергия экситона E_n будет определяться аналогично (1.10):

$$E_n = -\frac{m_r e^4}{8(\epsilon_r \epsilon_0 \hbar)^2} \frac{1}{n^2} = -\frac{R_{ex}}{n^2}, \quad (4.58)$$

где R_{ex} — энергия ионизации экситона (экситонный ридберг), а n — главное квантовое число ($n=1; 2; 3; \dots$).

Из сравнения (4.58) с (1.10) получаем, что экситонный ридберг связан с постоянной Ридберга R_∞ соотношением

$$R_{ex} = \frac{(m_r/m)}{\epsilon_r^2} R_\infty.$$

Это означает, что для полупроводника с $m_r \approx 0,1 m$ и $\epsilon_r = 10$ энергия ионизации экситона в 1000 раз меньше энергии ионизации атома водорода и составляет ~ 13 мэВ.

Из уравнения (4.57) видно, что экситонные волновые функции $\psi_n(r)$ аналогичны атомным волновым функциям для атома водорода, но вместо боровского радиуса $a_B = \frac{\hbar^2 \epsilon_0}{\pi m e^2} \approx 0,053$ нм необходимо ввести экситонный радиус

$$a_{ex} = \frac{\hbar^2 \epsilon_r \epsilon_0}{\pi m_r e^2} = \epsilon_r \left(\frac{m}{m_r} \right) a_B. \quad (4.59)$$

Здесь картина по существу аналогична рассмотренной в § 1.5 для водородоподобных примесей [ср. выражения (1.75) и (4.58), (1.77) и (4.59)], за исключением того, что экситон как квазичастица может свободно перемещаться по кристаллу.

Полная энергия экситона E состоит из его кинетической энергии $\frac{\hbar^2 K^2}{2M}$, определяемой движением центра масс, и внутренней энергии ($E_g + E_n$), так что можно записать:

$$E(K) = \frac{\hbar^2 K^2}{2M} + E_g + E_n, \quad (4.60)$$

где K — волновой вектор экситона; $M = m_n + m_p$, E_g — ширина запрещенной зоны, а спектр E_n определяется соотношением (4.58). Из (4.60) 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. 4.8. Каждая из зависимостей $E(K)$ с заданным n образует экситонную зону. Точка O есть энергия невозбужденного кристалла. Минимальная энергия $E_{gx} = E_g - E_{ex}$, необходимая для

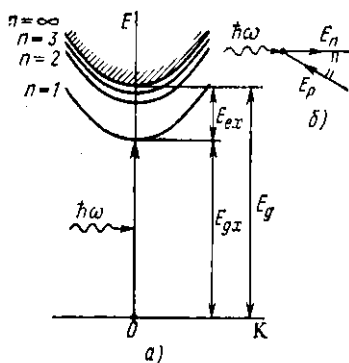


Рис. 4.8. Экситонные энергетические зоны и экситонное поглощение в кристаллах:

а — энергетическая диаграмма $E(K)$ (K — волновой вектор экситона);
б — диаграмма взаимодействия

4.3 является начало отсчета энергий, за которое принято состояние невозбужденного кристалла. Взаимодействие между электроном и дыркой обозначено на рис. 4.8, б штрихами.

Непосредственный оптический переход в экситонное состояние 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 состояние $K=0$. Поэтому спектр собственного поглощения кристалла с учетом экситонных эффектов должен состоять из серии линий при

$$\hbar\omega = E_g - (E_{gx}/n^2), \quad (4.61)$$

которые переходят в непрерывный спектр при $\hbar\omega > E_g$ (см. рис. 4.7). 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 .

создания экситона, называется экситонной шириной запрещенной зоны.

Экситонные эффекты и экситонные состояния не могут быть изображены на обычных энергетических диаграммах типа рис. 4.3 и 4.6, поскольку эти диаграммы справедливы в одноэлектронном приближении, а экситон состоит как минимум из двух частиц. Поэтому рис. 4.8 существенно отличается от рис. 4.3. В частности, по оси абсцисс на рис. 4.8, а отложен волновой вектор экситона K , а не электрона или дырки в отдельности. Общим для рис. 4.8 и

Каждая из экситонных линий уширена вследствие конечности времени жизни экситона, наличия дефектов и т. д. 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, как показано на рис. 4.7. Заметим, что экситонные эффекты существенно влияют на форму края собственного поглощения не только при низких, но и при комнатной температуре.

Рассмотрим теперь край собственного поглощения полупроводниковых кристаллов 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. 4.4, а законы сохранения будут определяться взаимодействием с фононами согласно соотношениям (4.47) и (4.48).

Для определенности рассмотрим процесс непрямого оптического поглощения фотона $\hbar\omega$ с одновременным поглощением фонона $\hbar\Omega_k$ при переходе через ближайшее виртуальное состояние l в зоне проводимости. Диаграмма взаимодействия для такого перехода изображена на рис. 4.4, в. Допустим, что оптический переход $j \rightarrow l$ в виртуальное состояние разрешен. Тогда, как и ранее, оптический матричный элемент $V_{lj}^{e, \text{фот}}$ можно считать не зависящим от волнового вектора, т. е. положить $V_{lj}^{e, \text{фот}}(k) = V_{lj}(0) = \text{const}$. 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 можно записать

$$N_\Omega = \frac{1}{e^{\frac{\hbar\Omega}{kT}} - 1}. \quad (4.62)$$

При непрямых оптических переходах (в отличие от прямых) возможны переходы из любого занятого j -состояния валентной зоны в любое свободное l -состояние зоны проводимости, лишь бы выполнялся закон сохранения энергии (4.47). Тогда комбинированная плотность состояний $N(\hbar\omega)$ при непрямых переходах будет определяться интегралом по всем возможным парам состояний, разделенных энергией $\hbar\omega + \hbar\Omega_k$, от произведения плотностей

начальных $\rho_v(E_f)$ и конечных $\rho_c(E_f)$ состояний. Для зон с параболическим законом дисперсии имеем

$$\rho_v(E_f) = \frac{1}{2\pi^2} \left(\frac{2m_p}{\hbar^2} \right)^{3/2} E_f^{1/2}, \quad (4.63)$$

$$\rho_c(E_f) = \frac{N_c}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} E_f^{1/2}, \quad (4.63a)$$

где N_c — число эквивалентных минимумов по различным симметричным кристаллографическим направлениям.

В отличие от (4.47) энергии E_f и E_g в (4.63) отсчитываются от разных точек. Поэтому вместо соотношения (4.47) можно написать

$$\hbar\omega = E_f + E_g + E_f - \hbar\Omega_k. \quad (4.64)$$

Тогда

$$N(\hbar\omega) = \int_0^\infty dE_f \int_0^\infty \rho_c \rho_v \delta(E_f + E_g + E_f - \hbar\omega - E_g + \hbar\Omega_k) dE_g = \int_0^\infty \rho_c \rho_v dE_f.$$

Подставляя (4.63) и интегрируя, находим

$$N(\hbar\omega) = \frac{N}{\pi^3} \left(\frac{2}{\hbar^2} \right)^3 (m_n m_p)^{3/2} (\hbar\omega - E_g + \hbar\Omega_k)^2. \quad (4.65)$$

Таким образом, спектральная зависимость показателя поглощения для непрямых оптических переходов с поглощением фонона (индекс a) имеем вид

$$k_\omega^{(a)} = \frac{B}{\frac{\hbar\Omega}{e^{\hbar\Omega/T}} - 1} (\hbar\omega - E_g + \hbar\Omega_k)^2. \quad (4.66)$$

В коэффициент пропорциональности B входят произведение квадратов оптического матричного элемента $V_{if}^{\text{фот}}$ и матричного элемента рассеяния $V_{fi}^{\text{фон}}$, 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 (4.64) — (4.66) изменится на противоположный. Вероятность эмиссии фонона с учетом спонтанного испускания пропорциональна $N_\Omega + 1$. Поэтому показатель по-

глощения для переходов с испусканием фононов (индекс e)

$$k_{\omega}^{(e)} = \frac{B}{1 - e^{-\frac{\hbar\Omega}{kT}}} (\hbar\omega - E_g - \hbar\Omega_k)^2. \quad (4.67)$$

Из выражений (4.66) и (4.67) 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 выражений (4.66) и (4.67), 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. 4.9). Непрямая ширина запрещенной зоны E_g лежит посередине между этими точками. В соответствии со статистикой распределения фононов (4.62) 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 $k_{\omega}^{(a)} \rightarrow 0$ при $T \rightarrow 0$.

На рис. 4.9 учтено взаимодействие только с одним типом фононов. На самом деле необходимо учесть все ветви колебаний: поперечную и продольную акустическую, а также оптические (см. рис. 1.17). При их учете показатель поглощения k_{ω} для непрямых переходов можно представить в виде суммы

$$k_{\omega} = \sum_i (k_{\omega}^{(a)} + k_{\omega}^{(e)})_i, \quad (4.68)$$

где индекс i относится к данному типу фононов, а индексы e и a — к компонентам с испусканием и поглощением фононов соответственно.

Кулоновское взаимодействие электрона и дырки, приводящее к экситонным эффектам, также существенно влияет на форму края поглощения, обусловленного непрямыми переходами. Непрямые экситонные переходы, как и электронные переходы, возможны лишь во втором приближении теории возмущений и для их осуществления необ-

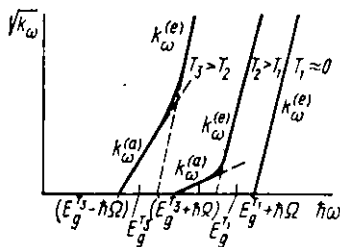


Рис. 4.9. Край собственного поглощения при не прямых электронных переходах в координатах $\sqrt{k_\omega}(\hbar\omega)$ при различных температурах

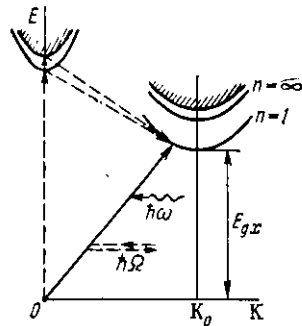


Рис. 4.10. Непрямые оптические переходы в экситонных состояниях при взаимодействии с фононом $\hbar\Omega$

ходимо рассеяние. Диаграмма, иллюстрирующая экситонное поглощение при не прямых переходах, представлена на рис. 4.10. Начальным состоянием здесь снова является состояние невозбужденного кристалла, т. е. точка $K=0$. При не прямых экситонных переходах (в отличие от прямых) в спектре поглощения не будут проявляться резкие линии, так как благодаря участию фононов возможны переходы не только в точку, например $K=K_0$, но и в любое состояние вблизи нее. 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 ρ_c , если вместо эффективной массы m_n подставить массу $M=m_n+m_p$:

$$\rho_{ex}(E) = \frac{1}{2\pi^2} \left(\frac{2M}{\hbar^2} \right)^{3/2} E^{1/2}. \quad (4.69)$$

Пороговая энергия $\hbar\omega_{\min} = E_g - E_{ex} \pm \hbar\Omega_k$. Поэтому для не прямых переходов в экситонные состояния имеем

$$k_\omega = k_\omega^{(e)} + k_\omega^{(a)} = \frac{c}{\hbar\Omega} [(\hbar\omega - E_{gx} + \hbar\Omega)^{1/2} + e^{\frac{\hbar\Omega}{kT}} (\hbar\omega - E_{gx} - \hbar\Omega)^{1/2}], \quad (4.70)$$

где $E_{gx} = E_g - E_{ex}$ — экситонная ширина запрещенной зоны (не прямой).

Соотношение (4.70) справедливо лишь в области $E_{gx} \leq (\hbar\omega \pm \hbar\Omega) < E_g$. При $(\hbar\omega - E_{gx}) \gg E_{ex}$ оно переходит в (4.66), (4.67). В качестве примера на рис. 4.11 изображен спектр собственного поглощения, обусловленный прямыми переходами с испусканием фонона $\hbar\Omega$ (сплошная кривая — с учетом, пунктирная — без учета экситонных эффектов). Здесь учтены про-

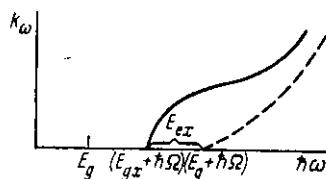


Рис. 4.11. Спектр собственного поглощения при не прямых переходах с учетом (сплошная кривая) и без учета (пунктирная кривая) экситонных эффектов

цессы поглощения фотонов при одновременном испускании лишь одного типа фононов $\hbar\Omega$. Реальный спектр поглощения будет определяться суммой аналогичных кривых, обусловленных взаимодействием с различными типами фононов. С повышением температуры на процессы с испусканием фононов будут накладываться процессы с их поглощением, как показано на рис. 4.12. Видно возрастание вклада компонент с поглощением фононов при повышении температуры.

В полупроводниковых твердых растворах с изменением их состава край собственного поглощения будет сдвигаться по энергии в соответствии с изменением ширины запрещенной зоны [см. формулу (1.70) и рис. 1.15]. В качестве примера на рис. 4.13 приведены экспериментальные данные по краю собственного поглощения кристаллов $\text{GaAs}_{1-x}\text{P}_x$. В области составов $x < x_c = 0,46$ край поглощения определяется прямыми переходами, а в области $x > x_c$ — непрямыми переходами в экситонные состояния. Кроме переходов с участием фононов наблюдаются не прямые переходы без участия фононов, пороговая энергия для которых равна экситонной ширине запрещенной зоны. Бесфононная компонента обозначена на рисунке сплошной линией. Ее интенсивность возрастает при уменьшении x .

Анализ многочисленных экспериментальных данных показал, что фундаментальные оптические свойства полупроводниковых твердых растворов, определяемые прямыми, вертикальными электронными переходами, в основном такие же, как и для идеальных кристаллов с эквивалентной зонной структурой. Форма края собственного поглощения в области прямых межзонных переходов не претерпевает существенных изменений по сравнению с бинар-

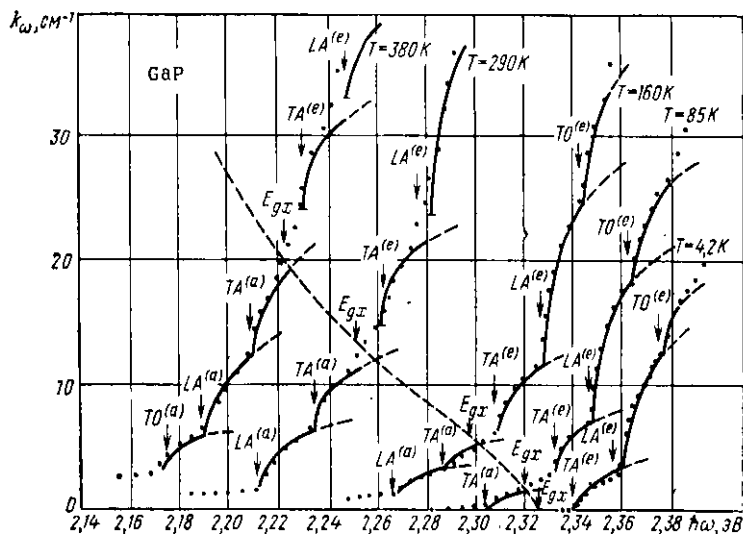


Рис. 4.12. Край собственного поглощения фосфида галлия при различных температурах:

точки обозначают экспериментальные данные, сплошные кривые — зависимости (4.66) и (4.67) для непрямых оптических переходов в экситонные состояния с испусканием (индекс e) и поглощением (индекс a) фононов: TA — поперечного акустического, LA — продольного акустического и TO — поперечного оптического; E_{gx} — экситонная ширина запрещенной зоны.

ными соединениями. Определяющую роль в формировании края собственного поглощения в твердых растворах высокого качества полупроводников групп $A^{III}B^V$ и $A^{II}B^{VI}$, как и в составляющих их бинарных соединениях, играют экситонные эффекты. Уширение экситонных линий, вызванное наличием флуктуирующей части $V_1(r)$ кристаллического потенциала (1.69), сравнительно невелико и обычно не превышает одного или нескольких миллиэлектронвольт. Это связано с тем, что борковский радиус «прямого» экситона в десятки раз превышает размеры элементарной ячейки, и статистические локальные неоднородности в объеме 10^4 — 10^5 элементарных ячеек усредняются и не приводят к заметному уширению. Все это указывает, что неперIODическая добавка $V_1(r)$ в (1.69) не оказывает существенного влияния на процессы оптического поглощения, определяемые прямыми переходами. Поэтому форма края собственного поглощения в твердых растворах с прямой структурой энергетических зон (при $x < x_c$ на рис. 4.13) аналогична таковой для эквивалентного идеального кристалла.

На оптические свойства, определяемые непрямymi переходами, неперIODическая добавка $V_1(r)$ в (1.69), наоборот,

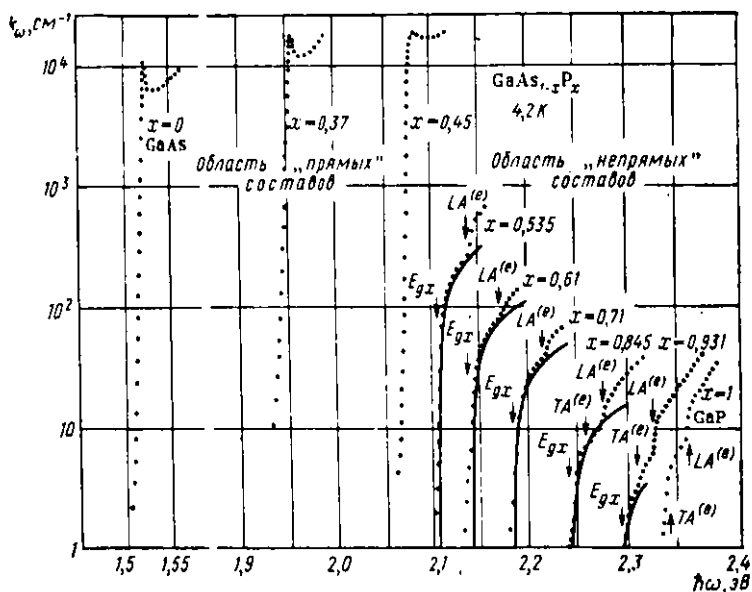


Рис. 4.13. Край собственного поглощения твердых растворов $\text{GaAs}_{1-x}\text{P}_x$ при низкой температуре для различных составов x

может сильно влиять. Прежде всего нарушение трансляционной симметрии гамильтониана в твердых растворах может вызвать «нарушение» закона сохранения волнового вектора (4.42) при оптических переходах [напомним, что соотношения (4.39) и (4.42) 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 наинизшее реальное состояние при $k=0$, а рассеяние осуществляется на флуктуирующем потенциале, вызванном статистически неупорядоченным распределением компонентов твердого раствора по узлам соответствующей подрешетки. При этом вместо концентрации рассеивающих центров необходимо ввести коэффициент, учитывающий степень разупорядочения твердого раствора A_xB_{1-x} . В отсутствие упорядочения этот коэффициент равен $x(1-x)$.

Возможность непрямых переходов без участия фононов

в твердых растворах обуславливает появление бесфононной компоненты в спектре. Ее пороговая энергия равна непрямой 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. Форма спектра поглощения без учета экситонных эффектов при непрямых бесфононных переходах будет определяться квадратичной функцией вида (4.67), а для переходов в экситонные состояния — соотношением, аналогичным (4.70), которое можно переписать в виде

$$k_{\omega}^{\text{бф}} = A(x) (\hbar\omega - E_{gx})^{1/2}. \quad (4.71)$$

Коэффициент $A(x)$ характеризует интенсивность бесфононных переходов. Он зависит от состава x твердого раствора и пропорционален $x(1-x)$. Кроме того, в матричный элемент для переходов через виртуальные состояния входит энергетический множитель $(1/\Delta)^2$, где $\Delta = E_i - \hbar\omega \approx E_i - E_g$ [см. формулы (3.59) и (4.46)]. В твердых растворах этот множитель зависит от состава x , поэтому

$$A(x) = A_0 x(1-x) \left| \frac{1}{\Delta(x)} \right|^2. \quad (4.72)$$

Поскольку для бесфононных переходов $E_i + \hbar\omega = E_f$ (см. рис. 4.4), величина $\Delta(x)$ в (4.72) есть энергетический зазор между виртуальным E_i и конечным E_f состояниями, т. е. для $\text{GaAs}_{1-x}\text{P}_x$ — зазор между Γ_1^c и X_1^c -минимумами (см. рис. 4.4). При уменьшении этого зазора величина $A(x)$, а следовательно, интенсивность непрямых бесфононных переходов возрастает.

В силу указанных причин спектр собственного поглощения твердого раствора $\text{GaAs}_{1-x}\text{P}_x$ 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. 4.13). Переходы с поглощением фононов не проявляются, так как температура мала. Бесфононная компонента в GaP отсутствует. По мере образования твердого раствора $\text{GaAs}_{1-x}\text{P}_x$ 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. 4.13 эта компонента обозначена сплошной линией). Быстрый рост $k_{\omega}^{\text{эф}}$ при увеличении содержания арсенида галлия в $\text{GaAs}_{1-x}\text{P}_x$ обусловлен двумя причинами: 1) увеличивается коэффициент разупорядочения $x(1-x)$; 2) уменьшается энергетический зазор $\Delta \approx \approx E_g^{\Gamma_1} - E_g^X$ (см. рис. 1.15).

Кроме увеличения интенсивности бесфононной компоненты, как видно из рис. 4.13, при увеличении $(1-x)$ происходит быстрое (резонансное) увеличение интенсивности ЛА-компоненты, что также определяется уменьшением энергетического зазора Δ между виртуальным Γ_1^f и конечным реальным X_1^f состояниями в $\text{GaAs}_{1-x}\text{P}_x$.

Эффекты, связанные с увеличением вероятности непрямых 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 $\text{GaAs}_{1-x}\text{P}_x$ эффект зонной структуры не проявляется, поскольку оптические переходы через наименьшее Γ_1^f -состояние в X_1^f -минимум зоны проводимости с участием поперечных акустических (ТА) фононов запрещены правилами отбора по симметрии.

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 ее колебания. Другими словами, при решеточном поглощении уничтожение фотона приводит к рождению фонона. Энергия электронов при этом в адиабатическом приближении остается неизменной.

Законы сохранения энергии и волнового вектора при взаимодействии фотона с N фононами записываются в виде

$$\hbar\omega = \sum_{i=1}^N \pm \hbar\Omega_i(k_{i\text{ фон}}), \quad (4.73)$$

$$0 \approx k_{\text{фот}} = \sum_{i=1}^N \pm k_{i\text{ фон}}. \quad (4.73a)$$

Знак плюс соответствует испусканию, а минус — поглощению i -го фонона. При взаимодействии с одним фононом

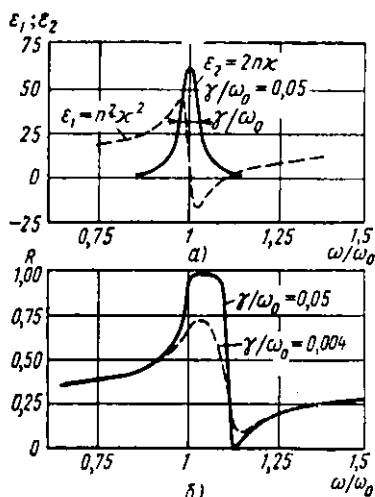


Рис. 4.14. Действительная ϵ_1 и мнимая ϵ_2 части комплексной диэлектрической проницаемости как функции частоты для модели классического осциллятора (а) и спектральная зависимость коэффициента отражения (б)

спектрами гармонического осциллятора с собственной частотой $\omega_0 = \Omega_{\text{ТО}}$ и затуханием γ (рис. 4.14). Показатель поглощения в максимуме может достигать очень больших значений ($k_\omega > 10^5 \text{ см}^{-1}$), 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 остаточных лучей. В кристаллах с ионной связью коэффициент отражения R в этой области близок к 100%. В кристаллах с чисто 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:

$$\hbar\omega = \hbar\Omega_1 + \hbar\Omega_2, \quad (4.75)$$

$$k_{\text{фон } 1} = k_{\text{фон } 2}. \quad (4.75a)$$

(однофононное поглощение) имеем

$$\hbar\omega = \hbar\Omega(0), \quad (4.74)$$

$$k_{\text{фон } 1} = k_{\text{фон } 2} \approx 0. \quad (4.74a)$$

Это означает, что в процессе однофононного поглощения могут участвовать только оптические фононы (см. рис. 1.17). Точнее, фотон может поглощаться при возбуждении лишь поперечных оптических колебаний решетки (ТО-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

Учет законов сохранения (4.75) 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 ($5\text{--}500\text{ см}^{-1}$).

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), и поглощение, обусловленное возбуждением локализованных колебаний примесных атомов.

Таким образом, можно выделить следующие процессы, приводящие к примесному поглощению в твердых телах: а) фотоионизацию примесей; б) фотовозбуждение примесей; в) фотонейтрализацию примесей; г) фотогенерацию связанных экситонов; д) возбуждение локализованных колебаний примесных атомов. Электронные переходы, обуславливающие примесное поглощение, соответствующие пунктам (а—г), схематически изображены на энергетической диаграмме рис. 4.15. Рассмотрим сущность каждого из этих процессов.

1. **Фотоионизация мелких примесей.** Этот процесс обусловлен переходом электрона (дырки) из

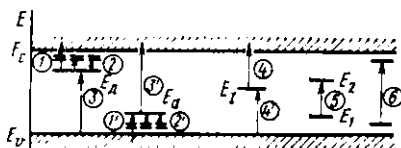


Рис. 4.15. Электронные переходы, ответственные за примесное поглощение в кристаллах:

1, 1' — фотоионизация; 2, 2' — фотовозбуждение; 3, 3' — фотонейтрализация мелких водородоподобных примесей; 4, 4' — фотоионизация и фотонейтрализация глубокого центра; 5 — внутрицентровые переходы; 6 — фотогенерация связанного экситона

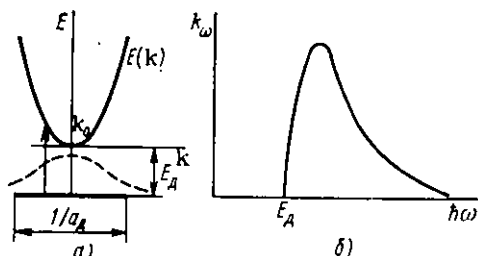


Рис. 4.16. Оптическое поглощение, обусловленное фотоионизацией мелкого водородоподобного донора: а — электронный переход на диаграмме $E(k)$, б — спектр поглощения

связанного состояния в свободное при поглощении фотона $\hbar\omega$ (переходы 1 и 1' на рис. 4.15). 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, т. е. 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 § 1.5, хорошо описываются в приближении эффективной массы с помощью водородоподобной модели. Из выражения (1.79) видно, что волновая функция электрона на водородоподобной примеси локализована в области k -пространства с характерными размерами порядка обратной величины борковского радиуса. На диаграмме $E(k)$ процесс фотоионизации мелкого водородоподобного донора изображает так, как показано на рис. 4.16, а. Минимальная энергия, при которой возможна фотоионизация (пороговая энергия), равна E_d . При $\hbar\omega = E_d$ 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. (4.63)), то при увеличении $\hbar\omega$ будет наблюдаться резкий рост показателя поглощения (рис. 4.16, б). При $\hbar\omega > E_d$ плотность конечных состояний в разрешенной зоне будет монотонно возрастать, а плотность начальных состояний — убывать в соответствии с формулой (1.79) (пунктирная кривая на рис. 4.16, а). Поэтому в спектре поглощения $k_\omega(\hbar\omega)$ будет наблюдаться характерный максимум и спад, как видно из рис. 4.16, б:

Для простых водородоподобных примесей спектр поглощения, вызванный процессами фотоионизации, определяется формулой

$$k_{\omega} = A \frac{(\hbar\omega E_d - 1)^{3/2}}{(\hbar\omega E_d)^3}, \quad (4.76)$$

где коэффициент пропорциональности A связан с концентрацией доноров N_d , энергией ионизации E_d и эффективной массой электронов в зоне m_n соотношением

$$A = \frac{128\pi e^2 \hbar}{3cm_n E_d} N_d. \quad (4.77)$$

Максимум функции (4.76), т. е. максимум спектра фотоионизации водородоподобной примеси, приходится на энергию $\hbar\omega = \frac{10}{7} E_d$.

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. 4.15). Поскольку такие переходы осуществляются 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. 4.17 представлены экспериментальные данные по спектру примесного поглощения в кристаллах кремния, легированного водородоподобным акцептором — бором ($E_a = 44,3$ мэВ). При $\hbar\omega < E_a$ наблюдаемая серия узких дискретных линий обусловлена переходами в возбужденные состояния примеси, т. е. процессами фотовозбуждения дырок, а широкая полоса при $\hbar\omega > E_a$ 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 $\lambda = 10 \div 200$ мкм.

3. **Фотонейтрализация мелких примесей.** Этот процесс вызван переходами электронов из

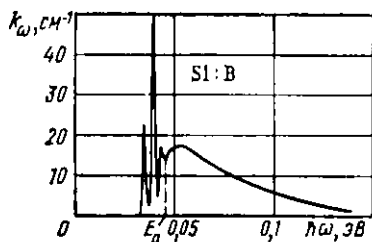


Рис. 4.17. Спектр примесного поглощения акцептора — бора в кремнии

валентной зоны на уровне ионизованных доноров вблизи дна зоны проводимости или с уровнями ионизованных акцепторов в зону проводимости (переходы 3 и 3' на рис. 4.15). 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 $\hbar\omega \geq (E_g - E_d(E_a))$. Так как для мелких примесей $E_d(E_a) \ll E_g$, то спектры фотонейтрализации обычно сливаются со спектрами собственного поглощения и проявляются лишь в виде дополнительной «ступеньки» на краю основной полосы поглощения.

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. 4.15). Но в отличие от мелких примесей возмущающий потенциал для глубоких центров локализован вблизи примеси. Боровский 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 $E_g - E_I^{\text{opt}}$) примеси до края собственного поглощения. Форма спектра поглощения для процесса фотоионизации глубокого центра с энергией ионизации E_I^{opt} может быть аппроксимирована выражением

$$k_{\omega} = B \frac{(\hbar\omega E_I^{\text{opt}} - 1)^{1/2}}{(\hbar\omega/E_I^{\text{opt}})^3}, \quad (4.78)$$

где E_I^{opt} — оптическая энергия ионизации центра, а коэффициент B пропорционален концентрации центров. Формула (4.78) получена для некулоновского (короткодействующе-

го) 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. 4.18 сплошной линией показан спектр поглощения, рассчитанный по формуле (4.78), т. е. без учета электрон-фононного взаимодействия. Пунктирные кривые соответствуют данным, полученным с учетом этого взаимодействия. Учет электрон-фононного взаимодействия приводит к тому, что вместо резкого длинноволнового края, описываемого соотношением (4.78), 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. **Внутривозбужденные 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 внутривозбужденными переходами, являются элементы с незастро-

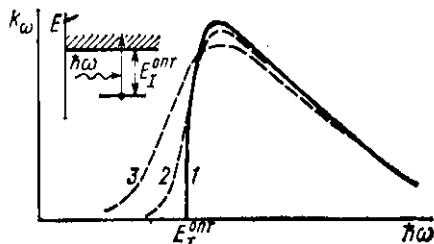


Рис. 4.18. Спектры примесного поглощения, обусловленного фотоионизацией глубокого центра, без учета (кривая 1) и с учетом (кривые 2,3) электрон-фононного взаимодействия

енными 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 6 на рис. 4.15.

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, т. е. 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. 4.15 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 :

$$\sigma = \frac{Ne^2}{m^*} \left\langle \frac{\tau(E)}{1 - \omega^2 \tau^2} \right\rangle, \quad (4.79)$$

где N — концентрация носителей заряда, а m^* — их эффективная масса. Усреднение в (4.79) проводится по энергии E .

Показатель поглощения k_ω связан с главным показателем поглощения κ и удельной электрической проводимости

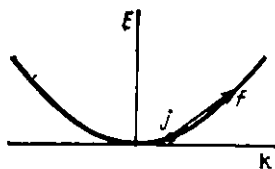


Рис. 4.19. Электронный переход, ответственный за оптическое поглощение свободными носителями заряда

стью σ соотношениями (4.9) и (4.126), откуда находим

$$k_{\omega} = \frac{\sigma}{c\epsilon_0 n} = \frac{Ne^2}{m^* c \epsilon_0 n} \left\langle \frac{\tau}{1 + \omega^2 \tau^2} \right\rangle. \quad (4.80)$$

При $\langle \tau \rangle \omega \gg 1$ имеем

$$k_{\omega} = \frac{Ne^2}{m^* c \epsilon_0 n \omega^2} \left\langle \frac{1}{\tau} \right\rangle. \quad (4.81)$$

Если $\langle 1/\tau \rangle$ не зависит от энергии E носителей заряда в зоне, то из (4.81) получаем $k_{\omega} \sim \lambda^2$, т. е. показатель поглощения на свободных носителях заряда монотонно возрастет 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. 4.19. Это означает, что для его осуществления необходимо рассеяние квазиимпульса, 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 коэффициент 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

$$k_{\omega} \sim N \lambda^r, \quad (4.82)$$

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 § 1.5, каждая из энергетических зон

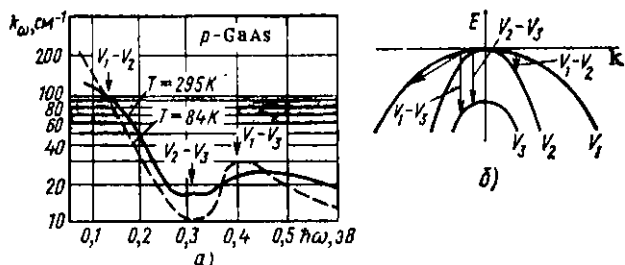


Рис. 4.20. Оптическое поглощение свободными дырками в GaAs p -типа: a — спектры поглощения арсенида галлия, легированного цинком до $N_A = 2,7 \times 10^{17} \text{ см}^{-3}$ при различных температурах; b — переходы, ответственные за поглощение

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 (4.82), определяемый переходами носителей заряда в пределах одной подзоны.

В качестве примера на рис. 4.20, a изображен спектр поглощения кристаллов арсенида галлия p -типа, содержащих $2,7 \cdot 10^{17} \text{ см}^{-3}$ свободных дырок. Стрелками указана интерпретация соответствующих особенностей в спектре. На диаграмме $E(k)$ (рис. 4.20, b) показаны переходы, ответственные за поглощение. В отличие от аналогичных диаграмм рис. 4.16 и 4.19 здесь изображены не электронные, а дырочные переходы.

§ 4.5. 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. Эмиттированные 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 в одинаковой мере 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.

Поглощение, обусловленное переходами электронов между долинами одной и той же зоны, и поглощение свободными носителями заряда практически не вызывают изменения концентрации и подвижности носителей заряда. Решеточное поглощение также не влияет непосредственно на электрические свойства кристалла, либо влияет косвенным образом (например, за счет разогрева кристалла). Поэтому указанные процессы поглощения называются фотоэлектрически неактивными.

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:

$$\frac{dp}{dt} = G_p - \frac{\delta p}{\tau_p} - \frac{1}{e} \operatorname{div} j_p, \quad (4.83)$$

$$\frac{dn}{dt} = G_n - \frac{\delta n}{\tau_n} + \frac{1}{e} \operatorname{div} j_n. \quad (4.83a)$$

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:

$$j_p = e\mu_p p \mathcal{E} - eD_p \text{grad } p, \quad (4.84)$$

$$j_n = e\mu_n n \mathcal{E} - eD_n \text{grad } n, \quad (4.84a)$$

а плотность полного тока

$$j = j_p + j_n. \quad (4.85)$$

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. 4.21.

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 подобных эффектов приконтактную область при проведении исследований часто затемняют. Изменение сопротивления образца под воздействием излучения регистрируется амперметром.

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:

$$\begin{aligned} \Delta\sigma = \sigma_{\text{св}} - \sigma_{\text{т}} &= [e\mu_n (n_0 + \delta n) + e\mu_p (p_0 + \delta p)] - \\ &- [e\mu_n n_0 + e\mu_p p_0] = e\mu_n \delta n + e\mu_p \delta p. \end{aligned} \quad (4.86)$$

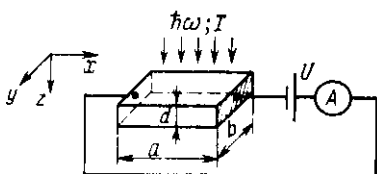


Рис. 4.21. Схема опыта для измерения фотопроводимости

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 (4.83) при условии $j_n = j_p \approx 0$. В стационарном состоянии $dn/dt = dp/dt = 0$ и уравнения (4.83) имеют простое и наглядное решение:

$$\delta p = p - p_0 = G_p \tau_p, \quad (4.87)$$

$$\delta n = n - n_0 = G_n \tau_n. \quad (4.87a)$$

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

$$\delta \sigma = e \mu_p (G_p \tau_p + b G_n \tau_n) \quad (4.88)$$

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 $\tau(I)$.

Скорость генерации G определяется интенсивностью света и показателем поглощения. Если I — интенсивность света, падающего на единицу поверхности в единицу времени, то из соотношения (3.3) следует, что в единице объема поглощается энергия $-dI/dz = k_\omega I$.

Поток фотонов q , измеряемый в $1/(m^2 \cdot s)$, легко выразить через интенсивность света: $q = I/\hbar \omega$. Тогда число поглощенных фотонов в единицу времени в единице объема равно

$$-\frac{1}{\hbar \omega} \frac{dI}{dz} = \frac{k_\omega I}{\hbar \omega} = k_\omega q. \quad (4.89)$$

Пусть каждый поглощенный фотон с вероятностью β

порождает свободный носитель заряда — электрон или дырку. Тогда скорость генерации носителей заряда, имеющая размерность $\text{см}^{-3} \cdot \text{с}^{-1}$, равна

$$G_n = \beta_n k_\omega q, \quad (4.90)$$

$$G_p = \beta_p k_\omega q. \quad (4.90a)$$

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 (4.88) и (4.90)

$$\delta\sigma = ek_\omega \frac{l}{\hbar\omega} (\beta_n \mu_n \tau_n + \beta_p \mu_p \tau_p). \quad (4.91)$$

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

$$\delta\sigma = e\mu\tau\beta k_\omega \frac{l}{\hbar\omega}. \quad (4.92)$$

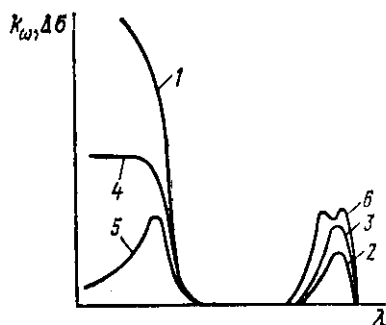


Рис. 4.22. Сравнение спектров фотопроводимости $\Delta\sigma(\lambda)$ и фотоактивного поглощения $k_\omega(\lambda)$:

1, 2 — спектры собственного и примесного поглощения; 3 — спектры примесной фотопроводимости при $(k_\omega)_{\max} d \ll 1$; 4, 5 — спектры собственной фотопроводимости при малой и большой скорости поверхностной рекомбинации; 6 — спектр примесной фотопроводимости при $(k_\omega)_{\max} d \ll 1$ и большой скорости поверхностной рекомбинации

Из выражений (4.91), (4.92) 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. Рекомбинацией 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 от 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 (3.40) и рис. 3.2, 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. 4.22 (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 $k_\omega \approx 10^3 \div 10^5 \text{ см}^{-1}$ 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

$$\frac{1}{\tau'} = \frac{1}{\tau} + \frac{2S}{d}, \quad (4.93)$$

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, скоростью которой в значительной мере определяется состоянием поверхности. При $k_\omega d \gg 1$ интенсивность возбуждения, а следовательно, скорость генерации, как видно из выражения (3.40), быстро уменьшается по экспоненциальному закону:

$$G = G_0 e^{-k_\omega z}. \quad (4.94)$$

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 (4.83) 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 $\hbar\omega$ фототок будет сначала возрастать вслед за увеличением k_ω , а затем достигать насыщения (кривая 4 на рис. 4.22). 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. 4.22). Причина уменьшения $\tau_{p,n}$ с ростом коэффициента поглощения состоит в том, что с увеличением k_ω все 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. 4.22).

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. Интенсивность рекомбинации γ , имеющая размерность $1/(m^3 \cdot c)$, пропорциональна первой степени концентрации неравновесных носителей заряда. Время жизни фотоносителя $\tau_n = \frac{\delta n}{\gamma_n p \delta n}$ (или $\tau_p = \frac{\delta p}{\gamma_p p \delta p}$) не зависит от их концент-

рации. Это случай *линейной рекомбинации*. Он реализуется, например, при наличии одного типа дырок, с которыми рекомбинируют неравновесные электроны, причем концентрация равновесных дырок p_0 довольно значительна и их полная концентрация практически не зависит от освещения. Коэффициенты γ_n и γ_p , называемые коэффициентами рекомбинации для электронов и дырок, характеризуют интенсивность рекомбинации при концентрациях p и δn (или n и δp), равных единице.

2. Интенсивность рекомбинации пропорциональна квадрату концентрации неравновесных носителей заряда. Время жизни фотоносителей обратно пропорционально их концентрации: $\tau \sim (\delta n)^{-1}$ ($\tau \sim (\delta p)^{-1}$). Это случай *квадратичной рекомбинации*. Он реализуется, например, когда концентрация равновесных носителей заряда равна нулю и наблюдаются только межзонные переходы.

При линейной рекомбинации уравнение непрерывности (4.83а) имеет вид

$$\frac{d(\delta n)}{dt} = \beta k_{\omega} \frac{I}{\hbar \omega} - \frac{\delta n}{\tau_n}. \quad (4.95)$$

В стационарном режиме $\frac{d(\delta n)}{dt} = 0$, откуда получаем

$$\delta n_{ст} = \beta k_{\omega} \frac{I}{\hbar \omega} \tau = G \tau, \quad (4.96)$$

т. е. при линейной рекомбинации стационарная фотопроводимость пропорциональна интенсивности света.

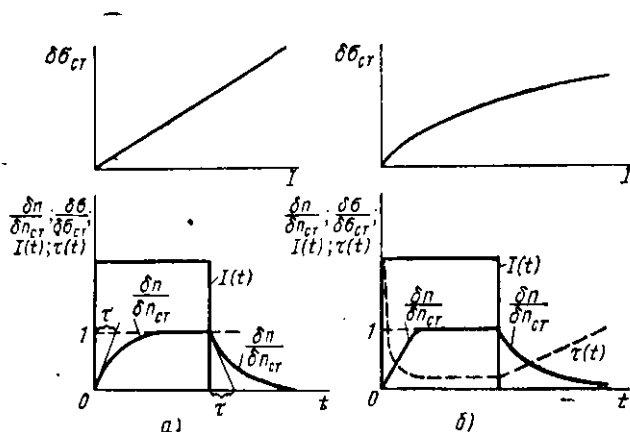


Рис. 4.23. Зависимости стационарной фотопроводимости $\delta\sigma_{ст}$ от интенсивности света I (люкс-амперные характеристики) и кривые релаксации фотопроводимости при воздействии прямоугольных импульсов света:

а — в случае линейной рекомбинации; б — в случае квадратичной рекомбинации

При квадратичной рекомбинации $1/\tau = \gamma \delta n$ и

$$\frac{d(\delta n)}{dt} = \beta k_{\omega} \frac{I}{\hbar \omega} - \gamma (\delta n)^2, \quad (4.97)$$

откуда для стационарного режима получаем

$$\delta n_{ст} = \sqrt{\frac{G}{\gamma}} = \sqrt{\frac{\beta k_{\omega} I}{\gamma \hbar \omega}}, \quad (4.98)$$

т. е. стационарная фотопроводимость пропорциональна корню из интенсивности света.

Законы нарастания и спада фотопроводимости также определяются характером рекомбинации. При линейной рекомбинации интегрирование уравнения (4.95) с начальными условиями $\delta n = 0$, при $t = 0$ дает закон нарастания фотопроводимости:

$$\delta n = G\tau (1 - e^{-t/\tau}). \quad (4.99)$$

Закон спада фотопроводимости после прекращения возбуждения легко определяется из уравнения (4.95) с начальными условиями $G = 0$, $\delta n = \delta n_{ст}$ при $t = 0$ и имеет вид

$$\delta n = G\tau e^{-t/\tau}. \quad (4.100)$$

Таким образом, релаксация неравновесной концентрации, а следовательно, фотопроводимости в случае линейной рекомбинации происходит по экспоненциальному закону

(4.99) или (4.100), при этом «постоянная времени» экспонент совпадает с временем жизни τ (рис. 4.23, а).

В случае квадратичной рекомбинации, используя начальные условия, как и для линейной рекомбинации, из интегрирования уравнения (4.97) получаем закон нарастания

$$\delta n = \sqrt{\frac{G}{\gamma}} \operatorname{th} [t \sqrt{G\gamma}] \quad (4.101)$$

и закон спада

$$\delta n = \sqrt{\frac{G}{\gamma}} \frac{1}{t \sqrt{G\gamma + 1}} \quad (4.102)$$

неравновесной концентрации носителей заряда, а значит, и фотопроводимости.

Как видно из выражений (4.101) и (4.102), при квадратичной рекомбинации законы нарастания и спада фотопроводимости различны. Нарастание определяется гиперболической тангенсоидой, а спад — значительно более медленной при больших t гиперболой. В противоположность линейной рекомбинации здесь не может быть введено понятие времени жизни как *постоянной величины*, не зависящей от интенсивности света и сохраняющей свое значение в течение всего релаксационного процесса. При квадратичной рекомбинации вводят понятие мгновенного времени жизни — величину, обратную средней вероятности рекомбинации для отдельного электрона. Среднее время жизни электронов

$$\tau = \frac{1}{\gamma \delta p} \quad (4.103)$$

Так как в рассматриваемом случае концентрация неравновесных дырок δp (или электронов δn) зависит от интенсивности света и времени, то и τ зависит от этих величин (рис. 4.23, б).

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-э. д. с. в однородном до освещения материале, т. е. они могут наблюдаться в отсутствие внешнего источника питания.

Фотодиффузионный эффект (эффект Дембера). Этот эффект связан с возникновением э. д. с. в однородном полупроводнике при освещении его сильно поглощаемым светом. Если однородный полупроводник освещать светом в области собственного поглощения, то при $k_{\omega}d \gg 1$ поглощение будет

происходить в основном вблизи поверхности образца. Возникающий градиент концентрации вызывает появление диффузионных токов электронов и дырок в направлении, перпендикулярном освещаемой поверхности. Эти диффузионные токи взаимно компенсируются при совпадении подвижностей носителей заряда двух типов. В случае неравенства подвижностей электронов и дырок в образце возникает электрическое поле, направленное перпендикулярно освещаемой поверхности. Действительно, воспользовавшись соотношением Эйнштейна $\left(\mu = \frac{e}{kT} D\right)$ и условием $\delta n = \delta p$, для z -компоненты тока из (4.84) получим

$$j_z = j_{nz} + j_{pz} = e(bn + p)\mu_p \mathcal{E}_z + (b-1)kT\mu_p \frac{\partial(\delta p)}{\partial z}, \quad (4.104)$$

где $b = \mu_n / \mu_p$ — отношение подвижностей носителей.

Если учесть граничное условие для «разомкнутой» цепи $j_z = 0$, то из выражения (4.104) получим, что в образце возникает электрическое поле, напряженность которого

$$\mathcal{E}_z = -\frac{kT}{e} \left(\frac{b-1}{nb+p} \right) \frac{\partial(\delta p)}{\partial z}$$

отлична от нуля при $b = \mu_n / \mu_p \neq 1$.

Фотомангнитный эффект (эффект Кикоина — Носкова). Этот эффект связан с возникновением э. д. с. при освещении сильно поглощаемым светом полупроводника, находящегося в магнитном поле. Если параллельно освещаемой поверхности образца, например в направлении y на рис. 4.21, 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.

Эффект увлечения носителей заряда фотонами. Этот эффект 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 все 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 э. д. с. *фотонного увлечения*. При $\omega\tau_p \gg 1$, где τ_p — время релаксации импульса, для образца толщиной d с показателем поглощения k_ω и подвижностью свободных носителей заряда μ разность потенциалов U может быть оценена с помощью соотношения

$$U = \frac{\mu\mu_0}{(\omega\tau_p)^2 k_\omega} \frac{I(1-R)[1 - \exp(-k_\omega d)]}{1 + R \exp(-k_\omega d)}. \quad (4.105)$$

Для интенсивности падающей световой волны $I \approx 10^8$ 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, $R = 0,3$ э. д. с. 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 даже субмиллиметровой областях спектра. Важно, что такие приемники могут работать при комнатной температуре.

§ 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. Из этого определения можно заключить, что тепловое излучение нагретых тел, различного рода рассеяния электромагнитного излучения, тормозное излучение, синхротронное излучение, излучение Вавилова — Черенкова и другие не относятся к люминесценции. 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

* От лат. lumen (luminis) — свет и escent — слабое действие.

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 *флюоресценцией*, а большой длительностью (более 10^{-8} с) — *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 люминесценций 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 ее 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 ее разогреву (выделяется в виде квантов колебаний решетки — фононов), то такая рекомбинация называется *безызлучательной*.

Для характеристик излучательных переходов определенного типа в общем числе рекомбинационных переходов вводят понятие *квантового выхода люминесценции*, под кото-

* Интервал времени 10^{-8} с выбран потому, что по порядку величины он соответствует времени жизни атома в возбужденном состоянии для переходов, разрешенных в дипольном приближении (см. § 2.4).

рым понимают отношение числа испущенных в заданной спектральной области фотонов к числу созданных в результате возбуждения (генерируемых) носителей заряда или электронно-дырочных пар. Различают внутренних и внешний квантовые выходы люминесценции. При определении внутреннего квантового выхода учитывают все испущенные фотоны. Его обычно находят расчетным путем из экспериментально измеряемого внешнего квантового выхода, где учитываются лишь фотоны, выведенные наружу, за пределы образца или прибора. За счет явлений внутреннего отражения, самопоглощения в материале, поглощения 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 возбуждение люминесценции. Как и квантовый выход, энергетический выход может быть внутренним и внешним. Последний по существу определяет коэффициент полезного действия прибора.

В зависимости от способа возбуждения люминесценции различают: 1) фотолюминесценцию (возбуждение светом); 2) катодолюминесценцию (возбуждение быстрыми электронами, предварительно ускоренными в электронной пушке); 3) радиолюминесценцию и рентгенолюминесценцию (возбуждение радиоактивным и рентгеновским излучением); 4) хемилюминесценцию (возбуждение при химических реакциях); 5) электролюминесценцию (возбуждение электрическим полем). Способы возбуждения электролюминесценции в свою очередь могут быть различными, а именно: а) возбуждение путем инжекции неосновных носителей заряда через гомо- или гетеропереход; б) возбуждение ударной ионизацией в сильном электрическом поле; в) 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 воз-

бужденные 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^{-10} — 10^{-12} 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 заряда 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 *. 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

$$\begin{aligned} f_n &= \frac{1}{\exp\left(\frac{E_n - F_n^*}{kT}\right) + 1}, & n &= N_c \frac{1}{\exp\left(\frac{E_c - F_n^*}{kT}\right) + 1}, \\ f_p &= \frac{1}{\exp\left(\frac{F_p^* - E_p}{kT}\right) + 1}, & p &= N_v \frac{1}{\exp\left(\frac{F_p^* - E_v}{kT}\right) + 1}. \end{aligned} \quad (4.106)$$

* Напомним, что статистика Ферми—Дирака справедлива лишь для термодинамически равновесных систем.

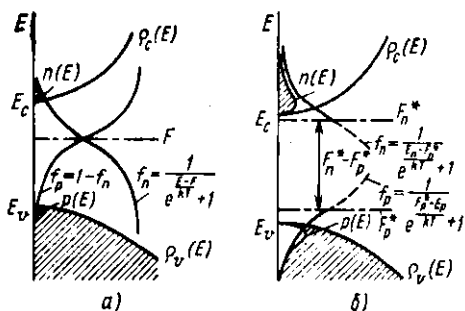


Рис. 4.24. Распределение носителей заряда по энергии:

а — в равновесном состоянии; б — в неравновесном состоянии при наличии возбуждения

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 плотнос-

ти состояний в зоне проводимости и в валентной зоне.

Физический смысл введенных квазиуровней Ферми для электронов и дырок поясняет рис. 4.24. 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 невырождено, то вместо распределения Ферми, как и для равновесных состояний, можно записать распределение Больцмана. Тогда

$$n = N_c \exp\left(-\frac{E_c - F_n^*}{kT}\right), \quad p = N_v \exp\left(-\frac{F_p^* - E_v}{kT}\right) \quad (4.107)$$

и произведение неравновесных концентраций

$$np = N_c N_v e^{-\frac{E_g}{kT}} e^{\frac{F_n^* - F_p^*}{kT}} = n_i^2 \exp\left(\frac{F_n^* - F_p^*}{kT}\right). \quad (4.108)$$

Расстояние между квазиуровнями Ферми определяется уровнем возбуждения и, как нетрудно получить из (4.108),

$$F_n^* - F_p^* = kT \ln \frac{np}{n_i^2}.$$

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.

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 для невзаимодействующих или слабо взаимодействующих атомов или молекул, когда каждому из активных центров можно приписать «свой» энергетические уровни, а каждый электрон рассматривать связанным с конкретным атомом или молекулой, т. е. локализованным в пространстве. В полупроводниках (см. § 1.5) состояния электронов в зоне проводимости и в валентной зоне описываются не локализованными волновыми функциями; а блоховскими функциями, «размазанными» в пространстве. Соответственно каждый из электронов в кристалле нельзя считать принадлежащим отдельному атому. Поэтому обычное условие инверсии для полупроводников неприемлемо, так как приходится иметь здесь дело не с дискретными уровнями, а с непрерывными полосами разрешенных значений энергии, поэтому неясно, что следует подразумевать под $E_{m,n}$ и $N_{m,n}$.

Условие усиления электромагнитной волны в полупроводнике можно получить, рассматривая полное число переходов с поглощением и испусканием фотонов также, как это было сделано в § 3.1.

Для определенности возьмем собственный полупроводник и допустим, что процессы взаимодействия с фотонами определяются только переходами зона — зона. Тогда число переходов $\Delta N_{\text{пол.з}}$ 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 час-

тоте ω . При наличии возбуждения и установлении квазиравновесия в каждой из зон вероятности заполнения состояний определяются функциями Ферми (4.106), при этом вероятность того, что состояние в валентной зоне будет занято электроном, $f_v = (1 - f_p)$. Следовательно, число поглощенных фотонов в единице объема

$$\Delta N_{\text{погл}} = \omega_{vc} N (\hbar\omega) (1 - f_p) (1 - f_n) \rho (\hbar\omega).$$

Число индуцированных переходов с испусканием фотонов будет пропорционально вероятности перехода ω_{cv} , вероятности того, что состояние в зоне проводимости занято (f_n), а в валентной зоне свободно (f_p), т. е.

$$\Delta N_{\text{инд}} = \omega_{cv} N (\hbar\omega) f_p f_n \rho (\hbar\omega).$$

Тогда условие превышения числа испущенных фотонов над поглощенными за счет переходов зона — зона запишется в виде

$$\Delta N_{\text{инд}} > \Delta N_{\text{погл}}.$$

Вероятности вынужденных переходов с поглощением ω_{vc} и испусканием ω_{cv} фотонов на основании принципа детального равновесия равны, т. е. $\omega_{vc} = \omega_{cv}$ (аналогично равенству коэффициентов Эйнштейна B_{mn} и B_{nm} для вынужденных переходов). Тогда из последнего соотношения имеем

$$\begin{aligned} f_p f_n &> (1 - f_p)(1 - f_n) \\ f_n &> (1 - f_p), \end{aligned}$$

откуда с учетом (4.106) получаем условие

$$F_n^* - F_p^* > E_n - E_p.$$

Поскольку минимальное энергетическое расстояние между E_n и E_p равно ширине запрещенной зоны E_g , это неравенство можно записать в виде

$$F_n^* - F_p^* > E_g. \quad (4.109)$$

Соотношение (4.109) по существу определяет условие инверсии населенностей в собственном полупроводнике для переходов зона — зона. Его физический 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 излучательных переходов с испусканием фотона $\hbar\omega$:

$$F_n^* - F_p^* > \hbar\omega. \quad (4.109a)$$

Условие (4.109 а) является необходимым, но не достаточным для получения усиления или генерации в системе в целом. Как и ранее (см. § 3.4), 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 величины 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.

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.

Процессы, приводящие к рекомбинационному излучению в полупроводниках, могут быть довольно многочисленными. Они начинаются с возникновения неравновесных свободных носителей заряда одного или двух типов (электронов и дырок) под воздействием возбуждения. Непосредственно

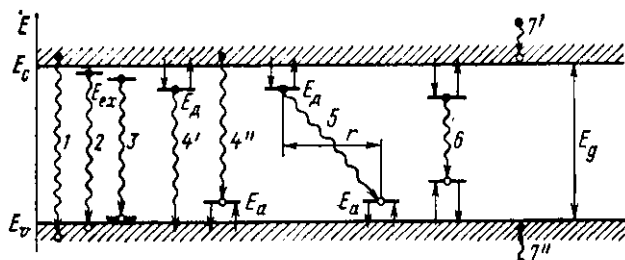


Рис. 4.25. Схематическое изображение электронных переходов, ответственных за различные механизмы излучательной рекомбинации в полупроводниках:

1 — рекомбинация 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 заряда

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: 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. 4.25. Механизмы 1 и 2 излучательной рекомбинации являются собственными, а механизмы 3, 4, 5 и 6 — примесными. Поскольку процессы испускания и поглощения фотона определяются аналогичными электрон-

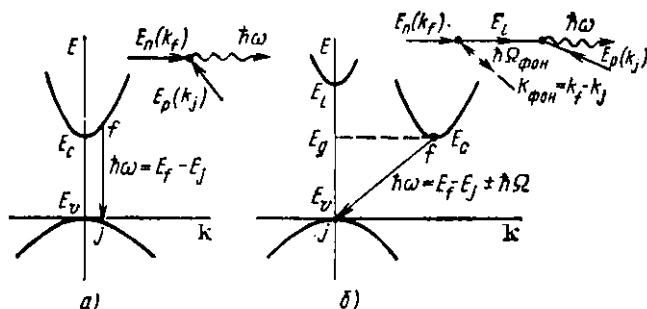


Рис. 4.26. Прямые (а) и непрямые (б) межзонные излучательные переходы

ными переходами, каждому из процессов излучательной рекомбинации можно сопоставить такой же процесс оптического поглощения, в чем легко убедиться, сравнивая рис. 4.25 и 4.6. Естественно, что основные закономерности, присущие квантовым переходам при взаимодействии с оптическим излучением, рассмотренные в § 2.2 и 4.3, справедливы и для излучательной рекомбинации в полупроводниках.

Рассмотрим особенности указанных механизмов излучательной рекомбинации.

Собственная излучательная рекомбинация. Собственная излучательная рекомбинация (излучательная рекомбинация зона — зона) 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. 4.26, а) и непрямыми (рис. 4.26, б) в зависимости от расположения экстремумов соответствующих зон в k -пространстве. При непрямых переходах, как и ранее, для выполнения законов сохранения часть квазиимпульса 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 :

$$\Phi = Bnp. \quad (4.110)$$

Постоянную B можно вычислить из принципа детального равновесия по результатам измерения показателя погло-

щения 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 $\tau_{изл}$, 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 $\omega \div \omega + d\omega$ равно числу электронно-дырочных пар, генерируемых термодинамически равновесным излучением той же частоты. Скорость генерации пар

$$G(\omega) d\omega = w(\omega) \frac{\rho(\omega)}{\hbar\omega} d\omega, \quad (4.111)$$

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.

$$w(\omega) = k_\omega \frac{c}{n}. \quad (4.112)$$

При термодинамическом равновесии для любого интервала $d\omega$ скорость генерации равна скорости рекомбинации. Обозначим скорость рекомбинации в условиях равновесия через $R_0(\omega)$. Тогда

$$R_0(\omega) d\omega = w \frac{\rho(\omega)}{\hbar\omega} d\omega = k_\omega \frac{c}{n} \frac{\rho(\omega)}{\hbar\omega} d\omega. \quad (4.113)$$

Плотность электромагнитной энергии $\rho(\omega)$ в условиях термодинамического равновесия определяется функцией Планка (2.77). Подставляя (2.77) в (4.113) и учитывая, что скорость света в веществе равна c/n , получаем выражение для скорости рекомбинации:

$$R_0(\omega) = \frac{\omega^2 n^3}{\pi^2 c^2} \frac{k_\omega}{\frac{\hbar\omega}{e^{kT}} - 1}, \quad (4.114)$$

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:

$$R_0 = \int_0^{\infty} R_0(\omega) d\omega = \frac{k^3 T^3}{\pi^2 c^2 \hbar^3} \int_0^{\infty} \frac{k \omega n^2 y^2}{e^y - 1} dy, \quad (4.115)$$

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

$$R = \Phi = \frac{n p}{n_i} R_0, \quad (4.116)$$

где $n_i^2 = n_0 p_0$ — квадрат собственной концентрации носителей заряда в материале при заданной температуре T .

Отсюда получаем постоянную B в соотношении (4.110):

$$B = R_0 / n_i^2. \quad (4.117)$$

Характерные значения B при комнатной температуре для некоторых полупроводников с непрямыми (Ge, Si, GaP) и прямыми (GaAs, GaSb, InP, InSb, PbS) межзонными переходами приведены в табл. 4.1. Естественно, что в случае прямых переходов вероятность излучательной рекомбинации на три-четыре порядка выше, чем в случае непрямых. В этой же таблице представлены необходимые для расчета

Таблица 4.1
Величины, характеризующие скорость излучательной рекомбинации в полупроводниках

Переходы	Вещество	E_g , эВ	n_i , см ⁻³	B , см ³ ·с ⁻¹	$\tau_{изл. макс}$, с	$\tau_{изл. макс}$ при $p(n) = 10^{18}$ см ⁻³
Непрямые	Ge	0,665	$2,3 \cdot 10^{13}$	$5 \cdot 10^{-14}$	0,4	2 000
	Si	1,086	$8 \cdot 10^9$	$2 \cdot 10^{-15}$	$3 \cdot 10^4$	50 000
	GaP	2,27	2	$5 \cdot 10^{-14}$	$5 \cdot 10^{12}$	2 000
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	$3 \cdot 10^{15}$	$2 \cdot 10^{-11}$	$6 \cdot 10^{-8}$	5
	InSb	0,18	$2 \cdot 10^{16}$	$4 \cdot 10^{-11}$	$6 \cdot 10^{-7}$	2,5

значения ширины запрещенной зоны E_g и концентрации собственных носителей заряда n_i при комнатной температуре.

Зная n_i и B , 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

$$\Phi = R_0 + \delta R = R_0 \frac{(n_0 + \delta n)(p_0 + \delta p)}{n_0 p_0}. \quad (4.118)$$

Отбрасывая член $\delta n \delta p / n_0 p_0$ второго порядка малости, получим

$$\frac{\delta R}{R_0} = \frac{\delta n}{n_0} + \frac{\delta p}{p_0}. \quad (4.119)$$

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:

$$\tau_{изл. \max} = \frac{\delta n}{\delta R} = \frac{\delta p}{\delta R} = \frac{n_0 p_0}{R(n_0 + p_0)} = \frac{1}{n_i B}. \quad (4.120)$$

Вычисленные по выражению (4.120) для различных полупроводников значения $\tau_{изл. \max}$ приведены в табл. 4.1. Как видно из этих данных, $\tau_{изл. \max}$ представляет собой быстро меняющуюся функцию от ширины запрещенной зоны.

Рассмотрим теперь легированный полупроводник, для определенности с электропроводностью p -типа. Тогда $p_0 \gg n_0$ и из (4.119) имеем $\delta R / R_0 \approx \delta n / n_0$. Излучательное время жизни, определяемое переходами зона — зона:

$$\tau_{изл} = \frac{\delta n}{\delta R} \approx \frac{n_0}{R_0} = \frac{n_i^2}{p_0 R_0} = \frac{1}{p_0 B} = \frac{2n_i}{p_0} \tau_{изл. \max}. \quad (4.121)$$

Соответствующие значения $\tau_{изл.}$, вычисленные для $p_0 = 10^{16} \text{ см}^{-3}$, также представлены в табл. 4.1. Если концентрация основных носителей заряда поддерживается постоянной, то $\tau_{изл.}$ сравнительно слабо зависит от E_g . При этом в полупроводниках с непрямой структурой зон $\tau_{изл.}$ при прочих равных условиях примерно на четыре порядка больше, чем в полупроводниках с прямой структурой зон.

Кроме процессов излучательной рекомбинации зона — зона, как отмечалось, возможны многочисленные процессы излучательной и безызлучательной рекомбинации, которые существенно уменьшают времена жизни неравновесных

носителей заряда. Если через $\tau_{изл}$ и $\tau_{би}$ обозначить времена жизни, определяемые излучательными и конкурирующими безызлучательными переходами в данном материале, то внутренний квантовый выход люминесценции

$$\eta = \frac{1/\tau_{изл}}{1/\tau} = \frac{\tau}{\tau_{изл}} = \frac{\tau_{би}}{\tau_{би} + \tau_{изл}}, \quad (4.122)$$

где τ — экспериментально измеряемое суммарное время жизни избыточных носителей заряда ($1/\tau = 1/\tau_{изл} + 1/\tau_{би}$).

Очевидно, для получения максимальной эффективности люминесценции необходимо, чтобы значение $\tau_{би}$ было велико и время жизни неравновесных носителей заряда определялось излучательными переходами, что возможно лишь в кристаллах высокого качества.

Из табл. 4.1 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 $\tau_{изл, \max} \approx 0,4$ 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.

Спектр люминесценции с точностью до постоянного множителя может быть определен с помощью соотношения (4.114), если известен спектр поглощения k_ω . Это справедливо в случае, когда функции распределения носителей заряда в зонах с точностью до постоянных множителей не отличаются от равновесных, т. е. при не очень сильном возбуждении, когда электронный и дырочный газы невырождены.

Для прямых разрешенных переходов спектр поглощения зона — зона определяется соотношением (4.52), откуда с помощью (4.114) получаем спектр люминесценции для прямых межзонных переходов:

$$\Phi_\omega = A\omega^3 (\hbar\omega - E_g)^{1/2} \exp\left(-\frac{\hbar\omega - E_g}{kT}\right). \quad (4.123)$$

Здесь учтено, что $\hbar\omega \gg kT$, поэтому оказалось возможным

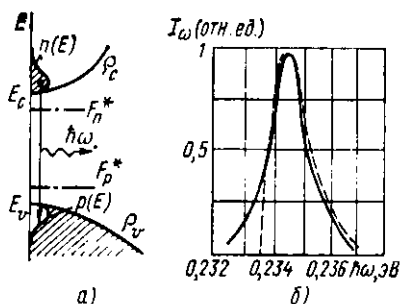


Рис. 4.27. Излучательная рекомбинация, обусловленная межзонными электронными переходами:

а — энергетическая диаграмма; б — спектр излучательной рекомбинации InSb при $T = 4,2$ К

$= E_g + kT/2$), а ширина спектра пропорциональна kT . Форма спектра излучения для переходов зона — зона показана на рис. 4.27, б на примере антимолида индия. Сплошная кривая соответствует эксперименту, пунктирная — теории [расчету по формуле (4.123)]. На рис. 4.27, а даны соответствующие пояснения.

Спектр собственного излучения при непрямых переходах (при тех же условиях) получается из соотношений (4.67), (4.68) и (4.114) в виде

$$\Phi_{\omega} = B_{\pm} \omega^2 (\hbar\omega - E_g \pm \hbar\Omega)^2 \exp\left(-\frac{\hbar\omega - E_g \pm \hbar\Omega}{kT}\right), \quad (4.124)$$

где знак плюс соответствует поглощению фонона с энергией $\hbar\Omega$, а минус — испусканию.

Излучательная рекомбинация свободных экситонов является процессом, обратным экситонному поглощению, и ее можно рассматривать аналогично тому, как это было сделано для переходов зона — зона, но с учетом кулоновского взаимодействия носителей заряда. Рекомбинация экситонов может быть «прямой» и «непрямой» (с участием фононов). Из принципа детального равновесия по спектрам экситонного поглощения можно вычислить вероятность излучательной рекомбинации через экситонные состояния, которая оказывается выше вероятности переходов зона — зона. Физически это можно объяснить тем, что электрон и дырка в экситоне связаны и «подготовлены» к рекомбинации. Однако следует помнить,

пренебречь единицей в знаменателе (4.114). Коэффициент пропорциональности A включает в себя матричный элемент перехода $V_{vc}(0)$, эффективные массы носителей заряда и зависит от интенсивности возбуждения, но не зависит от энергии фотонов $\hbar\omega$.

Из (4.123) видно, что максимум спектра люминесценции расположен при энергии $\hbar\omega$, несколько большей E_g ($\hbar\omega_{\max} =$

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} .

Спектр излучательной рекомбинации свободных экситонов более узкий по сравнению с переходами зона — зона и сдвинут в сторону меньших энергий на величину энергии связи экситона от ширины запрещенной зоны. Отметим, что на рис. 4.25 экситонный переход 2 изображен условно.

Излучательная рекомбинация свободных экситонов и межзонная рекомбинация называются собственными и проявляются в полупроводниках с малой концентрацией примесей либо при сильном возбуждении, когда концентрации неравновесных примесных электронов и дырок существенно больше концентрации дефектов и последние заполнены.

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.

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 § 4.5, 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 захва-

тывается 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 (4.39), (4.40) 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. 4.28, а). С повышением температуры проявляется температурное уширение и отдельные линии сливаются.

Бесфононная линия излучения (А-линия на рис. 4.28, а) расположена резонансно соответствующей линии поглощения. Остальные линии являются фононными повторениями А-линии с испусканием характеристических фононов: ТА — поперечного акустического, LA — продольного акустического; ТО — поперечного оптического и LO — продольного оптического. Символы в скобках у обозначений типов фононов указывают на их принадлежность точкам

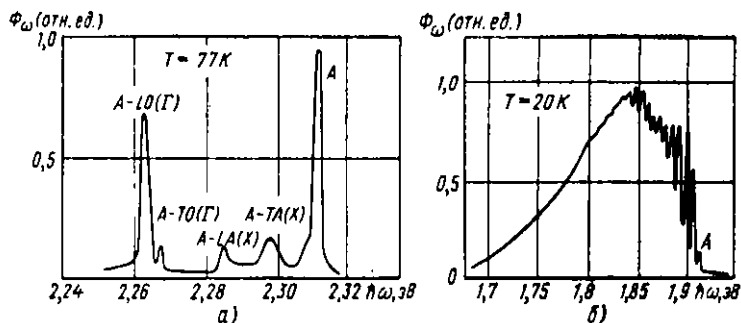


Рис. 4.28. Спектры излучательной рекомбинации связанных экситонов:

a — экситоны, связанные на изоэлектронной ловушке N в GaP (*A* — бесфононная линия и ее повторения с испусканием различных фононов); *б* — экситоны, связанные на молекулярном изоэлектронном комплексе Cd—O в GaP (*A* — бесфононная линия)

зоны Бриллюэна: X — край зоны в направлении (100), Γ — центр зоны (000).

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 (~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 ее 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. 4.28, б. С повышением

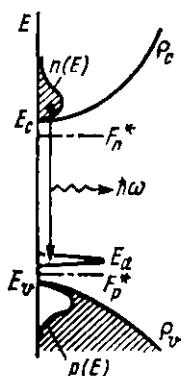


Рис. 4.29. Излучательная рекомбинация свободного электрона с дыркой на акцепторе

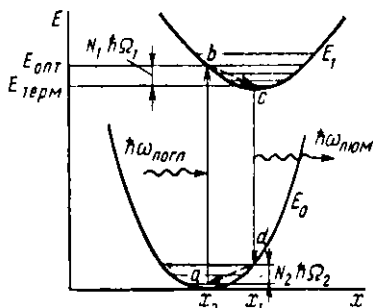


Рис. 4.30. Конфигурационная диаграмма:

x_0 и x_1 — конфигурационные координаты, отвечающие равновесному положению глубокого центра в основном E_0 и возбужденном E_1 состояниях; $E_{\text{терм}}$ и $E_{\text{опт}}$ — термическая и оптическая энергии активации

температуры или в несовершенных кристаллах эти линии уширяются и, накладываясь друг на друга, образуют широкую, бесструктурную полосу люминесценции.

Излучательная рекомбинация свободных носителей заряда на связанных состояниях. Излучательная рекомбинация свободных носителей заряда на связанных состояниях, называемая также излучательной рекомбинацией зона — примесь (или примесь — зона), обусловлена переходами носителей из одной энергетической зоны на примесные уровни вблизи противоположной зоны, как показано на рис. 4.29 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 межзонными переходами обеспечивают внутренний квантовый выход люминесценции при комнатной температуре, близкий к 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 (4.114).

Спектр излучения, обусловленного переходами элект-

ронов 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 межзонной излучательной рекомбинации и определяется выражением

$$\Phi_{\omega} = C\omega^2 (\hbar\omega - E_g + E_a)^{1/2} \frac{1}{e^{\frac{(\hbar\omega - E_g + E_a)}{kT}} - 1}. \quad (4.125)$$

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. 4.30, где энергия основного E_0 и возбужденного E_1 состояний центра представлены как функции его геометрического положения в решетке (ср. с рис. 1.7). Равновесным положениям центра в основном и возбужденном состояниях соответствуют различные конфигурационные координаты x_0 и x_1 . При оптическом возбуждении (поглощении фотона с энергией $\hbar\omega_{\text{полг}}$) 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 d сопровождается испусканием фотона с энергией $\hbar\omega_{\text{люм}}$, а затем снова происходит релаксация в наинизшее состояние a путем испускания N_2 фононов с энергией $\hbar\Omega_2$. При этом $\hbar\omega_{\text{люм}} < \hbar\omega_{\text{полг}}$. Такое смещение спектра называется *стоксовым смещением*. Оно аналогично стоксову смещению при комбинационном рассеянии света (см. § 2.6).

Атомы решетки и примесного центра колеблются вблизи равновесных положений, и энергия термической ионизации

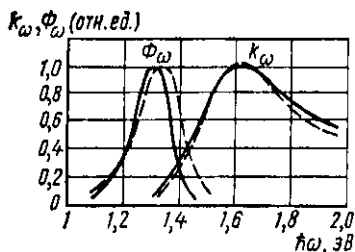


Рис. 4.31. Спектры поглощения k_ω и люминесценции Φ_ω с участием глубокого донора — кислорода в фосфиде галлия (сплошные кривые — теория, пунктирные — эксперимент)

$E_{\text{терм}}$ будет определяться минимальным энергетическим расстоянием между кривыми E_1 и E_0 на конфигурационной диаграмме. Как видно из рис. 4.30, 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. 4.31 изображены теоретические и экспериментальные спектры поглощения и люминесценции с участием глубокого донора — кислорода ($E_d = 0,9$ эВ) в фосфиде галлия. Отчетливо видно стоковое 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.

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 расстояния 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. Неравновесные электроны и дырки захватываются такими донорами и акцепторами. При температурах, достаточно низких для того,

чтобы термическая ионизация была несущественной, единственным значительным релаксационным процессом является межпримесная рекомбинация электронов и дырок.

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. 4.28, 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.

До момента рекомбинации и донор и акцептор в паре нейтральны: донор захватил электрон, а акцептор — дырку. После рекомбинации и донор и акцептор ионизованы, и между ними возникает взаимодействие, в первом приближении кулоновское. Энергия взаимодействия равна $q^2/(4\pi\epsilon_0\epsilon_r r)$, где ϵ_r — относительная статистическая диэлектрическая проницаемость среды, а r — расстояние между атомами пары. Тогда энергия фотона, излучаемого при подобном переходе:

$$\hbar\omega = E_g - (E_d + E_a) + \frac{q^2}{4\pi\epsilon_0\epsilon_r r}. \quad (4.126)$$

Учет более слабых взаимодействий типа ван-дер-ваальсовых приводит к более точной формуле

$$\hbar\omega = E_g - (E_d + E_a) + \frac{q^2}{4\pi\epsilon_0\epsilon_r r} - \frac{q^2 b^5}{4\pi\epsilon_0\epsilon_r r^6}, \quad (4.127)$$

где b — константа диполь-дипольного поляризационного взаимодействия между нейтральными примесными центрами.

Теория возмущений дает соотношение

$$b \approx (6,5)^{1/5} a_B, \quad (4.128)$$

где a_B — боровский радиус носителя из пары с большей энергией связи.

Вероятность излучательного перехода электрона с донора на акцептор пропорциональна квадрату дипольного матричного элемента $|V_{e-p}|^2$, содержащего, как обычно,

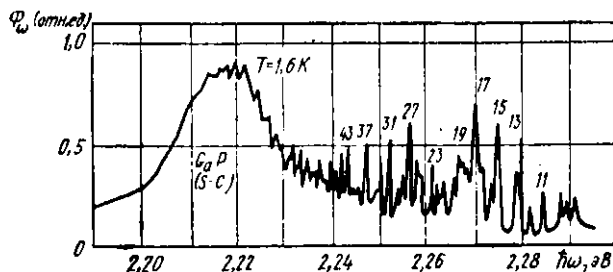


Рис. 4.32. Спектр излучательной рекомбинации электрона, захваченного мелким донором — серой (S) с дыркой, захваченный мелким акцептором — углеродом (C) в фосфиде галлия при низкой температуре (номера у линий люминесценции соответствуют номерам оболочек с данным разделением γ между донором и акцептором)

интеграл перекрытия волновых функций электрона и дырки. Так как расстояние r велико, то это перекрытие мало и экспоненциально убывает с увеличением расстояния r :

$$V_{e-p} \sim \exp(-r/a_B),$$

где a_B — боровский радиус носителя из пары с меньшей энергией связи.

Вероятность рекомбинации изменяется с расстоянием между донором и акцептором пары в соответствии с выражением

$$\omega = \omega_0 \exp(-2r/a_B), \quad (4.129)$$

где ω_0 — константа для доноров и акцепторов определенного типа в данном полупроводнике.

Все сказанное определяет ряд особенностей, наблюдаемых в излучательной рекомбинации донорно-акцепторных пар.

а. Спектр излучения с коротковолновой стороны состоит из большого числа узких линий, плотность которых на единицу энергии фотонов увеличивается с уменьшением $\hbar\omega$, как показано на рис. 4.32. Наличие дискретных линий в спектре объясняется соотношениями (4.126), (4.127), в которых все величины, за исключением r , имеют строго фиксированные значения, а r — расстояние между донором и акцептором, принимающее только дискретные значения. Последние зависят от параметров решетки и от того, какое положение данная примесь занимает в ней. С увеличением r $\hbar\omega$ уменьшается в соответствии с выражениями (4.126), (4.127). Расстояние между ближайшими линиями

при этом также уменьшается, так что при очень больших r структура в спектре излучения постепенно исчезает из-за наложения большого числа линий, как видно из рис. 4. 32. Кроме того, с увеличением r резко уменьшается вероятность рекомбинации «далеких» пар, поэтому с длинноволновой стороны в спектре наблюдается плавный спад.

б. Замена донора или акцептора приводит к сдвигу всего спектра излучения по шкале энергий на величину, в точности равную разности энергий ионизации заменяемого и заменного доноров (акцептора).

в. Следствием малого перекрытия волновых функций являются большие времена жизни носителей заряда на большинстве пар, и поскольку время жизни и энергия фотона зависят от r , спад люминесценции во времени будет неэкспоненциален. Внутри полосы люминесценции время спада зависит от энергии фотонов, уменьшаясь с увеличением $\hbar\omega$ в соответствии с выражениями (4.126) и (4.129).

г. Положение максимума в спектре излучательной рекомбинации донорно-акцепторных пар зависит от интенсивности возбуждения, сдвигаясь к большим энергиям при ее увеличении. Это связано с тем, что при больших интенсивностях возбуждения доноры и акцепторы в значительной мере заполнены, и большой вклад в излучение вносят пары с меньшим значением r , так как их вероятность рекомбинации в соответствии с (4.129) выше.

Анализ тонкой структуры спектров излучательной рекомбинации на донорно-акцепторных парах с помощью соотношений (4. 126) и (4.129) позволяет очень точно определить сумму $E_d + E_a$ из чисто оптических измерений.

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

Таблица 4.2

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	\bigcirc \pm	\bigcirc	$E_{gx} - E_{связи}$
Экситона, связанного на заряженном доноре	\ominus \mp	\oplus	$E_{gx} - E_{связи}$
Экситона, связанного на заряженном акцепторе	\ominus \pm	\ominus	$E_{gx} - E_{связи}$
Внутрицентральной переход на доноре	\ominus $-$ $*$	\oplus $-$	$E_2 - E_1$
Внутрицентральной переход на акцепторе	\ominus $+$ $*$	\ominus $+$	$E_2 - E_1$
Экситона, связанного на нейтральном доноре	\ominus $-$ \pm	\oplus $-$	$E_{gx} - E_{связи}$
Экситона, связанного на нейтральном акцепторе	\ominus $+$ \mp	\ominus $+$	$E_{gx} - E_{связи}$
Биекситона на нейтральном центре	\bigcirc \pm	\bigcirc	$E_{gx} - E_{связи}$
Электрона на доноре с дыркой на акцепторе	\oplus $-$ $+$ \ominus	\oplus \ominus	$E_g - (E_d + E_a) + E_{взаимод}$
Электрона на нейтральном центре с дыркой на акцепторе	\bigcirc $-$ $+$ \ominus	\bigcirc \ominus	$E_g - E_a - E_{связи}$
Дырки на нейтральном центре с электроном на доноре	\oplus $-$ $+$ \bigcirc	\oplus \bigcirc	$E_g - E_d - E_{связи}$

Обозначения: \bigcirc — нейтральный изоэлектронный центр; \oplus — заряженный донор; \ominus — заряженный акцептор; $-$ — электрон; $+$ — дырка; $*$ — возбужденное состояние.

уширения энергетических уровней центра за счет их расщепления и сдвига под воздействием внутрикристаллических полей, так и вследствие возможного сильного взаимодействия центра с колебаниями решетки. Уширение линии люминесценции в отдельных случаях может изменяться от до-

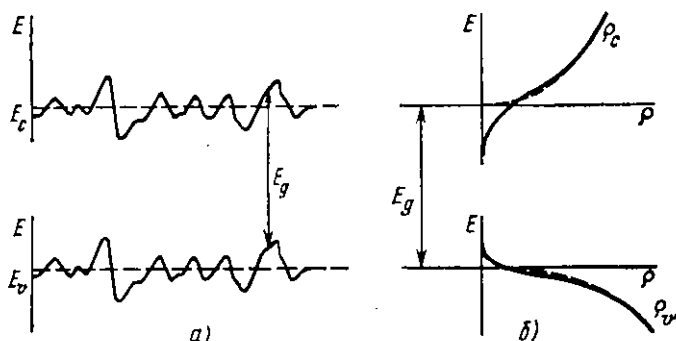


Рис. 4.33. Образование «хвостов» плотности состояний при сильном легировании полупроводников (с учетом компенсации):

a — схематическое изображение границ зон, искривленных потенциалом флуктуаций; *б* — «хвосты» плотности состояний (пунктиром показана плотность состояний в отсутствие возмущения)

лей до сотен миллиэлектронвольт в зависимости от свойств конкретного центра и матрицы. При наличии сильного электрон-фононного взаимодействия для внутризонных переходов, как и для переходов с участием глубоких центров, характерны стоксово смещение и сдвиг Франка—Кондона.

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. 4.2. Указанные переходы наблюдались в рекомбинационном излучении различных полупроводников.

Влияние сильного легирования и внешних воздействий на люминесценцию в полупроводниках. С увеличением степени легирования взаимодействие примесей приводит к изменению энергии локальных уровней и появлению зоны примесных состояний. Локальные флуктуации вызывают

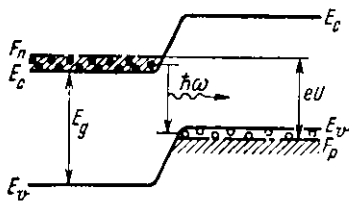


Рис. 4.34. Туннелирование испусканием фотона в вырожденном p - n -переходе (диагональное туннелирование)

также «размазывание» границ зоны проводимости и валентной зоны, запрещенная зона при этом как бы эффективно сужается. При достаточно больших концентрациях примеси примесная зона сливается с основной, образуя «хвост» плотности состояний (рис. 4.33).

Такие изменения энергетического спектра существенно влияют на вид спектров излучения: полосы уширяются, на них появляются длинноволновые «хвосты», тонкая структура «размаывается» и в конце концов исчезает. Кроме того, вследствие вырождения максимум заполненных состояний сдвигается в глубь разрешенных зон энергии. Соответственно максимумы спектров излучения для переходов типа зона — зона или зона — примесь с увеличением степени легирования смещаются в коротковолновую область. Этот эффект по существу обратен эффекту Бурштейна — Мосса в поглощении.

В p - n -переходах с сильно легированными областями ширина слоя объемного заряда может быть достаточно малой, чтобы в такой структуре наблюдались эффекты туннелирования. Помимо обычных процессов туннелирования из вырожденной зоны проводимости n -области в вырожденную валентную зону p -области без изменения энергии туннелирующего носителя заряда возможно так называемое диагональное туннелирование, при котором туннелирование сопровождается излучением фотона. Когда прямое смещение сравнительно велико, так что зоны не перекрываются, электрон может рекомбинировать с дыркой через двухступенчатый процесс, состоящий из туннелирования с испусканием фотона (рис. 4.34). Такой механизм рекомбинации дает вклад в «избыточный ток» на участке прямой ветви вольт-амперной характеристики туннельного диода. Из рис. 4.34. видно, что спектр излучения при диагональном туннелировании должен сдвигаться к более высоким энергиям по мере увеличения прямого смещения. Это является характерной особенностью процессов, обусловленных туннелированием с участием фотонов.

Диагональное туннелирование может сопровождаться не только рекомбинацией зона — зона, как изображено на

рис. 4.34. но и другими механизмами излучательной рекомбинации, в частности с участием глубоких уровней.

Влияние давления и магнитного поля на спектры излучательной рекомбинации определяется теми изменениями, которые происходят в энергетическом спектре вещества под их воздействием. Под действием деформаций полосы излучения должны сдвигаться вследствие изменения ширины запрещенной зоны. Действие одноосной деформации сводится не только к сдвигу по энергии, но также к снятию вырождения состояний и расщеплению дискретных уровней энергии. Это дает возможность по экспериментальным данным определить тип симметрии и идентифицировать различные примесные состояния.

Влияние магнитного поля на спектры люминесценции связано с зеемановским расщеплением уровней и появлением уровней Ландау в разрешенных зонах.

Внешнее электрическое поле также приводит к расщеплению энергетических уровней (эффект Штарка) и к различным туннельным эффектам (распад экситонов под воздействием электрического поля, эффект Келдыша — Франца), что соответствующим образом проявляется в излучательной рекомбинации.

§ 4.7. БЕЗЫЗЛУЧАТЕЛЬНАЯ РЕКОМБИНАЦИЯ

Очевидно, для получения высокой эффективности люминесценции необходимо знание не только ее механизмов, но и конкурирующих с ними процессов безызлучательной рекомбинации. Процессы безызлучательной рекомбинации в полупроводниках (как и в других веществах) изучены слабо и в деталях до настоящего времени остаются непонятными. 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 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 выделяться 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 то-

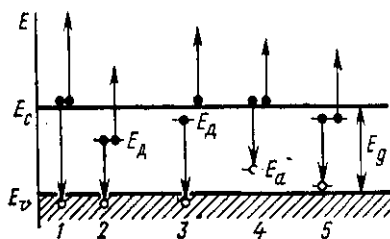


Рис. 4.35. Схематическое изображение оже-процессов для безызлучательной рекомбинации в полупроводнике n -типа:

1 — межзонная оже-рекомбинация; 2 — оже-рекомбинация электрона, захваченного на двухзарядный донор, 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

рона 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. 4.35.

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. 4.30. Если сдвиг Франка—Кондона большой и соответственно велико значение $\Delta x = x_1 - x_0$, а энергия $E_{\text{терм}}$ сравнительно небольшая, то кривая $E_1(x)$ и $E_0(x)$ могут пересечься вблизи точки c . Тогда из возбужденного состояния b центр может перейти в основное состояние путем каскадных переходов по колебательным подуровням, т. е. безызлучательным путем.

чечных 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 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. 4.35.

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 (процесс I на рис. 4.35) вероятность безызлучательной оже-рекомбинации будет пропорциональна n^2p , в то время как вероятность излучательной рекомбинации $\sim np$.

При сильном легировании в полупроводниках возможно образование различных макроскопических комплексов, преципитатов и т. д., которые служат центрами безызлучательной рекомбинации. В арсениде галлия, например, легирование серой или теллуром 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. 4.33) также облегчает условия безызлучательной рекомбинации. Все это требует в каждом конкретном случае определять оптимальную концентрацию легирующих центров, при которой эффективность люминесценции будет максимальной.

§ 4.8. ОПТИЧЕСКИЕ ЭФФЕКТЫ В НЕОДНОРОДНЫХ СТРУКТУРАХ

Под *неоднородными структурами* понимают такие системы, основные параметры которых определенным образом изменяются в пространстве. Изменяться могут, например, ширина запрещенной зоны материала, концентрация примесей, диэлектрическая проницаемость (показатель преломления). Наличие неоднородностей в кристалле приводит к появлению внутренних электрических полей и потенциальных барьеров. Типичным примером неоднородной структуры является *p-n*-переход. В квантовой электронике и оптоэлектронике наряду с «обычными» *p-n*-переходами, создаваемыми на базе одного гомогенного материала (их иногда называют гомопереходами), часто применяют *гетеропереходы*, представляющие собой контакт двух различных по химическому составу полупроводников, осуществленный в одном кристалле. Гомопереход, как известно, образуется в полупроводниковом монокристалле за счет искусственно созданного распределения примесей. В гетеропереходах помимо управления концентрацией и типом носителей заряда путем легирования появляется новая возможность управлять такой важной величиной, как ширина запрещенной зоны материала. Это имеет важное значение для создания ряда приборов оптической электроники и интегральной оптики (фотоприемников, фотопреобразователей, светодиодов, полупроводниковых лазеров, оптронов, транзисторов с оптической связью — оптотранзисторов и т. д.). 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. Последние называют также *варизонными структурами*, подчеркивая тем самым плавно изменяющуюся в пространстве ширину запрещенной зоны.

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, геометрических параметров решетки, эффективной массы носителей заряда и их подвижности. На границе раздела возможно появление различных дефектных состояний, 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

$$\frac{\Delta a}{a} = \frac{2(a_1 - a_2)}{a_1 + a_2} \quad (4.130)$$

обычно не должно превышать 0,1 %. Дополнительным условием является совпадение кристаллохимической природы контактирующих материалов.

Наиболее полно перечисленным требованиям удовлетворяют трех- и четырехкомпонентные твердые растворы в группе полупроводниковых соединений $A^{III}B^V:GaAs-AlAs$ ($Ga_xAl_{1-x}As$), $GaSb-AlSb(Ga_xAl_{1-x}Sb)$, $GaP-AlP$ ($Ga_xAl_{1-x}P$), $GaAs-GaP(GaAs_{1-x}P_x)$, $Ga-In-As-P(Ga_xIn_{1-x}As_yP_{1-y})$. Возможность плавно и в широких пределах изменять основные характеристики твердого раствора путем изменения его состава (см. § 1.5) при минимальном изменении параметров решетки и определило 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 $p-n$ -пере-

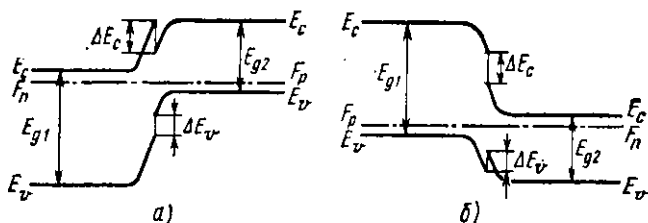


Рис. 4.36. Энергетические диаграммы *n-p*-(а) и *p-n*-гетеропереходов (б) в условиях теплового равновесия

хода было меньше влияния, оказываемого примесями, находящимися в слое объемного заряда. В трехкомпонентных твердых растворах изменение состава приводит одновременно к изменению периода идентичности решетки и ширины запрещенной зоны. В четырехкомпонентных твердых растворах типа $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ 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 твердых растворов, в котором ширина запрещенной зоны будет изменяться при неизменном периоде идентичности решетки. Такой ряд твердых растворов называют *изорешеточным*. Гетеропереходы, изготовленные на основе изорешеточных твердых растворов, по своим характеристикам наиболее близки к идеальным.

В отличие от *p-n*-переходов, изготовленных в однородном по составу полупроводнике, в гетероструктурах различают *p-n*- и *n-p*-гетеропереходы. При их обозначении на первом месте принято указывать тип электропроводности более широкозонного материала. Энергетические диаграммы *n-p* и *p-n*-гетеропереходов представлены на рис. 4.36, а, б. 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). Наличие потенциальных барьеров, различных для электронов и дырок, обуславливает некоторые особенности в электрических, фотоэлектрических и оптических свойствах таких структур.

Фотовольтаические эффекты. Наличие потенциальных барьеров и внутренних электрических полей в неоднородных структурах создает благоприятные условия для пространственного разделения носителей заряда, созданных в результате фотоактивного поглощения света. При освеще-

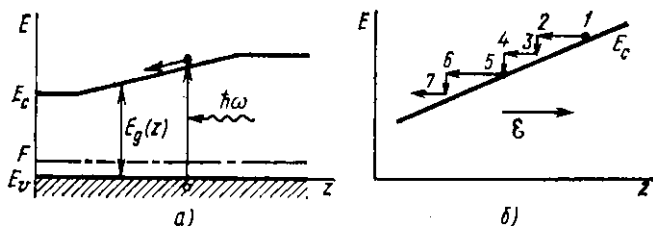


Рис. 4.37. Возникновение фото-э. д. с. в структуре с изменяющейся шириной запрещенной зоны:

а — общий вид энергетической диаграммы; б — траектория движения электрона на диаграмме $E(z)$

нии такой неоднородной структуры в ней появляется фото-э. д. с. Этот эффект, называемый *фотовольтаическим*, или *вентильным фотоэффектом*, наиболее сильно проявляется в *p-n*-переходах.

Для появления фото-э. д. с. должны существовать причины, приводящие к разделению в пространстве неравновесных носителей заряда разных знаков. Возникновение фото-э. д. с. в структуре с изменяющейся шириной запрещенной зоны показано на рис. 4.37. Это изменение может быть вызвано различными причинами, например изменением химического состава твердого раствора или наличием локальной механической деформации. В последнем случае можно наблюдать фотопьезоэлектрический эффект, обусловленный внешним локальным одноосным давлением.

В структуре с плавно изменяющейся шириной запрещенной зоны возникает локальное поле. В примере, приведенном на рис. 4.37, это поле ускоряет неосновные носители заряда (электроны) и они будут «скатываться» по потенциальному барьеру. Заметим, что стрелка на рис. 4.37, а условно отображает движение электрона. На самом деле траектория его движения в координатах $(E; z)$ представляет собой ломаную линию (рис. 4.37, б). При свободном движении электрона в кристалле под воздействием электрического поля \mathcal{E} его полная энергия E 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. 4.37, б). Электрическое поле не может изменить полную энергию электрона. Приобретая кинетическую энергию, электрон при своем движении может рассеивать ее, например, передавая решетке при

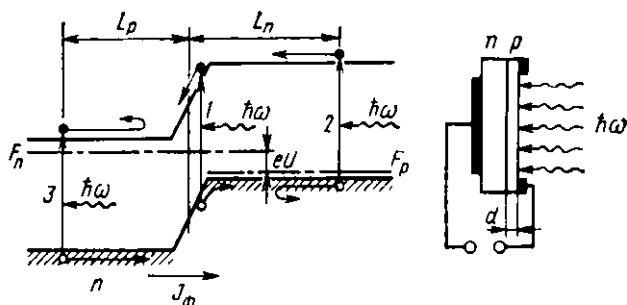


Рис. 4.38. Фотовольтаический эффект в p - n -переходе

столкновении с фононами. При этом полная энергия электронной системы уменьшается. Этим процессам рассеяния соответствуют участки 2—3, 4—5 и 6—7 на рис. 4.37, б.

Как указывалось, наиболее эффективное разделение носителей заряда происходит в слое объемного заряда p - n -перехода. Поэтому в таких структурах 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. 4.38. Если поглощение фотона происходит в области объемного заряда p - n -перехода (процесс 1 на рис. 4.38), 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 фото-э. д. с.

При этом p -область будет заряжаться положительно, а n -область отрицательно.

Таким образом, под действием света в p - n -переходе возникает фототок J_{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

$$J_y = J_s \left(e^{\frac{eU}{kT}} - 1 \right). \quad (4.131)$$

Здесь J_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

$$J = J_s \left(e^{\frac{eU}{kT}} - 1 \right) - J_{\Phi}. \quad (4.132)$$

Это общее уравнение фотодиода. Вольт-амперные характеристики p - n -перехода в отсутствие освещения ($G=0$) и при воздействии света ($G>0$) представлены на рис. 4.39. В случае разомкнутой цепи (режим холостого хода) фото э. д. с. определяется из (4.132) при $J=0$:

$$U_x = \frac{kT}{e} \ln \left(\frac{J_{\Phi}}{J_s} + 1 \right). \quad (4.133)$$

Из (4.132) при $U=0$ находим, что ток короткого замыкания равен фототоку:

$$J_k = J_{\Phi}.$$

В свою очередь фототок для p - n -перехода с ос. той поверхностью A определяется скоростью генерации избыточных носителей заряда G и диффузионными длинами L_n и L_p :

$$J_{\Phi} = J_k = Ae (L_n + L_p) G. \quad (4.134)$$

В промежуточном случае, когда p - n -переход нагружен на сопротивление R , ток и напряжение находят из общего уравнения фотодиода (4.132):

$$U_R = \frac{kT}{e} \ln \left(\frac{J_R + J_{\Phi}}{J_s} + 1 \right), \quad (4.135)$$

где $J_R = U_R/R$ (рис. 4.39).

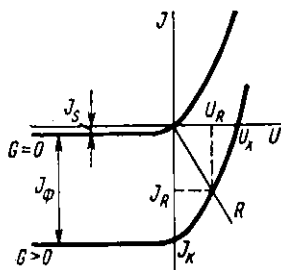


Рис. 4.39. Вольт-амперные характеристики p - n -перехода в отсутствие освещения ($G=0$) и при воздействии света ($G>0$)

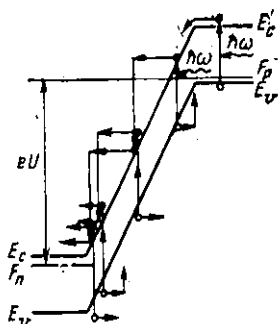


Рис. 4.40. Лавинное умножение генерированных светом носителей заряда в p - n -переходе, смещенном в обратном направлении

При увеличении интенсивности света, т. е. скорости генерации G , возрастают ток короткого замыкания и напряжение холостого хода. Ток J_K возрастает линейно с интенсивностью света, а напряжение U_x — по логарифмическому закону. Предельное значение U_x приблизительно соответствует ширине запрещенной зоны.

Если к p - n -переходу приложить обратное напряжение, близкое к напряжению лавинного пробоя, то в такой системе возможно усиление фототока, связанное с лавинным умножением генерированных светом носителей заряда, как показано на диаграмме рис. 4.40. Основным требованием при этом является очень высокая однородность материала p -перехода во избежание появления самопроизвольных пробоев в дефектных областях, пробивное напряжение должно быть ниже усредненного по площади p - n -перехода.

Рассмотрим спектральные характеристики. Заряды, определяющие возникновение фото-э. д. с. в p - n -переходе (или в других неоднородных структурах), генерируются в процессе поглощения фотона в слое толщиной $L_n + L_p$. К появлению фото-э. д. с. приводят только те процессы поглощения, в результате которых создаются неосновные носители заряда. Доминирует здесь собственное поглощение. Процессы фотонионизации мелких примесей не дают вклада в фото-э. д. с., а процессы их фотонейтрализации могут вызвать фото-э. д. с.

Рассмотрим спектральную зависимость фоточувствительности p - n -перехода. Для определенности освещение

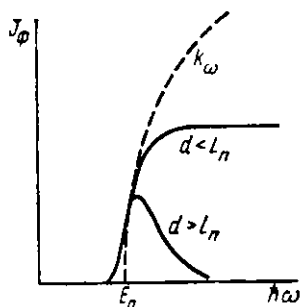


Рис. 4.41. Спектральные зависимости фототока короткого замыкания $J_\phi(\hbar\omega)$ p - n -перехода (пунктиром изображен спектр собственного поглощения)

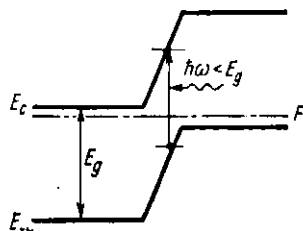


Рис. 4.42. Фотовольтаический эффект в p - n -переходе, возникающий в результате туннелирования носителей заряда с поглощением фотона ($\hbar\omega < E_g$)

будем проводить с p -стороны, как на рис. 4.38. 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 фотоэффект. При малых значениях $\hbar\omega - E_g$ фототок пропорционален показателю поглощения, поскольку скорость генерации G пропорциональна k_ω [см. выражение (4.90)]. 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. 4.41).

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 достигнет p - n -перехода. Поэтому при $d > L_n$ в спектрах фото-э. д. с. и фототока будет наблюдаться максимум, как показано на рис. 4.41.

Отметим, что при малых энергиях фотонов ($\hbar\omega \leq E_g$) спектральная зависимость фотовольтаического эффекта не будет точно следовать спектральной зависимости показателя собственного поглощения однородного материала. Это отличие может быть обусловлено двумя причинами. Во-первых, в спектр фотоответа могут вносить вклад переходы

с акцепторных состояний в зону проводимости или из валентной зоны на донорные состояния (процессы фотонейтрализации мелких примесей). Во-вторых, в области p - n -перехода имеется сильное электрическое поле. Поэтому электронно-дырочные пары в слое объемного заряда могут быть созданы фотонами с энергиями, меньшими E_g , что объясняется эффектом Келдыша — Франца: в сильном электрическом поле за счет туннелирования носителей заряда в запрещенную зону возможно поглощение фотона с энергией $\hbar\omega < E_g$ (рис. 4.42). Поскольку вероятность туннелирования в запрещенную зону экспоненциально спадает с расстоянием, это приведет к появлению экспоненциального «хвоста» на длинноволновом краю спектра фотоответа. Этот эффект тем больше, чем больше напряженность электрического поля в p - n -переходе.

Из приведенных результатов 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. Это можно достигнуть в гетеропереходе, используя его широкозонную часть в качестве «окна», через которое производится освещение. Такое «окно» прозрачно для света, поглощаемого в узкозонной части p - n -перехода (рис. 4.43, 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. 4.43, в). Используя твердые растворы, можно в заданных пределах изменять спектральную характеристику. В частности, используя гетеропереход, энергетическая диаграмма которого приведена на рис. 4.43, б можно получить очень узкую область спектральной чувствительности (селективный фотоэлемент), характеристика которого соответствует кривой 2 на рис. 4.43, в.

Фотовольтаические эффекты могут наблюдаться и на контакте полупроводника с металлом. В месте такого контакта происходит перераспределение зарядов, в результате чего в полупроводнике возникает слой, обедненный носителями заряда, и появляется энергетический барьер, называемый барьером Шоттки. Высота барьера равна разности работ выхода для металла и полупроводника. В такой структуре возможно также эффективное разделение носителей заряда противоположных знаков, как показано на

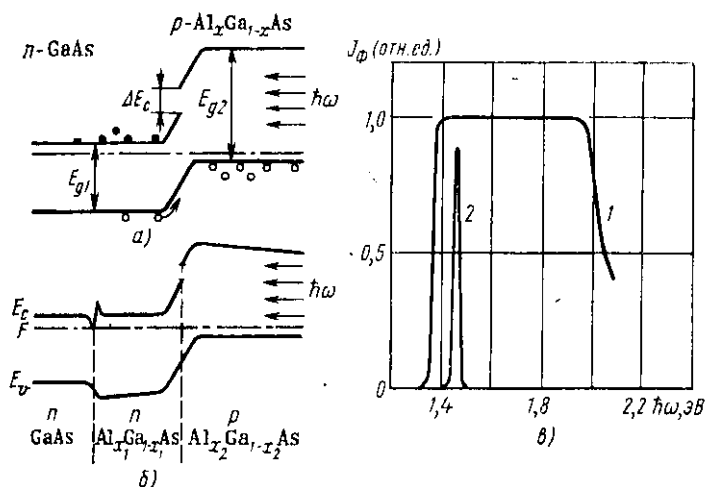


Рис. 4.43. Энергетические диаграммы (а, б) и спектральные характеристики (в) p - n -гетеропереходов при освещении через широкозонное «окно»

рис. 4.44, 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 Шоттки работает как фотодиод с p - n -переходом. Кроме того, на барьере Шоттки возможно наблюдать фотовольтаический эффект для $\hbar\omega_2 < E_g$, но $\hbar\omega_2 > e\phi$ (рис. 4.44, б). В этом случае электроны, возбужденные в металлической пленке, могут преодолеть

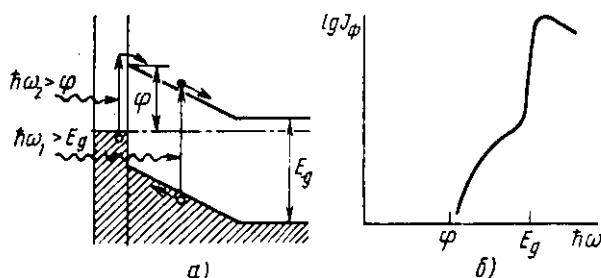


Рис. 4.44. Фотоэффект на контакте металл — полупроводник (на барьере Шоттки):

а — энергетическая диаграмма; б — спектральная характеристика

потенциальный барьер и перейти в полупроводник. При подаче больших напряжений в запирающем направлении в диодах Шоттки возможно лавинное умножение, как в p - n -переходах.

Помимо рассмотренных фотовольтаических эффектов, при которых генерируемая фото-э. д. с. не превышает ширины запрещенной зоны полупроводника, в тонких пленках некоторых полупроводников (Se, PbS, Si, Ge, GaAs, GaP, HgTe, ZnS и др.), прошедших специальную обработку, наблюдался фотовольтаический эффект, при котором генерируемое напряжение существенно превышает ширину запрещенной зоны и может достигать несколько киловольт (до 5 кВ). Этот эффект называют эффектом аномальных фотонапряжений. Хотя в деталях причина этого эффекта окончательно не ясна, наиболее вероятно, что он является следствием суммирования элементарных фото-э. д. с., генерируемых в микропереходах, на границах зерен или в результате эффекта Дембера в микроячейках пленки, имеющей зернистую структуру.

Электролюминесценция. Как отмечалось в § 4.6, электролюминесценция — это люминесценция, возбуждаемая внешним электрическим полем. При электролюминесценции осуществляется непосредственное преобразование электрической энергии в световую, что часто является наиболее удобным для различного рода практических приложений. Электрическое возбуждение люминесценции в твердых телах в подавляющем большинстве случаев возможно осуществить только в неоднородных структурах.

Наиболее эффективным методом электрического возбуждения является инжекция неосновных носителей заряда через p - n -переход в случае приложения к нему напряжения U в прямом направлении. Соответствующие энерге-

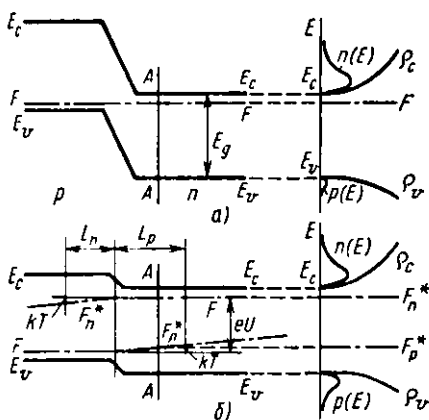


Рис. 4.45. Энергетические диаграммы и распределение носителей заряда по энергиям (по сечению $A-A$) для p - n -перехода в отсутствие внешнего напряжения (а) и в случае приложения напряжения U в прямом направлении (б)

тические диаграммы при $U=0$ и $U>0$ приведены на рис. 4.45, а, б. При смещении p - n -перехода в прямом направлении потенциальный барьер снижается на eU , электроны инжектируются в p -область, а дырки — в n -область (рис. 4.45, б).

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 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. 4.45, 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 A — A показаны на рис. 4.45. Они по существу не отличаются от соответствующего распределения, приведенного на рис. 4.24.

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 зна-

чение 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 не может превышать концентрацию этих же носителей в эмиттере, где они являются основными. По этой причине для создания состояний с инверсной населенностью при инжекции неосновных носителей заряда через такой p - n -переход, т. е. для удовлетворения условия (4.109), необходимо, чтобы минимум одна (обычно обе) из областей p - n -перехода была вырожденной. Это проиллюстрировано на рис. 4.46, где приведены диаграммы p - n -перехода в отсутствие смещения (рис. 4.46, а) и при максимально возможном смещении в прямом направлении (рис. 4.46, б), соответствующем полному спрямлению энергетического барьера. Распределение носителей заряда по энергиям на рис. 4.46 для простоты показано при температуре $T=0$. «Хвосты» плотности состояний появляются вследствие сильного легирования материала.

Применение гетеропереходов здесь также дает ряд преимуществ по сравнению с p - n -переходами на основе гомогенных материалов. Они определяются в основном эффектами широкозонного «окна», односторонней инжекции и «сверхинжекции».

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 квазимежзонными переходами с участием мелких примесей. В «обыч-

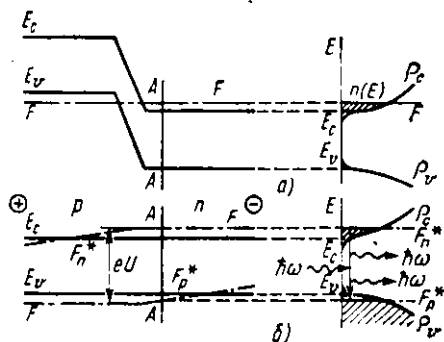


Рис. 4.46. Энергетические диаграммы и распределение носителей заряда по энергиям (при $T=0$) для вырожденного p - n -перехода в отсутствие смещения (а) и при максимально возможном смещении в прямом направлении (б)

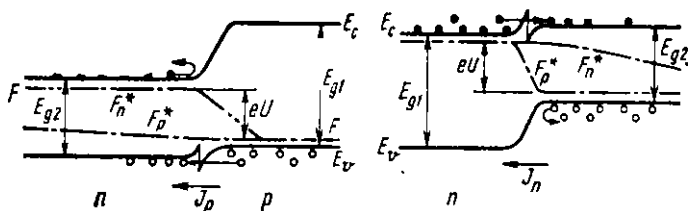


Рис. 4.47. Эффект односторонней инжекции в p - n - и n - p -гетеропереходах

ном 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 $\hbar\omega \approx E_g$, свободно проходит через широкозонную область, где $E_{g1} > E_{g2}$.

2. Эффект односторонней инжекции (рис. 4.47). В гетеропереходах в отличие от гомопереходов за счет наличия дополнительного энергетического барьера для электронов (или дырок) всегда будет осуществляться преимущественная инжекция носителей зарядов из широкозонной части в узкозонную. Так, для p - n -гетероперехода барьер в зоне проводимости препятствует инжекции электронов в широкозонную область, а для n - p -гетероперехода инжекции дырок в n -область препятствует барьер в валентной зоне. Токи инжектированных через гетеропереход 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 электронвольт, 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. 4.48). В гомопереходе концентрация неосновных носителей заряда, инжектированных из эмиттера, не может превышать их концентрации в эмиттере (см. рис. 4.45 и 4.46). В гетеропереходе за счет наличия разрывов в зоне проводимости ΔE_c (для n - p -гетероперехода) или в валентной зоне ΔE_v (для p - n -гетероперехода) при достаточно большом напряжении смещения возможно как бы образование «отрицательного»

барьера для электронов (дырок). Это проиллюстрировано на рис. 4.48 для n - p -гетероперехода. Узкий положительный 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. 4.48 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 n - p -гетеропереходе к их равновесной концентрации в эмиттере n_n (в n -области) равно

$$\frac{n_p}{n_n} \rightarrow \exp \frac{\Delta E_c}{kT}. \quad (4.136)$$

Аналогично в p - n -гетеропереходе

$$\frac{p_p}{p_p} \rightarrow \exp \frac{\Delta E_v}{kT}. \quad (4.137)$$

Эта особенность инжекции в гетеропереходе делает его уникальным по эффективности инжектором, что очень важно для полупроводниковых лазеров.

Указанные свойства гетеропереходов позволяют существенно облегчить достижение условия инверсии в инжекционных лазерах. В качестве примера рассмотрим гетероструктуру $(n\text{-Al}_x\text{Ga}_{1-x}\text{As}) - (p\text{-GaAs}) - (p^+\text{-Al}_x\text{Ga}_{1-x}\text{As})$. На рис. 4.49, а, б для этой структуры приведены энергетические диаграммы в отсутствие внешнего напряжения (в условиях равновесия) и при прямом смещении. На рис. 4.49, в представлено распределение концентрации электронов и дырок.

Центральная область (p -GaAs) является активной. Из широкозонной n -области в нее инжектируются электроны. При этом за счет эффекта «сверхинжекции» концентрация электронов в активной области может оказаться больше, чем в n -эмиттере (скачок концентрации на диаграмме рис. 4.49, в). Одновременно потенциальный барьер в валентной зоне на гетерогранице $n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ ограничи-

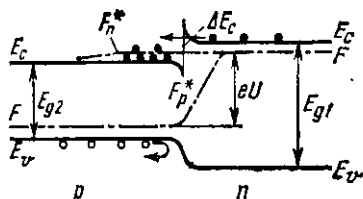


Рис. 4.48. Эффект «сверхинжекции» в n - p -гетеропереходе

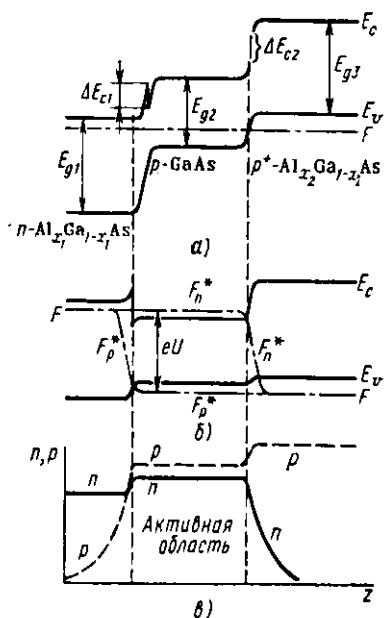


Рис. 4.49. Энергетические диаграммы гетероструктуры ($n\text{-Al}_x\text{Ga}_{1-x}\text{As}$) — ($p\text{-GaAs}$) — ($p^+\text{-Al}_{x_2}\text{Ga}_{1-x_2}\text{As}$) в условиях равновесия (а) и при прямом смещении (б); распределение концентрации носителей заряда как функции координаты (в)

вает распространение дырок из p -области, «запирающая» их в пределах активной области.

Из широкозонной сильно легированной p^+ -области происходит инжекция дырок в активную область. Для гетеропереходов в системе $\text{GaAs—Ga}_x\text{Al}_{1-x}\text{As}$ разрыв в валентной зоне ΔE_v обычно мал, так что вся разность энергий $E_{g1} - E_{g2}$ приходится на зону проводимости. Поэтому в данном случае эффект «сверхинжекции» из p -

$\text{Al}_x\text{Ga}_{1-x}\text{As}$ в $n\text{-GaAs}$ отсутствует и в качестве эмиттера дырок применен легированный до вырождения p^+ -материал. Барьер в зоне проводимости ограничивает распространение электронов из активной области в p^+ -эмиттер. Как видно, используя свойства гетеропереходов, возможно, создать в активной области наиболее благоприят-

ные условия для получения инверсной населенности. Возбуждение люминесценции электрическим полем в неоднородных структурах может быть осуществлено не только путем инжекции неосновных носителей заряда с приложением к $p\text{-}n$ -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. 4.50. Здесь в качестве примера представлена энергетическая диаграмма для структуры металл — диэлектрик — полупроводник n -типа. Слой диэлектрика, который является достаточно тон-

ким, 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. 4.50, 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, из рис. 4.40 видно, что при лавинном пробое p - n -перехода, смещенном в обратном направлении, происходит генерация неосновных носителей заряда, которые в последующем могут рекомбинировать с основными носителями излучательным путем. Такая люминесценция называется предпробойной. Она может возникать не только в p - n -переходах, но также на границах зерен и прочих неоднородностях, где возможно локальное повышение напряженности электрического поля и развитие процесса локального микропробоя.

Волноводные свойства. Управление электронными потоками осуществляется довольно просто с помощью внешних электрических или магнитных полей. Управлять световым потоком таким же способом невозможно, поскольку фотоны не реагируют на внешние электрические и магнитные поля. По существу все применяемые в настоящее время методы управления световыми потоками основаны на различных оптических явлениях, происходящих в неоднородных структурах. Например, фокусирующее действие оптической линзы основано на преломлении световых лучей при прохождении границы раздела двух сред с различными значениями показателя преломления.

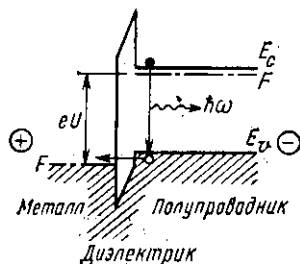


Рис. 4.50. Возбуждение электролюминесценции в структуре металл — диэлектрик — полупроводник n -типа с использованием процессов туннелирования электрона из полупроводника в металл через тонкий слой диэлектрика

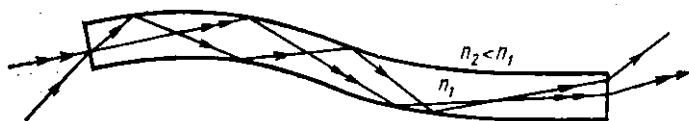


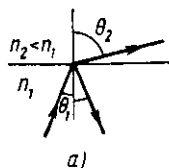
Рис. 4.51. Схематическое изображение распространения световых лучей в диэлектрическом волноводе (световоде)

Явление полного внутреннего отражения может быть использовано для создания различного рода волноводных структур, в которых при соблюдении определенных условий световая волна будет распространяться по траектории, определяемой геометрией структуры. Наглядным примером такой волноводной структуры является тонкий *диэлектрический волновод*, называемый также *световодом*. В простейшем исполнении он представляет собой тонкое волокно, изготовленное из оптически прозрачного материала, в торец которого вводится оптическое излучение. Распространение этого излучения по оптическому волокну, имеющему диаметр порядка нескольких микрометров, осуществляется за счет многократных процессов полного внутреннего отражения, как схематически показано на рис. 4.51. Если показатель преломления материала, из которого изготовлено волокно, в $\sqrt{2}$ раз превышает показатель преломления окружающей среды (это соответствует критическому углу 45°), то весь свет, падающий на один плоский торец световода, будет выходить из другого его конца; при этом необходимо, чтобы диэлектрический стержень (волокно) был прямым, его материал абсолютно однородным и прозрачным, а поверхность гладкой и чистой. Если волокно изогнуто, как на рис. 4.51, то наблюдаются потери света, которые зависят от отношения радиуса кривизны изгиба к диаметру волокна. Если это отношение больше 100, то потери незначительны. В случае если разница между показателями преломления волокна n_1 и среды n_2 мала, так что $\Delta n = (n_1 - n_2) \ll n_1$, то максимальный угол отклонения от оси световода, при котором еще имеется полное внутреннее отражение, определяется из соотношения

$$\sin \theta = \sqrt{n_1^2 - n_2^2} \approx \sqrt{2n_1 - \Delta n}. \quad (4.138)$$

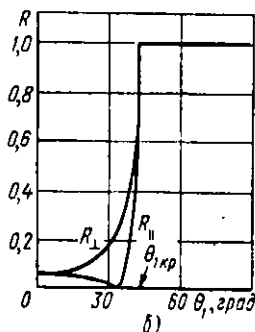
Тонкие гибкие волокна изготавливают из стекла, кварца, пластмассы или другого оптически совершенного материала и собирают в гибкие жгуты, форма и размеры которых определяются их функциональным назначением. Количество волокон в жгуте может достигать сотен тысяч и более.

Каждое волокно представляет собой оптический волновод, входное и выходное сечения которого линейно связаны с изменением масштаба передаваемого изображения. Для уменьшения влияния волокон друг на друга за счет нарушения полного внутреннего от-



а)

Рис. 4.52. Прохождение светового луча из оптически более плотной в оптически менее плотную среду (а); зависимость коэффициента отражения от угла падения θ_1 для двух поляризаций (б)



б)

ражения при их взаимном контакте в жгуте, волокна покрывают оболочкой, показатель преломления которой меньше показателя преломления самого стержня. Важным является тот факт, что уже в настоящее время созданы и освоены промышленностью световоды, потери оптического излучения в которых могут быть менее 0,5 дБ/км, т. е. при прохождении через такой световод длиной 20 км интенсивность световой волны уменьшается в 5 раз и менее.

Используя свойства световодов, с помощью волоконной оптики можно решить многие задачи геометрического и пространственного преобразования оптических сигналов. Свойствами световодов обладают не только волокна цилиндрической формы, но и тонкие пленки, а также p - n -переходы. Эти свойства широко используют при создании приборов и устройств оптоэлектроники и интегральной оптики.

На самом деле процесс распространения оптической волны по световоду является значительно более сложным, чем это представлено в чисто геометрическом приближении на рис. 4.51.

Рассмотрим более подробно явление полного внутреннего отражения от границы раздела двух сред, не имеющих потерь, т. е. $\sigma_1 = \sigma_2 = 0$, $\kappa_1 = \kappa_2 = 0$. Пусть электромагнитная волна проходит из оптически более плотной среды с показателем преломления $\hat{n}_1 = n_1$ в оптически менее плотную среду с $\hat{n}_2 = n_2 < n_1$ и падает на границу раздела под углом θ_1 , как показано на рис. 4.52, а. На рис. 4.52, б представлены зависимости коэффициента отражения от угла падения θ_1 для двух поляризаций. Если угол падения больше

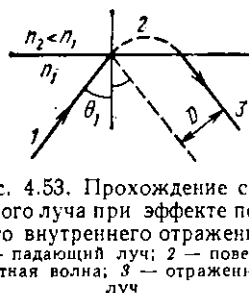


Рис. 4.53. Прохождение светового луча при эффекте полного внутреннего отражения: 1 — падающий луч; 2 — поверхностная волна; 3 — отраженный луч

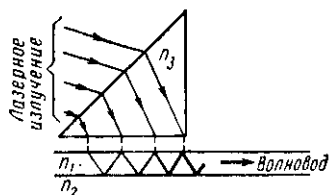


Рис. 4.54. Ввод лазерного излучения в плоскоплечный волновод с помощью элемента связи — призмы

некоторого критического угла, определяемого из выражения (4.15):

$$\theta_{\text{кр}} = \arcsin n_2/n_1, \quad (4.139)$$

то для любой поляризации коэффициент отражения равен единице, а коэффициент пропускания — нулю: электромагнитная волна испытывает полное отражение от границы раздела.

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: $\theta_1 > \theta_{\text{кр}}$, то он отражается не точно в месте падения, как на рис. 4.52, а, 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 более подробно показано на рис. 4.53. При этом амплитуда прошедшей волны экспоненциально затухает по мере удаления от границы раздела с показателем затухания

$$\alpha = \frac{2\pi}{\lambda_1} \sqrt{\sin^2 \theta_1 - \left(\frac{n_2}{n_1}\right)^2}. \quad (4.140)$$

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. 4.53 связано с α соотношением

$$D = 0,52 n_2 \frac{2\pi}{\alpha}. \quad (4.141)$$

Напомним, что рассматривается прохождение волны

через границу раздела двух непоглощающих сред в предположении, что каждая из сред (в частности, вторая среда) имеет бесконечную протяженность от границы раздела. Это предположение вполне корректно, если толщина оптически менее плотной второй среды, по крайней мере, в несколько раз превышает глубину проникновения волны. Если данное условие не выполняется, то полное внутреннее отражение будет отсутствовать и часть энергии будет передаваться преломленной поверхностной волне. Это является довольно важным для ряда технических применений в устройствах интегральной оптики. В частности, на таком эффекте основаны призматические устройства ввода и вывода излучения из тонкопленочных волноводов (рис. 4.54).

Призма с высоким показателем преломления n_3 располагается на малом расстоянии (порядка λ) от поверхности плоского диэлектрического волновода (пленки). Луч света входит в призму и падает на ее боковую грань под углом полного внутреннего отражения (для границы призма — воздух). Вследствие рассмотренного эффекта «просачивания» на границе призма — волновод возникает вытекающая волна, которая проходит тонкий слой воздуха и входит в волновод.

Аналогично можно осуществить вывод энергии из тонкопленочного волновода с помощью обратного процесса.

Рассмотрим прохождение света по простейшему пленочному волноводу, представляющему собой нанесенную на подложку полосу тонкой пленки, показатель преломления n_1 , которой больше показателей преломления подложки n_2 и окружающей среды n_3 (например, воздуха). В направлении y пленка имеет бесконечную протяженность. Распространение светового луча в таком волноводе показано на рис. 4.55. 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. 4.55, необходимо решить уравнения Максвелла (4.2) 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 на-

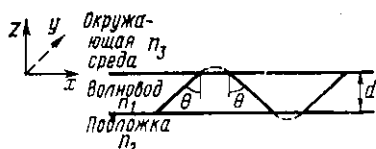


Рис. 4.55. Распространение светового луча в плоском планарном волноводе

кладывается сама на себя. Результирующая волна в волноводе представляет собой сумму полей первоначальной и отраженных волн. Очевидно, интенсивность волны будет нарастать только в том случае, если основная и отражен-

ные волны складываются в фазе. Это условие приводит к резонансному эффекту. Он означает, что в оптических волноводах (как и в 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 θ и эффективный показатель преломления $n_{эфф}$, определяющий скорость движения волноводной моды вдоль оптической оси.

В световой волне, распространяющейся в оптическом волноводе, как и в СВЧ-волне в волноводе, лишь одно из полей \mathbf{E} или \mathbf{H} может быть чисто поперечным, т. е. перпендикулярным направлению распространения волны. Соответственно этому различают магнитные H (или TE) и электрические E (или TH) волноводные моды. В оптическом волноводе, изображенном на рис. 4.55, могут распространяться различные волны $E_{m,n}$ или $H_{m,n}$. Они будут отличаться числом полуволин m и n первичной волны, укладывающихся в направлениях z и y (см. § 3.3). В качестве примера на рис. 4.56 показано распределение амплитуды 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. переносится вытекающими волнами, расположенными вне волноводной пленки.

Для лазерной структуры на основе p - n -гетероперехода, изображенной на рис. 4.49, показатель преломления активной узкозонной части всегда больше показателя преломления широкозонных областей. Поэтому в такой структуре

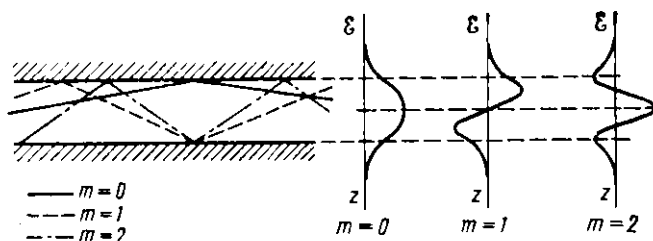


Рис. 4.56. Волноводные моды в плоском оптическом волноводе (справа — распределение электрического поля для наинизших мод с $m=0, 1, 2$; слева — условное геометрическое изображение хода лучей)

хорошо проявляются волноводные свойства. Отметим также, что для проявления волноводных свойств иногда достаточно очень незначительного изменения показателя преломления ($\Delta n \leq 10^{-3}$). Это изменение не обязательно должно быть скачкообразным, оно может быть и плавным.

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* *.

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.

* Первые работы по нелинейной оптике были выполнены С. И. Вавиловым еще в 30-х годах. В частности, в 1925 г. наблюдалось уменьшение поглощения света при увеличении его интенсивности (эффект просветления) в урановом стекле. В дальнейшем значительный вклад в развитие нелинейной оптики внесли советские (Р. В. Хохлов, С. А. Ахманов) и американские (П. Франкен, Н. Бломберген) ученые.

§ 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 \mathcal{P} , вызванная электрическим полем световой волны, пропорциональна напряженности этого поля \mathcal{E} . Это означает, что предполагалась линейная связь между \mathbf{D} и \mathcal{E} , \mathbf{B} и \mathcal{H} , \mathbf{j} и \mathcal{E} в уравнениях Максвелла (2.4) и (4.2). Величины ϵ , μ и σ в (2.5) и (4.1), отражающие реакцию среды, считались не зависящими от \mathcal{E} и \mathcal{H} . Только при этом решении уравнений Максвелла получаются в виде (2.15), (2.23) и (4.4). 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 \mathbf{D} и \mathcal{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 лазеров дало возможность получать световые пучки, напряженность электрического поля в которых достигает 10^7 В/см и более, т. е. сравнима с внутриатомными полями. Естественно, что в таких полях линейная связь между \mathbf{D} и \mathcal{E} должна нарушаться.

Так как в оптическом диапазоне $\mu \approx 1$, то нелинейные оптические эффекты могут определяться нелинейной связью между \mathbf{D} и \mathcal{E} (нелинейной поляризацией) или между \mathbf{j} и \mathcal{E} (нелинейным поглощением).

Диэлектрическая проницаемость среды ϵ определяется вектором ее поляризованности \mathcal{P} и связана с относительной диэлектрической восприимчивостью χ соотношениями

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathcal{E} = \epsilon_0 \mathcal{E} + \mathcal{P}, \quad \mathcal{P} = \epsilon_0 \chi \mathcal{E}, \quad \epsilon_r = 1 + \chi. \quad (5.1)$$

В общем случае $\chi = \chi(\mathcal{E})$. Ограничиваясь для простоты скалярными соотношениями, можно записать

$$\mathcal{P} = \epsilon_0 (\chi_1 \mathcal{E} + \chi_2 \mathcal{E}^2 + \chi_3 \mathcal{E}^3 + \dots) \quad (5.2)$$

или

$$\mathcal{P} = \mathcal{P}_л + \mathcal{P}_{нл},$$

где первое слагаемое $\epsilon_0 \chi_1 \mathcal{E}$ определяет линейную поляризацию $\mathcal{P}_л$, а остальные слагаемые — нелинейную поляризацию $\mathcal{P}_{нл}$. Все коэффициенты χ_n не зависят от \mathcal{E} и характеризуют свойства вещества.

Отношение каждого последующего члена ряда к предыдущему имеет порядок величины $\mathcal{E}/\mathcal{E}_a$, где \mathcal{E}_a для электронной поляризации определяется напряженностью внутриатомного электрического поля: $\mathcal{E}_a \approx 10^8$ В/см. Именно поэтому в «обычном» световом поле, создаваемом «классическими» источниками, где \mathcal{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 \mathcal{E} . Действительно, для такой среды при любом \mathcal{E} должно выполняться равенство $\mathcal{P}(-\mathcal{E}) = -\mathcal{P}(\mathcal{E})$, что возможно лишь при $\chi_2 = \chi_4 = \dots = 0$. По этой причине в нелинейной оптике обычно используют кристаллы, не обладающие центром инверсии.

Рассмотрим для определенности среду, в которой существенна квадратичная нелинейность, т. е. ограничимся первым нелинейным членом в разложении (5.2):

$$\mathcal{P}^{(2)} = \epsilon_0 \chi_2 \mathcal{E}^2. \quad (5.3)$$

Пусть в среде распространяются две электромагнитные волны, начальные фазы которых для простоты считаем одинаковыми:

$$\begin{aligned} \mathcal{E}_1(\mathbf{r}, t) &= \mathcal{E}_0^{(1)} \cos(\omega_1 t - \mathbf{k}_1 \mathbf{r}), \\ \mathcal{E}_2(\mathbf{r}, t) &= \mathcal{E}_0^{(2)} \cos(\omega_2 t - \mathbf{k}_2 \mathbf{r}). \end{aligned} \quad (5.4)$$

Подставляя (5.4) в (5.3), получаем

$$\mathcal{P}^{(2)} = \epsilon_0 \chi_2 [\mathcal{E}_0^{(1)^2} \cos^2(\omega_1 t - \mathbf{k}_1 \mathbf{r}) + \mathcal{E}_0^{(2)^2} \cos^2(\omega_2 t - \mathbf{k}_2 \mathbf{r}) + 2\mathcal{E}_0^{(1)} \mathcal{E}_0^{(2)} \cos(\omega_1 t - \mathbf{k}_1 \mathbf{r}) \cos(\omega_2 t - \mathbf{k}_2 \mathbf{r})]. \quad (\text{ар}) \rightarrow$$

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:

$$\mathcal{P}(2\omega_1) = \frac{1}{2} \epsilon_0 \chi_2 \mathcal{E}_0^{(1)^2} \cos(2\omega_1 t - 2\mathbf{k}_1 \mathbf{r}), \quad (5.5)$$

$$\mathcal{P}(2\omega_2) = \frac{1}{2} \epsilon_0 \chi_2 \mathcal{E}_0^{(2)^2} \cos(2\omega_2 t - 2\mathbf{k}_2 \mathbf{r}), \quad (5.5a)$$

$$\mathcal{P}(\omega_1 + \omega_2) = \epsilon_0 \chi_2 \mathcal{E}_0^{(1)} \mathcal{E}_0^{(2)} \cos[(\omega_1 + \omega_2)t - (\mathbf{k}_1 + \mathbf{k}_2)\mathbf{r}], \quad (5.56)$$

$$\mathcal{P}(\omega_1 - \omega_2) = \epsilon_0 \chi_2 \mathcal{E}_0^{(1)} \mathcal{E}_0^{(2)} \cos[(\omega_1 - \omega_2)t - (\mathbf{k}_1 - \mathbf{k}_2)\mathbf{r}] \quad (5.5в)$$

и постоянного слагаемого, описывающего статическую поляризацию:

$$\mathcal{P}(0) = \frac{1}{2} \epsilon_0 \chi_2 (\mathcal{E}_0^{(1)2} + \mathcal{E}_0^{(2)2}). \quad (5.5г)$$

В § 4.1. отмечалось, что уменьшение скорости света в веществе и появление показателя преломления вызвано поляризацией среды под действием электрического поля световой волны. В линейном 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 $\chi_3 \mathcal{E}^3$, 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 уравнения (4.1), приводящие к нелинейности всей системы уравнений, описывающих распространение световых волн в веществе. 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, представлена 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 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 форме, как в (5.5), а в показательной. Будем считать, что волна распространяется в направлении z . Тогда

$$\mathcal{P}(2\omega) = \mathcal{P}_0 \exp[i(2\omega t - 2k(\omega)z)], \quad (5.6)$$

где

$$k(\omega) = \frac{2\pi}{\lambda_\omega} = \frac{\omega n(\omega)}{c} \quad (5.7)$$

— величина волнового вектора на частоте ω ; λ_ω — длина световой волны с частотой ω в веществе; $n(\omega)$ — показатель преломления на частоте ω ; \mathcal{P}_0 — амплитуда поляризации при $z=0$.

Из соотношений (5.5) имеем $\mathcal{P}_0 = \epsilon_0 \chi_2 \mathcal{E}_0^2$.

Генерируемую волной поляризации (5.6) электромагнитную волну можно записать в виде

$$\begin{aligned} \mathcal{E}(2\omega) &= \\ &= \mathcal{E}_0(z) \exp[i(2\omega t - \\ &\quad - k(2\omega)z)], \quad (5.8) \end{aligned}$$

где волновой вектор для этой волны (на частоте 2ω)

$$k(2\omega) = \frac{2\pi}{\lambda_{2\omega}} = \frac{2\omega n(2\omega)}{c}. \quad (5.9)$$

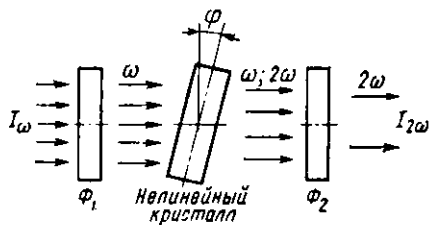


Рис. 5.1. Схема опыта для наблюдения удвоения частоты световой волны

Здесь $\mathcal{E}_0(z)$ — 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, т. е. пренебрегая потерями, положим $\sigma=0$. Включая поляризационный член в уравнения Максвелла (2.4), имеем

$$\frac{\partial^2 \mathcal{E}}{\partial z^2} - \epsilon_0 \epsilon_r \mu_0 \frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathcal{P}}{\partial t^2}. \quad (5.10)$$

Это соотношение отличается от (4.3) учетом движения связанных зарядов, т. е. учетом поляризации. Здесь $\epsilon_r = n^2(2\omega)$ — относительная диэлектрическая проницаемость на частоте 2ω . Вводя величину $\Delta n = n(2\omega) - n(\omega)$, подставляя (5.6) и (5.8) в (5.10), получаем

$$\frac{\partial^2 \mathcal{E}_0(z)}{\partial z^2} - i \frac{2\omega n(2\omega) \partial \mathcal{E}_0(z)}{c \partial z} = \mu_0 \omega^2 \mathcal{P}_0 \exp\left(i 2\omega z \frac{\Delta n}{c}\right). \quad (5.11)$$

Поскольку амплитуда $\mathcal{E}_0(z)$ слабо меняется на расстоянии длины волны, то в уравнении (5.11) можно пренебречь второй производной $\partial^2 \mathcal{E}_0(z)/\partial z^2$. Тогда решение этого уравнения, удовлетворяющее граничному условию $\mathcal{E}_0(0) = 0$, приобретет вид

$$\mathcal{E}_0(z) = \frac{\mathcal{P}_0}{2\epsilon_0 n \Delta n} \left[\exp\left(i \frac{2\omega \Delta n}{c} z\right) - 1 \right]. \quad (5.12)$$

Амплитуда вектора напряженности электрического поля определяется модулем $\mathcal{E}_0(z)$, и из (5.12) получаем

$$|\mathcal{E}_0(z)| = \frac{\mathcal{P}_0}{\epsilon_0 n \Delta n} \sin\left(\frac{\omega \Delta n}{c} z\right). \quad (5.13)$$

Интенсивность волны на удвоенной частоте

$$\begin{aligned} I(2\omega) &= \rho(2\omega) c = \epsilon_0 c |\mathcal{E}_0|^2 = \frac{\mathcal{P}_0^2 c}{\epsilon_0 n^2 (\Delta n)^2} \sin^2\left(\frac{\omega \Delta n}{c} z\right) \sim \\ &\sim I(\omega)^2 \chi^2 \frac{1}{(\Delta n)^2} \sin^2\left(\frac{\omega \Delta n}{c} z\right). \end{aligned} \quad (5.14)$$

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:

$$v_n = v(\omega) = c/n(\omega), \quad (5.15)$$

а фазовая скорость световой волны на второй гармонике определяется показателем преломления на частоте 2ω :

$$v_{\phi} = v(2\omega) = c/n(2\omega). \quad (5.18)$$

Волна второй гармоники приобретает свою энергию от основного излучения через волну поляризации (5.6). Обмен энергией между основной волной и ее второй гармоникой будет максимален, если разность фаз $\Delta\varphi$ между волной поляризации и генерируемой на частоте 2ω электромагнитной волной сохранится на достаточно больших расстояниях. Для этого необходимо, чтобы $v_n = v_{\phi}$, 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

$$\mathcal{E}_0(z) = \frac{\mu_0 c \omega \mathcal{P}_0}{n} z. \quad (5.18)$$

Тогда с учетом $\mathcal{P}_0 \sim \chi_2 \mathcal{E}_0^2(\omega)$ 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 χ_2 :

$$I(2\omega) \sim \chi_2^2 k^2(\omega) l^2 I^2(\omega). \quad (5.19)$$

В обычных случаях условие (5.17) не удается выполнить вследствие дисперсии показателя преломления света в веществе. Вне области поглощения $n(2\omega) > n(\omega)$, что соответствует так называемой нормальной дисперсии. Поэтому между волной поляризации и световой волной на частоте 2ω будет наблюдаться фазовый сдвиг, который на длине l равен

$$\Delta\varphi = l[k(2\omega) - 2k(\omega)]. \quad (5.20)$$

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* *:

$$l_{\text{ког}} = \frac{\pi}{2[k(2\omega) - 2k(\omega)]} \approx \frac{\lambda_{2\omega}}{2\Delta n}. \quad (5.21)$$

Из выражений (5.13), (5.14) видно, что при $z = l_{\text{ког}}$ интенсивность второй гармоники достигает максимума. При

* Это понятие не следует путать с аналогичным понятием, введенным в § 3.6 при определении когерентности электромагнитного излучения.

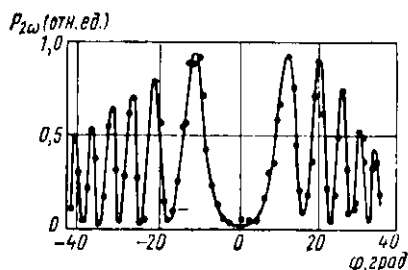


Рис. 5.2. Зависимость мощности второй гармоники от длины оптического пути s , пройденного световой волной в кристалле кварца (величина s изменялась за счет изменения угла наклона φ кварцевой пластинки относительно оси пучка лазера)

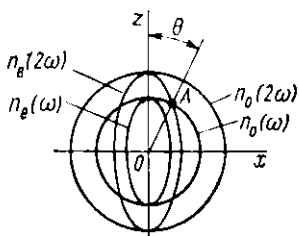


Рис. 5.3. Сечения поверхностей показателя преломления: сферы для обыкновенной волны n_0 и эллипсоида для необыкновенной волны n_e в кристалле KDP для частот рубинового лазера ω и его второй гармоники 2ω

дальнейшем увеличении $l > l_{\text{кор}}$ происходит уменьшение \mathcal{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 (5.21) 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 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. § 4.1). В одноосном кристалле для обыкновенной волны, показатель преломления n_o не зависит от направления распространения. Сечения поверхностной 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. 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 § 4.1, 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 A $n_e(2\omega) = n_o(\omega)$. 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 $\text{NH}_4\text{H}_2\text{PO}_4$ (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 волнового 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, равный 50%. Используя нелинейности более высокого порядка, можно получить излучение на 3-й, 4-й, . . . гармониках, т. е. осуществить эффективное преобразование (умножение) частоты лазерного излучения.

§ 5.3. ПАРАМЕТРИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ И ПАРАМЕТРИЧЕСКИЙ ГЕНЕРАТОР 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:

$$\mathcal{E}_n = \mathcal{E}_0^{(n)} \cos(\omega_n t - k_n z) \quad (5.24)$$

и одновременно две слабые волны

$$\mathcal{E}_1 = \mathcal{E}_0^{(1)} \cos(\omega_1 t - k_1 z), \quad (5.24a)$$

$$\mathcal{E}_2 = \mathcal{E}_0^{(2)} \cos(\omega_2 t - k_2 z) \quad (5.246)$$

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) и (5.3), получим компоненты нелинейной поляризации на частотах ω_1 и ω_2 :

$$\mathcal{P}(\omega_1) = \frac{1}{2} \epsilon_0 \chi_2 \mathcal{E}_0^{(n)} \mathcal{E}_0^{(2)} \cos[\omega_1 t - (k_n - k_2) z], \quad (5.26)$$

$$\mathcal{P}(\omega_2) = \frac{1}{2} \epsilon_0 \chi_2 \mathcal{E}_0^{(n)} \mathcal{E}_0^{(1)} \cos[\omega_2 t - (k_n - k_1) z]. \quad (5.26a)$$

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 $\mathcal{P}(\omega_1)$ со световой волной этой частоты ω_1 необходимо, чтобы сдвиг фаз между ними сохранялся с расстоянием z . На основании соотношений (5.26) 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 преобразова-

ния 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 света 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, т. е. 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 осуществлять например, поворотом кристалла, как показано на рис. 5.4. При повороте кристалла, как видно из 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 индикатрисы и направлений синхронизма. Кроме поворота кристалла возможна температурная перестройка. Можно также использовать электрооптические и фотоупругие эффекты.

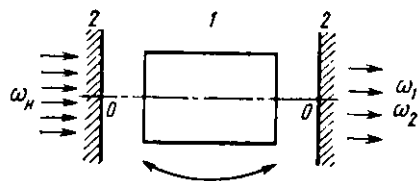


Рис. 5.4. Принцип действия параметрического генератора света:
1 — нелинейный кристалл; 2 — зеркала резонатора

Поскольку эффективное преобразование происходит лишь в мощном

монохроматическом световом потоке, накачка параметрических генераторов обычно осуществляется лазерным излучением. К. п. д. 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 инфракрасные изображения (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 [$\chi_3 \mathcal{E}^3$ (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 § 2.6 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.

Кроме рассмотренных эффектов, связанных с нелинейной поляризуемостью среды, существуют нелинейные оптические эффекты, обусловленные нелинейным поглощением, т. е. эффекты, определяемые зависимостью показателя оптического поглощения k_{ω} от интенсивности света. В частности, уменьшение показателя поглощения света при увеличении его интенсивности называется *оптическим просветлением среды*. На основе этого эффекта работают так называемые пассивные модуляторы добротности лазеров.

К нелинейным эффектам относятся также рассмотренное в § 2.6 *двухфотонное* (в общем случае *многофотонное*) *поглощение* и связанный с ним *двухфотонный (многофотонный) фотоэффект*.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. *Страховский Г. М., Успенский А. В.* Основы квантовой электроники.— М.: Высшая школа, 1979.
2. *Тарасов Л. В.* Физические основы квантовой электроники.— М.: Советское радио, 1976.
3. *Ищенко Е. Ф., Климов Ю. М.* Оптические квантовые генераторы.— М.: Советское радио, 1968.
4. *Квантовая электроника.* Маленькая энциклопедия.— М.: Советская энциклопедия, 1969.
5. *Справочник по лазерам/Под ред. А. М. Прохорова.*— М.: Советское радио, 1978, т. 1, 2.
6. *Файн В. М., Ханин Я. И.* Квантовая радиофизика, т. 1. Фотоны и нелинейные среды.— М.: Советское радио, 1972.
7. *Ярич А.* Квантовая электроника.— М.: Советское радио, 1980.
8. *Звелто О.* Физика лазеров.— М.: Мир, 1979.
9. *Мэйтланд А., Данн М.* Введение в физику лазеров.— М.: Наука, 1978.
10. *Калитеевский Н. И.* Волновая оптика.— М.: Высшая школа, 1978.
11. *Носов Ю. Р.* Оптоэлектроника.— М.: Советское радио, 1977.
12. *Мосс Т., Баррел Г., Эллис Б.* Полупроводниковая оптоэлектроника.— М.: Мир, 1976.
13. *Панков Ж.* Оптические процессы в полупроводниках.— М.: Мир, 1973.
14. *Алферов Ж. И.* Гетеропереходы в полупроводниковой электронике близкого будущего.— В кн.: Физика сегодня и завтра.— Л.: Наука, 1973.
15. *Милнс А., Фойхт Д.* Гетеропереходы и переходы металл—полупроводник.— М.: Мир, 1975.
16. *Богданкевич О. В., Дарзбек С. А., Елисеев П. Г.* Полупроводниковые лазеры.— М.: Наука, 1976.
17. *Бонч-Бруевич В. Л., Калашников С. Г.* Физика полупроводников.— М.: Наука, 1977.
18. *Зессер К.* Физика полупроводников.— М.: Мир, 1977.
19. *Шалимова К. В.* Физика полупроводников.— М.: Энергия, 1976.
20. *Свечников С. В.* Элементы оптоэлектроники.— М.: Советское радио, 1971.
21. *Рыбкин С. М.* Фотоэлектрические явления в полупроводниках.— М.: Физматгиз, 1963.
22. *Берг А., Дин П.* Светодиоды.— М.: Мир, 1979.
23. Излучательная рекомбинация в полупроводниках.— Сб. статей/Под ред. Я. Е. Покровского.— М.: Наука, 1972.
24. *Фотоника.*— Сб. статей/Под ред. М. Балкански и П. Лалемана.— М.: Мир, 1978.
25. Введение в интегральную оптику.— Сб. статей/Под ред. М. Барноски.— М.: Мир, 1977.
26. *Интегральная оптика/Под ред. Т. Тамира.*— М.: Мир, 1978.
27. *Основы волоконно-оптической связи/Под ред. М. Барноски.*— М.: Советское радио, 1980.
28. *Чернике Ф., Мидвинтер Дж.* Прикладная нелинейная оптика.— М.: Мир, 1976.

О Г Л А В Л Е Н И Е

Предисловие	3
Введение	5
§ В.1. Основные понятия и определения	5
§ В.2. Особенности оптической электроники	7
§ В.3. Историческая справка	9
Глава 1. Энергетические состояния квантовых систем	13
§ 1.1. Квантовые состояния атома водорода и квантовые числа	14
§ 1.2. Электронные состояния атомов	21
§ 1.3. Энергетические состояния молекул	30
§ 1.4. Эффект Зеемана и эффект Штарка	34
§ 1.5. Энергетические состояния кристаллов	40
Глава 2. Взаимодействие электромагнитного излучения с атомными системами	56
§ 2.1. Способы описания электромагнитного излучения	57
§ 2.2. Квантовые переходы	72
§ 2.3. Спонтанное и вынужденное излучение. Коэффициенты Эйнштейна	80
§ 2.4. Дипольное излучение	89
§ 2.5. Уширение спектральных линий	99
§ 2.6. Рассеяние света и двухфотонное поглощение	107
Глава 3. Усиление и генерация электромагнитного излучения	112
§ 3.1. Принцип работы квантовых усилителей и генераторов	112
§ 3.2. Возбуждение активного вещества (накачка)	118
§ 3.3. Оптические резонаторы	129
§ 3.4. Условие самовозбуждения и насыщение усиления	153
§ 3.5. Нестационарная генерация, модуляция добротности и синхронизация мод	158
§ 3.6. Свойства лазерного излучения	166
Глава 4. Оптические явления в твердых телах	179
§ 4.1. Оптические характеристики вещества и элементы кристаллооптики	179
§ 4.2. Электрооптические, магнитооптические и упругооптические эффекты	189
§ 4.3. Оптические переходы в полупроводниках	194
§ 4.4. Оптическое поглощение в кристаллах	201
§ 4.5. Фотоэлектрические эффекты в однородных кристаллах	227
§ 4.6. Люминесценция	238
§ 4.7. Безызлучательная рекомбинация	265
§ 4.8. Оптические эффекты в неоднородных структурах	267
Глава 5. Нелинейные оптические эффекты	289
§ 5.1. Общие понятия	290
§ 5.2. Генерация гармоник и условие фазового синхронизма	293
§ 5.3. Параметрическое преобразование и параметрический генератор света	298
§ 5.4. Прочие нелинейные эффекты	301
Рекомендуемая литература	303