

Министерство образования и науки Российской Федерации

**ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ**

В.Н. Давыдов

ФИЗИЧЕСКИЕ ОСНОВЫ ОПТОЭЛЕКТРОНИКИ

Учебное пособие

**Пособие имеет гриф СибРУМЦ
«Для межвузовского использования»**

2013

Давыдов В.Н.

Физические основы оптоэлектроники: Учебное пособие. Учебное пособие имеет гриф СибРУМЦ «Для межвузовского использования».

– Томск: ТУСУР, 2013. – 139 с.

В данном учебном пособии изложены основные физические явления в твердых телах, положенные в основу работы базовых элементов современной оптоэлектроники: фоторезисторов, фотоприемных элементов на основе фотоэда различной природы, твердотельных и полупроводниковых лазеров, светодиодов, устройств отображения информации на основе жидкокристаллических матриц. Описаны их основные характеристики, приведены типичные значения эксплуатационных параметров. Описаны также физические причины происхождения шумов в твердотельных приборах, математические методы их описания, а также строение и фундаментальные свойства жидких кристаллов. В заключение пособия указаны перспективные направления развития элементной базы оптоэлектроники, принципов построения оптоэлектронных устройств.

Пособие ориентировано на студентов технических вузов и не требует специальных знаний, кроме основ высшей математики и физики в объемах, осваиваемых на первых курсах вузов. Пособие будет полезно студентам дневной и вечерней форм обучения, испытывающим трудности в освоении курса «Физические основы оптоэлектроники», а также студентам, изучающим близкие по направлению дисциплины.

© Давыдов Валерий Николаевич, 2013

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ.....	5
2. ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ.....	8
2.1 Общие положения.....	8
2.2 Модельные представления.....	9
2.3 Зонная диаграмма и электропроводность.....	11
2.4. Квазиимпульс электрона. Долины энергии и зона Бриллюэна	17
2.5. Положительно заряженные частицы – дырки	22
2.6. Движение электронов и дырок под действием электрического поля.....	23
3. ВЗАИМОДЕЙСТВИЕ ОПТИЧЕСКОГО ИЗЛУЧЕНИЯ ТВЕРДЫМИ ТЕЛАМИ.....	32
3.1.Основные параметры процесса поглощения излучения в полупроводниках.....	32
3.2. Основные механизмы поглощения излучения.....	36
3.3. Собственное поглощение. Прямые и непрямые переходы.....	38
3.4. Примесное поглощение излучения.....	42
3.5. Поглощение свободными носителями. Внутризонное поглощение.....	43
3.6. Решеточное поглощение.....	45
4. ФОТОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ ПОЛУПРОВОДНИКОВЫХ ПРИБОРАХ.....	45
4.1 Основные понятия и параметры.....	46
4.2 Фотопроводимость полупроводников. Собственная и примесная фотопроводимость.....	49
4.3 Фотовольтаические эффекты в полупроводниках.....	58
4.4 Квазиуровни Ферми в полупроводниках.....	62
4.5 Фотоэдс в однородных полупроводниках (фотоэдс Дембера).....	65
4.6 Фотоэдс в неоднородных полупроводниках	

(объемная фотоэдс).....	67
4.7 Электронно-дырочные переходы.....	70
4.8. Барьерная фотоЭДС.....	74
4.9. Применение фотоэлектрических явлений в оптоэлектронике.....	77
5. ЭМИССИЯ ИЗЛУЧЕНИЯ ИЗ ТВЕРДЫХ ТЕЛ.....	78
5.1. Описание излучательных процессов в полупроводниках.....	78
5.2 Спектр излучения твердого тела.....	82
5.3 Спонтанное и вынужденное излучение атома.....	86
5.4 Стимулированное излучение твердых тел.....	91
5.5 Полупроводниковые лазеры	93
5.6 Светодиоды.....	97
6. ФЛУКТУАЦИОННЫЕ ПРОЦЕССЫ В ПОЛУПРОВОДНИКАХ ПОЛУПРОВОДНИКОВЫХ ПРИБОРАХ.....	98
6.1 Основные положения теории шумов.....	98
6.2 Математическое описание случайных процессов.....	104
6.3 Основные виды шумов в полупроводниковых приборах. Метод Ланжевена.....	109
6.4 Численные значения шумов различных видов.....	114
7. ЖИДКИЕ КРИСТАЛЛЫ В ОПТОЭЛЕКТРОНИКЕ.....	116
7.1 Общие сведения о жидких кристаллах.....	116
7.2 Основные свойства жидких кристаллов.....	117
7.3 Типы и структура жидких кристаллов.....	119
7.4 Ориентационные эффекты в жидких кристаллах. Переход Фредерикса.....	123
7.5 Оптические свойства жидких кристаллов.....	128
7.6 Применение жидких кристаллов в оптоэлектронике.....	134
8. ПЕРСПЕКТИВЫ РАЗВИТИЯ ОПТОЭЛЕКТРОНИКИ.....	137
9. ЛИТЕРАТУРА.....	139

1. ВВЕДЕНИЕ

При решении технических задач, в своей повседневной деятельности мы широко используем различные приборы и устройства электроники. Применение электронных явлений в различных сферах деятельности человека неслучайно: они отличаются высокой чувствительностью к различным внешним воздействиям (оптическому и корпускулярному излучению, электрическому и магнитному полям, упругим напряжениям, адсорбции-десорбции примесных молекул и т.д.), простотой и гибкостью управления, высоким быстродействием и универсальностью выполняемых функций. Фактически все задачи по функциональному преобразованию электрических, магнитных и других сигналов могут быть решены с помощью электроники.

Как и все объекты природы, электронная техника имеет свою диалектику развития, сопровождающегося накоплением количественных изменений и последующего качественного перехода с сохранением в новом состоянии элементов предыдущего своего состояния. Чтобы представить себе путь развития электроники и понять изменение требований к ней в будущем, обратимся к истории. В шестидесятые годы XX века в электронике возникла кризисная ситуация, связанная с низкой надежностью сверхсложных электронных схем (прежде всего электронно-вычислительных машин и их периферии), их большими габаритами и высоким энергопотреблением. Эта кризисная ситуация была преодолена с помощью идей и технологий интегральной микроэлектроники – в результате их развития появились микросхемы различного назначения, содержащие в себе сотни тысяч транзисторов. Однако последующее развитие и совершенствование техники, ориентированное преимущественно на решение информационных задач, создание современной вычислительной техники, показало, что многие их проблемы решаются средствами микроэлектроники неоптимальным способом. В первую очередь это касалось принципов функционирования

устройств ввода-вывода информации, увеличения памяти компьютера, отображения информации и, что особенно важно, увеличения емкости соединительных проводов. Постановка этих задач потребовала принципиально новых подходов к разработке информационных систем.

Среди новых подходов к созданию будущих информационных систем центральное место принадлежит оптоэлектронике, которая вобрала в себя наиболее перспективные направления микроэлектроники и элементы классической оптики. В оптоэлектронике используются как оптические, так и электрические явления в веществах, их взаимные связи и преобразования, а также изучаются и создаются на их основе новые приборы и информационные системы.

Совместное использование оптических и электрических явлений сулит следующие преимущества для создателей новой техники:

- высокая частота оптических колебаний позволяет создавать исключительно быстродействующие и широкополосные устройства;
- имеется возможность проведения с передаваемым сигналом чисто оптических преобразований (например, интерференции света на созданных неоднородностях кристалла), что в ряде случаев значительно упрощает схему функционального преобразования входного сигнала;
- использование оптического канала для связи блоков, например, вычислительного устройства позволяет получить высокую помехозащищенность и высокую емкость оптических соединительных линий;
- ввиду того, что параметры оптических сигналов зависят не только от времени, но и от пространственных координат, имеется возможность в устройствах оптоэлектроники работать одновременно со всем оптическим изображением, а не с отдельными его участками, как это делается в микроэлектронике.

Важно отметить, что, кроме чисто количественных преимуществ, сочетание оптических и электрических явлений дает качественно новые свойства (их называют «сверхэффектами»), более широкие возможности обработки сигналов. Эти сверхэффекты также могут быть использованы для решения технических задач. Для оптоэлектроники характерно совместное использование только таких элек-

тронных и оптических явлений, которые физически глубинным образом связаны между собой. Особенно наглядно эту связь можно обнаружить в твердых телах на примере взаимопревращения оптического излучения в электрический ток и обратно. Действительно, генерация излучения в полупроводниковом лазере или светодиоде – это электронное явление. Но хорошо известно и обратное явление – получение электрической энергии в полупроводниках при их освещении.

Список явлений, которые могут быть использованы для создания приборов оптоэлектроники, далеко не исчерпан. В настоящее время существует огромный арсенал оптоэлектронных явлений, связанных со взаимным превращением света в электричество и обратно, многие из которых до недавнего времени не имели широкого применения. К таковым можно отнести электрооптический эффект, эффект Фарадея, эффект Франца-Келдыша, оптические свойства квантовых систем: квантовых ям, квантовых точек и т.д. Их использование в будущих оптоэлектронных разработках сулит многие выгоды в создании устройств функциональных преобразований.

Объединяя сказанное выше, можно заключить: оптоэлектроника представляет собой направление науки и техники, характеризующееся следующими тремя основными положениями:

- во-первых, это использование явлений, для которых характерна неразрывность и принципиальное сочетание электрических и оптических явлений. При этом определяющим является использование оптических сигналов – именно этим достигается то качественно новое, что отличает оптоэлектронику от микроэлектроники. Здесь микроэлектроника выполняет вспомогательные, обслуживающие функции, или, другими словами, оптоэлектроника представляет собой оптику, управляемую электроникой;
- во-вторых, оптоэлектроника нацелена на решение задач информатики, а именно: генерирование информационных сигналов, их обработку и хранение, передачу и отображение;
- в третьих, оптоэлектроника целиком базируется на конструктивно-технологической базе современной микроэлектроники.

Итак, **оптоэлектроника** – это научно-техническое направление, связанное с явлениями, в которых неразрывны оптические и электронные процессы, ориентированное на прием, передачу, обработку, хранение, отображение информации и реализуемое на основе технологической базы микроэлектроники.

Целью данного курса является изучение основных физических явлений и закономерностей, которые используются или могут быть использованы для решения задач оптоэлектроники.

2. ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ

2.1. Общие положения

Ввиду того, что подавляющее число приборов и устройств оптической электроники используют свойства полупроводников, начнём изучение данного курса с рассмотрения фундаментальных свойств твердых тел. Нас интересуют преимущественно их электрические и оптические свойства.

Если рассматривать какое-либо электрическое или оптическое явление в твердом теле, то для понимания физических процессов, которые формируют данное явление, необходимо знать поведение в заданных условиях элементарной частицы: электрона - именно он определяет свойства твердого тела при действии на него различных физических полей. Исторически первой средой, в которой были изучены электронные явления, был газ. Последующее развитие науки и техники привело к необходимости изучения электронных процессов в твердых телах. Однако уже первые эксперименты с электропроводностью твердых тел показали, что она не может быть объяснена развитой к тому времени классической электронной теорией, хорошо объясняющей поведение электронов в вакууме и газах. Было непонятно: как, имея огромное количество электронов (находящихся у атомов), твердые тела различного физико-химического состава имеют различающиеся на многие порядки электропроводности; почему их электропроводности по-разному зависят от температуры, освещения и т.д.?

На эти и другие вопросы классическая электронная теория ответить не могла, и потому были предприняты попытки построения новой теории - теории

твёрдого тела.

2.2. Модельные представления

Модель Зоммерфельда. Исторически первой моделью, объясняющей электрические свойства твёрдых тел, была модель, разработанная немецким физиком Зоммерфельдом. В ее основе лежали ряд упрощающих допущений о физическом строении кристалла и механизмах взаимодействия его частиц с между собой и решеткой.

1. Поскольку атом является квантовой системой, то строгий подход к объяснению электропроводности твёрдых тел должен строиться на основе квантовых представлений об их строении. На момент создания теории Зоммерфельда в физике была хорошо разработана спектроскопическая модель атома, описывающая его состояния с помощью энергетического спектра атома в виде дискретного набора энергетических уровней. Однако в то время механизмы взаимодействия атомов в твёрдых телах не были известны, и Зоммерфельду ничего не оставалось, как считать атомы твёрдого тела невзаимодействующими между собой. Следовательно, каждый уровень энергии спектра изолированного атома в кристалле оказывался вырожден N раз, где N - число атомов в кристалле (N имеет порядок 10^{23}).

2. Из этой модели следовало, что для того, чтобы оторвать электрон от атома и перевести его в межатомное пространство, необходимо атому сообщить энергию, равную энергии ионизации изолированного атома E_{ion} . Обычное его значение составляет большую величину, достигающую несколько электронов-вольт. Этот электрон в межатомном пространстве Зоммерфельд считал уже классической частицей, способной менять свою энергию на сколь угодно малую величину как это происходит у частицы в классической (ньютоновской) механике. Предполагалось также, что свободные электроны не принадлежат отдельным атомам, а могут свободно перемещаться по кристаллу, не испытывая влияния кулоновского поля его ионов. Это допущение было основано на том умозаключении Зоммерфельда, что суммарное поле ионов внутри кри-

стала точно компенсироваться суммарным полем свободных электронов (их количества должны быть одинаковыми). Таким образом, заряд каждого иона должен быть компенсирован находящимся рядом зарядом свободного электрона. Поэтому движущийся через кристалл свободный электрон не испытывал электрического взаимодействия ни с ионами, ни с другими электронами. Потенциальный барьер на поверхности кристалла удерживает электроны в нем, затрудняя их выход в вакуум. Зависимость энергии E свобод-

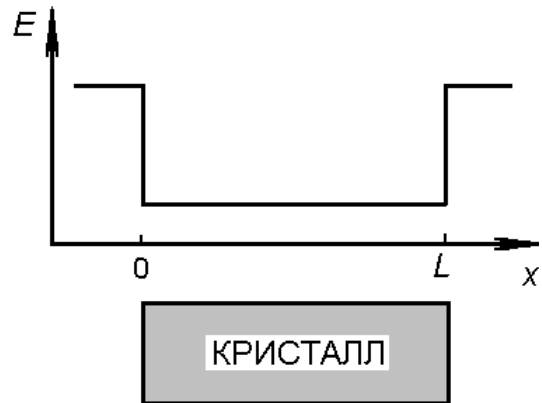


Рис.1.

ного электрона в кристалле от координаты x по модели Зоммерфельда показана на рис.1. Здесь потенциальный барьер на поверхности равен работе выхода электрона из кристалла в вакуум.

Модель электропроводности твердых тел по Зоммерфельду оказалась мало полезной для практики. Важным вопросом, на который она не давала ответ, являлся вопрос о причинах различия электропроводности различных твердых тел разного физико-химического состава. В этой связи Зоммерфельд считал, что в металлах валентные электроны легко отщепляются от атома, становясь свободными, и участвуют в электропроводности (E_{ion} - мала), тогда как у диэлектриков это сделать крайне трудно (E_{ion} - велика). Однако это объяснение не согласуется с тем фактом, что численные значения энергий ионизации изолированных атомов E_{ion} металлов и диэлектриков слабо различаются между собой, а иногда у металлов могут оказаться даже больше, чем у диэлектриков.

Модель Блоха. Следующий шаг в уточнении модельных представлений о поведении электрона в кристалле сделан другим немецким физиком Блохом, который при решении квантово-механической задачи об энергетических уровнях электронов в объеме твердого тела учёл взаимодействие атомов между собой за счет перекрытия электронных оболочек соседних атомов. Он также учел, что свободный электрон, двигаясь в межатомном пространстве, испытывает ку-

лоновское взаимодействие с электрическим полем ионов, находящихся в узлах кристаллической решетки.

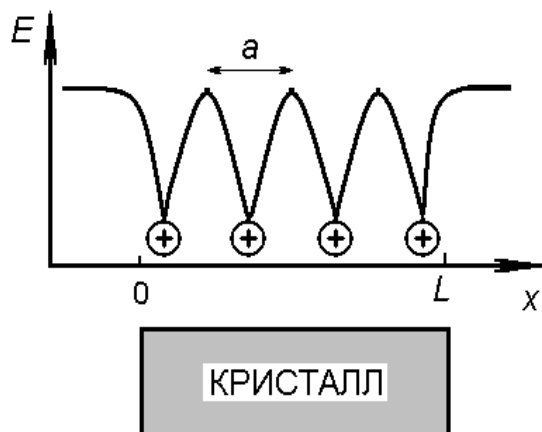


Рис.2.

Характер изменения энергии свободного электрона, движущегося вдоль ионных узлов решётки, имеет вид, показанный на рис. 2. Как можно видеть из рисунка, изменение энергии свободного электрона носит периодический характер с периодом, равным решётки a . Исследуя поведение свободного электрона в таком поле с по-

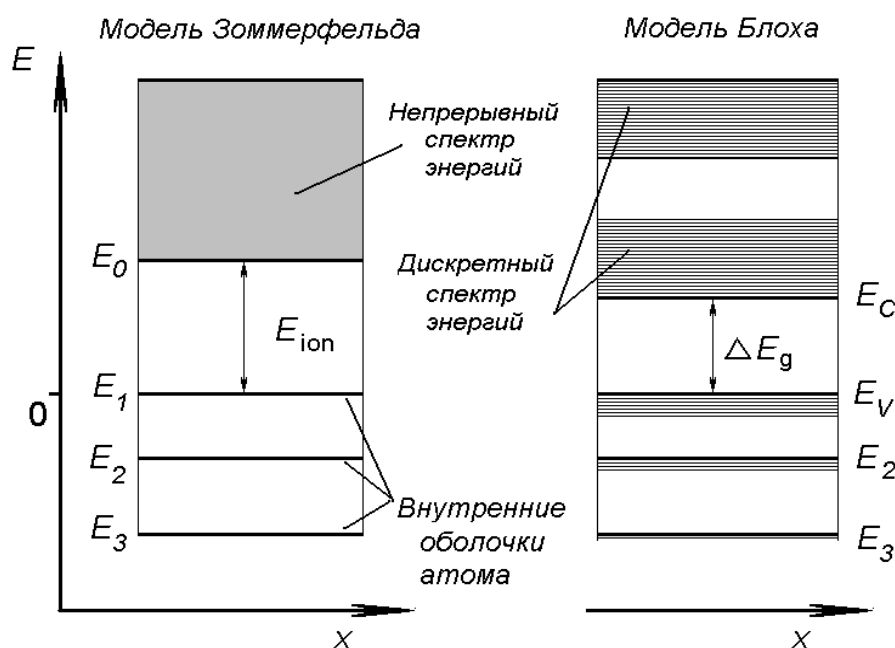
мощью аппарата квантовой механики, Блох показал, что в идеальной периодической решётке электрон движется свободно, как в вакууме, не испытывая столкновений с ионами кристалла - он их не “видит”! Но самое главное, оказалось, что свободный электрон в кристалле не может иметь произвольную энергию: весь диапазон ее значений оказался разделен на чередующиеся полосы разрешенных и запрещенных энергий. Их назвали **зонами** энергий. Более того, ширина зон запрещенных и разрешенных энергий определяется не только природой атомов решетки, но и степенью перекрытия электронных оболочек соседних атомов. В итоге ширина запрещенной зоны, определявшая энергию образования свободного электрона в кристалле оказалась отличной от энергии ионизации изолированного атома E_{ion} , и это также отличает модель Блоха от модели Зоммерфельда, где ширина запрещенной зоны отождествлена с E_{ion} .

2.3. Зонная диаграмма и электропроводность

Для объяснения электронных процессов в твердых телах удобно пользоваться так называемой **зонной диаграммой** твердого тела (полупроводника, диэлектрика). Она представляет собой зависимость энергии электрона от координаты в твердом теле. Чтобы нарисовать его правильную зонную диаграмму твердого тела, начнем с простейшей модели, описывающей поведение электро-

на в твердом теле, модели Зоммерфельда. Согласно ей, разрешённые уровни валентных электронов (т.е. электронов, находящихся на внешних орбитах атомов, именно они в первую очередь взаимодействуют с внешними полями, так как являются внешними электронами, как и человек, общается с окружающей средой через свою внешнюю оболочку - кожу) в кристалле расположены близко друг от друга и образуют систему уровней, простирающихся от дна потенциальной ямы до больших значений энергии. Эту систему уровней можно назвать **“валентной зоной”**, поскольку она описывает энергии электронов на валентных орбитах атома. Уровни энергий электронов с внутренних оболочек остаются невозмущёнными (напомню, что эти электроны не взаимодействуют ни друг с другом, ни с ионами решётки) и совпадают с уровнями энергии изолированного атома. Эти уровни одни и те же для каждого атома кристалла, и потому они оказываются N - кратно вырождены, где N - число атомов в кристалле, т.е. одному значению энергии соответствует N физически различных состояний. Разрешённые уровни энергии электронов в кристалле по модели Зоммерфельда показаны в левой части рис. 3. Здесь за нулевую энергию выбрана энергия валентного электрона.

В теории Блоха вследствие введения периодического потенциала решётки энергетические уровни группируются в определённые полосы, называемые



“зонами разрешённых энергий”, разделённые областями, в которых нет разрешённых значений энергий, “запрещённые зоны”. Для внутренних электронов разрешённые зоны чрезвычайно узки и соответствуют атомным уровням энергии. Зоны энергий для внешних (валентных) электронов оказываются широкими. Их расположение показано в правой части рис.3. В модели Блоха зоны состоят из множества близко расположенных уровней, так что во многих практических приложениях их можно рассматривать как непрерывный спектр.

К аналогичной картине энергетических зон в твердом теле можно прийти и другим путем. Для этого рассмотрим два одинаковых атома, расположенных на большом удалении друг от друга, и предположим, что каждый из них обладает одним невырожденным уровнем энергии (это равносильно тому, что из спектра энергий изолированного атома выбирается только один уровень). В этом случае спектр энергии системы из двух атомов содержит один уровень энергии, вырожденный дважды. Теперь будем сближать эти атомы. В результате усиливающегося взаимодействия между атомами вырожденный уровень энергии расщепится на два уровня, энергетический зазор между которыми будет увеличиваться с уменьшением расстояния (рис.4). Этот результат известен в классической теории из задачи по взаимодействию двух осцилляторов: при сближении двух одинаковых колебательных контуров образующаяся колебательная система будет иметь две резонансные частоты, отличающиеся от собственной частоты изолированных контуров на некоторую величину в сторону больших и меньших значений.

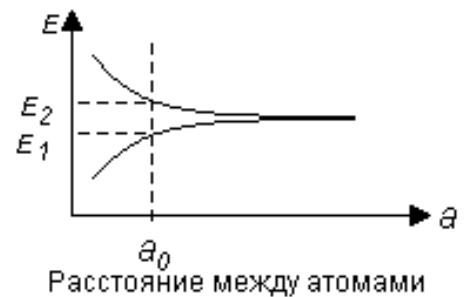


Рис. 4

Если расстояние между соседними атомами равно a_0 , то верхний уровень энергии валентной зоны будет иметь энергию, равную E_1 , а нижнее значение энергии зоны проводимости будет E_2 .

Данный подход может быть применен к системе из N атомов. Предположим, что рассматриваемый кристалл растяну настолько, что все межатомные расстояния в нем велики. Тогда разрешенные уровни энергии электронов в та-



Рис. 5

ком кристалле совпадут со спектром отдельного атома, но каждый уровень при этом N - кратно вырожден. Если уменьшать расстояние между атомами, то из-за межатомного взаимодействия (из-за перекрытия электронных оболочек сближаемых атомов) каждый уровень расщепится на серию из N одиночных, невырожденных уровней. В результате

вместо вырожденного уровня получим зону разрешённых энергий для электрона из N плотно расположенных уровней (рис.5). Для внутренних электронов из-за их экранировки внешними электронами возмущение оказывается слабым по сравнению с взаимодействием их с ядром и потому расщепление этих уровней будет малым. Для валентных электронов оно может быть вплоть до перекрытия зон (сравните зазоры между E_1 и E_2 на рис.5 для случаев: $a = a_1$ и $a = a_2$).

Энергетическая диаграмма состояний электрона в твердом теле показывает возможности размещения электронов (состояния подобны креслам в зрительном зале). Но будут ли они заняты и как – это другой вопрос.

Проанализируем случай, когда валентная зона и зона проводимости не перекрываются. Чтобы ответить на вопрос о распределении электронов по состояниям в зонах, необходимо воспользоваться принципом Паули, согласно которому на каждом уровне в зоне может находиться не более двух электронов с противоположно ориентированными спинами. Начнём заполнять уровни атома электронами с глубоко лежащих уровней. В любом атоме каждый такой уровень должен быть занят двумя электронами, так что в кристалле полностью заполненным оболочкам атомов можно поставить в соответствие полностью заполненные зоны. Частично

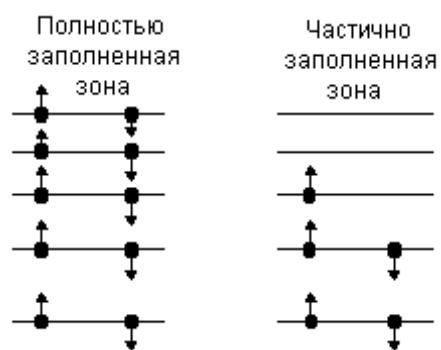


Рис.6.

заполнены зоны, соответствующие внешним (валентным) электронам. Различие между полностью и частично занятыми зонами иллюстрирует рис.6. Здесь

пустые уровни соответствуют возбужденным состояниям валентных электронов; спин электрона показан стрелкой.

С помощью этих модельных представлений можно понять, почему электроны в низших зонах, соответствующих атомным остовам, не могут принять участия в электропроводности. Для участия в электропроводности электрону необходимо начать двигаться, т.е. он должен приобрести дополнительную энергию за счёт действия приложенного электрического поля. На языке квантовой механики это означает, что электрон должен перейти на более высокий энергетический уровень. Если же все уровни в данной зоне заняты, то электрон не может ускориться в электрическом поле до тех пор, пока в результате возбуждения не перейдёт в лежащую выше зону, в которой имеются свободные уровни. Процесс возбуждения, однако, зависит от величины энергии, требуемой для возбуждения электрона и в случае ее большого значения (доли и единицы эВ) маловероятен.

Поэтому, чтобы электроны приняли участие в электропроводности (а значит, обнаружили себя в электрических измерениях), необходимо наличие близлежащих пустых энергетических уровней, на которые рассматриваемые электроны могли бы перейти при действии приложенного электрического поля. Именно это обстоятельство составляет принципиальное различие в величине электропроводности металлов, полупроводников и диэлектриков.

Рассмотрим теперь наивысшую валентную зону, в которой имеются как занятые электронами, так и свободные уровни. В щелочном металле, например, каждый атом содержит по одному валентному электрону. В невозбуждённом состоянии N электронов займут $N/2$ наиболее глубоких уровней в этой валентной зоне. Зона окажется заполненной лишь наполовину и в ней останется ещё много свободных уровней энергии, необходимых для возникновения электропроводности. Таким образом, при наличии лишь одного валентного электрона у атома ситуация напоминает модель твёрдого тела по теории Зоммерфельда.

Обратимся к кристаллу, построенному из атомов, содержащих на внешней электронной оболочке два электрона с противоположно ориентированными

спинами. В этом случае валентная зона кристалла, т.е. наивысшая из всех содержащих электроны зон, оказывается полностью занятой. Если к тому же между валентной зоной и следующей за ней верхней пустой зоной имеется энергетический зазор - запрещённая зона, то такой кристалл не будет обладать электропроводностью и соответствующее вещество будет диэлектриком, если запрещённая зона широкая. Однако если она невелика, то существует возможность теплового возбуждения электронов, приводящего к их забросу из валентной зоны в пустую зону проводимости, после чего эти электроны могут принимать участие в электропроводности. Число возбуждённых электронов будет увеличиваться с ростом температуры, что приведёт к росту и электропроводности. Так ведут себя кристаллические полупроводники и диэлектрики, различие между которыми заключается только в величине запрещённой зоны. Обычно полупроводник - это кристалл, у которого в химически чистом состоянии величина ΔE_g является положительной и не превышает 2-3 эВ. У диэлектриков ширина запрещённой зоны выше 3 эВ.

Сказанное иллюстрируется рис. 7, на котором показано энергетическое состояние элементных полупроводников (типа кремний и германий) при абсолютном нуле температуры. Сообщив электрону энергию, равную ширине запрещённой зоны, можно перевести его из валентной зоны в зону проводимости и тем самым заставить принять участие в электропроводности. В зоне проводимости много пустых близкорасположенных уровней. Поэтому любое сколь малое электрическое поле даст прирост энергии электрону зоны проводимости и в этой зоне обязательно найдет пустой уровень энергии, соответствующий величине поля.

Однако можно добиться высокой электропроводности полупроводника и без столь значительных затрат энергии: введение малого количества специальных примесей может заметно повлиять на число свободных электронов в зоне проводимости и дырок в валентной зоне, так что материалы даже с $\Delta E_g < 3$ эВ ведут себя как высокопроводящие полупроводники. Эти примеси называются **донорными** (элементы V группы таблицы Менделеева) или **акцепторными**

(элементы III группы) в зависимости от того, отдаёт или забирает атом примеси электрон у атома полупроводника, соответственно.

Пользуясь энергетической диаграммой, можно объяснить электропроводность диэлектриков, полупроводников и металлов. Так, у диэлектриков энергетическая диаграмма аналогична приведенной на рис. 7, но у них ширина

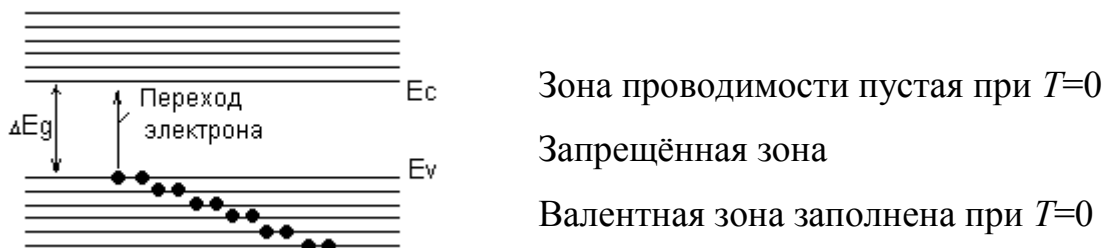


Рис. 7

запрещенной зоны велика: $\Delta E_g > 2 \div 3$ эВ. У полупроводников ширина запрещенной зоны лежит в диапазоне $0,1 < \Delta E_g < 2$ эВ. У металлов $\Delta E_g = 0$ или трицательное, т.е. имеет место перекрытие валентной зоны и зоны проводимости.

2.4. Квазиимпульс электрона. Долины энергии и зона Бриллюэна

Напомним, что наше задачей является объяснение токопротекания в кристаллических телах. Но движение свободных электронов в твердых телах отличается от их движения в вакууме наличием большого числа столкновений с дефектами решетки кристалла. Как это учесть?

Из классической механики известно, что для корректного описания результата столкновения двух тел необходимо введение нового физического параметра – импульса частицы. Его определяют как произведение массы частицы на ее скорость: $p = mv$. Для характеристики движения со столкновениями квантовой частицы (электрона в зоне проводимости), которая является одновременно частицей и волной и потому относится к квазичастицей, также вводят понятие импульса. Однако, учитывая двойственную природу квантовой частицы, его называют **квазиимпульсом**, т.е. «почти импульсом».

Двойственная природа квантовой частицы приводит к двум определениям ее квазиимпульса p . Квазиимпульс электрона-частицы вычисляется через ее кинетическую энергию E :

$$E = \frac{m^* v^2}{2} = \frac{p^2}{2m^*},$$

где m^* – эффективная масса электрона, которая отличается от массы электрона в вакууме, так как инерционность одной и той же частицы в классическом и квантовом проявлениях может сильно различаться. В квантовой механике определение квазиимпульса электрона-волны таково:

$$\bar{p} = \bar{k} \cdot \hbar, \quad (2.1)$$

где \hbar – постоянная Дирака, равная $h/2\pi$; h – постоянная Планка; \bar{k} – волновой вектор, модуль которого определяется длиной волны: $k = 2\pi/\lambda$, λ – длина волны де Бройля. Если рассматривать кристалл в виде прямоугольной потенциальной ямы длиной L и с барьером на ее концах бесконечной высоты (см. ри.1), то стационарное состояние частицы-волны в кристалле будет описываться в виде пакета стоячих волн. Это возможно, если по длине кристалла укладывается целое число полуволен, входящих в волновой пакет частицы-волны: $L = n\lambda/2$ или $\lambda = 2L/n$, где число $n = \pm 1, \pm 2, \dots, \pm N$ указывает, сколько раз на длине кристалла $L = N \cdot a$ укладывается полуволен. Подставив это условие в выражение для волнового вектора, получим, что его модуль может принимать только дискретные значения:

$$k = n(\pi / L), \quad n = \sqrt{n_x^2 + n_y^2 + n_z^2}. \quad (2.2)$$

Здесь n_x, n_y, n_z – целые числа, показывающие количество полуволен, укладываемых по размерам кристалла L_x, L_y, L_z вдоль осей x, y, z (положительные значения n_x, n_y, n_z) и осей $-x, -y, -z$ (отрицательные значения n_x, n_y, n_z) соответственно. Все возможные значения и направления вектора \bar{k} в трехмерном

пространстве (k_x, k_y, k_z) заполняют некоторый объем, образуя симметричную трехмерную фигуру, которую называют **зоной Бриллюэна**.

Существование максимального значения $n_0 = N$ связано с тем, что на самой короткой длине волны λ_0 , описывающей рассматриваемую квантовую частицу, должно укладываться не менее двух атомов – один в максимуме, а другой в минимуме волны:

$$L = n_0 \lambda_0 / 2 = Na, \quad \lambda_0 = 2a.$$

При меньшем числе атомов гармоническое колебание не будет распознано как волна. В этом случае $k = \pi/a$ есть максимальное значение волнового вектора. Этому максимальному значению соответствует край зоны Бриллюэна. В случае изотропной среды, параметры которой не зависят от направления их измерения, эта зона представляет собой шар радиусом $k = \pi/a$. В многокомпонентных кристаллах расстояние между соседними атомами a может зависеть от направления, и потому в них зона Бриллюэна имеет более сложный вид.



Рис. 8

Используя волновое определение квазиимпульса, кинетическую энергию электрона можно представить в виде

$$E = \frac{\hbar^2 k^2}{2m^*},$$

где k принимает дискретные значения, определяемые по (2.2). Тогда в кристалле, имеющем форму куба с ребром длиной L и рассматриваемом как потенциальная яма, разрешенные значения энергии частицы-электрона будут равны:

$$E = \left(\frac{\hbar^2}{8m^*L^2} \right) \left(n_x^2 + n_y^2 + n_z^2 \right). \quad (2.3)$$

Поскольку квантовые числа n_x, n_y, n_z могут принимать только целые значения, то и величина E тоже может принимать только дискретные значения. Это означает, что спектр энергий свободного электрона как квантовой частицы в кристалле,

строга говоря, дискретен. Однако в реальных кристаллах получающиеся энергетические расстояния между соседними уровнями энергии столь малы ($\sim 10^{-18}$ эВ для кристалла с $L = 1$ см), что спектр разрешенных значений энергий свободного квантового электрона в кристалле с высокой степенью точности можно считать непрерывным, что позволяет рассматривать его как классическую частицу.

Рассмотрим зависимость энергии от волнового вектора для одного направления в пространстве волновых векторов. Обычно в качестве начала отсчета энергии принимают энергию потолка валентной зоны. Тогда дну зоны проводимости отвечает более высокая потенциальная энергия, соответствующая ширине запрещенной зоны (рис. 8). Как это следует из выражения (2.2), зависимость E от k в пределах зоны разрешенных энергий является параболической. На рис. 8 показана зависимость E от k реальных полупроводников, у которых эффективная масса электронов в валентной зоне больше эффективной массы электрона в зоне проводимости. Следовательно, крутизна зависимости E от k валентной зоны оказывается меньше, чем зоны проводимости. Такое распределение состояний называется **параболической долиной**. Смысл этого «образного» названия становится более ясным при трехмерном представлении зависимости E от k_x и k_y – получающаяся при этом пространственная фигура $E(\vec{k})$ напоминает собой долину среди гор.

Отрицательная кривизна валентной зоны на рис. 8 означает, что если бы электроны, находящиеся в ней, могли двигаться, то они приобретали бы ускорение в направлении, противоположном направлению действующей силы. Если действующей силой является кулоновская, то отрицательная кривизна валентной зоны означает положительный заряд подвижной частицы, движущейся в этой зоне. Вскоре мы рассмотрим эти частицы и убедимся, что они действительно заряжены положительно.

Показанное на рис. 8 взаимное положение экстремумов зон разрешенных энергий электрона в кристалле не является единственно возможным. Из-за

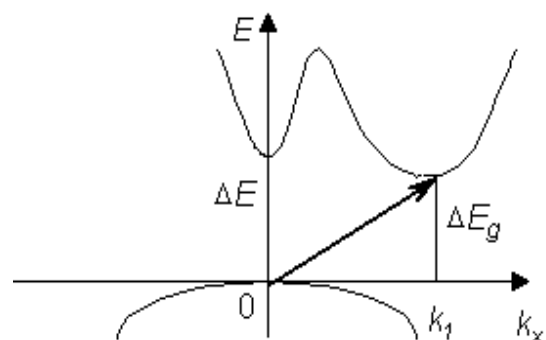


Рис. 9

сильного взаимодействия с ближайшими соседями, со следующими за ними атомами и с атомами, более удаленными по отношению к рассматриваемому атому решетки полупроводника, минимум долины может быть смещен относительно точки $k = 0$ в некотором кристаллографическом направлении, например, k_x (рис. 9). Полупроводники, у которых минимум зоны проводимости имеет место при том же значении квазиимпульса, что и максимум валентной зоны, называют **прямозонными**. При их несовпадении полупроводники называют **непрямозонными**. В природе встречаются оба типа полупроводников.

Наличие разрешенных энергетических уровней не является достаточным условием возникновения электропроводности кристалла. Для того чтобы кристалл проводил электрический ток, наряду с зонами разрешенных энергий необходимо наличие в них подвижных (свободных) носителей заряда.

Рассмотрим заполнение электронных состояний в зонах разрешенных энергий носителями заряда. Концентрация электронов на уровне с данной энергией E определяется плотностью состояний (их количеством в единичном энергетическом интервале) в зоне и функцией их заполнения. Она именуется **распределением Ферми-Дирака** и записывается в виде

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}$$

Вид этой функции показан на рис. 10, а. Она показывает вероятность заполнения электронами состояния с энергией E . Здесь k – постоянная Больцмана; T – абсолютная температура; E_F – уровень Ферми, т.е. уровень энергии, вероятность за-
пол-
нения

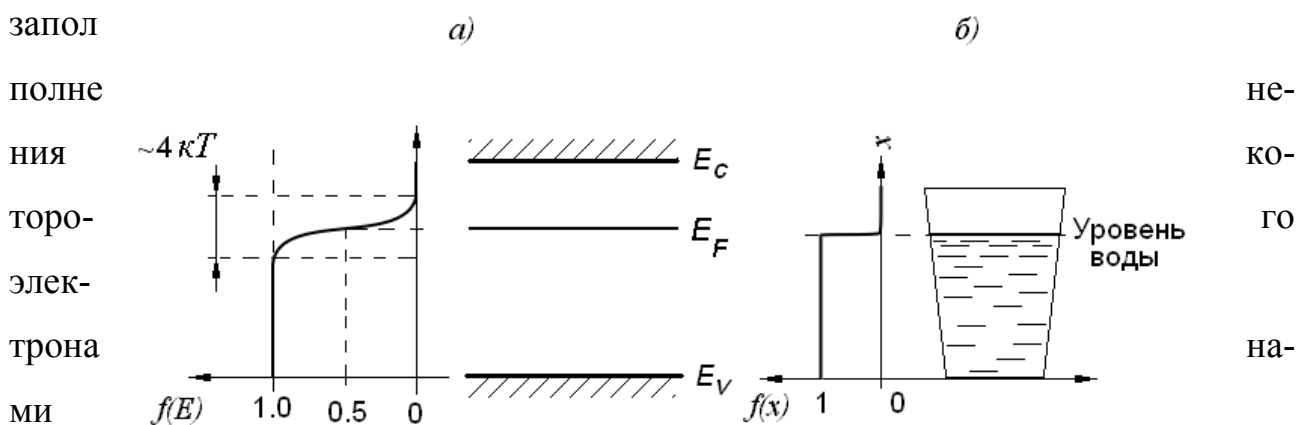


Рис. 10

равна $1/2$. Если рассматриваемый уровень энергии на несколько единиц kT лежит выше уровня Ферми, то вероятность заполнения этого уровня электронами равна нулю. Если же уровень энергии находится ниже уровня Ферми на несколько единиц kT , то вероятность его заполнения электронами равна единице.

Для того чтобы лучше уяснить смысл распределения Ферми-Дирака, обратимся к аналогии. Если имеем стакан с водой, заполненный лишь частично, то функция заполнения стакана водой $f(x)$ (где x - координата вдоль стакана, отсчитываемая от его дна) имеет значение $f(x) = 1$ при значениях координаты x , соответствующих заполненной части стакана. Если же значения x находятся выше уровня воды, то $f(x) = 0$ (рис.10, б). Такое распределение воды по высоте стакана аналогично заполнению электронами уровней энергии в полупроводнике. Если уровень энергии лежит ниже уровня Ферми, то он заполнен электронами (в рассмотренном примере - водой). Если же уровень энергии лежит выше, то он свободен от электронов. Уровень Ферми разделяет эти области заполнения, в приведенном примере в качестве такой разделительной линии – уровня Ферми выступает уровень воды в стакане.

Согласно квантово-механическому принципу Паули, на каждом энергетическом уровне может находиться не более двух электронов. Однако данному энергетическому уровню может соответствовать более чем одно состояние (говорят, что уровень вырожден). Тогда при определении количества электронов на данном уровне функцию Ферми-Дирака умножают на фактор вырождения – число, показывающее, сколько состояний имеют данное значение энергии.

2.5. Положительно заряженные квазичастицы – дырки

Всякая система стремится занять состояние с минимальной энергией – это закон равновесия действует в любой физической системе. В беспримесном полупроводнике при низкой температуре и в отсутствии электрического поля, когда все электроны находятся в наинизших энергетических состояниях, в зоне проводимости нет электронов. Однако такое положение теоретически возможно лишь при абсолютном нуле температуры. При реальных температурах, соот-

ветствующих применению полупроводников (например, при комнатной температуре), в зоне проводимости всегда находится некоторое количество электронов, заброшенных туда из валентной зоны в результате передачи им тепловой энергии решетки. Это происходит следующим образом.

Все атомы решетки находятся в хаотическом движении в пределах, определенных силами межатомных связей. Из-за хаотичности и возможных столкновений атомов в какой-то момент отдельный атом может получить энергию теплого колебания большую, чем окружающие. Реализовать новое энергетическое состояние атом может, передав энергию электронам. Но внутренние электроны принять эту энергию не могут – все уровни энергии около их заняты, а валентные электроны могут. Забирая избыток тепловой энергии, валентные электроны переходят на более высокий энергетический уровень. Таковыми является уровни зоны проводимости.

В результате появления свободных электронов в зоне проводимости в полупроводнике начинает течь электрический ток. Мгновенная плотность тока \vec{j} , обусловленная движением какого-либо электрона, пропорциональна его скорости \vec{v} и совпадает с ним по направлению. Величину \vec{j} можно вычислить следующим образом.

Рассмотрим электрон в кристалле единичного объема, и пусть этот электрон движется со скоростью \vec{v} . Поскольку скорость движения электрона определяется его квазиимпульсом, то будем считать, что квазиимпульс k рассматриваемого электрона равен p . Тогда движение этого электрона будет создавать ток с плотностью

$$\vec{j} = -e \cdot \vec{v}_p .$$

Плотность электрического тока, создаваемого многими движущимися электронами, равна сумме плотностей токов, создаваемых отдельными движущимися электронами с разными значениями квазиимпульса:

$$\vec{j} = -e \cdot \sum_S \vec{v}_S ,$$

где вектор скорости \bar{v}_S соответствует состояниям зоны Бриллюэна, которые заняты движущимися электронами. Из данного выражения следует, что если электроны полностью заполняют валентную зону, то соответствующий ток $\bar{j} = 0$, так как каждому значению \bar{k}_S в зоне обязательно найдется равный по величине и обратный по направлению волновой вектор « $-\bar{k}_S$ », порождающий равный по величине и обратный по направлению ток. Это подтверждается симметричностью зоны Бриллюэна относительно положительных и отрицательных значений квазиимпульсов.

Предположим теперь, что вся зона полностью заполнена электронами, за исключением единственного состояния, характеризуемого волновым вектором \bar{k}_i , которому соответствует скорость \bar{v}_i . Суммарный ток всех электронов в зоне записывается в виде

$$\bar{j} = -e \cdot \sum_{S \neq i} \bar{v}_S = -e \cdot \sum_S \bar{v}_S + e \cdot \bar{v}_i .$$

В данном выражении первый член равен нулю, так как в нем суммирование ведется по всем состояниям зоны. Поэтому полный ток равен

$$\bar{j} = e \cdot \bar{v}_i . \quad (2.4)$$

Из данного выражения следуют три вывода: во-первых, если один электрон убрать из валентной зоны, переместив в зону проводимости, то суммарный ток всех электронов в валентной зоне будет эквивалентен току одной свободной частицы в валентной зоне; во-вторых, эта свободная частица имеет положительный заряд $+e$; в-третьих, эта частица на физическом уровне представляет собой вакантное место, откуда ушел электрон, которое способно перемещаться в валентной зоне (в выражении (2.4) фигурирует ее скорость). Поскольку эта частица представляет собой вакансию электрона на валентной оболочке атома вещества полупроводника, то ее называют **дыркой**. Ей можно приписать эффективную массу, равную эффективной массе электрона, который занимал это валентное состояние.

Ясно, что при наличии некоторого количества дырок плотность создаваемого ими тока равна

$$\bar{j} = e \sum_i \bar{v}_i .$$

При наличии и электронов, и дырок в полупроводнике мгновенное значение плотности тока определяется так:

$$\bar{j} = -e \left(\sum_m \bar{v}_m - \sum_i \bar{v}_i \right),$$

где сумма по m относится к электронам, а сумма по i – к дыркам.

Для описания процессов рождения и уничтожения электронов и дырок удобно пользоваться той же энергетической диаграммой,

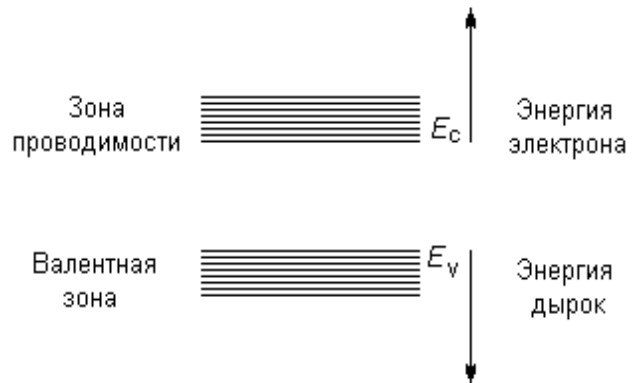


Рис.11

которую рисовали для электронов, с той лишь разницей, что для дырок их энергию следует считать возрастающей вниз (рис. 11). Действительно, из физических соображений ясно, что создание вакансии на внутренней оболочке в атоме потребует значительно больше энергии, чем вакансии на внешней (валентной) орбите. Поэтому чем глубже находится вакансия (на энергетической диаграмме ниже), тем ее энергия больше.

Рассмотрим в качестве примера вид энергетических зон, показанный на рис. 9. Пусть E_0 обозначает энергию электронов в полностью заполненной зоне, в которой зависимость E от \bar{k} определяется соотношением $E = E_v(k)$, за начало отсчета взята вершина заполненной зоны. При малых \bar{k} функцию $E_v(k)$ можно представить в виде:

$$E_v(\bar{k}) = -\frac{\hbar^2}{2} \cdot \left(\frac{k_x^2}{m_{px}^*} + \frac{k_y^2}{m_{py}^*} + \frac{k_z^2}{m_{pz}^*} \right).$$

Зависимость E от \bar{k} в зоне проводимости, можно представить в виде: $E = \Delta E_g + E_c(\bar{k})$. У дна зоны проводимости имеем:

$$E_c(\bar{k}) = \frac{\hbar^2}{2} \cdot \left(\frac{k_x^2}{m_{nx}^*} + \frac{k_y^2}{m_{ny}^*} + \frac{k_z^2}{m_{nz}^*} \right).$$

Здесь m_{nx}, m_{ny}, m_{nz} – эффективные массы электрона, характеризующие его инерциальные свойства при движении по осям x, y, z .

Предположим теперь, что один электрон изъят из валентной зоны, где он имел волновой вектор \bar{k} , и переведен в зону проводимости с тем же вектором \bar{k} . При этом будет создана электронно-дырочная пара, и энергия всей системы станет равной:

$$E = E_0 + \Delta E_g + E_c(\bar{k}) - E_v(\bar{k}), \quad (2.5)$$

где E_0 – энергия системы до образования пары. В этом выражении третье слагаемое правой части представляет собой энергию электрона в зоне проводимости, а четвертое – энергию дырки в валентной зоне.

Таким образом, согласно (2.5), можно объединить в одном выражении энергии электронов и дырок, что при переходе к графическому представлению энергетических преобразований в полупроводнике означает: можно использовать одну и ту же схему энергетических уровней для электронов и для дырок, если энергию дырки отсчитывать в направлении, обратном направлению отчета энергии электрона.

2.6. Движение электронов и дырок под действием электрического поля

Напомним, что основной задачей, решаемой в данной главе, является объяснение механизма электропроводности кристаллов. В предыдущих параграфах выяснено, что частицами, способными создать электрический ток в кристаллах, являются электроны и дырки, находящиеся в зонах разрешенных энергий. Но как заставить их двигаться по своим зонам и тем самым создать элек-

трический ток, из энергетической диаграммы полупроводника - неясно. Для решения этого вопроса рассмотрим, как влияет наличие внешнего электрического поля на величину средней энергии электрона, находящегося в зоне разрешенных значений энергии.

Работа, совершаемая электрическим полем $\vec{\varepsilon}$ над свободным электроном или дыркой в единицу времени, равна изменению их энергии, определяемому кулоновским взаимодействием поля и движущегося заряда:

$$\frac{\partial E}{\partial t} = \pm e(\vec{\varepsilon} \cdot \vec{v}) = \pm e \left(\vec{\varepsilon} \cdot \frac{\partial \vec{x}}{\partial t} \right).$$

Здесь знак «+» соответствует энергии свободного электрона, а знак «-» - свободной дырки. Для простоты будем рассматривать одномерный случай, когда электрическое поле направлено по оси x и движение частиц также происходит вдоль этой оси. Тогда данное выражение можно переписать следующим образом:

$$\frac{\partial E}{\partial x} \frac{\partial x}{\partial t} = \pm e \varepsilon \frac{\partial x}{\partial t}. \quad (2.6)$$

откуда следует, что

$$\frac{\partial E}{\partial x} = \pm e \varepsilon. \quad (2.7)$$

Отсюда следует, что энергия электронов и дырок во внешнем электрическом поле изменяется с расстоянием в полупроводнике по линейному закону. Скорость изменения энергии определяется величиной приложенного электрического поля. Значит, в присутствии постоянного электрического поля потенциальную энергию электрона можно записать в виде:

$$E_C = E_C(\bar{p}) + e \cdot (\vec{\varepsilon} \cdot \vec{x}). \quad (2.8)$$

Данное выражение показывает, что в этих условиях представление о зонном энергетическом спектре электрона может быть сохранено. Однако зонная диаграмма полупроводника в поле отличается от случая в отсутствие электриче-

ского поля: кроме того, что спектр энергий наклонен по отношению к координатной оси, при больших значениях x изменение второго слагаемого в (2.8) может оказаться сравнимым с первым слагаемым, т.е. с энергией электрона в невозмущенном полупроводнике (см. рис. 12). Согласно выражению (2.8), тангенс угла наклона энергетических зон определяется величиной приложенного электрического поля. Из рисунка следует интересный вывод: при действии электрического поля $\vec{\varepsilon}$ в полупроводнике как таковой запрещенной зоны нет. Действительно, для любого значения энергии, показанного на рис. 12 пунк-

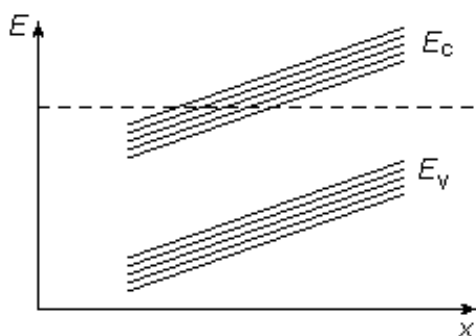


Рис. 12

тирной линией, можно найти область пространства, в которой это значение попадает в зону проводимости или в валентную зону. Представление же о границах зон, тем не менее, сохраняет свой смысл.

Из сказанного следует, что во внешнем поле электрон может, не совершая работы, перейти из одной разрешенной зоны в другую, преодолев барьер в виде запрещенной зоны. Вероятность такого перехода зависит от расстояния, разделяющего разрешенные зоны энергий, и величины запрещенной зоны: чем они меньше, тем больше вероятность перехода между зонами без энергетических затрат. Это явление аналогично известному в квантовой механике туннельному переходу электрона сквозь потенциальный барьер. Соответственно и в теории полупроводников данный эффект, вызванный действием электрического поля, называют **туннельным** эффектом. На этом эффекте основано действие туннельных диодов. Его также следует учитывать при электрическом пробое полупроводниковых приборов – пробое **Зинера**.

Токперенос. Дрейфовый ток. При помещении полупроводника p - или n -типа в электрическое поле величиною ε_0 в зонах разрешенных значений энергии возникает направленное движение электронов и дырок: электроны, находясь в зоне проводимости, перемещаются навстречу полю, а дырки в валентной зоне движутся по полю. Поэтому полный ток в полупроводнике равен сум-

ме этих токов:

$$j = j_n + j_p.$$

Каждый из этих токов можно рассчитать исходя из следующих соображений. Если концентрация электронов равна n_0 и они перемещаются навстречу полю со средней скоростью v_n , то переносимый ими заряд за время t через единичную площадку, перпендикулярную направлению их скорости, будет

$$Q_n = qn_0v_n \cdot t,$$

где $q = 1.6 \cdot 10^{-19}$ Кл - заряд электрона. Следовательно, протекающий электронный ток будет иметь плотность

$$j_n = \frac{dQ_n}{dt} = qn_0v_n.$$

Этот ток должен подчиняться закону Ома

$$j_n = \sigma \cdot \varepsilon_0 = qn_0 \frac{v_n}{\varepsilon_0} \varepsilon_0 = q\mu_n n_0 \varepsilon_0, \quad (2.9)$$

где обозначено: $\mu_n = \frac{v_n}{\varepsilon_0}$ - подвижность электронов в полупроводнике. В соответствии с законами классической механики скорость электрона должна линейно возрастать как с ростом величины электрического поля, так и времени его воздействия t :

$$v_n = at = \frac{F}{m_n} t = \frac{e\varepsilon_0}{m_n} t.$$

В действительности же из-за множественных столкновений скорость движения свободных частиц в кристалле с ростом продолжительности действия электрического поля быстро насыщается и перестает зависеть от времени

$$v_n = \frac{e\langle\tau_n\rangle}{m_n} \varepsilon_0,$$

где $\langle\tau_n\rangle$ - среднее время между соседними соударениями электрона. В резуль-

тате она остается только линейной функцией величины электрического поля. Поэтому подвижность электронов, как и дырок, оказывается практически **постоянной величиной** для данного полупроводника:

$$\mu_n = \frac{e\langle\tau_n\rangle}{m_n}.$$

Подвижность свободных носителей заряда связана с коэффициентами их диффузии D_n , D_p соотношениями:

$$D_n = \frac{kT}{q} \mu_n, \quad D_p = \frac{kT}{q} \mu_p.$$

Воспользовавшись законом Ома ($j_n = \sigma_n \varepsilon_0$), из выражения (2.9) можно найти электропроводность полупроводника, созданную движением электронов:

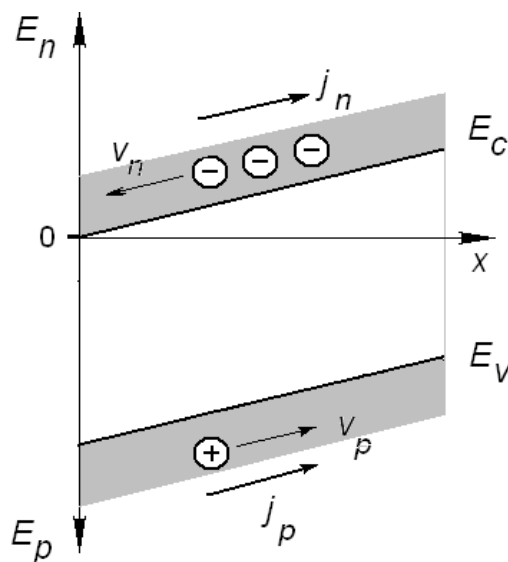
$$\sigma_n = q\mu_n n_0.$$

Аналогичным образом находится дырочная составляющая электропроводности кристалла, в которой фигурируют подвижность дырок μ_p и их концентрация в валентной зоне p_0 . Поэтому полная электропроводность полупроводника равна сумме электронной и дырочной электропроводностей:

$$\sigma = q\mu_n n_0 + q\mu_p p_0. \quad (2.10)$$

Теперь обратимся к объяснению на основе энергетической диаграммы возникновения тока проводимости в полупроводниках при приложении к нему электрического поля. Пусть имеется полупроводник n -типа, донорная примесь которого полностью ионизована.

В результате в зоне проводимости находятся свободные электроны с концентрацией n_0 , равной концентрации введенной донорной примеси N_d . В валентной зоне также имеются свобод-



ные носители заряда – дырки, но их концентрация меньше концентрации свободных электронов и равна

$$p_0 = \frac{n_i^2}{N_d} .$$

Пусть теперь к этому полупроводнику приложили электрическое поле, причем его положительный полюс находится на левом торце полупроводникового образца, а отрицательный – на правом. Наложение электрического поля приводит к наклону энергетической диаграммы полупроводника на угол, определяемый величиной поля. Энергетическая диаграмма полупроводника в поле показана на рис.13. Здесь E_n , E_p - направления отсчитывания энергий электронов и дырок, x - координата в объеме полупроводника в направлении действия электрического поля, v_n , v_p - скорости дрейфа электронов и дырок, j_n , j_p - дрейфовые токи, создаваемые движением электронов и дырок соответственно. Стремясь понизить свою энергию, электроны будут двигаться налево, а дырки – направо. Имея противоположно направленные скорости и противоположные заряды, создаваемые ими токи будут направлены одинаковы. Поэтому полный ток дрейфового движения носителей заряда будет равен сумме дрейфовых токов электронов и дырок.

Диффузионный ток. Электрический ток в твердых телах может возникнуть не только под действием электрического поля. Направленное движение электронов и (или) дырок может возникнуть при неоднородном их распределении в пространстве. В результате из мест с высокой концентрацией свободные носители устремятся в места, где их концентрация меньше. Возникающий при этом ток называется **диффузионным**, т.к. причина возникновения движения - диффузия носителей заряда. Величина диффузионного тока пропорциональна градиенту концентрации свободных носителей заряда. В одномерном случае диффузионные токи электронов и дырок равны:

$$j_n^{(диф)} = qD_n \frac{\partial n}{\partial x}, \quad j_p^{(диф)} = qD_p \frac{\partial p}{\partial x}, \quad (2.11)$$

где D_n , D_p – коэффициенты диффузии электронов и дырок соответственно. Поскольку коэффициенты диффузии являются константами вещества, то, зная градиенты концентраций, по выражениям (2.11) можно найти величины диффузионных токов.

3. ВЗАИМОДЕЙСТВИЕ ОПТИЧЕСКОГО ИЗЛУЧЕНИЯ С ТВЕРДЫМИ ТЕЛАМИ

3.1. Основные параметры процесса поглощения излучения в полупроводниках

При рассмотрении поглощения электромагнитного излучения в полупроводниках будем вести речь только о слабых потоках энергии. Слабая электромагнитная волна не изменяет энергетический спектр носителей заряда, а лишь создает новые пары «электрон-дырка» или вызывает перераспределение носителей по состояниям зоны проводимости или валентной зоны. Поэтому параметры полупроводника, характеризующие его оптические свойства, не зависят от интенсивности падающего излучения.

1. При падении излучения интенсивностью I_0 на полупроводник (см. рис. 13) в его объем проникает не весь поток, т.к. часть излучения отражается от поверхности. Для определения степени отражения света вводится параметр, именуемый **коэффициентом отражения**. Обозначим интенсивность излучения через I – количество энергии света, переносимое лучом через единичную площадку в единицу времени. Если падающая интенсивность I_0 , а отраженная – I_R , то коэффициент отражения определяется как

$$R = \frac{I_R}{I_0}.$$

Это первый параметр, характеризующий взаимодействие оптического излучения с твердым телом. Удобство данного параметра заключается в том, что диапазон его значений ограничен ($0 \leq R \leq 1$), что позволяет сравнивать между собой по отражающей способности различные поверхности.

Зависимость коэффициента отражения R от длины волны падающего излучения $\lambda - R(\lambda)$ называется **спектром отражения** твердого тела. Это первая характеристика, описывающая взаимодействие света и твердое тело.

2. Экспериментальные исследования показывают, что часть излучения, проникающая в объем полупроводника, по мере распространения в его объеме уменьшается. Причиной уменьшения интенсивности излучения с координатой в объеме является взаимодействие излучения с атомами и решеткой полупроводника, со свободными носителями и т.д. Это уменьшение описывается законом **Бугера-Ламберта**:

$$I(x) = I_0(1 - R) \cdot \exp(-\alpha \cdot x),$$

где α – некоторый параметр среды, который имеет размерность обратной длины и называется **коэффициентом поглощения**. Это второй параметр, характеризующий взаимодействие света и твердого тела.

Для того чтобы выяснить физический смысл коэффициента поглощения, будем считать показатель экспоненты в законе равным единице: $\alpha \cdot x_0 = 1$. Отсюда $x_0 = 1/\alpha$ – это расстояние от освещаемой поверхности полупроводника, при прохождении которого интенсивность излучения уменьшается в e - раз (рис. 1). Это расстояние обозначим как $l_\phi = x_0$ и назовем **длиной свободного пробега фотона** в полупроводнике, т.е. это то расстояние, которое проходит фотон в

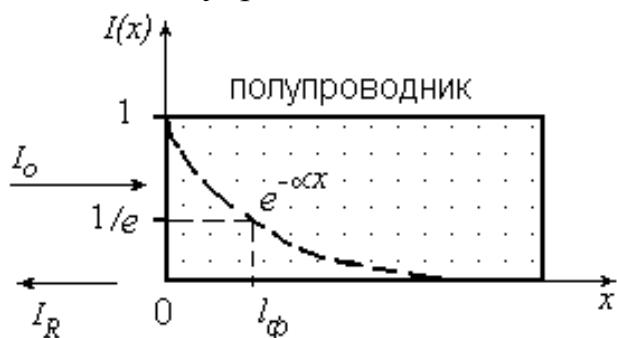


Рис.14

твердом теле до момента своего поглощения: $l_\phi = \alpha^{-1}$.

Почему же интенсивность уменьшается по экспоненциальному закону, а не по линейному? Чтобы ответить на этот вопрос, начнем с построения физической

картины поглощения. Для этого мысленно разделим весь полупроводник на тонкие слои и будем определять долю поглощенной интенсивности света в выбранном слое по отношению к падающей на этот слой интенсивности. Коэффици-

коэффициент поглощения в каждом слое предполагаем одинаковым (полупроводник считается однородным по свойствам). Тогда можно считать, что в каждом слое поглощается одинаковая доля мощности, предположим 20%, падающей интенсивности. Поскольку к каждому очередному слою полупроводника подходит интенсивность по величине меньшая, чем к предыдущему, то и абсолютное значение поглощенной в рассматриваемом слое интенсивности будет меньше, чем в предыдущем. Значит, с увеличением расстояния от облучаемой поверхности величина поглощенной интенсивности будет уменьшаться. Если бы в каждом слое поглощалась одна и та же интенсивность, то закон уменьшения интенсивности излучения с координатой был бы линейным. В данном случае уменьшение интенсивности от слоя к слою непостоянно. Следовательно, закон уменьшения интенсивности с расстоянием должен быть нелинейным.

Для количественного ответа на поставленный вопрос выведем математический закон уменьшения интенсивности излучения по мере распространения в полупроводнике, исходя из квантовой теории излучения. Для этого выразим интенсивность I через число фотонов в падающем излучении. Пусть n – число фотонов в единичном объеме светового потока, тогда $n \cdot c$ есть поток фотонов (количество фотонов, переносимых через единичную площадку за единицу времени). Здесь c – скорость света. Поэтому интенсивность света равна:

$$I = n \cdot c \cdot \hbar \omega. \quad (3.1)$$

Из наблюдений известно, что интенсивность излучения с координатой x уменьшается. Зададимся вопросом: за счет какого параметра это может происходить? Из выражения (4.1) следует, что таковым может быть только концентрация фотонов в потоке n , т.к. скорость света c , частота светового потока ω и постоянная Дирака \hbar не могут зависеть от координаты. Поэтому выражение (3.1) можно переписать в виде:

$$I(x) = n(x) \cdot c \cdot \hbar \omega. \quad (3.2)$$

Из выражения (3.2) следует достаточно важный вывод: с точки зрения фотонной теории света, уменьшение его интенсивности при распространении в полу-

проводнике возможно только за счет уменьшения числа фотонов. Уменьшение числа фотонов может происходить только за счет их поглощения, т.к. фотоны не могут покинуть полупроводник (эта возможность в данной главе не рассматривается). Поглощение фотонов, вероятно, осуществляется некими центрами поглощения. Как показывает эксперимент, существование их в чистых и структурно совершенных кристаллах наталкивает на мысль, что этими центрами с большой долей вероятности могут выступать атомы основного вещества твердого тела. Примем эту гипотезу о возможной природе центров поглощения как одну из возможных.

Чтобы сделать какие-либо выводы относительно причин и механизма поглощения фотонов твердым телом, рассмотрим случай, когда фотон попадает в площадь поперечного сечения атома полупроводника. Пусть σ – площадь поперечного сечения поглощающего центра, т.е. это вероятность исчезновения фотона при его попадании в одиночную мишень сечением σ . Если N – число центров, поглощающих фотоны в единичном объеме, то уменьшение концентрации фотонов на величину ∂n при прохождении им слоя полупроводника толщиной ∂x должно быть пропорционально площади поперечного сечения поглощающего центра σ , концентрации N , величине потока и толщине слоя:

$$-\partial n = \sigma N \cdot n(x) \partial x.$$

Здесь $\sigma \cdot N$ – вероятность поглощения фотонов всеми центрами в единичном объеме. Из этого уравнения найдем ($q = nc$)

$$\frac{\partial q}{q(x)} = -\sigma N \cdot \partial x.$$

Интегрирование этого уравнения дает:

$$q(x) = q(0) \cdot \exp(-\sigma N \cdot x).$$

Умножив обе части этого выражения на энергию кванта падающего излучения, получим закон изменения интенсивности света по мере его распространения в твердом теле, в котором имеются поглощающие центры:

$$I(x) = I(0) \cdot \exp(-\sigma N \cdot x).$$

Сравнивая это выражение с выражением (4.1) находим, что коэффициент поглощения равен: $\alpha = \sigma N$. Нетрудно убедиться, что он имеет размерность обратной длины.

Можно считать поперечное сечение поглощения равным поперечному сечению самих атомов: $\sigma \approx S_{am}$. Это означает, что достаточно фотону попасть в размер атома, чтобы быть поглощенным им. Поэтому $\sigma \approx (10^{-16} \div 10^{-17}) \text{ см}^2$, $N \approx 10^{22} \text{ см}^{-3}$. Тогда

$$\alpha \approx (10^5 \div 10^6) \text{ см}^{-1},$$

$$l_{\phi} \approx (10^{-5} \div 10^{-6}) \text{ см} = (0.1 \div 0.01) \text{ мкм}.$$

Отсюда следует важный физический вывод: излучение поглощается в тонком приповерхностном слое, а объем полупроводника в поглощении излучения никак не участвует. Поэтому поглощение излучения очень чувствительно к состоянию приповерхностного слоя и поверхности полупроводника.

Если в твердом теле обнаруживается несколько механизмов поглощения со своими значениями σ_i и N_i , действующими независимо друг от друга, то результирующий коэффициент поглощения излучения будет равен алгебраической сумме коэффициентов поглощения различных механизмов

$$\alpha(\lambda) \approx \sum_i \alpha_i(\lambda),$$

т.к. вероятности независимых событий складываются. Здесь сумма берется по различным сортам поглощающих центров.

3.2. Основные механизмы поглощения излучения

В приведенном выше объяснении физической сущности закона Бугера-Ламберта дана феноменологическая картина поглощения фотона, показавшая зависимость коэффициента поглощения от концентрации поглощающих атомов и их площади. Ясно, что реальное значение коэффициента поглощения и его зависимость от длины волны поглощаемого излучения определяются конкрет-

ным механизмом взаимодействия фотона и полупроводника. В реальном полупроводнике возможны различные механизмы поглощения. Перечислим и кратко охарактеризуем их.

1. Собственное (фундаментальное) поглощение. Этот вид поглощения обусловлен взаимодействием квантов света с атомами основного вещества, например, кремния или германия, и потому называется собственным. Он имеет место при энергии кванта, равной или больше ширины запрещенной зоны полупроводника: $\hbar\omega \geq \Delta E_g$.

2. Примесное поглощение. Данный вид поглощения имеет место, если энергия кванта равна или несколько превышает энергетические зазоры между зоной разрешенных энергий и уровнями примеси донорного (E_d) или акцепторного (E_a) типа: $\hbar\omega \geq E_c - E_d, E_a - E_v$. При этом будет наблюдаться переход электронов с примесного донорного уровня в зону проводимости или из валентной зоны на примесный акцепторный уровень. Таким образом, в случае примесного поглощения энергия кванта излучения идет не на возбуждение атомов основного вещества, а на ионизацию примесей в нем.

3. Поглощение свободными носителями заряда. В этом виде поглощения квант сталкивается со свободным электроном или дыркой и, как в случае соударения с атомами, передает подвижной частице свою энергию и квазиимпульс. Приобретенный электроном при соударении квазиимпульс ускоряет носители заряда в направлении падающего света. В результате действия этого вида поглощения создается направленный поток свободных носителей заряда. Этот тип поглощения называют еще внутризонным. Оно наблюдается только в полупроводниках со сложной структурой валентной зоны или зоны проводимости и связано с переходом электронов (дырок) с одной зоны $E(\vec{k})$ в другую зону, отличающуюся кривизной указанной зависимости.

4. Решеточное поглощение. В этом случае квант света взаимодействует с колебаниями решетки. Результатом этого является увеличение количества акустических и оптических фононов кристаллической решетки, что выражается

в повышении температуры освещаемого твердого тела..

5. **Экситонное поглощение** света характерно тем, что энергия поглощаемого излучения идет на образование специфического комплекса – экситона: связанной пары «электрон – дырка», когда рожденные светом частицы не разлетаются в разные стороны, а остаются рядом до момента их рекомбинации.

6. **Плазменное поглощение** света. Этот вид поглощения создается совокупностью электронов и дырок ($n_{\text{электр}} = n_{\text{дырок}}$). Энергия поглощаемых фотонов идет на возбуждение плазменных колебаний.

Из приведенного рассмотрения физических процессов, определяющих взаимодействие оптического излучения с твердым телом, следует, что спектр поглощения $R(\lambda)$ должен зависеть от внешних воздействий на полупроводник, которые способны изменить состояние основных и примесных атомов полупроводника, спектр колебаний кристаллической решетки. Такими воздействиями являются температура тела, давление, электрическое и магнитное поле и т.д.

3.3. Собственное поглощение. Прямые и не прямые переходы

Поглощение кванта происходит, если он попадает в площадь поперечного сечения атома в любое его место: в центр или периферию. Размеры ядра атома малы и поэтому основная его площадь занята электронами, вращающимися по своим орбитам. Поэтому взаимодействие кванта осуществляется в основном электронами. Первыми на пути кванта встречаются внешние (валентные) электроны – они вращаются по внешним орбитам и защищают внутренние орбиты, проникнуть к которым может только высоко энергетичный квант (это особый случай, который не рассматриваем). Поэтому взаимодействие оптического излучения с веществом определяется валентными электронами. Если частота кванта находится в резонансе с частотой перехода валентных электронов ($\hbar\omega = \Delta E_g$), то столкновение кванта и атома, различающихся по массе на несколько порядков, носит характер неупругого соударения. Отдав атому всю

свою энергию и весь квазиимпульс, фотон исчезает, т.к. в состоянии покоя он имеет нулевую массу ($m_{\text{фот}} = 0$). Описывая этот процесс, говорят, что квант «поглотился» атомом.

Получив энергию кванта, валентный электрон переходит в новое энергетическое состояние: он отрывается от атома и переходит в межатомное пространство, где имеет нулевую скорость. Если энергия кванта больше ширины

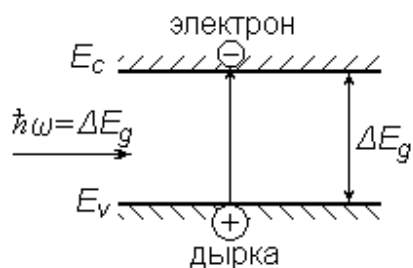


Рис. 14

запрещенной зоны, то избыток энергии над ΔE_g идет на приобретение свободным электроном кинетической энергии.

Что же происходит с квазиимпульсом атома? Полученный от кванта квазиимпульс по своей величине обычно пренебрежимо мал по сравнению с собственным квазиимпульсом

атома и потому поглощение кванта практически никак не сказывается на кинетических параметрах теплового движения атома. Хотя «удар» от кванта света есть, но он становится заметным при поглощении излучения легкой подвижной частицей, например, электроном, или при поглощении атомом частицы с высокой энергией частицы.

Суммируя описанное выше, на энергетической диаграмме процесс поглощения кванта излучения и выход электрона в межатомное пространство изображают так (см. рис. 14): исчезновение кванта и переход электрона из валентной зоны в зону проводимости.

При собственном поглощении излучения необходимо различать два случая переходов электронов из валентной зоны в зону проводимости:

- переход электрона осуществляется только под действием фотона. Такой переход называется **прямым**;
- переход в зону проводимости совершается под действием фотона и частично под действием колебаний решетки – фононов. Такой переход называется **непрямым**.

Первый тип переходов реализуется в полупроводниках, у которых минимум энергии в зоне проводимости совпадает с максимумом энергии электрона в валентной зоне (рис. 8): при $k = 0$ имеет место совпадение экстремумов зоны проводимости и валентной зоны. Второй тип переходов имеет место, если эти экстремумы не совпадают (рис. 10). Рассмотрим подробнее оба типа переходов.

Прямые переходы. При прямых переходах импульс электрона фактически не изменяется. Из закона сохранения квазиимпульса следует, что разность импульсов электрона в разных зонах должна равняться импульсу фотона, т.к. при них электрон взаимодействует только со светом. Действительно:

$\bar{k}_C - \bar{k}_V = \bar{k}_{\text{фотон}} = \frac{\hbar\omega}{c} \approx 0$, т.к. c велико, т.е. $k_V \approx k_C$ – импульсы в зонах примерно равны. Значит, на энергетической диаграмме прямые переходы можно показать так: переход из максимума валентной зоны в минимум (см. рис. 8) зоны проводимости без изменения квазиимпульса. Для получения прямого перехода необходима энергия фотона не менее ΔE_g . Поэтому граничную длину волны поглощаемого излучения данным способом можно найти так: $\hbar\omega_{кр} = \Delta E_g$ или $hc/\lambda_{кр} = \Delta E_g$, откуда находим: $\lambda_{кр} = hc/\Delta E_g$. Здесь $\omega_{кр}$ и $\lambda_{кр}$ – круговая частота, и длина волны излучения, соответствующего красной границе фотоэффекта. Подставив численные значения входящих сюда констант, получим:

$$\lambda_{кр} (\text{мкм}) = \frac{1,24}{\Delta E_g (\text{эВ})}.$$

Так, в кремнии при комнатной температуре $\Delta E_g = 1,12$ эВ, поэтому $\lambda_{кр} \approx 1,1$ мкм. Это означает, что если на образец кремния будет падать излучение с длиной волны меньше $\lambda_{кр}$, то оно будет поглощаться за счет собственного поглощения. Но при длине волны падающего излучения большей $\lambda_{кр}$ поглощение прекратится – энергии кванта окажется недостаточно для перевода электрона из валентной зоны в зону проводимости.

Непрямые переходы. Если переход не прямой, то разность квазиимпульсов конечного и начального состояний электрона (рис. 9) должна быть скомпенсирована участием в процессе поглощения дополнительной частицы, например, фонона. **Фононом** называют квант теплового колебания решетки, так что ее тепловые колебания представляют как хаотичное движение совокупности фононов с различными частотами и квазиимпульсами. Тогда

$$\bar{k}_C - \bar{k}_V = \bar{k}_{\text{фотон}} \pm \bar{k}_{\text{фонон}}.$$

Но при поглощении кванта света, кроме квазиимпульса, электрон должен получить (или отдать) от фонона и энергию $-E_{\text{фонон}}$, и закон сохранения энергии будет выглядеть так:

$$\hbar\omega \pm E_{\text{фонон}} = \Delta E,$$

Вероятность таких переходов с поглощением или излучением фонона будет меньше, чем вероятность прямых переходов, т.к. при непрямых переходах необходима встреча трех частиц: электрона, фотона и фонона. В связи с этим непрямые переходы имеют место, когда прямые переходы по каким-либо причинам невозможны. На энергетической диаграмме непрямые переходы показываются стрелкой, выходящей из максимума валентной зоны и заканчивающийся в абсолютном минимуме зоны проводимости (рис. 9). Если на такой полупроводник падает свет с энергией кванта $\hbar\omega = \Delta E_g$, то прямые переходы в нем невозможны, но будут происходить непрямые переходы, при которых электрон должен изменить свой квазиимпульс k от 0 до k_1 , а также увеличить свою энергию E на величину ΔE_g .

Непрямые переходы могут иметь место и в прямозонном полупроводнике, если его зона проводимости в значительной степени занята электронами, что возможно в полупроводниках с высокой концентрацией свободных частиц в зоне разрешенных энергий.

3.4. Примесное поглощение излучения

Этот тип поглощения света реализуется в полупроводниках, в которых есть донорные (E_d) или акцепторные (E_a) примеси, имеющие энергетические уровни в запрещенной зоне, локализованные вблизи ее краев. Если будет па-

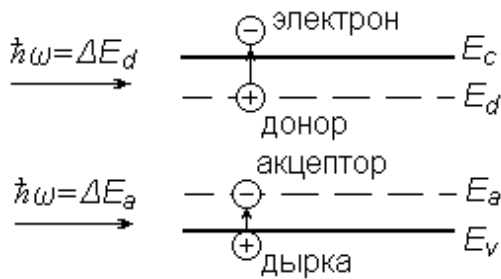


Рис. 15

дать свет с $\hbar\omega \geq E_c - E_d$, $\hbar\omega \geq E_a - E_v$ то электроны с уровня E_d будут переходить в E_c или из валентной зоны на E_a (см. рис. 15). Как следствие этого, при поглощении света атомами примеси в полупроводнике будут образовываться неподвижные ионы с фотоиндуцированным положительными

или отрицательным зарядом. Заметим, что под действием света возможны не только переходы с образованием ионов, но и обратные: электронов из E_c на E_d дырок с E_a в валентную зону. Это поглощение будет наблюдаться при низких температурах – акцепторные уровни не должны быть полностью заполнены электронами, а донорные – полностью ионизованными. В остальном процесс примесного поглощения протекает аналогично описанному выше в п. 3.3 для собственного поглощения, но с заменой атомов основного вещества на атомы примеси. Более того, в результате примесного поглощения в полупроводнике образуется один тип свободного носителя: электрон при поглощении донорной примесью или дырка – при поглощении кванта акцепторной примесью, тогда как при собственном поглощении генерируется пара «электрон – дырка». Этот тип поглощения имеет непрерывный спектр, т.е. при изменении длины волны света от значения, соответствующего переходу «зона – примесный уровень» и выше, коэффициент примесного поглощения остается постоянным.

Существует важное количественное различие между собственным и примесным видами поглощения: ввиду того, что концентрация примесей (обычно от 10^{11} до 10^{18} см^{-3}) значительно меньше концентрации атомов основного вещества (10^{22} см^{-3}), вероятность столкновения кванта с атомом примеси во много

раз ниже вероятности столкновения кванта с атомом основного вещества. Следовательно, коэффициент примесного поглощения имеет величину ($10^1 - 10^2 \text{ см}^{-1}$), т.е. на несколько порядков меньше коэффициента собственного поглощения. Значит, примесное поглощение проявляется при исключении других видов, маскирующих примесное поглощение.

3.5. Поглощение свободными носителями (внутризонное поглощение)

Поглощение свободными носителями заряда – это один из основных механизмов поглощения оптического излучения в полупроводниках, когда квант света

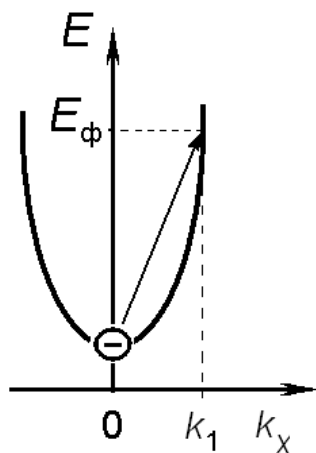


Рис. 16

взаимодействует с носителем, который свободно движется внутри зоны и реагирует на внешние воздействия. Физическая картина поглощения, например, свободным электроном кванта света такова. Поглощая энергию фотона E_ϕ , электрон на долинной диаграмме полупроводника (см. рис. 16) должен переместиться из начального состояния с $k = 0$ в новое состояние, которое выше по энергии на величину энергии кванта E_ϕ , но находится на параболической зависимости энергии

от волнового вектора (на рисунке переход электрона показан стрелкой). Но это возможно, если электрон наряду с энергией кванта электрона получит квазиимпульс величиной k_1 . Этот квазиимпульс электрон может получить либо от взаимодействия с фононами (акустическими или оптическими), либо от взаимодействия с ионизованными примесями.

Поглощение свободными носителями характеризуется монотонным спектром поглощения $\alpha(\lambda)$, описываемым законом $\alpha(\lambda) \sim \lambda^p$, где λ – длина волны фотона, а показатель p может принимать значения в пределах от 1,5 до 3,5. Рассеяние света на акустических фононах приводит к поглощению, меняющемуся по закону $\alpha_A(\lambda) \sim \lambda^{1.5}$. Рассеяние на оптических фононах дает зависимость $\alpha_\phi(\lambda)$

$\sim \lambda^{2.5}$, тогда как рассеяние на ионизованных примесях может дать зависимость $\alpha_{\text{п}}(\lambda) \sim \lambda^{3.0 \div 3.5}$. В общем случае реализуются все типы рассеяния, и результирующий коэффициент поглощения свободными носителями представляет собой сумму трех слагаемых:

$$\alpha_{\Sigma}(\lambda) = \alpha_A(\lambda) + \alpha_{\Phi}(\lambda) + \alpha_{\text{п}}(\lambda) = A\lambda^{1.5} + B\lambda^{2.5} + C(N)\lambda^{3.5}$$

где A, B, C – константы, характеризующие вклад того или иного механизма рассеяния избытка энергии в суммарный коэффициент поглощения $\alpha_{\Sigma}(\lambda)$. Вклад от рассеяния на примесях тем больше, чем выше их концентрация N . Поэтому $C = C(N)$. Спектр поглощения на свободных носителях с учетом собственного поглощения будет иметь вид, показанный на рис. 17.

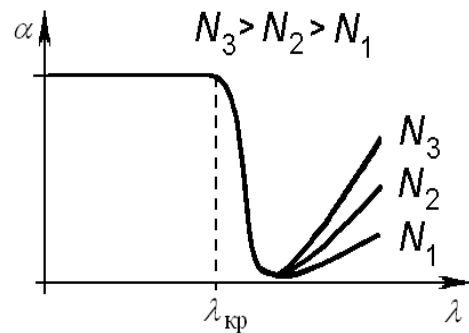


Рис.17.

Таким образом, в зависимости от типа доминирующего механизма рассеяния избыточной энергии кванта характер спектральной зависимости коэффициента поглощения на свободных носителях заряда будет различен: от $\lambda^{1.5}$ до $\lambda^{3.5}$.

В дополнение к этому заметим, что поглощение кванта света и изменение энергетического состояния свободной частицы происходит в соответствии с энергетической диаграммой той зоны разрешенных энергий, какая свободная частица

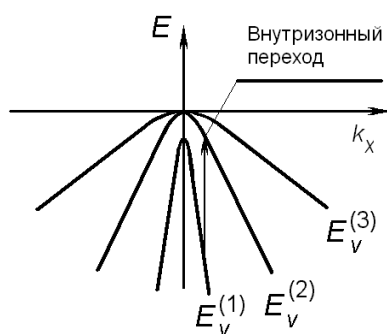


Рис. 18

поглощает квант. Например, поглощение происходит дыркой в полупроводнике, в котором имеется три валентные зоны: $E_v^{(1)}, E_v^{(2)}, E_v^{(3)}$. Переход дырки с одной параболы $E(\vec{k})$ на другую в спектре поглощения полупроводника проявится как пик величины коэффициента поглощения. Эти переходы называют **внутризонными** (рис. 18).

3.6. Решеточное поглощение

Этот тип поглощения реализуется при взаимодействии кванта света с тепловыми колебаниями решетки твердого тела, в результате которого энергия кванта переходит в колебательную энергию атомов или ионов. Поскольку важным является вопрос о связи атомов решетки, кратко рассмотрим их виды. Различают четыре основных типа связей: ионная, металлическая, ковалентная и связь Ван-дер-Ваальса. Ионная связь образуется кулоновским взаимодействием положительных и отрицательно заряженных ионов. Металлическая связь создается кулоновским взаимодействием положительных ионов в решетке и свободных электронов, легко проникающих во все части металла и тем самым «цементирующих» его. Ковалентная связь создается обобществленными валентными электронами. Связь Ван-дер-Ваальса создается за счет взаимной поляризации соседних атомов. Механизмом взаимодействия кванта света и ионов решетки является кулоновское взаимодействие заряженной частицы (иона решетки) и электрического поля кванта с силой, равной $F = e\varepsilon_{\phi}$, где ε_{ϕ} – напряженность электрического поля в световой волне, e – заряд иона. Когда частоты колебания в кванте света и теплового колебания иона в узле решетки совпадают, энергия кванта передается решетке, увеличивая амплитуду колебания атома соответствующей частоты.

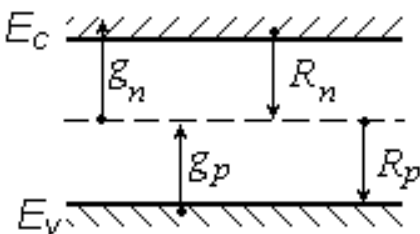
В случае же атомной решетки взаимодействие кванта света и атомов осуществляется за счет взаимодействия электрического поля световой волны с ионной компонентой связи между атомами решетки.

4. ФОТОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

4.1. Основные понятия и параметры

В полупроводниках, в отличие от металлов, под влиянием внешних воздействий (освещения, электрического поля и т.д.) концентрации электронов и дырок могут изменяться во много раз. Это приводит к ряду специфических явлений, на которых основана работа многих полупроводниковых приборов.

При нарушении термодинамического равновесия концентрации электронов и дырок в зонах (n и p) изменяются по сравнению с их равновесными значениями n_0 и p_0 , т.е. в зонах появляются неравновесные носители заряда с концентрациями $\delta n = n - n_0$ и $\delta p = p - p_0$. Установление концентраций в зонах определяется процессами генерации и рекомбинации электронов и дырок. Существует несколько разновидностей процессов генерации (световая, тепловая и т.д.) и рекомбинации (тепловой заброс носителей заряда «зона-зона», тепловой заброс носителей заряда с участием примесного уровня «зона – уровень – зона» и т.д.). Суммарные скорости этих процессов (количество генерируемых или рекомбинируемых частиц в единичном объеме в единицу времени) принято обо-



значать как g и R , соответственно. Это два противоположно направленных процесса, равенство скоростей которых создает термодинамическое равновесие в зонах разрешенных энергий полупроводника.

Рис. 19

Время жизни неравновесных носителей.

Пусть под влиянием внешнего воздействия в единице объема полупроводника в единицу времени возникает g_n электронов проводимости и соответственно g_p дырок в валентной зоне. Скорости генерации g_n и g_p будем считать постоянными по всему объему полупроводника, хотя и не обязательно равными друг другу. Пусть R_n и R_p – скорости исчезновения электронов и дырок в результате их рекомбинации. Если в полупроводнике нет электрического тока, то сум-

марная скорость изменения неравновесных концентраций электронов и дырок в зонах будет определяться скоростями их генерации и рекомбинации (рис. 19):

$$\frac{\partial \delta n}{\partial t} = g_n - R_n; \quad \frac{\partial \delta p}{\partial t} = g_p - R_p .$$

Напомним, что g_n и g_p описывают генерацию за счет внешнего воздействия и не учитывают переходы, вызванные тепловым движением. Эти переходы учитываются в величинах R_n и R_p .

Для описания кинетики неравновесных электронных процессов (развития процессов во времени) вводят понятие **среднего времени жизни неравновесных электронов в зоне проводимости τ_n и дырок в валентной зоне τ_p** , которые определяются через скорости рекомбинации электронов и дырок:

$$R_n = \frac{n - n_0}{\tau_n}, \quad R_p = \frac{p - p_0}{\tau_p}$$

или иначе: $1/\tau_n$ – это вероятность исчезновения одного избыточного электрона из зоны проводимости в единицу времени в результате рекомбинации с дыркой, а $1/\tau_p$ – вероятность рекомбинации одной дырки в единицу времени.

Пользуясь понятиями времени жизни носителей заряда, уравнения кинетики неравновесных концентраций электронов и дырок в однородном образце, в котором нет тока проводимости, можно переписать в виде:

$$\frac{\partial \delta n}{\partial t} = g_n - \frac{\delta n}{\tau_n}; \quad \frac{\partial \delta p}{\partial t} = g_p - \frac{\delta p}{\tau_p}. \quad (4.1)$$

Стационарные концентрации неравновесных носителей заряда $(\delta n)_s$ и $(\delta p)_s$, устанавливающиеся после длительного воздействия внешней генерации, можно найти, если в (4.1) приравнять к нулю левые части. Действительно, стационарное состояние характеризуется неизменностью концентраций свободных носителей в зонах, что требует в выражениях (4.1) положить все производные по времени обратить в нуль. В результате этого упрощения из (4.1) можно най-

ти выражения для подсчета стационарных концентраций свободных носителей заряда в зонах:

$$(\delta n)_s = g_n \tau_n; (\delta p)_s = g_p \tau_p. \quad (4.2)$$

Теперь рассмотрим кинетику изменения концентраций носителей заряда. В простейшем случае, когда τ_n и τ_p не зависят от n и от p , интегрирование кинетических уравнений (4.1) с учетом выражений (4.2) дает:

$$\begin{aligned} \delta n(t) &= g_n \tau_n - C_n \exp\left(-\frac{t}{\tau_n}\right); \\ \delta p(t) &= g_p \tau_p - C_p \exp\left(-\frac{t}{\tau_p}\right). \end{aligned} \quad (4.3)$$

Здесь C_n, C_p – постоянные интегрирования, определяемые из начальных условий: если в начале полупроводник находился в термодинамическом равновесии и затем в момент времени $t = 0$ включено внешнее воздействие (создающее генерацию носителей заряда), то при $t = 0$ будем иметь $\delta n = 0$. Использование данного начального условия дает возможность найти постоянные интегрирования:

$$\begin{aligned} C_n &= g_n \tau_n = (\delta n)_s; \delta n(t) = (\delta n)_s \cdot \left[1 - \exp\left(-\frac{t}{\tau_n}\right)\right]; \\ C_p &= g_p \tau_p = (\delta p)_s; \delta p(t) = (\delta p)_s \cdot \left[1 - \exp\left(-\frac{t}{\tau_p}\right)\right]. \end{aligned} \quad (4.4)$$

Ход этих зависимостей показан на рис. 20 сплошной линией. Если в некоторый момент времени $t = t_1$ генерация выключается, то для времен $t \geq t_1$, $g_n = 0$ и начальное условие принимает вид: $t = t_1, \delta n = (\delta n)_1$. Тогда:

$$C_n = -(\delta n)_1 \cdot \exp\left(-\frac{t_1}{\tau_n}\right)$$

и далее

$$\delta n(t) = (\delta n)_1 \cdot \exp\left(-\frac{(t-t_1)}{\tau_n}\right). \quad (4.5)$$

По аналогии можно записать:

$$\delta p(t) = (\delta p)_1 \cdot \exp\left(-\frac{(t-t_1)}{\tau_p}\right). \quad (4.6)$$

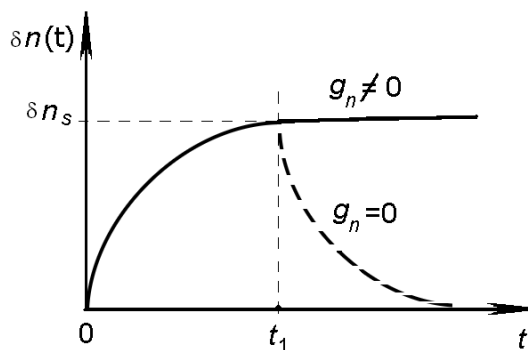


Рис. 20

Ход этой зависимости показан на рис. 20 пунктирной линией. Как следует из рисунка, выключение генерации носителей заряда приводит к тому, что избыточные концентрации электронов и дырок за счет процессов рекомбинации будут стремиться к своим равновесным значениям. Согласно выражениям (4.4), (4.5) и (4.6) при постоянных значениях параметров τ_n и τ_p установление избыточных концентраций электронов и дырок в зонах разрешенных энергий, а также их исчезновение описываются экспоненциальным законом. Поэтому по аналогии с определением физического смысла коэффициента поглощения (см. п. 3.1) можно определить физический смысл констант τ_n и τ_p так: τ_n (или τ_p) есть промежуток времени, в течение которого неравновесная концентрация электронов (дырок) при включении или выключении источника генерации носителей заряда увеличивается или уменьшается в e -раз. В большинстве практических случаев $\tau_n \approx \tau_p$.

4.2. Фотопроводимость полупроводников.

Собственная и примесная фотопроводимость

Простейший способ создания неравновесных носителей заряда состоит в освещении полупроводника электромагнитным излучением. Зарегистрировать изменение концентрации свободных носителей заряда во всем объеме полупроводника проще всего, измерив величину электрического тока, протекающего через полупроводник, при действии на него постоянного напряжения фиксированной величины. В таких случаях говорят о возникновении фотопроводимости. Под **фотопроводимостью** полупроводника понимают изменение его проводимости, вызванное освещением полупроводника.

Электронные переходы при оптической генерации могут быть различными. Если энергия фотонов $\hbar\omega > \Delta E_g$, то неравновесные электроны и дырки образуются в результате возбуждения электронов, переходящих из валентной зоны в зону проводимости, а также дырок, остающихся в валентной зоне. Это так называемая собственная оптическая генерация и, соответственно, **собственная фотопроводимость**. Обратный процесс есть прямая рекомбинация «свободный электрон – свободная дырка» (рис. 19). В результате оптической генерации в области собственного поглощения в полупроводнике образуется пара свободных носителей заряда противоположных знаков. Поэтому данный тип фотопроводимости еще называют **биполярной**.

Однако при наличии достаточного количества примесей в полупроводнике фотопроводимость может возникать и при $\hbar\omega < \Delta E_g$. На рис. 15 показана энергетическая диаграмма полупроводника, содержащего глубокую акцепторную примесь. Тогда при падении света с энергией $\hbar\omega > E_C - E_d$ происходит возбуждение электронов, связанных с примесью, в зону проводимости и имеем случай примесной оптической генерации и, соответственно, **примесной проводимости**. В случае, когда $\hbar\omega \geq E_a - E_v$, возможна генерация дырок в валентной зоне. Оба случая – случаи **монополярной** проводимости. Поглощение в собственной полосе длин волн на несколько порядков больше поглощения в примесной области.

Скорость оптической генерации связана с коэффициентом поглощения света. Пусть $I(x)$ есть интенсивность монохроматического светового потока на расстоянии x от освещаемой поверхности полупроводника, а α – коэффициент поглощения. Тогда количество энергии, поглощаемой в единицу времени на единичной площади в слое, расположенном между слоями x и $x + dx$, есть:

$$\begin{aligned} -dI &= I(x+dx) - I(x) = I_0 \exp(-\alpha(x+dx)) - I_0 \exp(-\alpha x) = \\ &= I_0 \exp(-\alpha x) \exp(-\alpha \cdot dx) - I_0 \exp(-\alpha x) = I_0 \exp(-\alpha x) \cdot [\exp(-\alpha \cdot dx) - 1] = \\ &= -I(x) \cdot [1 - \alpha \cdot dx - 1] = -I(x) \cdot \alpha \cdot dx, \end{aligned}$$

где произведено разложение экспоненциальной функции в ряд Тейлора по степеням dx вблизи точки x . Следовательно, число поглощаемых фотонов за еди-

ницу времени в единичном объеме, равно $\alpha \cdot I(x)/\hbar\omega$. Поэтому скорость оптической генерации g можно представить в виде:

$$g = \frac{\eta(\omega) \cdot \alpha(\omega) \cdot I(x)}{\hbar\omega}.$$

Здесь обозначено: $\eta(\omega)$ – **квантовый выход внутреннего фотоэффекта**, который равен числу пар свободных носителей, рождаемых в среднем одним поглощенным фотоном. Квантовый выход может быть больше единицы, если при поглощении одного фотона высокой энергии возникает две или более электронно-дырочные пары. Обычно, квантовый выход меньше единицы, поскольку часть фотонов поглощается либо решеткой (фононами), либо свободными носителями в зонах без образования дополнительных электронов и дырок. В общем случае g различно в разных точках полупроводника – случай **неоднородной генерации**. Если $\alpha \cdot d \ll 1$ (где d – толщина образца), то $I(x) \approx \text{const}$ (слабо поглощаемый свет) и g можно считать постоянным по объему. Этот случай называют случаем **однородной генерации** носителей заряда, т.к. при таком коэффициенте поглощения число генерируемых светом электронно-дырочных пар примерно постоянно по всей длине полупроводника в направлении падающего света.

Для того чтобы определиться в количестве возможных фотоэффектов в полупроводнике при его освещении излучением из собственной полосы поглощения, обратимся к выражению для проводимости полупроводника, которая в общем случае представляет собой сумму проводимостей по зоне проводимости, создаваемой свободными электронами – σ_n , и проводимости по валентной зоне, создаваемой дырками – σ_p :

$$\sigma = \sigma_n + \sigma_p = e\mu_n n + e\mu_p p. \quad (4.7)$$

Экспериментально установлено, что при падении на полупроводник излучения с интенсивностью I проводимость полупроводника изменяется на величину $\delta\sigma(I)$. Так как рассматриваются только слабые световые потоки, то, со-

гласно правилам дифференциального исчисления, без каких-либо дополнительных предположений можно вычислить полный дифференциал выражения (4.7):

$$\delta\sigma(l) = e n \cdot \delta\mu_n + e p \cdot \delta\mu_p + e \mu_n \cdot \delta n + e \mu_p \cdot \delta p, \quad (4.8)$$

где $\delta\mu_n$, $\delta\mu_p$, δn , δp представляют собой полные дифференциалы соответствующих величин. Таким образом, согласно выражению (4.8), освещение полупроводника может изменить его проводимость двумя способами: изменив концентрации свободных носителей в зонах, а также изменив их подвижности. Однако относительное влияние этих двух причин в реальности оказывается различным. Действительно, возникающая при поглощении фотона пара «электрон – дырка» получает некоторый квазиимпульс и энергию ($\hbar\omega - \Delta E_g$), которая идет на кинетическую энергию электрона. Если кинетическая энергия велика, то электрон быстро перемещается в междоузлии кристалла и испытывает множественные столкновения, происходящие чаще, чем у свободного электрона, находящегося внизу зоны проводимости. Поэтому его избыточная энергия быстро растрчивается. В результате через некоторое время, именуемое **временем релаксации энергии** – $\tau_E \sim 10^{-12}$ с и которое много меньше τ_n , τ_p , средняя энергия фотоэлектронов, имевших в момент своего рождения большую кинетическую энергию, принимает значение, соответствующее температуре решетки. В таком случае говорят, что электрон **термолизировался** – его кинетическая энергия стала соответствовать температуре кристалла.

Таким образом, за время своей жизни фотоэлектроны в зоне успевают быстро термолизироваться (прийти в состояние термодинамического равновесия с решеткой), так что все их параметры в течение практически всего времени жизни будут такими же, как и у равновесных носителей. В этом случае подвижности фотоэлектронов равны подвижности темновых (равновесных) электронов. Значит, даже если освещать полупроводник светом с $\hbar\omega \gg \Delta E_g$, то избыток энергии вызывает изменение и концентрации носителей, и изменение их подвижностей, но через очень короткое время $\tau_E \ll \tau_n$, τ_p подвижности фотоноси-

телей становятся такими же, что и у темновых электронов и дырок. Следовательно, с высокой степенью точности можно считать, что в результате поглощения света в полупроводниках изменяются только концентрации свободных носителей заряда: электронов и (или) дырок. Тогда выражение (4.8) для фотопроводимости освещаемого полупроводника с высокой степенью точности можно записать так:

$$\delta\sigma = e(\mu_p \cdot \delta p + \mu_n \cdot \delta n).$$

В дальнейшем будут рассматриваться только такие ситуации.

Будем считать, что происходит однородная генерация пар ($g_n = g_p = g = const$) и что в полупроводнике нет тока. Тогда, умножая кинетические уравнения для концентраций носителей в зонах (см. п. 4.1) на $e\mu_n$, $e\mu_p$ и складывая их, получим:

$$\frac{\partial \delta\sigma}{\partial t} = e(\mu_p + \mu_n)g - \frac{\delta\sigma}{\tau_{фп}}; \quad (4.9)$$

где обозначено:

$$\tau_{фп} = \frac{\mu_n \cdot \delta n + \mu_p \cdot \delta p}{\frac{\mu_n \cdot \delta n}{\tau_n} + \frac{\mu_p \cdot \delta p}{\tau_p}}. \quad (4.10)$$

Из уравнения (4.9) видно, что характеристическое время $\tau_{фп}$ – **время релаксации фотопроводимости** определяет время установления стационарного значения фотопроводимости $\delta\sigma$ при изменении условий освещения. В стационарном состоянии фотопроводимость равна

$$\delta\sigma(t \gg \tau_{фп}) = \delta\sigma_s = e(\mu_n + \mu_p) \cdot g \cdot \tau_{фп}. \quad (4.11)$$

Здесь под $\tau_{фп}$ следует понимать его значение в стационарном состоянии, т.е. при установившихся значениях δp и δn .

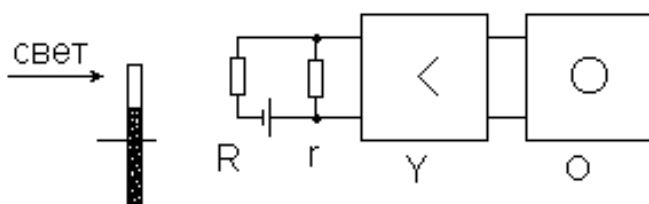


Рис. 21

Для измерения фотопроводимости и ее кинетики разработано много различных методик. Рассмотрим одну из них, блок-схема которой показана на рис. 21. Фотопроводник R , включенный последовательно с источником тока и нагрузочным сопротивлением r , освещается прерывистым светом. Модуляция интенсивности света может быть получена, например, с помощью вращающегося непрозрачного диска с прорезью. Вследствие изменения проводимости полупроводника в цепи возникает переменный ток, а на сопротивлении r появляется падение напряжения, в котором заключена информация об изменении проводимости полупроводника. Это напряжение усиливается широкополосным усилителем Y , а затем визуализируется с помощью осциллографа O . Если величина постоянного напряжения внешнего источника равна V , то напряжение на входе усилителя будет равно:

$$U_{\text{вх}} = V \frac{r}{R^0 + r} \frac{\delta R(I)}{R^0},$$

где R^0 - его величина в отсутствии освещения, $\delta R(I)$ - изменение сопротивления фоторезистора за счет его освещения

$$\delta R(I) = -\frac{\delta \sigma}{\sigma^2} \frac{L}{S} = -R^0 \frac{\delta \sigma}{\sigma^0}.$$

Здесь L , S - длина полупроводникового фоторезистора и его площадь, а σ^0 - темновая проводимость. В итоге напряжение на входе усилителя будет равно:

$$U_{\text{вх}} = -V \left(\frac{r}{R^0 + r} \right) \frac{\delta \sigma(I)}{\sigma^0}.$$

Отсюда видно, что чем больше V , тем больше сигнал на входе усилителя. Более того, данное выражение показывает, что для получения большого сигнала от фотопроводимости полупроводника необходимо выбирать сопротивление нагрузки r более темного сопротивления фоторезистора: $r \gg R^0$. Таким образом, оптимальной схема запитывания фоторезистора является схема в виде источника постоянного тока.

Из выражения (4.11) следует, что чем больше время релаксации фотопроводимости $\tau_{\text{фп}}$, тем больше $\delta \sigma_S$, т.е. тем выше чувствительность фоторезистора

к оптическому излучению. Однако при увеличении $\tau_{фп}$ одновременно с ростом фоточувствительности увеличивается время установления фотопроводимости, т.е. будет больше инерционность фоторезистора и им нельзя будет регистрировать быстрые оптические процессы. С этим противоречием приходится считаться при выборе фоторезисторов для технических применений.

Если $\tau_{фп}$ постоянно во времени, то $\delta\sigma$ изменяется во времени так, как это показано на рис. 22. Если T - длительность светового импульса, за которым следует затемнение фоторезистора такой же продолжительности, то общий период модуляции светового

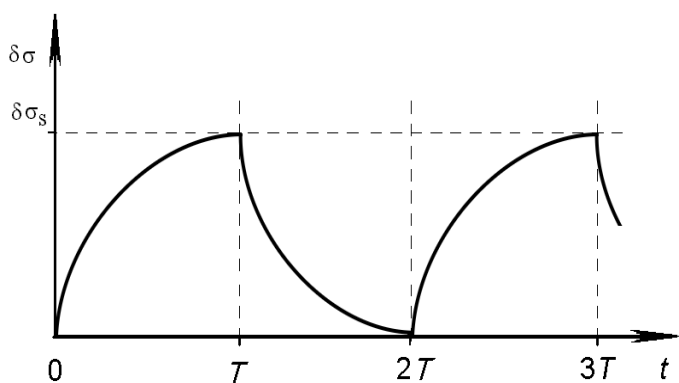


Рис. 22

потока составит $2T$. Будем считать, что импульсы света длинные: $T \gg \tau_{фп}$ и потому за время его действия фотопроводимость достигает своего максимального значения $\delta\sigma_s$. На отрезке времени

$(T; 2T)$ генерация неравновесных носителей заряда отсутствует, а доминирует рекомбинация ранее созданных фотоэлектронов и фотодырок. Потому фотопроводимость уменьшается по экспоненциальному закону:

$$\delta\sigma(t) = \delta\sigma_s \cdot \exp\left(-\frac{t}{\tau_{фп}}\right).$$

Кривые релаксации фотопроводимости могут быть использованы для определения параметров полупроводника. На начальном участке нарастания $\delta\sigma$ в уравнении (4.9) можно пренебречь вторым слагаемым в правой части (рекомбинация еще отсутствует, т.к. она проявится при временах наблюдения сравнимых или больших $\tau_{фп}$). Тогда:

$$\delta\sigma(t) \approx e(\mu_n + \mu_p)gt.$$

Линейное увеличение фотопроводимости со временем освещения характеризуется наклоном, тангенс угла которого равен

$$\operatorname{tg}(\alpha) = e(\mu_n + \mu_p)\eta \cdot \alpha \cdot I_0 / \hbar\omega,$$

где η – квантовый выход, α – коэффициент поглощения. Этот участок роста фотопроводимости не зависит от процессов рекомбинации, а определяется только процессами генерации. Поэтому исследование начальной стадии нарастания фотопроводимости используют для экспериментального определения квантового выхода η .

При выключении света спад фотопроводимости используют для определения времени релаксации фотопроводимости как время, за которое фотопроводимость уменьшается от своего максимального значения $\delta\sigma_s$ в e -раз.

Коэффициент усиления фотопроводимости. Будем считать, что рассматриваемый полупроводниковый образец с омическими контактами на концах имеет длину L . В цепи с нагрузочным сопротивлением R (рис. 23) возникающий фототок при действии электрического поля E будет равен:

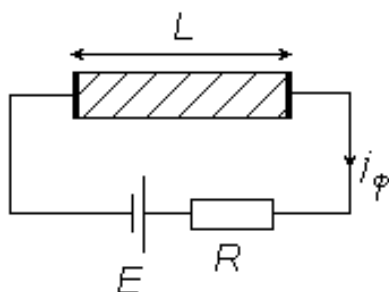


Рис. 23

$$j_\phi = \delta\sigma_s \cdot E,$$

$$j_\phi = (e\mu_n\delta n + e\mu_p\delta p)E.$$

Расписав концентрации неравновесных носителей заряда в виде: $\delta n = g\tau_n$, $\delta p = g\tau_p$, где g – скорость генерации электронов и дырок светом, получим:

$$j_\phi \approx e(\mu_n\tau_n + \mu_p\tau_p) \cdot g \cdot E.$$

Введем в рассмотрение время пролета образца электроном и дыркой, определив их следующим образом

$$t_n = \frac{L}{v_n} = \frac{L}{\mu_n \cdot E}; \quad t_p = \frac{L}{v_p} = \frac{L}{\mu_p \cdot E},$$

где v_n , v_p – скорости электронов и дырок. Отсюда находим:

$$j_{\phi} \approx e \cdot \left(\frac{\tau_n}{t_n} + \frac{\tau_p}{t_p} \right) \cdot L \cdot g.$$

Обозначим: $K = \frac{\tau_n}{t_n} + \frac{\tau_p}{t_p}$. Эта величина показывает, сколько раз за время своей жизни электронно-дырочная пара может пролететь через весь образец, т.е. участвовать в фотопроводимости. Тогда выражение для фототока примет вид:

$$j_{\phi} \approx e \cdot K \cdot L \cdot g.$$

Проанализируем его. Если $K < 1$, то это означает, что электрон и дырка успевают прорекомбинировать, не пролетев полностью образец. Значит, они не дали полный вклад в фотопроводимость. Увеличив прикладываемое электрическое поле, можно уменьшить t_n и t_p , а значит, увеличить j_{ϕ} .

Параметр K называют **коэффициентом усиления фотопроводимости**. С его помощью можно оптимизировать параметры входной цепи схемы регистрации для данного типа фоторезистора.

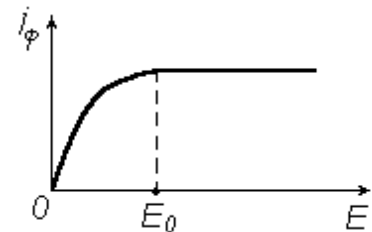


Рис. 24

При решении практических задач необходимо стремиться к тому, чтобы K было максимальным. Зависимость величины фототока от приложенного поля обычно имеет вид, показанный на рис. 24. Величину электрического поля E_0 , при котором достигается предельное значение K , можно найти из условия:

$$t_n \approx \tau_n, \quad (t_p \approx \tau_p)$$

$$E_0 \approx \frac{L}{\mu_n \cdot \tau_n} \quad \text{или} \quad V_0 \approx \frac{L^2}{\mu_n \cdot \tau_n}.$$

Увеличивать напряжение на образце V выше V_0 нецелесообразно, т.к. при $V = V_0$ неравновесные носители заряда обоих знаков дали максимальный вклад в фотопроводимость, и K достигает максимального значения (около двух для биполярной фотопроводимости). В случае примесной фотопроводимости, когда в ее формировании принимает участие только один тип подвижного носителя заряда, а второй носитель заряда (ион донорной или акцепторной примеси) не-

подвижен, ситуация с выбором напряжения смещения сложнее. Уход подвижного носителя заряда раньше времени его жизни нарушит условие электрической нейтральности в объеме полупроводника (там остается ионизированный атом примеси). Для восстановления нарушенного равновесия через отрицательный (или положительный, в зависимости от знака свободного носителя) контакт фоторезистора в полупроводник будет инжектирован с новый электрон (или дырка). И так будет происходить до тех пор, пока неравновесный электрон не прорекомбинирует в объеме полупроводника. Значит, в этом случае коэффициент усиления фотопроводимости может оказаться большой величиной.

4.3. Фотовольтаические эффекты в полупроводниках

Рассмотренная выше фотопроводимость полупроводников имеет положительные и отрицательные качества. К положительным качествам фотопроводимости можно отнести простоту изготовления полупроводникового образца, предназначенного для регистрации фотопроводимости – прямоугольный брусок полупроводника с омическими контактами на противоположных торцах, а также простота интерпретации физических процессов, формирующих явление фотопроводимости. К отрицательным качествам фотопроводимости относятся следующие: необходимость в схеме фотоприемника иметь периодически обновляемый источник постоянного напряжения, что ограничивает область применения данного фотоэффекта (исключаются системы космического базирования и т.п.), протекание постоянного тока через полупроводник приводит к дополнительным шумам (дробовым, контактным), что снижает обнаружительную способность фотоприемника. В дополнение к этому заметим, что схемотехнически регистрация фотопроводимости, как правило, осуществляется путем ее перевода в напряжение – падение напряжения на сопротивлении нагрузки радиотехнического тракта фотоприемника.

Эти обстоятельства побуждают искать фотоэффекты, не нуждающиеся в источниках постоянного напряжения, а сами вырабатывающие напряжения, пропорциональные интенсивности оптического излучения.

При освещении полупроводника изменяется не только его электропроводность, но также возникают электродвижущие силы. Общая причина возникновения фотоэлектродвижущей силы (или фотоэдс), по крайней мере в наиболее важных случаях, одна и та же: пространственное разделение неравновесных носителей заряда за счет их диффузии или за счет действия внутреннего электрического поля полупроводника, которое создается специально или возникает случайным образом. Двигаясь от места их генерации, неравновесные носители создают направленные потоки зарядов, которые продолжаются до тех пор, пока пространственное разделение электронов и дырок не создаст электрическое поле - фотоэдс, препятствующее дальнейшему перемещению неравновесных носителей заряда.

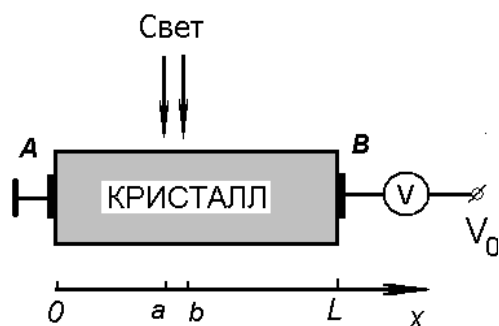


Рис. 25

Несмотря на общую картину происхождения фотовольтаических эффектов, оказывается удобным говорить о разных типах фотоэдс в зависимости от особенностей строения полупроводника и условий проведения регистрации излучения. Остановимся на общем условии, необходимом для возникновения фотоэдс.

Чтобы сделать рассуждения простыми и наглядными, рассмотрим полупроводник в виде прямоугольного параллелепипеда, часть которого « ab » освещается излучением из полосы собственного поглощения. Образец ограничен слева и справа одинаковыми металлическими электродами A и B для измерения фотоэдс (рис. 25). Полупроводник может быть неоднородным, т.е. примеси в нем могут быть распределены неравномерно по длине образца, а отдельные его части могут иметь даже различный тип проводимости (электронный или дырочный). Будем считать, что подвижности электронов и дырок при освещении не меняются. Неравновесные концентрации электронов и дырок δn и δp будем предполагать зависящими только от координаты, отсчитываемой x вдоль образца в направлении регистрации фотоэдс (контактов). Пусть к полупроводнику

приложено постоянное напряжение V_0 .

При освещении в образце появятся токи с плотностями, содержащими дрейфовую и диффузионную компоненты:

$$j_p = \sigma_p E - e D_p \frac{dp}{dx}; \quad j_n = \sigma_n E + e D_n \frac{dn}{dx}.$$

Полная плотность тока будет:

$$j = j_n + j_p = \sigma \cdot \left(E_0 + e \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\sigma} \right),$$

где $\sigma = e(\mu_n n + \mu_p p)$ – полная электропроводность в данном материале. С другой стороны, согласно закону Ома, можно записать:

$$j = \sigma \cdot (E_0 + E^*),$$

где E^* – напряженность поля, возникшая из-за освещения полупроводника. Сравнивая оба выражения для j , найдем

$$E^* = \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_n n + \mu_p p}.$$

Тогда полное фотонапряжение, измеренное на торцах образца A и B , будет

$$V^* = \int_0^L E^* dx = \int_0^L \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_n n + \mu_p p} dx, \quad (4.12)$$

где интегрирование проводится вдоль образца по направлению к контактам.

Из выражения (4.12) следует, что возникающая на торцах полупроводника фотоэдс тем больше, чем значительнее различие в величинах диффузионных потоков электронов и дырок, а также чем меньше темновая проводимость полупроводника. При этом важно, что регистрируется фотонапряжение, поэтому в выражении (4.12) в этот параметр дают вклады градиенты концентраций и равновесных, и неравновесных электронов и дырок.

Чтобы выяснить общие свойства фотоиндуцированного напряжения, исследуем более подробно полученное выражение.

Однородный полупроводник. Предположим, что

$$p = p_0 + \delta p; \quad n = n_0 + \delta n.$$

Поскольку исходно полупроводник считается однородным, то $\frac{dn_0}{dx} = \frac{dp_0}{dx} = 0$.

В случае одинакового градиента неравновесных электронов и дырок

$$D_n \frac{dn}{dx} - D_p \frac{dp}{dx} = \frac{kT}{e} \left[\mu_n \frac{d\delta n}{dx} - \mu_p \frac{d\delta p}{dx} \right] = \frac{kT}{e} [\mu_n - \mu_p] \frac{d\delta p}{dx},$$

поскольку градиенты избыточных концентраций носителей заряда обычно одинаковы. Тогда по выражению (4.12)

$$V^* = \frac{kT}{e} (\mu_n - \mu_p) \int_{\delta p(0)}^{\delta p(L)} \frac{d(\delta p)}{\mu_n n_0 + \mu_p p_0 + (\mu_n + \mu_p) \delta p}.$$

Здесь подынтегральное выражение представляет собой дифференциал функции

$$\ln \left[\frac{\mu_n n_0 + \mu_p p_0 + (\mu_n + \mu_p) \delta p}{\mu_n + \mu_p} \right],$$

и потому, вычислив ее значение в точках *A* и *B*, получим разность двух значений логарифмической функции, отличающихся друг от друга только небольшим значением $\delta p(x)$. Следовательно, в этом случае фотоэдс V^* будет сравнительно малой величиной.

Таким образом, в однородном полупроводнике с одинаковыми градиентами неравновесных электронов и дырок фотонапряжение на освещаемом полупроводнике оказывается малым.

Примесное освещение. Пусть свет создает только основные носители. Будем для определенности рассматривать полупроводник *n*-типа. Для него справедливы следующие неравенства:

$$D_n \frac{dn}{dx} \gg D_p \frac{dp}{dx}, \quad \mu_n n \gg \mu_p p.$$

Поэтому выражение (4.12) дает:

$$V^* = \frac{kT}{e} \int_{\delta n(0)}^{\delta n(L)} \frac{d\delta n}{n_0 + \delta n} = \frac{kT}{e} \int_{1+\delta n(0)/n_0}^{1+\delta n(L)/n_0} \frac{d(1+\delta n/n_0)}{1+(\delta n/n_0)} =$$

$$= \frac{kT}{e} \left[\ln(1+\delta n(L)/n_0) - \ln(1+\delta n(0)/n_0) \right] = \frac{kT}{e} \left[\frac{\delta n(L) - \delta n(0)}{n_0} \right].$$

Здесь в знаменателе стоит большое число: концентрация основных носителей. Следовательно, примесное освещение полупроводника, в результате которого создаются только основные неравновесные носители заряда, создает на концах полупроводника малую по величине фотоэдс.

Если продолжить этот анализ, то можно показать (приводится без полного математического доказательства), что для возникновения значительной по величине фотоэдс необходимо, чтобы полупроводник должен быть неоднородным по уровню легирования вдоль направления регистрации фотонапряжения (это обеспечит большой градиент концентраций) или же (и) свет должен создавать большое количество неосновных носителей заряда. Поясним этот вывод простым примером. Пусть в полупроводнике имеется 1000 основных носителей, например, электронов и 100 неосновных – дырок. Пусть освещение полупроводника создает одинаковое количество тех и других: например, по 10 штук. Освещение изменило число основных носителей на 1%, а неосновных – на 10%. Поэтому и вклад в фотоэдс от изменения числа неосновных носителей следует ожидать больший, чем от основных.

4.4. Квазиуровни Ферми в полупроводнике

Существует достаточно простой и наглядный способ оценки величины фотоэдс, возникающей в полупроводнике при его освещении. Для его рассмотрения обратимся к оценке вероятности заполнения уровней энергии. уровней энергии. В термодинамически равновесном состоянии (например, при отсутствии освещения полупроводника) вычислить концентрации электронов и дырок в зонах можно, если воспользоваться распределением Ферми-Дирака. При этом необходимо знать положение уровня Ферми F (см. п. 1.4). Если концентрация

электронов в полупроводнике (например, за счет его однородного легирования) будет увеличиваться, то уровень Ферми, оставаясь в запрещенной зоне, будет смещаться к дну зоны проводимости (см. рис. 26, а). В случае увеличения концентрации дырок он смещается к потолку валентной зоны. Поэтому уровни энергий, лежащие вблизи уровня Ферми, при изменении его энергетического положения меняют вероятность заполнения электронами. Такой подход к описанию заполнения энергетических состояний в равновесии прост и удобен.

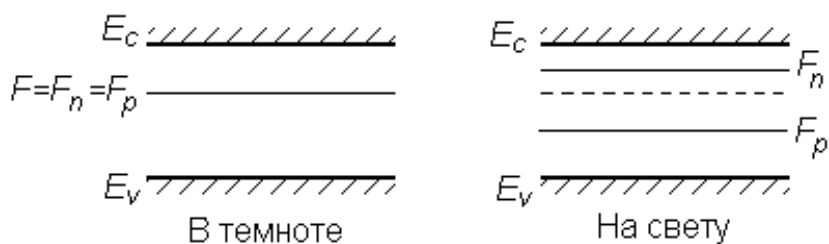


Рис.26

Однако часто необходимо знать концентрации носителей в валентной зоне и зоне проводимости в неравновесных условиях, например, при освещении полупроводника. Попробуем применить для этой цели равновесную функцию распределения Ферми-Дирака. Использовать ее будет возможно, если удастся определить положение уровня Ферми F в неравновесных условиях. При отсутствии освещения уровень Ферми в полупроводнике определяется концентрациями равновесных носителей заряда в зонах разрешенных энергий, т.е. в конечном счете – концентрацией легирующей донорной или акцепторной примеси (см. рис. 26, а). Освещение полупроводника увеличивает концентрацию электронов, и потому в неравновесных условиях уровень Ферми должен сместиться вверх относительно своего равновесного положения. Одновременно с освещением увеличивается и концентрация дырок и тогда, согласно рассуждениями первого абзаца этого параграфа, уровень Ферми должен переместиться вниз (см. рис. 26, б). Но он не может при освещении одновременно двигать вверх и вниз. Следовательно, введение в рассмотрение единичного уровня Ферми в этих условия невозможно. Поэтому распределение электронов и дырок по состояниям в неравновесных условиях нельзя описать равновесной функци-

ей Ферми-Дирака с одним уровнем Ферми.

Однако можно обобщить положения равновесной статистики на неравновесную ситуацию, если вместо одного уровня Ферми формально ввести два квазиуровня Ферми: один - для описания распределения электронов по энергии в неравновесных условиях, а другой – для описания распределения дырок. Тогда вероятность заполнения электроном состояния с энергией E в зоне проводимости можно представить в виде функции Ферми-Дирака:

$$f_n = \left(1 + \exp\left(\frac{E - F_n}{kT}\right) \right)^{-1}.$$

Здесь F_n есть **квазиуровень Ферми для электронов**.

Для вероятности обнаружения дырки на уровне с энергией E в валентной зоне положим

$$f_p = \left(1 + \exp\left(\frac{F_p - E}{kT}\right) \right)^{-1},$$

где по определению F_p есть **квазиуровень Ферми для дырок**. Введение квазиуровней для неравновесных состояний полупроводника допустимо, т.к. созданные светом носители заряда, как уже указывалось выше, через очень короткое время ($\sim 10^{-12}$ с) термолизуются.

Таким образом, появление в зонах неравновесных электронов и дырок можно описать как расщепление первоначального уровня Ферми F на два квазиуровня F_n и F_p , каждый из которых смещается по направлению к своей зоне: F_n – к зоне проводимости, а F_p – к валентной зоне. Нетрудно видеть, что разность квазиуровней Ферми на концах полупроводника определяет величину разности потенциалов на них при освещении:

$$V^* = \frac{F_n - F_p}{e}.$$

Следовательно, разность квазиуровней Ферми для электронов и для дырок есть та фотоэдс, которую можно измерить вольтметром на концах полупроводникового образца при его освещении.

4.5. Фотоэдс в однородных полупроводниках (фотоэдс Дембера)

Согласно выполненного в предыдущем параграфе анализа, фотоэдс в однородных полупроводниках может возникнуть, если градиенты неравновесных концентраций электронов и дырок будут различны. Для выяснения свойств фотоэдс в однородных полупроводниках рассмотрим полупроводник прямоугольной формы, одна из граней которого освещается светом из собственной полосы поглощения (рис. 27). Предполагаем, что интенсивность падающего излучения невелика, так что электропроводность образца

σ при освещении мало отличается от темновой проводности σ_0 . Ввиду большого значения коэффициента поглощения света, он полностью поглотится в тонком приповерхностном слое полупроводника. В результате возникнет большое различие концентраций неравновесных электронов и дырок у освещаемой поверхности и в объеме полупроводника, что вызовет их диффузия в глубь полупроводника. Обычно коэффициенты диффузии, характеризующие скорости разбегания электронов и дырок, различны: электроны движутся в междоузлии, а дырки перемещаются по вакансиям в атомах, что сделать сложнее. В результате этого различия через некоторое время после начала диффузии электроны уйдут вглубь дальше, чем дырки. Это приведет к разделению зарядов в пространстве, а значит, и к возникновению внутреннего электрического поля E^* . Оно направлено таким образом, чтобы воспрепятствовать диффузии, остановить ее. Стационарное значение возникающего внутреннего поля достигается, когда создаваемый им дрейфовый ток электронов и дырок уравнивает суммарный диффузионный ток, вызванный неравномерным распределением фотоносителей по глубине полупроводника. На рис. 27 показаны знаки заряда поверхностей образца для случая $D_n > D_p$ и ориентация поля E^* . Электроны,

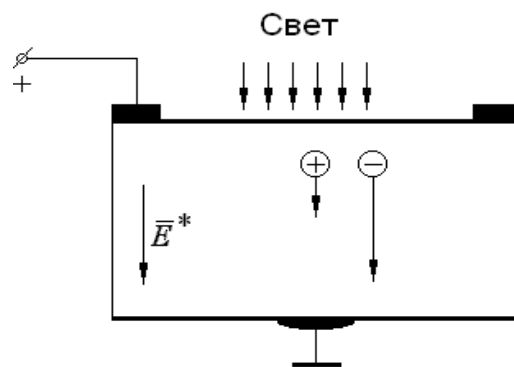


Рис.27

электроны движутся в междоузлии, а дырки перемещаются по вакансиям в атомах, что сделать сложнее. В результате этого различия через некоторое время после начала диффузии электроны уйдут вглубь дальше, чем дырки. Это приведет к разделению зарядов в пространстве, а значит, и к возникновению внутреннего электрического поля E^* . Оно направлено таким образом, чтобы воспрепятствовать диффузии, остановить ее. Стационарное значение возникающего внутреннего поля достигается, когда создаваемый им дрейфовый ток электронов и дырок уравнивает суммарный диффузионный ток, вызванный неравномерным распределением фотоносителей по глубине полупроводника. На рис. 27 показаны знаки заряда поверхностей образца для случая $D_n > D_p$ и ориентация поля E^* . Электроны,

диффундирующие быстрее дырок, заряжают нижнюю грань отрицательно, а на освещенной грани появляется положительный заряд.

Для того чтобы оценить величину возникающего за счет поля E^* фотоэдс на концах полупроводникового образца, рассмотрим полупроводник n -типа проводимости, в котором концентрации неравновесных (созданных светом) электронов и дырок равны ($\delta n = \delta p$). Использование выражения (4.12) дает формулу для подсчета величины фотоэдс в этих условиях:

$$V^* = \frac{e}{\sigma_0} (D_n - D_p) \int_{\delta p(0)}^{\delta p(d)} dp = \frac{e}{\sigma_0} (D_n - D_p) \cdot [\delta p(d) - \delta p(0)],$$

где d – толщина образца в направлении падения света, $\delta p(0)$ и $\delta n(d)$ – концентрации фотоносителей у освещенной и неосвещенной поверхностей соответственно. Здесь достаточно интегрировать по толщине пластинки, т.к. в остальной части цепи $E^* = 0$. Если толщина образца в направлении падающего света d превышает длину диффузии электронов и дырок, то $\delta p(d) \ll \delta p(0)$ и тогда фотоэдс будет

$$V^* \approx \frac{e}{\sigma_0} (D_p - D_n) \cdot \delta p(0). \quad (4.13)$$

Фотоэдс, возникающая вследствие различия коэффициентов диффузии неравновесных носителей заряда, называется **фотоэдс Дембера**. Согласно выражению (4.13), ее величина не зависит от геометрических размеров полупроводникового образца (если $d > L_n, L_p$) и определяется только различием в коэффициентах диффузии электронов и дырок, темновым сопротивлением полупроводника и уровнем его освещения. Ее величина обычно невелика: при комнатной освещенности образцов германия с темновой проводимостью $\sigma_0 \approx 1 \text{ Ом}^{-1}\text{см}^{-1}$ экспериментально измеряется фотоэдс $V^* \sim 10^{-3}$ Вольт. Фотоэдс Дембера можно использовать для регистрации оптического излучения из полосы собственного поглощения полупроводника, определения разности коэффициентов диффузии электронов и дырок и т.д.

4.6. Фотоэдс в неоднородных полупроводниках (объемная фотоэдс)

Объемная, или распределенная, фотоэдс возникает в неоднородных полупроводниках, в которых градиент удельного сопротивления отличен от нуля. Этот тип фотоэдс рассмотрим на примере тонкой пластинки, на поверхность которой падает свет в виде узкой полоски. Полоска света удалена от торцов образца по крайней мере на несколько диффузионных длин, так что фотоэлектроны и фотодырки рекомбинируют, не достигая торцов. Экспериментально доказано, что в такой ситуации между концами образца появляется напряжение, которое тем больше, чем больше градиент удельного сопротивления в месте освещения. Описываемая ситуация изображена на рис.28.

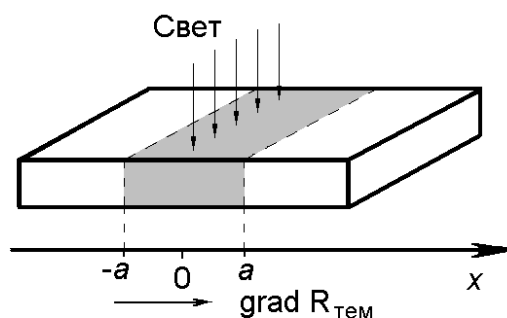


Рис.28

Рассмотрим физический механизм формирования фотоэдс на неоднородности легирования полупроводника n - типа. Неоднородность легирования означает, что сопротивление образца изменяется с координатой вследствие изменения концентрации темновых носителей. Это означает, что энергетические зоны будут наклонены относительно постоянного уровня Ферми, т.е. в образце будет существовать внутреннее электрическое поле. Действительно, в равновесных условиях уровень Ферми постоянен во всех частях рассматриваемой системы. Но если уровень легирования изменяется слева направо на рисунке, то это означает, что расстояние между уровнем Ферми F и дном зоны проводимости

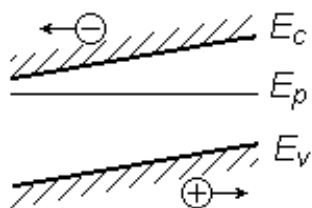


Рис. 29

должно увеличиваться при движении вдоль полупроводника слева направо. Это означает, что разрешенные зоны должны быть наклонены так, как это показано на рис.29 или что в полупроводнике в области неоднородности существует встроенное электрическое поле. При отсутствии освещения это поле не вызывает появления

тока, т.к. обусловленный ими дрейф в точности компенсируется током диффузии. Однако при освещении это равновесие нарушается и возникают потоки фотоэлектронов и фотодырок, направленные в разные стороны. Если на рисунке градиент сопротивления направлен слева направо, то под действием внутреннего электрического поля электроны будут двигаться к левому концу образца и заряжать его отрицательно, а дырки – к правому концу, создавая на нем положительный заряд (рис. 29). Образующийся в результате этого разделения заряда диполь изменит величину внутреннего поля на некоторую величину, что можно обнаружить на торцах полупроводника как дополнительное напряжение, величина которого определяется уровнем освещения.

Для того чтобы вычислить величину возникающей фотоэдс и определить характер ее зависимости от параметров полупроводника будем рассматривать полупроводник n -типа, генерация электронно-дырочных пар в котором происходит в пределах тонкого слоя толщиной $2a$, которая предполагается много меньше длин диффузии носителей заряда L_{diff} . Освещение считаем слабым ($\delta p/n_0 \ll 1$) и постоянным в пределах полосы освещения, причем при освещении $\mu_n n \gg \mu_p p$.

Тогда по формуле (4.12) имеем:

$$V^* \approx \int_0^L \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_n n} dx = V_{01} + V_{02}.$$

Первое слагаемое здесь

$$V_{01} = \frac{D_n}{\mu_n} \int \frac{\delta n(L)}{\delta n(0)} \frac{dn}{n} = 0,$$

т.к. подынтегральная функция есть полный дифференциал логарифмической функции. При вычислении второго слагаемого будем считать $n \approx n_0$, а также учтем то обстоятельство, что в темноте $p = p_0(x)$ и $V_{02} = 0$. Поэтому:

$$V^* \approx -\frac{D_p L}{\mu_n n_0(x)} \frac{d(\delta p)}{dx} dx.$$

Введем в это выражение темновое удельное сопротивление ρ_0 и плотность тока диффузии избыточных дырок j_p :

$$\rho_0(x) = \frac{1}{e\mu_n n_0(x)}; \quad j_p(x) = -eD_p \frac{d(\delta p)}{dx}.$$

Тогда выражение для определения объемной фотоздс можно переписать в виде:

$$V_0 = \int_{-\infty}^{-a} j_p(x) \rho_0(x) dx + \int_a^{\infty} j_p(x) \rho_0(x) dx.$$

Здесь совершен переход от интегрирования по конечному интервалу к интегрированию в бесконечных пределах, т.к. $j_p(x)$ быстро затухает с удалением от освещенного участка. Т.к. $\delta p(x) \sim \exp(-x/L_p)$, то по такому же закону изменяется и $j_p(x)$. Поэтому можно записать:

$$\begin{aligned} x > a; \quad j_p &= j_p(a) \cdot \exp\left(-\frac{x-a}{L_p}\right); \\ x < -a; \quad j_p &= -j_p(-a) \cdot \exp\left(-\frac{x+a}{L_p}\right), \end{aligned}$$

причем из симметрии задачи ясно, что $j_p(-a) = j_p(a)$. Тогда

$$V_0 \approx -j_p(a) \int_{-\infty}^{-a} \rho_0(x) \cdot \exp\left(\frac{x+a}{L_p}\right) dx + j_p(a) \int_a^{\infty} \rho_0(x) \cdot \exp\left(-\frac{x-a}{L_p}\right) dx.$$

Заменяя в первом интеграле x на $-x$, получим:

$$V_0 \approx j_p(a) \int_a^{\infty} \exp\left(-\frac{x-a}{L_p}\right) \cdot [\rho_0(x) - \rho_0(-x)] dx.$$

Неизвестную функцию $\rho_0(x)$ в пределах узкого освещенного слоя можно аппроксимировать прямой линией:

$$\rho_0(x) - \rho_0(-x) = 2 \left[\frac{\partial \rho_0}{\partial x} \right]_0 \cdot x.$$

Тогда возникающая за счет освещения объемная фотоэдс будет равна:

$$V_0 \approx 2j_p(a) \frac{d\rho_0}{dx} \int_a^\infty \exp\left(-\frac{x-a}{L_p}\right) x \cdot dx = 2j_p(a) \frac{d\rho_0}{dx} (L_p a + L_p^2) \approx 2j_p(a) \frac{d\rho_0}{dx} L_p^2, \quad a \ll L_p.$$

И, наконец, значение тока на границе освещенного слоя $j_p(a)$ можно найти из условия баланса для фотодырок. Если N – полное число электронно-дырочных пар, сгенерированных светом во всем слое $2a$ в одну секунду, то

$$2j_p(a)S = eN,$$

где S – площадь сечения образца. Поэтому окончательно имеем следующее выражение для объемной фотоэдс:

$$V^* \approx \frac{e}{S} \cdot N \cdot L_p^2 \cdot \left(\frac{d\rho_0}{dx}\right)_0.$$

Здесь $(d\rho_0/dx)_0$ – градиент темнового сопротивления образца в месте освещения.

Для дырочного полупроводника мы получили бы такую же формулу, но с заменой диффузионной длины дырок на диффузионную длину для электронов.

По порядку величины V^* для $(d\rho_0/dx) \approx 10$ Ом и $L_p \approx 0,1$ см составляет единицы милливольт. Объемную фотоэдс можно использовать для исследования неоднородностей удельного сопротивления полупроводниковых пластин. Этот способ значительно проще, чем способ, основанный на пропускании тока по полупроводнику и исследовании металлическим зондом картины распределения по поверхности электростатического потенциала. Он позволяет судить о степени неоднородности $\rho_0(x)$ не внося изменений в полупроводниковый образец, поскольку оптический метод по своей сути является бесконтактным.

4.7. Электронно-дырочные переходы

Из выражения для величины фотовольтаического эффекта (4.12) следует, что она тем больше, чем меньше проводимость полупроводника, т.е. чем более высокоомен полупроводник. Наибольшим сопротивлением обладает полупроводник с концентрациями носителей заряда, равными собственной концентра-

ции $n = p = n_i$. Однако изготовить такой полупроводниковый образец технологически сложно. Поэтому целесообразно использовать для регистрации оптического излучения образец, в котором имеется отдельная область собственной проводимости. Такую область можно получить, изготовив полупроводниковый $p-n$ -переход. Как известно, он образуется при соединении двух полупроводников, обладающих разными типами проводимости: электронной (n -тип) и дырочной (p -тип).

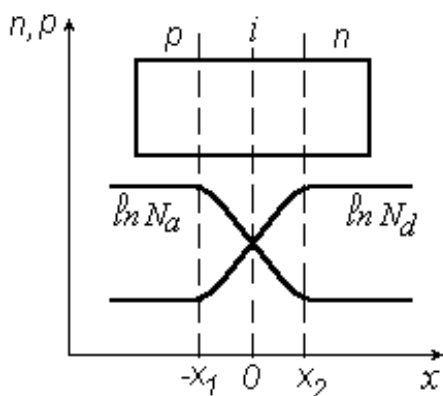


Рис. 30

Чтобы избежать неконтролируемого влияния поверхностей на свойства $p-n$ -перехода, его изготавливают из одного монокристалла легированием торцов образца донорной и акцепторной примесями, а затем вжиганием примесей в объем высокотемпературной обработкой. Рис. 30 поясняет сказанное. Здесь в левой части кристалла создана дырочная проводимость с концентрацией основных носителей (дырок) $p \approx N_a - N_d$, а в правой части – электронная ($n \approx N_a - N_d$). Между ними образуется переходной слой, в котором концентрация примесей быстро изменяется. В некоторой тонкой области этого слоя доноры и акцепторы компенсируют друг друга ($N_a \approx N_d$) и потому в ней имеет место собственная проводимость (i -проводимость).

Требуемое распределение доноров и акцепторов осуществляется с помощью специальных технологических приемов. Распределение концентраций электронов и дырок в $p-n$ -переходе показано на рис. 31 сплошными линиями. Причина формирования именно такого хода зависимостей концентраций равновесных носителей кроется в большом различии концентраций подвижных носителей заряда в n - и p -областях. В результате этого электроны из n -области устремляются туда, где их мало, т.е. в p -область, а дырки из p -области будут диффундировать в n -область. Эти два встречных потока будут продолжаться до тех пор, пока оставляемые без подвижного заряда положительные ионы в n -области и отрицательные в p -области не создадут электрическое поле

– \bar{E}_k (см. рис. 31), которое затормозит процесс диффузии равновесных носителей. Это электрическое поле направлено от n - к p -области.

Одним из важных параметров p - n -перехода является высота его потенциального барьера – φ_0 . Чтобы ее оценить, вспомним, что в равновесном состоянии положение уровня Ферми в n - и p -областях должно быть одинаковым (см. рис. 32). Это позволяет записать следующее равенство

$$e \cdot \varphi_0 + (F_p - E_V) = F_n - E_V.$$

Уровень Ферми в n - области перехода относительно середины запрещенной зоны E_i определяется концентрацией донорной примеси в ней:

$$F_n - E_i = kT \cdot \ln(N_d / n_i).$$

Положение уровня Ферми в p -области перехода определяется концентрацией акцепторной примеси:

$$F_p - E_i = -kT \cdot \ln(N_a / n_i).$$

Поэтому будем иметь:

$$e \cdot \varphi_0 = kT \cdot \left\{ \ln\left(\frac{N_d}{n_i}\right) + \ln\left(\frac{N_a}{n_i}\right) \right\} \quad \text{или}$$

$$\varphi_0 = \frac{kT}{e} \cdot \ln\left(\frac{N_a N_d}{n_i^2}\right).$$

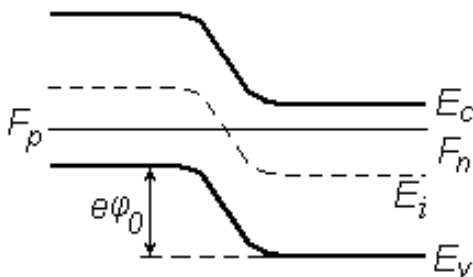


Рис. 32

Такова высота потенциального барьера между n - и p -областью в отсутствие освещения, электрического поля и других факторов, нарушающих состояние термодинамического равновесия в полупроводнике. Она оказывается тем выше, чем выше уровень легирования областей и ниже собственная концентрация полупроводника. Величина барьера p - n -перехода

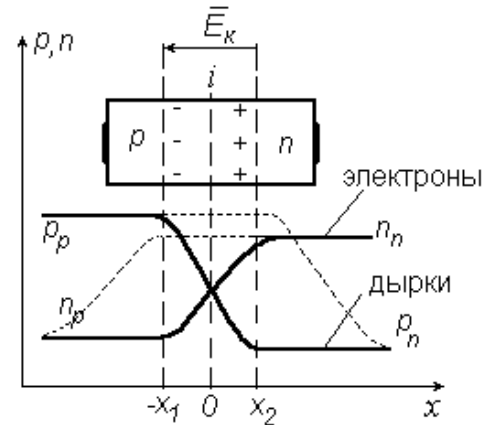


Рис. 31

не превышает ширину запрещенной зоны полупроводника, т.к. в этих условиях уровень Ферми занимает предельные положения в запрещенной зоне: в n -области совпадает с дном зоны проводимости, а в p -области – с потолком валентной зоны.

Важной характеристикой p - n -перехода является его вольт-амперная характеристика (ВАХ). Ее аналитическое выражение таково (без вывода):

$$j(V) = j_s \left(\exp\left(\frac{eV}{kT}\right) - 1 \right).$$

Данная характеристика дает экспоненциальный рост тока при приложении положительных напряжений («+» на p -области) и постоянство тока в области отрицательных напряжений: $j(V) = -j_s$, где ток j_s называется **обратным током** или **током насыщения**. Его величина определяется технологией изготовления p - n перехода, качеством исходного полупроводникового материала.

Пусть теперь к переходу приложено внешнее напряжение так, что на p -области находится «+» источника (рис. 32). Это приложенное напряжение частично компенсирует поле \bar{E}_k . В результате равновесие между диффузионным потоком носителей и током проводимости поля \bar{E}_k будет нарушено в пользу первого. В результате дырки из p -области устремятся в n -область, где они станут неосновными носителями заряда. Так как $p_p \gg p_n$, то эти дырки в n -области будут рекомбинировать с электронами. Однако вследствие конечного значения времени жизни дырок τ_p , рекомбинация произойдет не сразу, и поэтому в некоторой области за пределами перехода концентрация дырок будет больше p_n . По этим же причинам увеличится концентрация электронов в n -области, так как дополнительные электроны войдут из электрода для компенсации объемного заряда пришедших дырок: основные носители в n -области – электроны будут переходить в p -область, становясь там неосновными носителями, и постепенно рекомбинировать с основными носителями – дырками. Поэтому, слева от перехода концентрация электронов увеличивается, а также увеличивается концентрация дырок, которые войдут из левого электрода для компенсации объемного

заряда электронов. Это явление получило название инжекции неосновных носителей. Распределение концентраций неосновных носителей при инжекции показано на рис. 31 штриховыми линиями. В результате приложения напряжения V указанной полярности ток через переход будет расти по экспоненте с ростом величины V .

4.8. Барьерная фотоэдс

Обратимся к вопросу о влиянии освещения на свойства полупроводников. При освещении $p-n$ - перехода высота барьера понижается на величину, пропорциональную уровню освещения. Изменение высоты барьера на границе раздела двух типов проводимости называют **барьерной фотоэдс**. Она возникает в том случае, когда свет генерирует носители заряда вблизи $p-n$ -перехода. Следует различать две принципиально различные схемы включения освещаемого перехода. В первой из них $p-n$ - переход замкнут на внешнюю нагрузку без внешнего источника напряжения и является преобразователем световой энергии в электрическую (рис. 33, *a*). Такое включение называют **фотопентильный режим** работы.

Другая схема во внешней цепи содержит источник напряжения, включенный так, чтобы на $p-n$ -переходе было запирающее напряжение (рис. 33, *b*) – это так называемый **фотодиодный режим**. В схеме *b* существующий в цепи ток сильно изменяется с освещением. Поэтому изменяется падение напряжения на сопротивлении R . При правильном выборе напряжения источника и внешнего сопротивления величина сигнала в фотодиодном режиме может быть сделана больше, чем в фотопентильном. Величина вентильной фотоэдс на несколько порядков больше, чем эдс Дембера и объемная фотоэдс. Физиче-

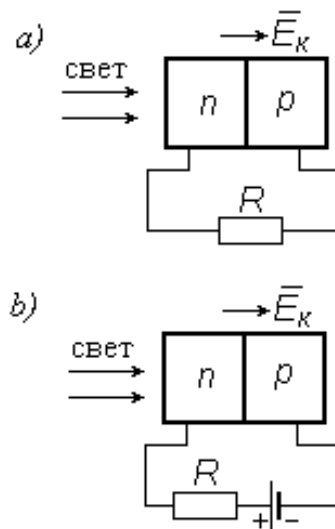


Рис. 33

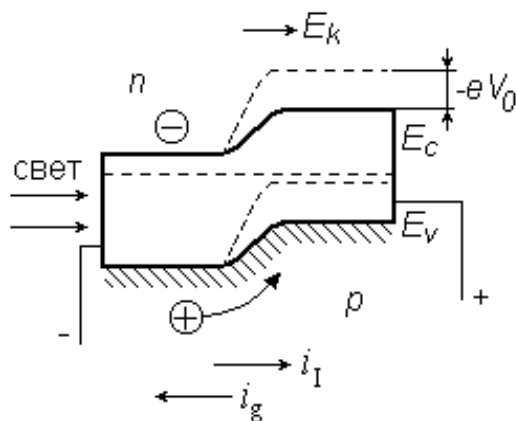


Рис. 34

соединении R . При правильном выборе напряжения источника и внешнего сопротивления величина сигнала в фотодиодном режиме может быть сделана больше, чем в фотопентильном. Величина вентильной фотоэдс на несколько порядков больше, чем эдс Дембера и объемная фотоэдс. Физиче-

ская причина ее появления заключается в том, что электрическое поле p - n – перехода разделяет фотодырки и фотоэлектроны, подошедшие к переходу. Поясним это, предположив, что сильно поглощаемое излучение падает на одну из граней кристалла, параллельную плоскости p - n -перехода. Для определенности будем считать, что освещается n -область, излучение поглощается в ней. Толщина этого слоя выбирается меньше длины диффузии неосновных носителей в нем (дырок). На концах прибора измеряется эдс холостого хода (см. рис. 34).

Фотоэлектроны и фотодырки в n - слое будут диффундировать в глубь слоя, и некоторая их доля, не успев прорекомбинировать, достигнет p - n - перехода. Однако для основных носителей (электронов) в p - n -переходе существует потенциальный барьер, и поэтому они не пройдут в p -область. Для неосновных носителей (дырок) потенциального барьера нет, и потому все достигшие переход дырки перенесутся полем перехода в p -область. При этом они создадут фототок $j_{\phi m}$. Если g – скорость световой генерации электронно-дырочных пар в n -области, а β – их доля, дошедшая до p - n - перехода, то $j_{\phi m} = eg\beta$. Вследствие появления фототока $j_{\phi m}$ p -область будет заряжаться положительно, а n -область – отрицательно, и между n - и p - областями появится разность потенциалов – барьерная фотоэдс V^* . Снижение высоты барьера между n - и p - областями перехода на величину фотоэдс согласно ВАХ приведет к росту прямого темнового тока, созданного основными носителями – j_{np} :

$$j_{np} = j_s \left[\exp\left(\frac{eV^*}{kT}\right) - 1 \right].$$

Он направлен обратно току $j_{\phi m}$. Его стационарное значение может быть найдено из условия равновесия, при котором полный ток через переход отсутствует: $j_{\phi m} = j_{np}$. Тогда, используя выражение для ВАХ p - n -перехода при его освещении, получим:

$$j_{\phi m} = j_s \cdot \left[\exp\left(\frac{eV^*}{kT}\right) - 1 \right],$$

где j_s – ток насыщения (обратный ток) p - n -перехода. Отсюда можно найти величину вентильной фотоэдс:

$$V^* = \frac{kT}{e} \cdot \ln \left[1 + \frac{j_{\phi m}}{j_s} \right]. \quad (4.18)$$

Как следует из данного выражения, вентильная фотоэдс тем больше, чем меньше обратный ток перехода и выше уровень освещенности. Следовательно, для получения высокочувствительного фотодиода необходимо сильное легирование n - и p -областей p - n перехода. Согласно выражению (4.18), при малых уровнях освещения барьерная фотоэдс равна

$$V^* = \frac{kT}{e} \cdot \ln \left[1 + \frac{j_{\phi m}}{j_s} \right] \approx \frac{kT}{e} \cdot \left(\frac{j_{\phi m}}{j_s} \right),$$

т.е. она прямо пропорциональна интенсивности падающего излучения:

$$j_{\phi m} = e\beta g = e\beta \alpha l_0 (1-R)/\hbar \omega.$$

Однако рост барьерной фотоэдс с увеличением уровня освещения не беспределен. Поскольку V^* направлено обратно к φ_0 – высоте барьера, то она не может быть больше высоты барьера. Следовательно, величина фотоэдс не может превышать контактную разность потенциалов между n - и p -областями, а ее максимальная величина сопоставима с шириной запрещенной зоны ΔE_g :

$$V_{max}^* \approx \frac{\Delta E_g}{e}.$$

Так, для кремния $V_{max}^* \approx 1$ В, а для германия $V_{max}^* \approx 0.6$ В при быстродействии, определяемом временем жизни неравновесных носителей заряда в p - n -переходе. В зависимости от физико-химического состава полупроводникового материала оно колеблется в диапазоне от 10^{-3} до 10^{-6} с.

В настоящее время p - n -переход – один из наиболее широко используемых фотодатчиков, т.к. он обладает высокой чувствительностью и приемлемым быстродействием, не требует посторонних источников напряжения (как фоторезистор), технологически совместим со схемой обработки сигналов, т.е. фотопри-

емник на основе барьерной фотоэдс можно в одном технологическом цикле изготовить со схемой усиления фотоэдс, детектирования и т.д. на одном полупроводниковом кристалле.

4.8. Применение фотоэлектрических явлений в оптоэлектронике

Фотоэлектрические свойства полупроводников широко используются в твердотельной электронике для создания фоточувствительных датчиков – приборов, реагирующих на освещение. К таковым относят фотодиоды различного назначения, фоторезисторы, приборы, сочетающие фотодиоды и излучатели света – оптронные пары, используемые для развязки схем и т.д. Важными параметрами фотодатчиков являются следующие: полоса пропускания, фоточувствительность, размеры фоточувствительной площадки, уровень шумового напряжения в полосе рабочих частот.

Фоточувствительность представляет собой величину выходного сигнала фотодатчика при освещении светом единичной мощности. Чем выше этот параметр, тем меньшие уровни освещения может зарегистрировать датчик. Типичным значением фоточувствительности для современных фотодиодов, используемых в режиме генераторов тока, является величина ($10^{-5} \div 10^{-6}$) В/А.

Полоса пропускания фотодатчика определяется областью частот модуляции принимаемого излучения, для которых уровень сигнала на выходе датчика уменьшается от максимального не более чем в $\sqrt{2}$. У кремниевых фотодиодов данный параметр может достигать ($10^6 \div 10^7$) Гц.

Уровень шумового напряжения в полосе рабочих частот определяется дробовой шум обратного тока перехода и в режиме генератора тока обычно составляет величину ($10^{-5} \div 10^{-6}$) А/Гц^{1/2}.

5. ЭМИССИЯ ИЗЛУЧЕНИЯ ИЗ ТВЕРДЫХ ТЕЛ

5.1. Описание излучательных процессов в полупроводниках

Термин «эмиссия излучения» означает испускание света, например, твердым телом в результате какого-либо воздействия. Излучение, связанное с рекомбинацией электронов и дырок в полупроводниках, известно с начала XX века. По своей физической сущности излучательная рекомбинация представляет собой процесс, обратный процессу фотовозбуждения, подробно рассмотренному в главе 3 данного учебного пособия. В простейшем случае эмиссия излучения возникает, когда электроны из зоны проводимости рекомбинируют с дырками, находящимися в валентной зоне. При этом рекомбинация может сопровождаться излучением высвобождающегося избытка энергии в виде кванта света (фотона), когда атом «выстреливает» частичку света, либо кванта тепла (фонона), передаваемого кристаллической решетке. В последнем случае рекомбинация заканчивается «вздрагиванием» атома, на валентную оболочку которого сел электрон из зоны проводимости. Излучение фотона создает эмиссию излучения из полупроводника.

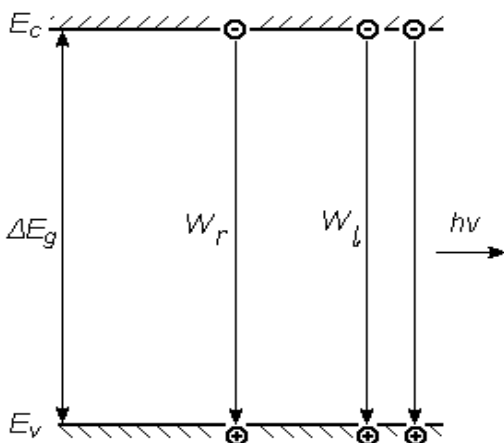


Рис.35

Аналогично может происходить излучательная рекомбинация через примесные уровни. В этом случае энергия эмитируемых фотонов будет меньше ширины запрещенной зоны и определяется энергетическим положением этого уровня в запрещенной зоне полупроводника.

Одним из основных условий наблюдения эмиссии излучения является преобладание процесса рекомбинации свободных носителей заряда над их генерацией (она может происходить за счет поглощения приповерхностными атомами фотонов, испущенных атомами объема полупроводника). Поэтому преобладание рекомбинации над генерацией может иметь место только в случае, если в зонах име-

ются неравновесные носители заряда, т.е. создана неравновесная ситуация. Для того чтобы рекомбинация неравновесных носителей заряда шла преимущественно с испусканием фотона, а не фонона, необходимо, чтобы в данном полупроводнике (или, в общем случае, твердом теле) вероятность излучательной рекомбинации $W_l = 1/\tau_l$ была бы больше вероятности безызлучательной – $W_r = 1/\tau_r$ (см. рис. 35). Поэтому эффективность испускания светового излучения полупроводником зависит от соотношения времен жизни неравновесных носителей, оканчивающихся излучательным τ_l или безызлучательным τ_r переходом электрона в валентную зону.

Для численной оценки эффективности излучательной способности твердого тела вводят два параметра. Первый из них называется **внутренней квантовой эффективностью** η . Он определяет число генерируемых фотонов на прорекомбинировавшую пару «электрон-дырка», характеризует способность атомов твердого тела (полупроводника) излучать фотоны при рекомбинации электронно-дырочных пар и определяется следующим образом:

$$\eta = \frac{\frac{1}{\tau_l}}{\frac{1}{\tau_l} + \frac{1}{\tau_r}} = \frac{\tau_r}{\tau_r + \tau_l}. \quad 0 \leq \eta \leq 1.$$

Данный параметр показывает долю переходов с генерацией оптического излучения в общем числе рекомбинационных процессов неравновесных носителей заряда в объеме полупроводника.

Второй параметр, называемый **внешней квантовой эффективностью** и обозначаемый как ξ , характеризует долю сгенерированных в объеме полупроводника фотонов, вышедших из полупроводника через его поверхность. Действительно, для того чтобы наблюдать эмиссию излучения из полупроводника, необходимо, чтобы сгенерированные фотоны вышли из полупроводника – ведь распространяясь в кристалле на пути к излучающей поверхности фотоны могут быть поглощены другими атомами, находящимися в основном энергетическом состоянии. Физически этот процесс поглощения происходит так, как это описа-

но в главе 3, по закону Бугера-Ламберта, который для данной ситуации будет записан в виде:

$$I(0) = (1 - R) \int_0^L I(x) \exp\left(-\frac{x}{\ell_{\phi}}\right) dx,$$

где интегрирование вызвано необходимостью суммирования интенсивностей от всех атомов с длины кристалла. Здесь $I(0)$ – интенсивность вышедшего из полупроводника излучения; $I(x)$ – интенсивность излучения, сгенерированного атомами, находящимися на удалении x от излучающей поверхности в слое толщиной dx ; R – коэффициент отражения излучения от границы раздела «полупроводник – внешняя среда», L – толщина полупроводника в направлении на излучающую поверхность; ℓ_{ϕ} – длина свободного пробега фотона в полупроводнике, в которой учтены все возможные механизмы поглощения света, изложенные в главе 3. Определение внешней квантовой эффективности в упрощенной форме таково:

$$\xi = \frac{\ell_{\phi}}{\ell_{\phi} + L}. \quad 0 < \xi \leq 1$$

Следовательно, условие достижения высокого выхода светового излучения заключается в том, чтобы сгенерированные фотоны могли выйти из полупроводника, не будучи поглощенными другими атомами на пути от излучающего атома до поверхности. Для этого необходимо, чтобы длина свободного пробега фотона была больше толщины образца: $\ell_{\phi} \gg L$.

Способы возбуждения полупроводника. Эмиссия излучения может быть вызвана различными воздействиями на твердое тело. Неравновесные носители заряда можно инжектировать в однородный полупроводник путем оптического возбуждения или бомбардировкой заряженными частицами. Возникающее при этом излучение называется **фотолюминесценцией** или **катодолюминесценцией** соответственно. Возможно также создание неравновесных носителей за счет ударной ионизации в сильном электрическом поле. Этот способ возбуждения приводит к эмиссии, называемой **электролюминесценцией**.

Наиболее же эффективный метод создания больших концентраций неравновесных носителей заряда в полупроводниках заключается в использовании *p-n*-перехода, включенного в прямом направлении. Именно на этом принципе основано действие полупроводниковых лазеров и светодиодов.

Другим видом эмиссии в полупроводниках является тепловое излучение. Этот излучательный процесс имеет место, когда между образцом полупроводника и окружающей средой существует разность температур. Тогда полупроводник будет излучать во всем спектре частот в соответствии со спектром черного тела. В результате установления динамического равновесия поглощения фотонов данной частоты и их излучения полупроводник будет излучать свет. Ввиду конечных размеров образца часть эмитируемого излучения выходит из кристалла и не уравнивается эквивалентным потоком, входящим в образец, если он не находится в тепловом равновесии с окружающей средой. Поэтому свойства полупроводника отличаются от свойств абсолютно черного тела, и изменение наблюдаемой излучательной способности определяется изменением коэффициента поглощения. Это утверждение справедливо независимо от природы процесса поглощения. Такую люминесценцию называют **термолюминесценцией**.

Флюоресценция – это люминесценция, которая происходит только во время действия источника возбуждения неравновесных пар. **Фосфоресценция** – это люминесценция, которая продолжается после действия возбуждения.

Определение люминесценции в общем случае таково. **Люминесценция** есть избыток излучения твердого тела над тепловым излучением, если это избыточное излучение обладает длительностью, значительно превышающей период испускаемых световых волн. Данное определение отделяет люминесценцию от теплового излучения в равновесных условиях и дает право отнести ее к разделу неравновесных явлений. Это определение выделяет люминесценцию из всех других видов неравновесного излучения, таких как отражение и рассеяние света, излучение Вавилова-Черенкова и т.д. В отличие от перечисленных излучений, в случае люминесценции акты поглощения и испускания света разделе-

ны промежуточными процессами, что и приводит к продолжительному существованию свечения после прекращения возбуждения.

5.2. Спектр излучения твердого тела

Кроме двух параметров, с помощью которых можно численно охарактеризовать эффективность эмиссии излучения твердым телом, вышедшее излучение характеризуют зависимостями его параметров от независимых переменных. Наиболее используемая из них – зависимость интенсивности свечения от энергии изученного фотона или его длины волны, называемая **спектром излучения** твердого тела. Спектры излучения твердого тела отличаются от соответствующих спектров изолированных атомов примеси двумя основными чертами. Во-первых, излучение твердого тела обычно представляет собой набор широких полос спектра, и, во-вторых, спектр люминесценции, как правило, смещен в сторону длинных волн относительно спектра поглощения того же полупроводника. Таким образом, излучаемая энергия всегда меньше поглощаемой. Разницу между этими энергиями называют **сдвигом Франка-Кордона**. Эти особенности обязаны своим происхождением взаимодействию центра свечения с полем кристаллической решетки.

При взаимодействии центра с ближайшим окружением его потенциальная энергия складывается из энергий кулоновского взаимодействия, обменного взаимодействия и ван-дер-ваальсовского притяжения. В результате действия этих сил атомные уровни основного (1) и возбужденного (2) состояний будут смещены друг относительно друга (см. рис. 36). Таким образом, сдвиг Франка-Кордона – это смещение, обусловленное сдвигом атомов в решетке после оптического возбуждения. Физически наличие сдвига Франка-Кордона можно представлять себе так: поглощение света производится атомами, а его излучение – ионами полупроводника. Ввиду их различного зарядового состояния различны и их диаметры: потеря валентный электрон, у иона размер электронного облака оказывается меньше, чем у атома того же вещества. Поскольку различны размеры атомов и ионов одного и того же вещества, то различны расстояния меж-

ду атомами и ионами. Поэтому ширина запрещенной зоны полупроводника, составленного из атомов и ионов, различна, т.к. она зависит от постоянной решетки (см. рис. 5 данного учебного пособия). Напомним, что ширина запрещенной зоны тем меньше, чем меньше степень перекрытия электронных оболочек соседних атомов или ионов. Если же ширины запрещенных зон для этих двух случаев различны, то и энергии фотонов, способных вызывать переходы «зона – зона», также будут различны.

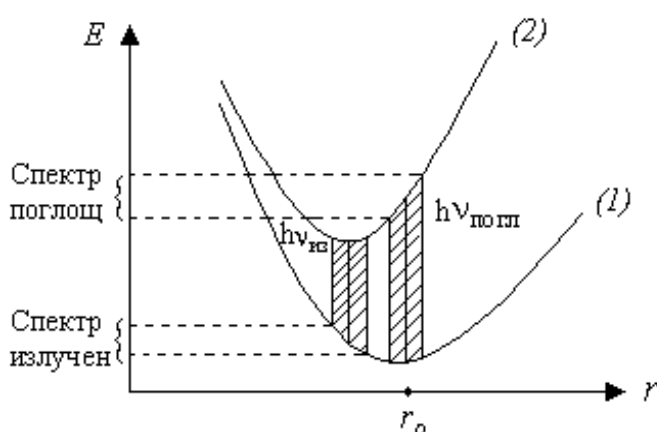


Рис. 36

Те же механизмы взаимодействия света и твердого тела, которые рассмотрены при изучении спектра поглощения оптического излучения, имеют место при рассмотрении спектра излучения. Это прямые и непрямые переходы «зона – зона», переходы с участием экситонов, примесных состояний и т.д. Однако

наиболее важными являются переходы «зона – зона».

Зависимость эмиссии от уровня легирования. Проанализируем зависимость времени излучательной рекомбинации в полупроводнике от его параметров. Для примера рассмотрим прямые переходы «зона проводимости – валентная зона» в полупроводнике, выведенном из равновесного состояния. Пусть n_i – собственная концентрация носителей. Из физических соображений следует, что скорость излучательной рекомбинации R_i в полупроводнике с произвольной концентрацией электронов пропорциональна концентрации электронов n и концентрации пустых состояний (дырок) p : чем больше электронов, тем выше скорость (есть чему рекомбинировать); но скорость рекомбинации растет и с увеличением концентрации дырок (есть с чем рекомбинировать). Чтобы упростить дальнейшие физические рассуждения и математические выкладки, будем считать, что избытки концентраций электронов и дырок над их равновесными значениями малы. Для того чтобы математически в первом приближении

описать прямо пропорциональную зависимость скорости излучательной рекомбинации R_l от n и p , необходимо взять произведение полных концентраций носителей заряда. Таким образом, в процесс излучательной рекомбинации включены не только неравновесные носители заряда Δn и Δp , но и равновесные – n_0 и p_0 . Тем самым учтено то обстоятельство, что рекомбинация неравновесной дырки возможна с равновесным электроном, а неравновесного электрона – с равновесной дыркой.

Пронормируем указанное произведение концентраций на квадрат собственной концентрации носителей заряда. Это позволит упростить физическую интерпретацию коэффициента пропорциональности в обсуждаемой зависимости. Тогда скорость излучательной рекомбинации можно записать в виде:

$$R_l = \frac{n \cdot p}{n_i^2} R_i ,$$

где R_i – скорость излучательной рекомбинации в собственном полупроводнике. Поскольку, по определению, скорость любой рекомбинации обратно пропорциональна времени жизни неравновесных носителей заряда, то из данного выражения следует, что время жизни носителей в зоне проводимости, заканчивающееся излучением фотона, уменьшается при возрастании концентраций и электронов, и дырок в соответствующих зонах.

Выразим концентрации носителей через отклонения от равновесных значений: $n = n_0 + \Delta n$, $p = p_0 + \Delta p$. Тогда скорость рекомбинации также может быть представлена как сумма скоростей в равновесии (будем ее обозначать с индексом «0») и ее отклонения (отметим ее символом Δ):

$$R_l^0 + \Delta R_l = \frac{(n_0 + \Delta n)(p_0 + \Delta p)}{n_i^2} R_i = \frac{n_0 p_0 + (n_0 + p_0) \Delta n}{n_i^2} R_i ,$$

т.к. считаем $\Delta n = \Delta p$. Поэтому имеем:

$$R_l^0 = \frac{n_0 p_0}{n_i^2} R_i , \quad \Delta R_l = \frac{(n_0 + p_0) \Delta n}{n_i^2} R_i .$$

По определению, скорость рекомбинации есть $\frac{\Delta n}{\tau_I}$ (см. гл. 3), где τ_I – время жизни электрона в зоне проводимости, заканчивающееся излучением фотона:

$$\Delta R_I = \frac{\Delta n}{\tau_I} = \frac{(n_0 + p_0)\Delta n}{n_i^2} R_i.$$

Поэтому можно записать выражение для определения времени излучательной рекомбинации как

$$\tau_I = \frac{1}{R_i} \cdot \frac{n_i^2}{n_0 + p_0} = \frac{1}{R_i} \frac{n_0 \cdot p_0}{n_0 + p_0}.$$

Для собственного полупроводника $n_0 = p_0 = n_i$, $\tau_I = \tau_I^i$ и тогда

$$\tau_I^{(i)} = \frac{n_i}{2R_i}.$$

Время излучательной рекомбинации в собственном полупроводнике τ_I^i является известным параметром для любого полупроводникового материала. Поэтому для времени излучательной рекомбинации в полупроводнике произвольного уровня легирования имеем:

$$\frac{\tau_I}{\tau_I^{(i)}} = \frac{2n_i}{n_0 + p_0}. \quad (5.1)$$

Численные значения времени излучательной рекомбинации в различных собственных (беспримесных) полупроводниках сильно отличаются друг от друга. Эксперименты показывают, что, например, для Si $\tau_I^{(i)} \approx 4,6$ часа, для Ge $\tau_I^{(i)} \approx 0,6$ с, InSb $\tau_I^{(i)} \approx 6 \cdot 10^{-7}$ с.

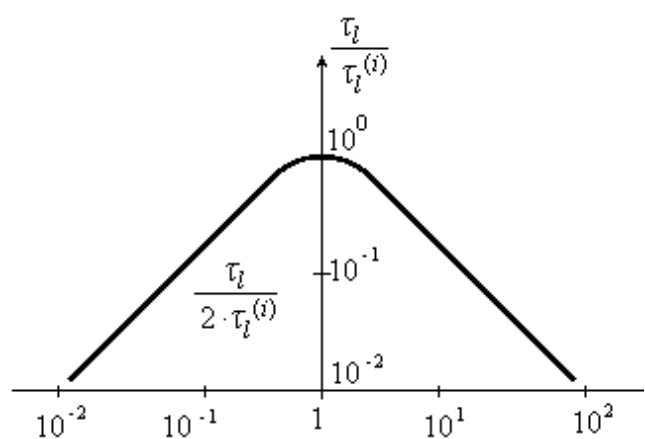


Рис. 37

Согласно выражению (5.1), зависимость времени излучательной рекомбинации «зона – зона» от концентрации носителей заряда в полупроводнике имеет вид, оказанный на рис. 37. По оси абсцисс отложено отношение концентрации

электронов n_0 к концентрации дырок p_0 . Поэтому правая половина рисунка соответствует полупроводнику n -типа проводимости, а левая – p -типу. Из рисунка видно, что с увеличением концентрации основных носителей время излучательной рекомбинации быстро уменьшается по величине, а скорость излучательной рекомбинации растет. Такое поведение времени излучательной рекомбинации объясняется увеличением вероятности рекомбинации неравновесного неосновного носителя за счет роста концентрации основных носителей заряда. Дополнительные расчеты показывают, что уменьшение ширины запрещенной зоны также снижает время излучательной рекомбинации.

Таким образом, если для изготовления источника излучения использовать объемный монокристалл, то для повышения интенсивности его эмиссии излучения необходимо брать его с высокой концентрацией основных носителей и малой шириной запрещенной зоны.

5.3. Спонтанное и вынужденное и излучение атома

Для облегчения понимания условий возникновения оптического излучения в твердых телах рассмотрим процесс поглощения и излучения света в атоме.

Из всей совокупности квантовых состояний, в которых может находиться атом, для простоты выберем два состояния с энергиями E_1 и E_2 . Это основное (невозбужденное) состояние и первое возбужденное (см. рис. 38). Пусть на такой атом падает излучение, имеющее спектральную плотность энергии U_ν . Не- возбужденный атом, поглощая энергию излучения $h\nu$, перейдет в возбужденное состояние. При этом переходы с поглощением кванта света совершаются тем чаще,

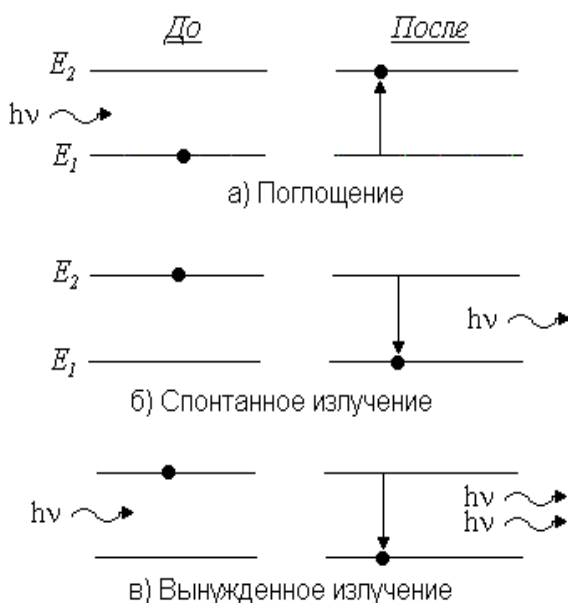


Рис.38

чем больше величина U_ν . Если N_1 – число атомов в невозбужденном состоянии, а $U_\nu B_{12}$ – вероятность перехода с поглощением, то число переходов с поглощением в единицу времени в единице объема равно:

$$\partial N_{12} = B_{12} U_\nu N_1$$

Спонтанное излучение. Атом, находясь в возбужденном состоянии в результате поглощения кванта света, через некоторый промежуток времени из возбужденного состояния перейдет в нижнее состояние, отдавая избыток энергии $E_2 - E_1$ в виде кванта $h\nu$. Переход с излучением может произойти произвольно, независимо от внешних воздействий. Такой переход называется **спонтанным**, а создаваемое при этом излучение носит название **спонтанного излучения**. Спонтанные переходы являются случайными, их вероятность обозначим через A_{21} . Если N_2 – число атомов в возбужденном состоянии, то число спонтанных переходов в единицу времени:

$$dN'_{21} = A_{21} N_2 .$$

Следовательно, уменьшение количества возбужденных атомов в результате спонтанных переходов за время dt составит величину:

$$-dN_2 = A_{21} N_2 dt .$$

Интегрируя это уравнение, получим закон убывания числа возбужденных атомов со временем при спонтанных переходах:

$$N_2 = N_{20} \exp(-A_{21} \cdot t) ,$$

где N_{20} – число возбужденных атомов в момент $t = 0$.

Так как за каждый спонтанный переход излучается энергия $E_2 - E_1 = h\nu$, то в результате dN'_{21} переходов в единицу времени выделяется энергия в количестве: $h\nu \cdot dN'_{21} = h\nu \cdot A_{21} \cdot N_2$. Тогда интенсивность спонтанного излучения (энергия, испускаемая в единицу времени) равна:

$$I = h\nu \cdot A_{21} \cdot N_{20} \cdot e^{-A_{21} \cdot t} = I_0 \cdot e^{-A_{21} \cdot t} ,$$

где $I_0 = h\nu \cdot A_{21} \cdot N_{20}$. Так как $A_{21}N_2dt$ есть число атомов, совершивших за время dt переход 2–1, то оно определяет количество атомов, «проживших» в возбужденном состоянии в течение времени t . Поэтому продолжительность пребывания в возбужденном состоянии всех атомов, испытавших спонтанные переходы за время от 0 до ∞ , составит величину:

$$T = \int_0^{\infty} t \cdot A_{21} \cdot N_2 \cdot dt = \int_0^{\infty} N_{20} \cdot A_{21} \cdot t \cdot e^{-A_{21} \cdot t} dt = \frac{N_{20}}{A_{21}} .$$

Здесь использовано выражение для определения среднего значения по известной плотности вероятности случайного процесса. Из приведенного вычисления следует, что средняя продолжительность жизни атома в возбужденном состоянии равна

$$\tau = \frac{T}{N_{20}} = \frac{1}{A_{21}} .$$

Поэтому релаксацию интенсивности свечения можно переписать в виде:

$$I = I_0 \cdot e^{-\frac{t}{\tau}} .$$

Таким образом, после прекращения возбуждения интенсивность спонтанного излучения во времени уменьшается по экспоненциальному закону с постоянной времени $\tau = 1/A_{21}$. Поскольку спонтанные переходы являются случайными, то спонтанное излучение имеет статистический характер, оно некогерентно, т.е. генерируются фотоны, распространяющиеся в произвольных направлениях и имеющие произвольные фазы, хотя частоты излучаемых квантов света близки. Это обусловлено тем, что в силу случайного характера акты спонтанного излучения происходят независимо друг от друга.

Вынужденное излучение. Однако возбужденный атом под действием излучения может перейти в невозбужденное состояние. Такой переход называется вынужденным, а излучение носит название **вынужденного, стимулированного** или **индуцированного**. Индуцированные переходы будут совершаться тем чаще, чем больше плотность энергии возбуждающего света U_ν . Поэтому веро-

ятность этих переходов в единицу времени будет $U_\nu B_{21}$. Здесь B_{21} – вероятность индуцированного перехода атома с уровня 2 на уровень 1 при единичной интенсивности падающего излучения. Если N_2 – число атомов в возбужденном состоянии, то число индуцированных переходов в единицу времени в единице объема составит:

$$dN_2'' = U_\nu B_{21} N_2 .$$

Следовательно, внешнее излучение с частотой ν вызывает не только поглощение, но и добавочное, индуцированное излучение.

Таким образом, среднее число переходов с излучением кванта света в единицу времени в единице объема определяется спонтанными и вынужденными переходами и равно:

$$dN_{21} = dN_{21}' + dN_{21}'' = (A_{21} + U_\nu B_{21}) N_2 .$$

В стационарном состоянии на основании принципа детального равновесия числа прямых и обратных переходов в единицу времени равны. Следовательно, $dN_{21} = dN_{12}$ или $(A_{21} + U_\nu B_{21}) N_2 = U_\nu B_{12} N_1$. Здесь B_{12} – вероятность перехода атома с уровня 1 на уровень 2 под действием излучения единичной интенсивности. Отсюда следует, что:

$$\frac{N_2}{N_1} = \frac{U_\nu B_{12}}{A_{21} + U_\nu B_{21}} . \quad (5.2)$$

Поскольку равновесие имеет статистический характер, то отношение N_2/N_1 можно выразить при помощи распределения Больцмана, согласно которому для системы, имеющей состояния с энергией E_1, E_2, \dots, E_m , вероятность пребывания в состоянии E_m равна

$$W_m = C g_m \cdot e^{-\frac{E_m}{kT}} ,$$

где g_m – фактор вырождения данного уровня, показывающий, сколько атомом находятся в данном состоянии. Поэтому можно записать:

$$\frac{N_2}{N_1} = \frac{g_2 \cdot e^{-\frac{E_2}{kT}}}{g_1 \cdot e^{-\frac{E_1}{kT}}} = \frac{g_2}{g_1} \cdot e^{-\frac{(E_2-E_1)}{kT}}. \quad (5.3)$$

Из сравнения равенств (5.2) и (5.3) следует, что:

$$\frac{U_\nu B_{12}}{A_{21} + U_\nu B_{21}} = \frac{g_2}{g_1} \cdot e^{-\frac{(E_2-E_1)}{kT}}.$$

С учетом того, что $E_2 - E_1 = h\nu$, получим:

$$U_\nu = \frac{g_2 A_{21}}{g_1 B_{12} \cdot e^{kT} - g_2 B_{21}}. \quad (5.4)$$

Для определения коэффициентов A и B используем предельное условие. При $T \rightarrow \infty$ плотность энергии U_ν должна стремиться к бесконечности (при сильном нагревании твердое тело излучает больше, чем при слабом нагреве), а это значит, что знаменатель в (5.4) стремится к нулю или $g_1 B_{12} = g_2 B_{21}$. С учетом этого выражение (6.4) переписывается в виде:

$$U_\nu = \frac{g_2 A_{21}}{g_1 B_{12}} \cdot \frac{1}{e^{\frac{h\nu}{kT}} - 1}.$$

При $\frac{h\nu}{kT} \ll 1$ в первом приближении имеем:

$$e^{\frac{h\nu}{kT}} \approx 1 + \frac{h\nu}{kT},$$

и тогда:

$$U_\nu = \frac{g_2 A_{21}}{g_1 B_{12}} \cdot \frac{kT}{h\nu}. \quad (5.5)$$

С другой стороны, при условии $\frac{h\nu}{kT} \ll 1$ справедлива формула Релея-Джинса, по которой:

$$U_\nu = \frac{8\pi\nu^2}{c^3} \cdot kT.$$

Из ее сопоставления с формулой (5.5) найдем:

$$\frac{g_2 A_{21}}{g_1 B_{12}} = \frac{8\pi\nu^2}{c^3} h\nu;$$

откуда следует: $A_{21} = \frac{8\pi\nu^2}{c^3} h\nu \cdot \frac{g_1}{g_2} B_{12}$. Окончательно имеем:

$$A_{21} = \frac{8\pi\nu^2}{c^3} h\nu \cdot B_{21}. \quad (5.6)$$

Таким образом, установлено, что все вероятности переходов атома между уровнями его энергии взаимно связаны между собой, что указывает на единый механизм атома по излучению или поглощению кванта света. При вынужденном переходе генерируется фотон, имеющий ту же частоту, направление и фазу, что и фотон, вызвавший этот переход. Поэтому генерируемое излучение называют **когерентным**.

5.4. Стимулированное излучение твердых тел

Рассмотрим условия, при которых будет наблюдаться стимулированное излучение. Пусть мы имеем пучок параллельного монохроматического света интенсивностью I . При прохождении слоя вещества толщиной Δx интенсивность света уменьшится на величину $-\Delta I = \alpha \cdot I \cdot \Delta x$. Это уменьшение интенсивности происходит из-за переходов с поглощением энергии в количестве $U_\nu B_{12} \cdot h\nu \cdot N_1 \Delta x$ и индуцированных переходов с излучением энергии в количестве $U_\nu B_{21} \cdot h\nu \cdot N_2 \Delta x$. В результате имеем:

$$-\Delta I = \alpha \cdot I \cdot \Delta x = (B_{12} N_1 - B_{21} N_2) U_\nu \cdot h\nu \cdot \Delta x.$$

Если учесть, что плотность излучения U_ν в параллельном пучке сечением 1 см^2 , распространяющемся со скоростью c , равна: $U_\nu = \frac{I}{c}$, то коэффициент поглощения приобретает вид:

$$\alpha = \frac{h\nu}{c} (B_{12} N_1 - B_{21} N_2) = \frac{h\nu}{c} \left(1 - \frac{g_1 N_2}{g_2 N_1}\right) B_{12} N_1.$$

Учитывая выражение (6.3), получим ($E_2 - E_1 = h\nu$):

$$\alpha = B_{12}N_1 \left(1 - e^{-\frac{h\nu}{kT}} \right) \frac{h\nu}{c}, \quad (5.7)$$

Поскольку показатель экспоненты всегда отрицателен, то круглые скобки дают положительную величину. Это означает, что в случае термодинамического равновесия при любой конечной температуре суммарный коэффициент поглощения положителен, т.е. имеет место поглощение света, а индуцированные переходы лишь уменьшают его. Причина этого заключается в том, что в равновесном состоянии число невозбужденных атомов всегда больше числа возбужденных ($N_1 > N_2$). Поэтому количество переходов с уровня E_1 будет больше, чем с E_2 . Как же добиться отрицательного значения коэффициента поглощения α ? Из (5.7) следует, что это возможно, если показатель экспоненты положителен, а для этого необходимо, чтобы абсолютное значение температуры кристалла было отрицательным: $T < 0$. Этого можно достичь следующим образом.

Из распределения Больцмана (5.3) следует, что температура тела связана с заполнением уровней энергии основного E_1 и возбужденного состояния E_2 :

$$T = - \frac{E_2 - E_1}{k \cdot \ln \left(\frac{g_1 N_2}{g_2 N_1} \right)}.$$

Поскольку $E_2 > E_1$, а в равновесии заполнение верхнего уровня всегда меньше, чем нижнего $g_1 N_2 < g_2 N_1$, то абсолютное значение температуры твердого тела T всегда положительно. Однако если выполняется условие:

$$g_1 N_2 > g_2 N_1, \text{ то } T < 0. \quad (5.8)$$

Таким образом, согласно выражению (5.8), для того чтобы создать среду с отрицательным поглощением (с усилительными свойствами), необходимо получить неравновесное состояние вещества, при котором число возбужденных атомов было бы больше числа невозбужденных. Заселенность энергетических состояний атомов, удовлетворяющих условию (5.8), носит название **инверсной**

заселенности, т.е. обратной заселенности по отношению к равновесному распределению атомов по уровням энергии.

5.5. Полупроводниковые лазеры

Как следует из названия, данный тип лазеров изготавливается из полупроводникового материала и использует фундаментальные свойства полупроводников и полупроводниковых приборов. Этот тип лазеров характеризуется наибольшим коэффициентом полезного действия, малыми габаритами, простотой накачки и модуляции излучения, а также наилучшей совместимостью с сопутствующими передатчику микроэлектронными устройствами, поскольку те и другие имеют полупроводниковую основу и потому создаются в едином технологическом процессе.

Первые полупроводниковые лазеры были созданы в 1962 г. В настоящее время применяются в основном полупроводниковые лазеры с инжекционной накачкой. Для этой цели используются люминесцирующие полупроводники с прямыми переходами, обмен носителями зарядов в которых между минимумом зоны проводимости и максимумом валентной зоны характеризуется высокой вероятностью. Иными словами, излучательная рекомбинация свободных электронов и дырок в таких материалах протекает наиболее интенсивно.

Излучение в полупроводниках, в отличие от лазеров на твердых телах с примесями активных ионов, характеризуется тем, что в процессе взаимодействия принимают участие не два энергетических уровня, а две зоны с конечным количеством уровней: зона проводимости и валентная зона.

Для работы лазера в нем необходимо создать инверсию населенности. Наиболее просто инверсию населенности получить в p - n -переходе, сместив его внешним напряжением в прямом направлении. В этом случае устраняется потенциальный барьер, существовавший на границе раздела двух областей полупроводника с разными типами проводимости. В результате этого электроны из n -области (где их больше, чем дырок) диффундируют в p -область (где электронов меньше, чем дырок), а дырки – в обратном направлении. Эти два

мощных потока равновесных носителей заряда встречаются в i -области, где и наблюдается наиболее интенсивная излучательная их рекомбинация, сопровождаемая выходом квантов оптического излучения с энергией, примерно равной ΔE_g (рис. 39, *a* и *б*). Излучающая область обычно очень тонкая – около 1 мкм.

Диаграмма энергетических зон инжекционного лазера показана на рис. 39, *б*. У сильно легированного p - n -перехода в отсутствие внешнего постоянного напряжения уровень Ферми лежит в зоне проводимости и валентной зоне.

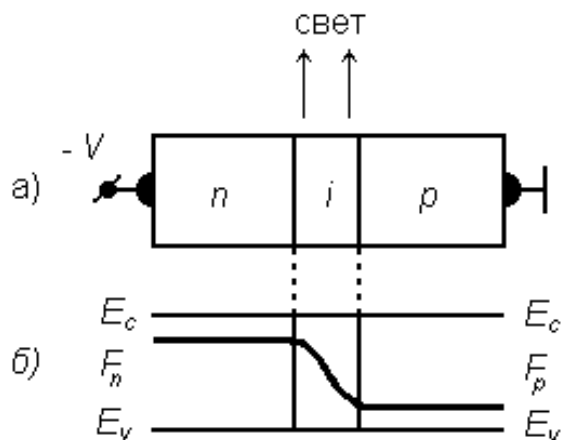


Рис. 39

Устранив внешним напряжением внутреннее контактное поле, мы распрямляем энергетические зоны в области собственно p - n -перехода. В этой ситуации заполнение электронами зоны проводимости слева превосходит ее заполнение справа от перехода. Обратная картина имеет место для дырок. Тогда электроны могут беспрепятственно перемещаться из n -области вправо и занимать состояния валентной зоны с испусканием фотонов. Подобные переходы могут также происходить между краем зоны проводимости и акцепторным уровнем (или зоной) вблизи потолка валентной зоны. Кроме того, дырки могут переходить в n -область и рекомбинировать с электронами с испусканием фотона. При этом в тип преобладающего процесса определяется концентрацией примеси, значениями подвижностей и времен жизни неравновесных носителей заряда. Максимальное увеличение интенсивности излучения будет происходить при многократном прохождении световой волны по кристаллу в результате отражения излучения от торцов кристалла. Многократно отражаться будут лишь те лучи, которые падают на торцевые поверхности кристалла перпендикулярно. В силу этого обстоятельства излучение, вызванное индуцированными переходами, сосредоточено в пределах малого телесного угла $\Delta\Omega$. В результате возникает узконаправленный поток излучения большой интенсивности, так как яркость его

возрастает в отношении $\frac{4\pi}{\Delta\Omega}$. Спонтанное же излучение распределяется равномерно во всех направлениях в пределах телесного угла 4π . Поэтому оно уходит из кристалла через боковые поверхности, равномерно окрашивая кристалл в цвет генерируемого излучения: красный, зеленый, желтый и т.д. Индуцированное излучение не является когерентным. Оно распространяется в том же

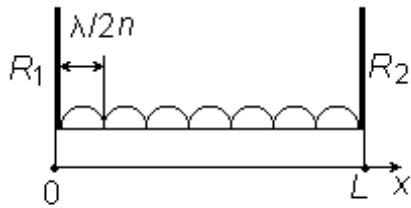


Рис. 40

направлении и имеет ту же фазу, что и первичное. Поэтому лазер можно рассматривать как оптический резонатор, внутри которого возникают стоячие волны. Их число должно быть кратным длине кристалла L (см. рис.40):

$$m \frac{\lambda}{2n} = L,$$

где m – любое целое число, кроме нуля; λ/n – длина волны излучения в полупроводнике, n – его показатель преломления.

Динамика излучаемого света при росте тока через переход такова. При малых токах, текущих через p - n -переход, возникает спонтанное излучение, распространяющееся во всех направлениях. Возрастание тока увеличивает скорость рекомбинации, что соответственно увеличивает плотность фотонов. Они в свою очередь индуцируют переходы «зона – зона», а значит, и излучательную рекомбинацию. Поскольку на максимум спектрального распределения спонтанного излучения приходится больше фотонов, на этой частоте вызывается больше переходов, чем на краях спектра. Это приводит к значительному сужению спектра излучения и, следовательно, усилению интенсивности излучения в его максимуме. Если интенсивность пика излучения растет сверхлинейно с уровнем возбуждения, то процесс называется «сверхсвечением». Возникающие при сверхсвечении фотоны, так же как и при спонтанном излучении, распространяются во всех направлениях с произвольными фазами. Для создания лазера необходимо, чтобы усиление, как минимум, равнялось потерям, а излучение было когерентным. Когерентность достигается помещением источника в

резонатор, обеспечивающий селективное усиление электромагнитных волн определенной частоты, на которой в резонаторе образуют стоячие волны.

Типичный характер зависимости интенсивности излучения лазера от уровня возбуждения представлен на рис. 41. Как видно из рисунка, при протекании тока через переход I ниже некоторого порогового уровня $I_{\text{пор}}$ излучение имеет спонтанный характер. При увеличении тока через переход интенсивность свечения также растет. Но когда достигнут порог генерации, интенсивность излучения резко возрастает и из p - n -перехода выходит когерентное излучение $P_{\text{свет}}$.

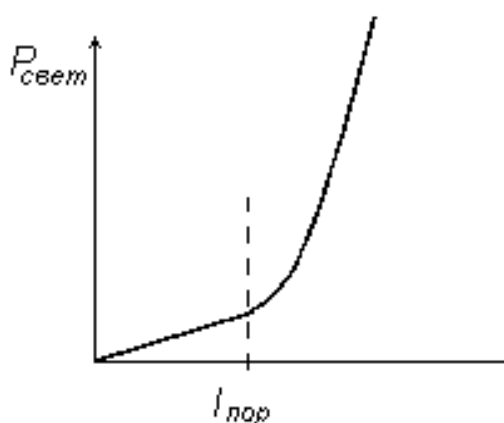


Рис. 41

лучения резко возрастает и из p - n -перехода выходит когерентное излучение $P_{\text{свет}}$.

Чтобы носители заряда получили от электрического поля достаточную энергию, необходимо выполнение неравенства:

$$V_{p-n} > \frac{\Delta E_g}{q},$$

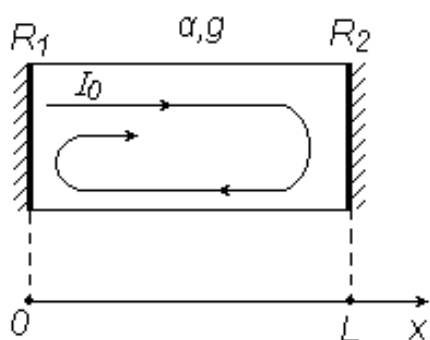
где V_{p-n} – прямое напряжение на p - n -переходе.

Если на такую систему не положены дополнительные условия, то спектр ее генерации будет достаточно широким. Поэтому для получения большего коэффициента усиления применяют оптический резонатор: его зеркалами служат противоположные плоско параллельные грани полупроводникового кристалла, обработанные так, чтобы обеспечить требуемые значения коэффициентов отражения.

Определим условия возникновения светового возбуждения в кристалле, помещенном в резонатор. Пусть резонатор имеет длину L , а R_1 , R_2 – коэффициенты отражения на гранях. Если α – коэффициент поглощения лазерного излучения посторонним механизмом, то излучение, имевшее на входе в резонатор интенсивность I_0 , прошедшее через резонатор и вернувшееся в ту же точку, будет иметь интенсивность (рис.42)

$$I = I_0 \cdot R_1 R_2 \cdot \exp(2gL - 2\alpha L).$$

Следовательно, условие возникновения лазерного эффекта, предполагающее нарастание интенсивности излучения после однократного прохождения активного вещества, заключается в превышении коэффициента усиления над коэффициентом потерь, будет



$$g > \alpha - \frac{1}{2L} \cdot \ln(R_1 \cdot R_2).$$

Рис. 42

Излучательная рекомбинация – это малоэффективный процесс, в котором только малая часть энергии, затраченной на создание инверсии населенности, выделяется в виде лазерного излучения. Остальная часть идет на нагревание кристалла и тем самым заставляет использовать импульсный режим накачки.

Выше обсуждался лазерный эффект на переходах «зона – зона». Однако он может быть реализован на процессах с участием примесных центров либо экситонов.

5.6. Светодиоды

Инжекция неосновных носителей в $p-n$ - переходе, включенном в прямом направлении, может привести к высокой эффективности преобразования электронов и дырок в фотоны даже в отсутствии лазерного эффекта. Процесс спонтанного излучения может происходить в полупроводниках как с прямыми, так и непрямыми переходами посредством одного из механизмов излучательной рекомбинации, о которых говорили ранее. Хотя внутренняя квантовая эффективность может быть близка к единице,

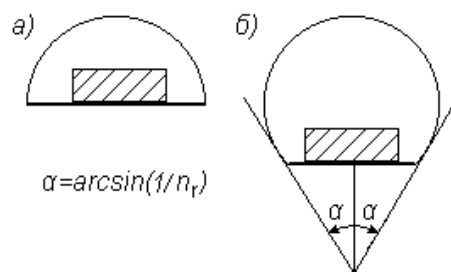


Рис. 43

внешняя эффективность у источников некогерентного излучения обычно мала: из-за высокого показателя преломления полупроводника большая часть света на поверхности испытывает полное внутреннее отражение, после чего возвра-

щается в полупроводник и там поглощается. Через излучающую поверхность проходит лишь та небольшая часть излучения, которая падает на поверхность излучения под углом, меньше, чем $\alpha = \arcsin(1/n_r)$, что, как правило, составляет угол меньше 20° . Действительно: $\sin\beta/\sin\alpha = n_r$, где β - угол падения света из вакуума на твердое тело, а α - угол преломления. Если луч света падает в полупроводнике на границу раздела под большим углом, то свет будет выходить в вакуум под углом $\beta = 90^\circ$. Это значит, что свет не выйдет в вакуум. Предельный угол падения находим из закона Снелиуса, положив $\sin\beta = 1$, получим выражение для предельного угла: $\alpha = \arcsin(1/n_r)$.

Для расширения апертуры излучения применяют различные технические решения. Так, использование полусферы (рис. 43, а) или сферы Вейерштрасса (рис. 43, б) позволяет до нескольких раз увеличить внешнюю квантовую эффективность светодиодов. Можно также поверхность полупроводника просветлить, т.е. нанести вещество толщиной $\lambda/4$ с показателем преломления, равным $\sqrt{n_r}$. Часто для увеличения коэффициента полезного действия светодиода применяют люминофор в качестве материала линзы.

6. ФЛУКТУАЦИОННЫЕ ПРОЦЕССЫ В ПОЛУПРОВОДНИКАХ И ПОЛУПРОВОДИКОВЫХ ПРИБОРАХ

6.1 Основные положения теории шумов

Изучение шумов в физических системах представляет собой важную проблему в науке и технике, поскольку они определяют наименьшие уровни сигналов, которые могут быть обработаны электронными средствами, а также точность измерений величин электрических параметров. Чтобы определить эти уровни, необходимо уметь корректно измерять основные параметры шумов, уметь минимизировать отношения «шум – сигнал» в устройствах приема и обработки сигналов. Решение указанных вопросов невозможно без понимания природы шумов.

Физические механизмы появления шумов во всех случаях связаны со случайным характером рождения-уничтожения свободных носителей заряда в активной области полупроводникового прибора, хаотическим движением заряженных частиц, разбросом значений его параметров, тепловыми колебаниями решетки другими случайными изменениями параметров объекта.

В полупроводниках и полупроводниковых приборах различают следующие основные виды шумов:

- тепловой шум;
- генерационно-рекомбинационный (ГР) шум;
- дробовой шум;
- избыточный шум.

Тепловой шум (шум Найквиста). Данный вид шума имеет место в любой среде, обладающей заметной омической проводимостью и находящейся при температуре, отличной от температуры абсолютного нуля.

Свободные носители заряда в проводнике находятся в беспорядочном тепловом движении. Это движение характеризуется тепловой мощностью, которая определяет интенсивность хаотического движения заряженных частиц. Поэтому на концах проводника, имеющего конечное сопротивление, существует флуктуационная эдс – $V(t)$. Этот вид шума называется **тепловым шумом**. Рассмотрим подробно физическую картину его формирования, поскольку в литературе по шумам ее описания нет.

Предположим, что имеется полупроводниковый образец, в котором при температуре T содержится достаточное количество свободных носителей заряда. Известно, что ионы и атомы решетки, а также свободные носители заряда в реальном образце участвуют в тепловом движении, совершая тепловые колебания. Это движение носит случайный характер: свободные частицы движутся с тепловыми скоростями в произвольных направлениях, а, сталкиваясь, обмениваются энергией и импульсом. В результате этого обмена одни свободные частицы ускоряются, а другие замедляются; и те, и другие изменяют направление своего движения. Следовательно, тепловое движение создает сопротивление

движению заряженных частиц, которое можно представить эквивалентным резистором – R .

Тепловое движение может быть рассмотрено с двух сторон: молекулярно-кинетической (тепловой) и электрической, т.к. движутся электрически заряженные частицы. Единственной мерой теплового движения является его энергия, определяемая «квантом колебания решетки» kT , где k – постоянная Больцмана. Но в соответствии с законами электричества, хаотическое движение заряженных частиц создает хаотически направленные микротоки (при этом среднее значение суммарного хаотического тока через полупроводник равно нулю). Согласно опять-таки законам электричества, движение заряженных частиц в твердом теле, при наличии сопротивления этому движению, сопровождается выделением джоулева тепла. В условиях термодинамического равновесия между решеткой полупроводника и его электронно-дырочной подсистемой температура решетки равна температуре, создаваемой за счет джоулева тепла, выделяемого при движении свободных заряженных частиц – хаотически направленных микротоков. Наличие хаотически направленных микротоков можно описать как результат действия хаотически направленных источников напряжения. Ввиду равновероятного характера распределения этих источников в пространстве в каждый момент времени это суммарное шумовое напряжение на концах полупроводника отлично от нуля (хотя среднее значение хаотических напряжений равняется нулю). Складываясь по длине полупроводника, хаотически направленные источники напряжения образуют на торцах образца некое результирующее хаотическое электрическое напряжение. Это напряжение можно зарегистрировать, если к противоположным торцам полупроводника подключить регистрирующую аппаратуру, например, осциллограф. Оно проявится на экране в виде случайно изменяющегося во времени по амплитуде, частоте и фазе биполярного напряжения.

Такова физическая картина появления теплового шума. Теперь нетрудно определить зависимость его величины от параметров материала. Для отыскания электрических характеристик теплового шума будем рассуждать следующим

образом. Энергия тепловых колебаний решетки вдоль координаты, совпадающей с направлением измерения шумового напряжения на образце, равна kT . В условиях термодинамического равновесия между решеткой и свободными носителями заряда точно такая же энергия выделяется с джоулевым теплом

$$E = \frac{\overline{U_w^2}}{2R} \cdot \tau,$$

где τ – время наблюдения за случайным процессом, $\overline{U_w^2}$ – квадрат амплитуды шумового напряжения на частоте наблюдения, R – омическое сопротивление всего образца. Согласно теореме Котельникова, для получения полной информации о процессе его измерение необходимо вести через промежутки времени τ , определяемые полосой пропускания измерительного прибора Δf : $\tau = 1/2 \Delta f$. Следовательно, приравняв тепловые энергии решетки и электронно-дырочной подсистемы, получим

$$kT = \frac{\overline{U_w^2}}{2R} \cdot \frac{1}{2 \cdot \Delta f},$$

откуда находим, что шумовое напряжение, измеряемое на концах полупроводникового образца, определяется соотношением для квадрата ее амплитуды вида

$$\overline{U_w^2} = 4kTR \cdot \Delta f. \quad (65.1)$$

Данное выражение называют **теоремой Найквиста**. Таким образом, для описания теплового шума удобно ввести эквивалентный генератор шумового напряжения величиной $\sqrt{\overline{U_w^2}}$ или эквивалентный генератор шумового тока – $\sqrt{\overline{i_w^2}}$. Выбор способа описания теплового шума определяется соотношением величин сопротивления образца R и сопротивления нагрузки R_n . Если $R \ll R_n$, то целесообразно пользоваться эквивалентным генератором шумового напряжения, а при обратном их соотношении – генератором шумового тока, дисперсии (квадраты амплитуд генерируемых ими сигналов) которых равны:

$$\overline{U_w^2} = 4kTR \cdot \Delta f, \quad \overline{i_w^2} = \frac{4kT}{R} \Delta f.$$

Генерационно-рекомбинационный шум. В реальном полупроводнике электроны и дырки могут появляться и исчезать случайным образом под влиянием процессов тепловой генерации и рекомбинации. Этот процесс происходит следующим образом: в результате тепловых колебаний решетки полупроводника в случайных ее местах решетка может получить большую амплитуду флуктуации. Находящийся в этом месте атом полупроводника, например кремния, получает от решетки большую энергию. Если эта дополнительная энергия превышает ширину запрещенной зоны, то атом направляет ее внешнему электрону. Тот отрывается от атома и переходит в междоузлие, оставляя в валентной зоне свободную дырку. Если тепловая энергия меньше ширины запрещенной зоны, то атом оставляет эту энергию в тепловом виде – сильно отклоняется от положения равновесия, передавая энергию соседним атомам. В результате происходят изменения концентраций электронов и дырок в зонах на величины δn , δp . Избыток носителей заряда в зонах с течением времени убывает за счет случайной рекомбинации электронно-дырочных пар, что также вносит свой вклад в случайное изменение числа носителей в зонах. Следствием этих двух противоположно направленных процессов изменения числа электронов и дырок является то, что сопротивление образца R испытывает флуктуации – $\delta R(t)$. Если теперь через такой образец пропустить постоянный ток I , то на концах образца возникнет флуктуационная ЭДС: $\delta V(t) = I \cdot \delta R(t)$, которая может быть обнаружена с помощью стандартных контрольно-измерительных приборов (осциллограф, вольтметр и т.д.). Этот вид шума называют **генерационно-рекомбинационным**.

Дробовой шум. Этот вид шума возникает вследствие случайного характера влета и вылета заряженных частиц в область регистрации этих частиц. Так, в электронной лампе акты вылета электронов из катода представляют собой последовательность независимых событий, происходящих в случайные моменты времени. Пусть весь промежуток времени по наблюдению за количеством элек-

тронов, вылетающих с поверхности катода, разбит на равные промежутки времени длительностью Δt . Пусть в течение первого промежутка с катода вылетело 105 электронов, в течение второго – 95, третьего – 90, четвертого – 110 и т.д. Можно видеть, что среднее количество вылетающих электронов примерно составляет 100 штук, а наблюдаемый разброс около этого значения представляет собой случайный процесс. Поэтому ток эмиссии катода $I(t)$ имеет постоянную составляющую и флуктуирующую. Последняя и называется **дробовым шумом**. То же самое происходит в транзисторе или диоде, так как пролет электронов и дырок через потенциальные барьеры происходит независимо и в случайные моменты времени.

Избыточный шум. Этот вид шума обнаружен практически во всех электронных приборах при протекании через них электрического тока. Отличием этого вида шума от дробового является характерная зависимость его спектральной плотности от частоты:

$$\overline{i_{\text{шф}}^2} = \alpha \frac{I_0}{f^n},$$

где α – коэффициент пропорциональности, $n = 1 \div 2$. Величина α сильно зависит от технологии изготовления прибора. Видно, что с понижением частоты шум растет и может превысить дробовой шум (что обычно и имеет место). Именно поэтому этот шум называют **избыточным**, хотя есть и другие названия. Так как величина шума зависит от качества изготовления контактов, то его иногда называют **контактным**. Характерная зависимость спектральной плотности от частоты привела к другому его названию – **шум типа $1/f$** . Наконец, поскольку этот шум возникает при протекании тока через прибор, то иногда его называют **токовым шумом**. Величина шумового напряжения избыточного шума в зависимости от частоты измерений показана на рис. 44.

Природа избыточного шума многообразна: он может возникнуть за счет флуктуаций температуры образца, протекания тока по структурным неоднородностям как в объеме полупроводника, так и на

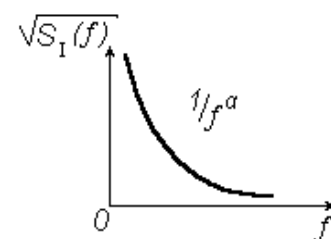


Рис. 44

его поверхности. Однако наиболее частой причиной этого вида шума является захват носителей заряда на ловушки в области контактов или на поверхности прибора, где всегда есть тонкая диэлектрическая пленка с большой концентрацией медленных поверхностных состояний. Каждое поверхностное состояние дает спектр шума, аналогичный спектру ГР-шума. Но действующих ловушек на поверхности очень много. Поэтому результирующий шум будет суммой ГР-спектров от каждой ловушки. В результате наложения многих спектров ГР-шума от действия отдельной ловушки, для которых характерно плавное изменение концентраций ловушек и времен жизни носителей заряда на них, их высокочастотные спады складываются в общий спектр, который наблюдается как $1/f$ -шум (см. рис. 44).

5.2 Математическое описание случайных процессов

Флуктуирующие напряжения, токи являются случайными переменными, поэтому изучение шумов целесообразно проводить, пользуясь методами теории вероятностей. Один из основных способов статистического описания случайной переменной $x(t)$ связан с вычислением ее среднего значения – $\overline{x(t)}$, которое обозна-

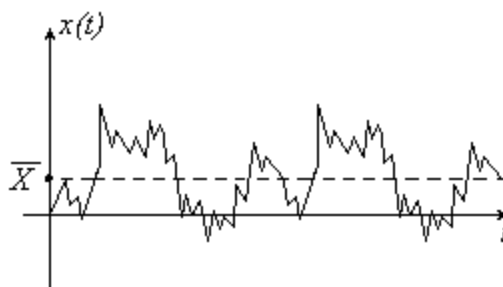


Рис. 45

чают как \overline{X} , и среднего квадрата случайной величины – $\overline{x(t)^2}$. Графическая иллюстрация смысла этих параметров случайного процесса показана далее на рисунках. Пусть имеется случайный процесс, показанный на рис. 45. Его среднее значение \overline{X} находится как та постоянная составляющая, вокруг которой происходят знакопеременные изменения переменной $x(t)$. Дисперсия случайного процесса $x(t)$ в общем случае вычисляется путем вычитания из каждого значения случайного процесса его среднего значения, возведения полученного результата в квадрат и усреднения полученной величины по времени – $\overline{(x(t) - \overline{X})^2}$. Графическая интерпретация данного параметра представлена на рис. 46. Как

можно видеть, дисперсия представляет собой квадрат средней амплитуды отклонения случайного процесса от своего среднего значения. Грубо говоря, дисперсия – это усредненный по времени квадрат амплитуды случайного процесса.

Часто \bar{X} равно нулю, и тогда наиболее значимой величиной становится $\overline{x(t)^2}$.

Таким образом, по аналогии с напряжением в электронном устройстве, где есть

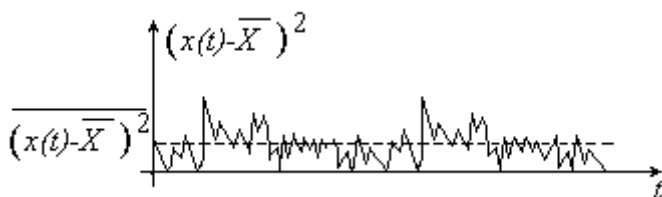


Рис.46

постоянное и переменное напряжение и потому полное напряжение представляет собой их сумму, случайный процесс также представляют как сумму среднего значения и квадратного корня из его дисперсии.

В электронике возможные источники шума имеют такой характер флуктуаций, что их средние значения и дисперсии не зависят от времени наблюдения. Такие случайные процессы называют **стационарными**.

Плотность вероятности. Случайные процессы могут быть описаны их плотностью вероятности, которая представляет собой вероятность попадания значения случайной величины в интервал между X и $X+dX$. Стационарные случайные переменные имеют плотности вероятности, которые не зависят от времени. Их средние значения можно рассчитать, если известна плотность вероятности.

Пусть рассматривается большое число идентичных систем, которые подвержены флуктуациям. Можно определить вероятность ΔP того, что случайная переменная X , описывающая флуктуации, принимает значения, заключенные между значениями случайной величины от X до $X + \Delta X$. Для этого представим ΔP в виде $\Delta P = \frac{\Delta N}{N}$, где ΔN – число систем в момент времени t , для которых эта переменная заключена в интервале ΔX , а N – число систем в ансамбле. В дифференциальной форме это соотношение записывается так:

$$dP = f(X)dX.$$

Функцию $f(X)$ называют **плотностью вероятности** случайной величины X . Она определяется экспериментально и удовлетворяет условию нормировки

$$\int f(X)dX = 1,$$

где интегрирование ведется по всем значениям X . Это соотношение выражает тот факт, что X обязательно лежит в диапазоне допустимых значений.

Если известна плотность вероятности $f(X)$, то средние значения $\overline{X^m}$ вычисляются следующим образом:

$$\overline{X^m} = \int f(X) \cdot X^m dX \quad (m=1, 2, \dots),$$

а среднее значение функции $g(X)$ равно:

$$\overline{g(X)} = \int g(X) \cdot f(X) dX.$$

В обоих случаях интегрирование ведется по всем допустимым значениям X .

Важными примерами дискретных плотностей вероятности являются **биномиальный закон**, **закон Пуассона** и **нормальный закон**. Обсудим их.

Биномиальный закон. Пусть некоторое событие имеет вероятность p реализации в форме A и $1-p$ вероятность реализации в форме B , и пусть отдельные события независимы. Если событие случается m раз, то вероятность $P_m(n)$ того, что n - раз оно реализуется в форме A , будет:

$$P_m(n) = \frac{m!}{n!(m-n)!} \cdot p^n (1-p)^{m-n}.$$

Здесь: $\bar{n} = m \cdot p$; $\sigma^2 = m \cdot p(1-p)$.

Закон Пуассона. Пусть отдельные события независимы и происходят случайно со средней частотой \bar{n} . Тогда вероятность $P(n)$ того, что n событий произойдут в течение временного интервала единичной длительности будет:

$$P(n) = \frac{[\bar{n}]^n \cdot \exp(-\bar{n})}{n!}.$$

Здесь: $\sigma^2 = \bar{n}$.

Нормальный закон. Пусть события случаются со средней частотой \bar{n} . Пусть также \bar{n} – велико и дисперсия σ определена обычным способом. Тогда вероятность того, что n событий произойдут в течение единичного временного интервала, будет

$$P(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(n-\bar{n})^2}{2\sigma^2}\right].$$

Можно показать, что биномиальный и пуассоновский законы сводятся к нормальному при больших значениях n . Случай, когда $\sigma^2 = \bar{n}$, называют **законом Гаусса**.

Автокорреляционная функция. В стационарных случайных процессах важным параметром является среднее значение произведения двух значений случайного процесса, сдвинутых по времени на промежуток s : $\overline{X(t) \cdot X(t+s)}$. Оно называется **автокорреляционной функцией** и является мерой продолжительности влияния значения случайной переменной в данный момент времени на последующие ее значения, т.е. описывает влияние настоящего случайного процесса на его будущее.

Если $\overline{X(t) \cdot X(t+s)} = A\delta(s)$, т.е. является δ -функцией параметра запаздывания s , то шум называют **белым**. Обычно стараются представить флуктуационные явления с помощью источников белого шума.

Метод Фурье. Одним из эффективных методов анализа случайных величин является метод Фурье, основанный на введении в рассмотрение спектральных плотностей случайного процесса – $S_V(f)$. С помощью этой величины флуктуационную эдс $V(t)$ в небольшом интервале частот можно представить в виде источника синусоидальной величиной шумового эдс $\sqrt{S_V(f)\Delta f}$. Достоинство такого подхода состоит в том, что с введением спектральной плотности для анализа шумов можно пользоваться теорией цепей переменного тока.

Теорема Винера-Хинчина. Важной теоремой спектрального анализа случайной величины является теорема Винера-Хинчина. Пусть $X(t)$ является ста-

ционарным случайным процессом. Разложим $X(t)$ в ряд Фурье в интервале времени $0 \leq t \leq T$, в пределах которого анализируется случайный процесс:

$$X(t) = \sum_{n=-\infty}^{\infty} a_n \cdot \exp(j \cdot \omega_n \cdot t),$$

где $n = 0, \pm 1, \pm 2, \dots$, а $\omega_n = \frac{2\pi n}{T}$ и

$$a_n = \frac{1}{T} \cdot \int_0^T X(t) \cdot \exp(-j\omega_n t) dt.$$

Спектральная плотность $S_X(f)$ случайного процесса $X(t)$ определяется следующим образом:

$$S_X(f) = \lim_{T \rightarrow \infty} \overline{2 \cdot T a_n a_n^*},$$

где знаком (*) отмечена комплексно-сопряженная величина. Тогда в соответствии с теоремой Винера-Хинчина спектральная плотность может быть найдена с помощью следующего выражения:

$$S_X(f) = 2 \cdot \overline{\int_{-\infty}^{\infty} X(t) \cdot X(t+S) \cdot \cos \omega S ds}.$$

При вычислениях обычно находят сначала автокорреляционную функцию, а потом определяют спектральную плотность $S_X(f)$. При измерениях сначала измеряют $S_X(f)$, а затем находят автокорреляционную функцию.

Далее, если временной сдвиг S равен нулю, то можно найти:

$$\overline{X^2(t)} = \int_0^{\infty} S_X(f) df, \quad (6.1)$$

т.е. средний квадрат можно определить путем интегрирования, если известна спектральная плотность случайного процесса $S_X(f)$.

Причину, по которой $S_X(f)$ можно легко измерить, поясним следующим образом. Пусть случайный сигнал $X(t)$ подается на вход линейной системы с передаточной функцией $g(f)$ и пусть $Y(t)$ – сигнал на выходе этой системы.

Если $S_x(f)$ и $S_y(f)$ являются спектральными плотностями, а X_n и Y_n – коэффициентами разложения их в ряды Фурье, то $Y_n = X_n \cdot g(f)$ так, что

$$S_y(f) = S_x(f) \cdot |g(f)|^2.$$

Используя соотношение (6.1), получаем:

$$\overline{Y^2(t)} = \int_0^{\infty} S_x(f) \cdot |g(f)|^2 df.$$

Конкретным выводом, который можно сделать из данного выражения, является следующий: если рассматриваемая линейная система является усилителем, то амплитуда случайного процесса $\overline{Y^2(t)}$ может быть измерена с помощью селективного выделения заданной частоты (определение спектральной плотности), квадратичного детектора и интегратора (взятие интеграла).

5.3 Основные виды шумов в полупроводниковых приборах.

Метод Ланжевена

Этот метод применяется для определения шумовых свойств различных электрических цепей. Его сущность заключается в том, что записывается макроскопическое уравнение, описывающее кинетические свойства, рассматриваемой системы. В правую часть этого уравнения вводится случайная возмущающая функция $H(t)$, которая описывает флуктуации в системе. Хотя $H(t)$ неизвестно, обычно можно получить достаточно информации о системе, чтобы рассчитать спектральные плотности флуктуаций. Покажем это на конкретных примерах.

Тепловой шум. Рассмотрим RL -цепочку с источником теплового шума напряжением $H(t)$, создаваемого сопротивлением R (см. рис. 47). Дифференциальное уравнение Ланжевена в этом случае будет иметь вид:

$$L \frac{di}{dt} + R \cdot i = H(t). \quad (6.2)$$

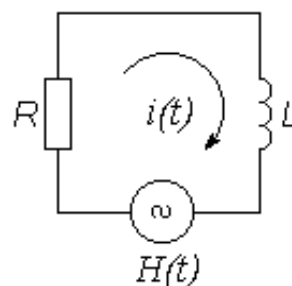


Рис. 47

Для диапазона времени $0 \leq t \leq T$ разложим $H(t)$ и $i(t)$ в ряды Фурье:

$$H(t) = \sum_{n=-\infty}^{\infty} a_n \cdot \exp(j\omega_n t);$$

$$i(t) = \sum_{n=-\infty}^{\infty} \beta_n \cdot \exp(j\omega_n t).$$

Здесь $\omega_n = \frac{2\pi n}{T}$. Подставляя Фурье-разложения в основное уравнение (6.2) и

учитывая, что $\frac{d}{dt} = j \cdot \omega_n$, получим

$$\beta_n = \frac{a_n}{R + j\omega_n L}. \quad (6.3)$$

По определению спектральной плотности

$$S_H(f) = \lim_{T \rightarrow \infty} \overline{2T a_n \cdot a_n^*};$$

$$S_i(f) = \lim_{T \rightarrow \infty} \overline{2T \beta_n \cdot \beta_n^*}. \quad (6.4)$$

Теперь подставим в выражение (6.4) значения коэффициентов Фурье-разложения тока по (6.3) и получим связь между спектральными плотностями шумового тока и шумового напряжения в рассматриваемой цепи в виде

$$S_i(f) = \frac{S_H(f)}{R^2 + \omega^2 L^2}.$$

Так как шумовое напряжение на сопротивлении $H(t)$ есть белый шум, то его значение должно быть одинаково на любой частоте, включая нулевую частоту:

$S_H(f) = S_H(0)$. Тогда можно записать

$$S_i(f) = \frac{S_H(f)}{R^2 + \omega^2 L^2} = \frac{S_H(0)}{R^2 + \omega^2 L^2}.$$

Теперь задача сводится к отысканию $S_H(0)$. Для этого рассчитаем $\overline{i^2}$:

$$\overline{i^2} = \int_0^{\infty} S_i(f) df = S_H(0) \cdot \int_0^{\infty} \frac{df}{R^2 + \omega^2 L^2} = \frac{S_H(0)}{4RL}. \quad (6.5)$$

Из теоремы о равномерном распределении энергии по элементам схемы известно, что запасенная энергия в индуктивности в точности равна тепловой энергии в сопротивлении

$$L \frac{\overline{i^2}}{2} = \frac{kT}{2} \quad \text{или} \quad \overline{i^2} = \frac{kT}{L}. \quad (6.6)$$

Подставляя в выражение (6.5) найденное выражение дисперсии шумового тока по (6.6), получим:

$$S_H(0) = 4kTR; \quad S_i(f) = 4kT \cdot \frac{R}{R^2 + \omega^2 L^2}. \quad (6.7)$$

Первое соотношение в (6.7) известно как теорема Найквиста – мы его получали из общих соображений в п. 6.1. Оно говорит о том, что уровень теплового шума (квадрат дисперсии шумового напряжения) зависит только от величины сопротивления и его температуры. К диэлектрикам это соотношение неприменимо.

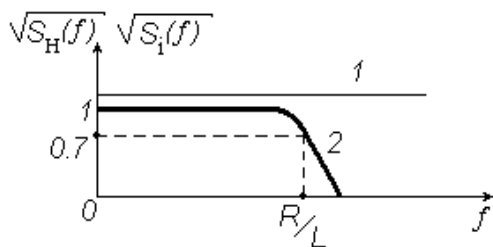


Рис. 48

Второе выражение (6.7) определяет спектральную плотность шумового тока, протекающего в рассматриваемой цепи. Оно легко выводится из теоремы Найквиста с использованием теории линейных цепей.

На рис. 48 показаны частотные зависимости напряжения теплового шума (кривая 1) и напряжения шума (кривая 2), измеряемого на активном сопротивлении R в цепи рис. 47.

Генерационно-рекомбинационный шум. Возникновение и исчезновение носителей заряда в образце полупроводника из-за процессов их случайной генерации и рекомбинации можно описать дифференциальным уравнением

$$\frac{d\Delta N}{dt} = -\frac{\Delta N}{\tau} + H(t), \quad (6.8)$$

где ΔN – общее число флуктуирующих носителей, $H(t)$ – случайное воздействие, τ – время жизни неравновесных носителей заряда.

Для того чтобы определить спектральную плотность ГР-шума, поступим следующим образом. Обратим внимание на то, что данное уравнение по своей структуре полностью аналогично уравнению (6.2). Поэтому можно не решать уравнение (6.8), а воспользоваться уже имеющимся решением уравнения (6.2). По аналогии с выражением (6.7) для ГР-шума будем иметь:

$$S_N(f) = S_H(0) \cdot \frac{\tau}{1 + \omega^2 \tau^2}. \quad (6.9)$$

Найдем спектральную плотность шума на нулевой частоте. Для этого определим дисперсию числа носителей заряда через $S_N(f)$. По определению дисперсии

$$\overline{\Delta N^2} = \int_0^{\infty} S_N(f) df = S_H(0) \cdot \int_0^{\infty} \frac{\tau \cdot df}{1 + \omega^2 \tau^2} = \frac{1}{4} S_H(0) \cdot \tau.$$

Отсюда найдем искомую спектральную плотность на нулевой частоте и подставим ее в выражение (6.9). Тогда для этого вида шума будем иметь

$$S_N(f) = \overline{\Delta N^2} \cdot \frac{4\tau}{1 + \omega^2 \tau^2}. \quad (6.10)$$

Таким образом, частотный спектр ГР-шума $S_N(f)$ можно считать известным, если определены: среднее время жизни неравновесных носителей заряда в данном полупроводнике τ и дисперсия флуктуаций числа носителей $\overline{\Delta N^2}$. Последняя величина определяется ГР-механизмом носителей заряда и уровнем легирования. Как следует из выражения (6.10), частотная зависимость спектральной плотности ГР-шума представляет собой полочку на низких частотах ($\omega \ll 1/\tau$) величиной, определяемой дисперсией числа носителей $\overline{\Delta N^2}$ и средним временем жизни носителей заряда в полупроводнике. На высоких частотах ($\omega \gg 1/\tau$) спектральная плотность убывает по закону $1/\omega^2$. В целом поведение кривой $S_N(f)$ аналогично показанному кривой 2 на рис. 46, но с заменой частоты среза на значение $1/\tau$.

Дробовой шум. Этот тип шума наблюдается во всех физических системах, где есть поток независимых частиц. Представим себе поток таких частиц (например, излучаемых нагретым катодом электронов), у которых вероятность испускания каждой последующей частицы не зависит от испускания предыдущей. Такие события называются **марковскими**. Пусть регистрация этих частиц производится прибором с полосой пропускания Δf . Тогда в соответствии с теоремой Котельникова (теорема отсчетов) для извлечения полной информации о потоке пролетающих частиц отсчеты должны производиться через интервалы времени $\tau = 1/2\pi\Delta f$. Разобьем ось времени на интервалы длительностью τ . Тогда величину потока частиц можно определить из соотношения:

$$F_0 = \bar{n}_\tau / \tau,$$

где \bar{n}_τ – среднее число частиц, проходящих через прибор за время τ . Если частица заряжена, то данным потоком создается электрический ток

$$I_0 = \frac{\bar{n}_\tau}{\tau} q.$$

Поскольку поток состоит из независимых частиц, то среднеквадратичное отклонение числа частиц от среднего значения за время наблюдения должно подчиняться закону Пуассона, для которого $\overline{(n_\tau - \bar{n}_\tau)^2} = \overline{\Delta n_\tau^2} = \bar{n}_\tau$. Следовательно,

$$\begin{aligned} \sqrt{i_{ш}^2} &= \frac{\sqrt{\overline{\Delta n_\tau^2}}}{\tau} \cdot q; \\ \overline{i_{ш}^2} &= \frac{\overline{\Delta n_\tau^2}}{\tau^2} \cdot q^2 = \frac{\bar{n}_\tau}{\tau^2} \cdot q^2 = F_0 \cdot \frac{q^2}{\tau} = I_0 \frac{q}{\tau} = 2q \cdot I_0 \cdot \Delta f. \end{aligned}$$

Таким образом, дробовой шум, представленный генератором тока, равен:

$$\overline{i_{ш}^2} = 2q \cdot I_0 \cdot \Delta f.$$

Из этого выражения следует, что дробовой шум является «белым». Однако, если частота измерений сравнима с величиной, обратной времени пролета частиц через область регистрации τ_{np} (вакуумный промежуток в лампе, область про-

странственного заряда в фотодиодах), то спектральная плотность дробового шума становится зависящей от частоты по закону:

$$\overline{i_{шф}^2} = 2q \cdot I_0 \cdot \left(\frac{\sin(\omega\tau_{пр}/2)}{(\omega\tau_{пр}/2)} \right)^2.$$

Множитель в круглых скобках описывает частотную характеристику межэлектродного промежутка и представляет собой частотный спектр прямоугольного импульса. В качестве такого выступает импульс тока, вызванный пролетом электрона через область регистрации.

5.4 Численные значения шумов различных видов

Для того чтобы «почувствовать» уровень шумов различной природы, сравним тепловые и дробовые шумы фоторезисторного приемника в единичной полосе частот: $\Delta f = 1$. Эквивалентная схема такого приемника имеет следующий

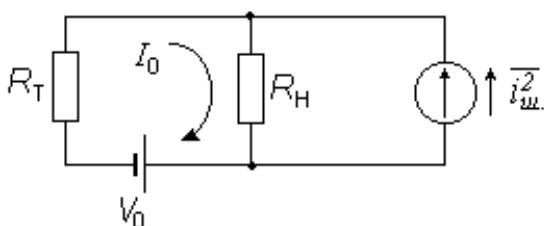


Рис. 49

вид (см. рис. 49). Для получения максимального сигнала фотопроводимости сопротивление нагрузки R_H выбирают равным темновому сопротивлению фоторезистора R_T . Поэтому тепловой шум суммарного сопротивления, равного $R_T / 2$, будет:

$$\overline{U_{ш\text{тепл}}^2} = 2 \cdot kT \cdot R_T.$$

Для дробового шума напряжение равно:

$$\overline{U_{шдр}^2} = \overline{i_{шдр}^2} \cdot \left(\frac{R_T}{2} \right)^2 = 2q \cdot I_0 \cdot \left(\frac{R_T}{2} \right)^2.$$

С другой стороны, можно записать: $I_0 = \frac{V_0}{2R_T}$. Тогда выражение для определения квадрата напряжения дробового шума будет иметь вид:

$$\overline{U_{ш.др}^2} = 2q \cdot \frac{V_0}{2R_T} \cdot \left(\frac{R_T}{2}\right)^2 = q \cdot V_0 \cdot \frac{R_T}{4}.$$

Тепловые и дробовые шумы фоторезисторного приемника будут равны друг другу при напряжении смещения $V_=\$, равном

$$2kT \cdot R_T = q \cdot V_ = \cdot \frac{R_T}{4}, \quad V_ = = \frac{8 \cdot kT}{q} = 200 \text{ mV}$$

для комнатной температуры.

Если сопротивление фоторезистора равно 10^4 Ом, то напряжение теплового шума на нем будет составлять $\sqrt{\overline{U_{ш.тепл}^2}} \approx 2 \cdot 10^{-8} \text{ В/Гц}^{1/2}$.

Таким же будет шумовое напряжение фоторезисторного приемника рис. 47, если к фоторезистору приложено постоянное напряжение $V_0 < V_ =$. При напряжениях смещения V_0 , превышающих $V_ =$, шумовое напряжение на сопротивлении нагрузки будет расти по закону $\sqrt{V_0}$ за счет того, что в этих условиях доминирующим шумом будет дробовой шум постоянного тока, протекающего через фоторезистор. Так, для напряжения $V_0 = 800 \text{ mV}$ будем иметь шумовое напряжение на сопротивлении нагрузки $\sqrt{\overline{U_{ш.др}^2}} \approx 4 \cdot 10^{-8} \text{ В/Гц}^{1/2}$. Приведенные выше рассуждения иллюстрируются рис. 50.

В дополнение к изображенным на рис. 48 шумам заметим, что если на практике напряжение смещения значительно превышает $V_ =$, то к дробовому шуму, как правило, добавляется избыточный шум типа $1/f$ приконтактных

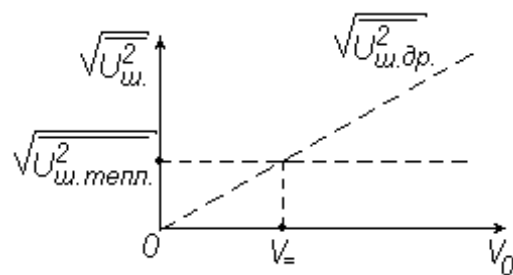


Рис. 50

областей фоторезистора, что резко увеличивает амплитуду шумового напряжения на сопротивлении нагрузки. Таким образом, превышение оптимального значения напряжения на фоторезисторе не увеличивает величину фотопроводимости (если коэффициент усиления фотопроводимости достиг насыщения), но резко увеличивает напряжение шума на сопротивлении нагрузки.

8. ЖИДКИЕ КРИСТАЛЛЫ В ОПТОЭЛЕКТРОНИКЕ

7.1. Общие сведения о жидких кристаллах

В 1888 г. австрийский ученый Рейнитцер получил новое органическое вещество – холестерилбензоат. Определяя температуру его плавления, он заметил, что вещество плавится в две стадии: вначале образуется мутный расплав, который при дальнейшем повышении температуры превращается в прозрачную изотропную жидкость. Необычными в этом процессе были свойства мутного раствора. Обладая всеми свойствами жидкости, мутный расплав проявлял оптические свойства одноосного кристалла: свет, проходя через него, испытывал

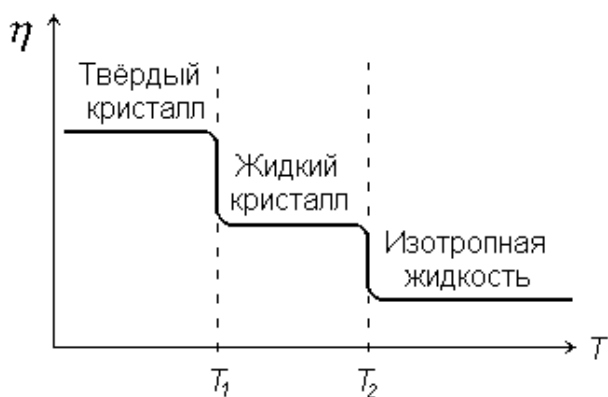


Рис. 51

двойное лучепреломление. Рейнитцер решил, что перед ним смесь двух веществ, плавящихся при разных температурах, причем в расплаве одного из веществ находятся кристаллиты другого. Он не смог разделить эту смесь и послал свой препарат немецкому физику Леману, занимавшемуся определением оптических

свойств кристаллов. Леман выяснил, что присланный ему холестерилбензоат представляет собой чистое вещество, а не смесь двух, которое при нагревании переходит в ранее неизвестное агрегатное состояние: между кристаллической фазой и изотропной жидкостью образовывался мутный расплав с анизотропными оптическими свойствами кристалла. При дальнейшем повышении температуры эта фаза переходила в изотропную жидкость. На рис. 51 показано изменение условного параметра – «параметра структурного порядка» η кристалла, проявляющего указанные температурные изменения своего строения.

Вскоре выяснилось, что холестерилбензоат не единственное вещество, обладающее такими необычными свойствами. Леман нашел подобные фазы при плавлении еще целого ряда органических веществ. Всем им он дал название

жидкие кристаллы, а оптически анизотропную фазу вскоре назвали **мезофазой** (от греческого слова «мезос» – промежуточный). Леман обнаружил, что у некоторых органических веществ аналогичные фазы возникают не только при нагревании, но и при растворении в спиртосодержащем растворе. Такие жидкие кристаллы называют **лиотропными** (от греческого слова «лио» – растворяю), в отличие от **термотропных**, которые образуются при нагреве. В настоящее время известно около 500 органических веществ, обладающих жидкокристаллической мезофазой.

7.2. Основные свойства жидких кристаллов

После открытия жидких кристаллов их свойствами заинтересовались химики и физики, и к 30-м годам XX века были выяснены все основные типы жидких кристаллов, особенности строения и их основные физические свойства, наиболее необычными из которых оказались следующие.

Двулучепреломление света. Явление двулучепреломления – это типично кристаллический эффект, состоящий в том, что скорость света в кристалле зависит от ориентации плоскости поляризации света. Она достигает максимального и минимального значений для двух взаимно ортогональных ориентаций плоскости поляризации. Поляризации, соответствующие экстремальным значениям скорости света, определяются анизотропией свойств кристалла и однозначно задаются направлением распространения света относительно кристаллографических координатных осей.

Линейно поляризованный свет, распространяясь в оптически активной среде, изменяет ориентацию плоскости поляризации. Угол поворота плоскости поляризации прямо пропорционален пути L , пройденному светом, т.е. выражается формулой $\varphi = \varphi_a L$, где величина φ_a определяет угол поворота на единичном отрезке пути и называется **удельной вращательной способностью**. Величина вращательной способности φ_a для жидких кристаллов в сотни и тысячи раз превосходит эту величину для наиболее оптически активных кристаллов, например,

кварца. Более того, угол поворота плоскости поляризации в жидких кристаллах зависит от длины волны света. В твердых телах, как, впрочем, и в обычных жидкостях, удельная вращательная способность φ_a имеет вполне определенный знак, не зависящий от длины волны света. Это означает, что вращение плоскости поляризации света в них происходит в определенном направлении. В жидких кристаллах для коротких длин волн величина φ_a , например, могла быть положительной, а для более длинных – отрицательной. Но в любом случае направление вращения плоскости поляризации света при прохождении им жидкокристаллической фазы изменялось в зависимости от длины волны. Такое поведение угла вращения плоскости поляризации не укладывалось в рамки существовавших представлений об оптической активности веществ.

Чувствительность к воздействиям. Удивительными оказались и другие свойства жидких кристаллов, такие как исключительно высокая чувствительность структурного состояния и оптических свойств жидкого кристалла к магнитным, электрическим полям, примесям, ионизирующему воздействию, температуре и другим воздействиям, что и определило практическое применение жидких кристаллов.

В первые годы жидкие кристаллы казались неперспективными для практического использования. Однако в 1963 г. американским инженером был запатентован метод регистрации тепловых излучений с помощью тонкой пленки жидкого кристалла, изменяющей свой цвет при нагревании. Он решал важную задачу квантовой электроники – отображение тепловых потоков. С этого момента начинается бурное развитие исследований физических свойств жидких кристаллов и их применение в оптике, твердотельной электронике, приборостроении и т.д., которое продолжается по настоящее время.

Итак, жидкие кристаллы – это специфическое агрегатное состояние органического вещества, в котором оно одновременно проявляет свойства жидкости и кристалла. При плавлении кристаллы этих веществ образуют жидкокристаллическую фазу, отличающуюся от обычных жидкостей: она существует в узком температурном диапазоне, вне которого имеем кристалл или обычную жид-

кость. Обязательным свойством жидких кристаллов является наличие порядка в пространственной ориентации молекул. Все известные в настоящее время жидкие кристаллы составлены из одного сорта молекул, которые, как правило, имеют вытянутую стержнеобразную форму либо вид вытянутой спирали. В последнем случае направление закручивания спирали может быть либо в одну сторону, либо в другую. Такие жидкие кристаллы называют **хиральными**.

7.3. Типы и структура жидких кристаллов

В случае стержнеобразных молекул в зависимости от характера упорядочения длинных осей молекул в слое жидкие кристаллы разделяют на **нематики**, **смектики** и **холестерики**. Рассмотрим их внутреннюю структуру и пространственное упорядочение молекул. Для наглядности будем представлять молекулы жидких кристаллов в виде стержней, т.к. их длина во много раз больше толщины.

Нематики. Структура нематика будет представлять собой «жидкость одинаково ориентированных стержней» (см. рис. 52, а, б). В твердом состоянии молекулы кристалла строго ориентированы в пространстве (рис. 52, а). В жидкокристаллическом состоянии эти молекулы разбросаны по жидкому кристаллу хаотически, они могут перемещаться под действием температуры в любую сторону. Относительный порядок существует только в ориентации длинных осей молекул. Эти стержни движутся хаотически в жидкости, но их преимущественная ориентация не изменяется. Под действием температуры они совершают ориентационные (угловые) колебания, амплитуда которых возрастает при приближении к температуре фазового перехода (рис. 52, б). После точки фазового перехода ориентационное упорядочение исчезает, и ориентация молекул становится полностью хаотической (рис. 52, в). В отношении оптических свойств нематики подобны одноосным кристаллам.

Для характеристики ориентационного порядка вводится вектор единичной длины \vec{n} , называемый **директором**, направление которого совпадает с направлением усредненной ориентации длинных осей молекул. По существу он представляет собой единичный вектор усредненного направления длинных осей молекул. Кроме того, вводится еще один параметр – скалярный **параметр порядка** Q , который характеризует степень ориентационного упорядочения молекул:

$$Q = \frac{3}{2} \left(\overline{\cos^2 \theta} - \frac{1}{3} \right),$$

где θ – угол между направлением директора и мгновенным направлением оси молекулы, а $\overline{\cos^2 \theta}$ – среднее значение по времени. Значение $Q = 1$ соответствует полной ориентации всех молекул нематика в одном направлении, а $Q = 0$ – полному беспорядку в ориентации молекул. По физическому смыслу параметр порядка есть дисперсия случайной переменной – ориентации длинных осей молекул жидкого кристалла.

До сих пор речь шла об однодоменном образце нематика, в котором ориентация директора одинакова во всех его точках. В реальности молекулы жидкого кристалла могут изменять свою ориентацию по поверхности слоя жидкого кристалла, а также с его толщиной, образуя области одинаковой ориентации длинных осей молекул жидкого кристалла – домены. Причины такого деления нематика заключены в случайном характере внешних воздействий, способных задавать ориентацию молекул. Преимущественные ориентации длинных осей молекул в различных доменах никак не связаны между собой. Поэтому ориентация директоров в разных доменах хаотична, и, следовательно, оптические индикатрисы в доменах также разориентированы – в каждом домене ориентация индикатрисы своя. Значит и показатель преломления для света, падающего

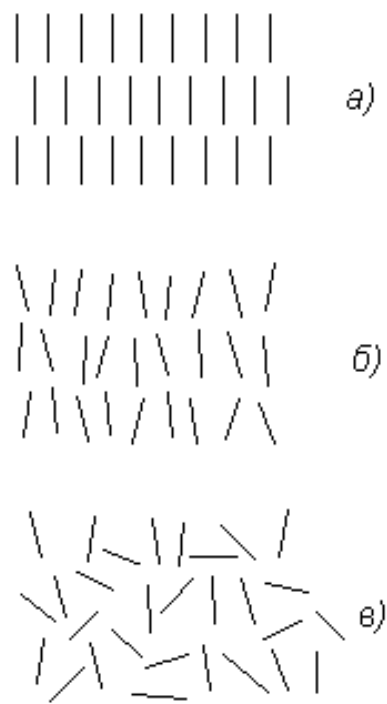


Рис. 52

на многодоменный жидкий кристалл, в каждом домене свой. На границе доменов имеет место скачок показателя преломления, который в поляризованном свете в скрещенных поляризаторах видится как светлая нить произвольной формы на темном фоне (нить по-гречески «*нема*», отсюда и название жидкого кристалла такого строения), которая указывает местоположение отдельных доменов в слое нематика и их размеры. Поэтому, чтобы сделать его одномерным, говорят, «однодоменным», необходимы специальные меры. К ним относятся любые способы, задающие ориентацию длинных осей молекул: механическая полировка внутренних поверхностей стекол, между которыми помещается жидкий кристалл; задание ориентации с помощью внешних электрических или магнитных полей и т.д.

Смектики. В них степень упорядочения молекул выше, чем в нематиках (см. рис. 53): помимо ориентационной упорядоченности молекул, существует частичное упорядочение центров их тяжести. Таким образом, смектики организованы в слои, расстояния между которыми фиксированы. Что же касается расположения центров тяжести молекул в пределах одного слоя, то оно хаотично, как в нематике. Поэтому для описания состояния смектика также используются понятия директора \bar{n} и параметра порядка Q .

Общим для всех смектиков является достаточно сильное взаимодействие молекул в пределах выделенного слоя и слабое взаимодействие между слоями. Поэтому слои легко скользят друг относительно друга и на ощупь все смектики мылоподобны (мыло по-гречески «*смегама*»). В отношении оптических свойств смектики ведут себя как двухосные кристаллы.



Рис.53

В зависимости от расположения молекул в слое, а также общего поведения директора по многим слоям смектики делятся на следующие типы: *A*, *B*, *C*, *D*, *F*, *G* и т.д. Так, в *A*-смектике оси молекул во всех его слоях перпендикулярны поверхности слоя (см. рис. 53); в *B*-смектике оси молекул находятся под небольшим углом к поверхности, одинаковым для всех слоев. В каждом слое *D*-

смектика отсутствует какое-либо упорядочение осей молекул – молекулы ориентированы хаотически.

Смектики обладают двулучепреломлением, но в исходном состоянии представляют собой совокупность хаотически ориентированных доменов.

Если растопить смектический кристалл, то сначала он дает жидкий смектик. При дальнейшем повышении температуры смектик переходит в нематик, а уже тот – в изотропную жидкость. Это и понятно: сравнительно слабое взаимодействие между слоями с повышением температуры разрушается раньше, чем взаимодействие молекул внутри каждого слоя – так появляется нематик. Впрочем, у отдельных смектиков нематическая фаза отсутствует.

Холестерики. Они устроены более сложно, чем нематики и смектики. Локально холестерический кристалл устроен так же, как и нематик: в малом объеме упорядочение молекул холестерика можно охарактеризовать директором и параметром порядка Q . Отличие холестерика от нематика проявляется в больших, по сравнению с молекулярными размерами, масштабах. Эти отличия связаны с тем, что в нем молекулы образуют **холестерическую спираль**: в нематике существует направление, называемое **холестерической осью**, вдоль которого ориентация директора регулярным образом изменяется. Директор перпендикулярен этой оси и меняет свое положение вдоль холестерической оси по спирали, оставаясь неподвижным (см. рис. 54). Угол поворота директора по мере удаления в объем холестерика линейно связан с расстоянием z вдоль холестерической оси и может быть представлен в виде: $\varphi = (2\pi / \rho_0) \cdot z$. Расстояние ρ_0 , на котором директор поворачивается на 360° , называется **шагом холестерической спирали**.

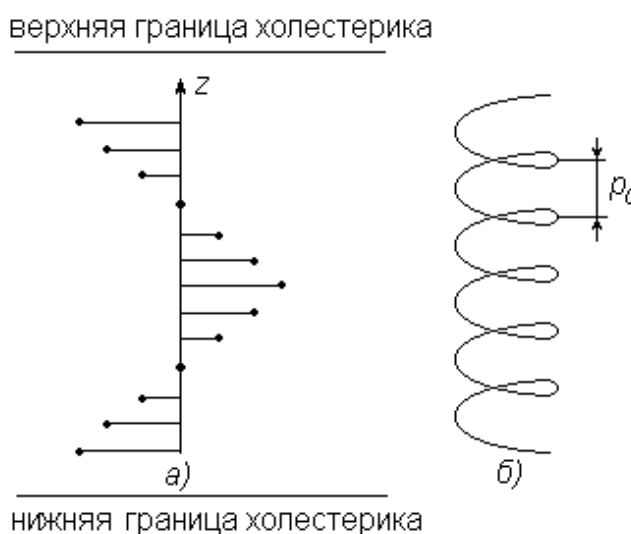


Рис. 54

Холестерическая фаза может наблюдаться сразу после плавления кристалла или после смектической фазы. При дальнейшем повышении температуры холестерик переходит в неатик, а затем в изотропную жидкость.

7.4. Ориентационные эффекты в жидких кристаллах

Нематики. Как указывалось выше, в исходном состоянии пленки жидких кристаллов по своим свойствам разбиты на отдельные домены, в пределах которых ориентация длинных осей молекул постоянна. Для научных исследований и практических применений необходимы однодоменные пленки. Это предполагает, что за счет специально принятых мер все

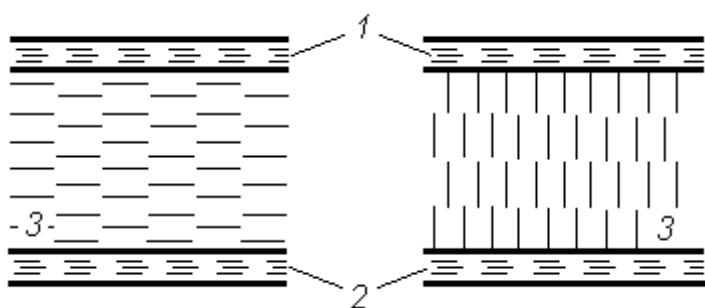


Рис.55

молекулы жидкокристаллического образца нематика ориентированы одинаковым образом. Существуют две основных ориентации молекул нематика: вдоль поверхности и перпендикулярно ей. В соответствии с этим используют два типа жидкокристаллических ячеек, показанные на рис. 55. Здесь 1 – верхняя опорная поверхность ячейки, 2 – нижняя опорная поверхность, которые обычно изготавливают из прозрачного стекла. Пленка нематика – 3 помещается между опорными поверхностями. Ячейку, показанную в левой части рис. 55, называют **планарной**, а показанную в правой части рис. 65 – **гомеотропной**. Для получения планарной ячейки внутренние поверхности стекол, контактирующие с жидким кристаллом, подвергают механической шлифовке в одном направлении, в результате которой на поверхности стекла остаются микроборозды. Молекулы нематика «ложатся» в эти борозды, образуя строго ориентированный поверхностный слой нематика планарной структуры. Ввиду действия сил связи молекулы соседних слоев перенимают ориентацию граничного слоя, передавая его в объем жидкого кристалла. Силы упругости между молекулами стремятся разрушить этот порядок. Поэтому существует некоторое расстояние, при уда-

молекулы жидкокристаллического образца нематика ориентированы одинаковым образом. Существуют две основных ориентации молекул нематика: вдоль поверхности и перпендикулярно ей. В соответствии с этим используют два типа жидкокристаллических ячеек, показанные на рис. 55. Здесь 1 – верхняя опорная поверхность ячейки, 2 – нижняя опорная поверхность, которые обычно изготавливают из прозрачного стекла. Пленка нематика – 3 помещается между опорными поверхностями. Ячейку, показанную в левой части рис. 55, называют **планарной**, а показанную в правой части рис. 65 – **гомеотропной**. Для получения планарной ячейки внутренние поверхности стекол, контактирующие с жидким кристаллом, подвергают механической шлифовке в одном направлении, в результате которой на поверхности стекла остаются микроборозды. Молекулы нематика «ложатся» в эти борозды, образуя строго ориентированный поверхностный слой нематика планарной структуры. Ввиду действия сил связи молекулы соседних слоев перенимают ориентацию граничного слоя, передавая его в объем жидкого кристалла. Силы упругости между молекулами стремятся разрушить этот порядок. Поэтому существует некоторое расстояние, при уда-

лении на которое от поверхности ориентирующее действие опорной поверхности и разориентирующее влияние упругих сил уравниваются друг друга. Это расстояние называется **когерентной длиной** и для жидких кристаллов составляет величину несколько микрон. Поэтому, если толщина слоя нематика невелика – до единиц микрон, то разориентирующее действие упругих сил, теплового движения слабое, и в объеме жидкого кристалла его молекулы (на рисунке они показаны в виде коротких линий между стеклами) сориентированы вдоль поверхности ячейки.

Для получения гомеотропной ячейки внутренние поверхности стекол обрабатывают в специальном химическом веществе – поверхностно-активном веществе. Молекулы этого вещества также имеют вытянутую форму и при обработке одним концом «приклеиваются» к обрабатываемым поверхностям (см. рис. 56). Молекулы жидкого кристалла входят в пространство между молекулами поверхностно-

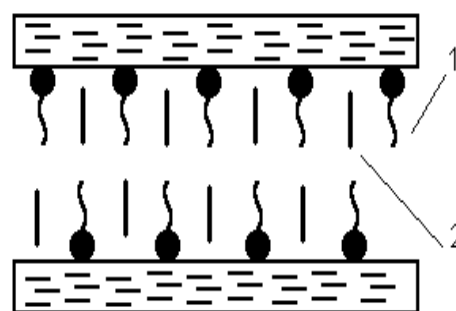


Рис. 56

активного вещества, ориентируясь перпендикулярно поверхности ячейки. Следующие слои молекул нематика, стремясь минимизировать свою потенциальную энергию по отношению к молекулам соседнего слоя жидкого кристалла, выстраиваются также перпендикулярно стеклам.

Переход Фредерикса. Для технических применений жидких кристаллов важной является возможность изменять ориентацию молекул путем приложения, например, электрического поля, которое обычно прикладывают перпендикулярно поверхностям ячеек. Важной особенностью нематиков является различное значение его диэлектрической проницаемости ϵ , измеренной вдоль осей молекул и перпендикулярно им. Поэтому при помещении нематика в постоянное электрическое поле из-за анизотропии диэлектрической проницаемости его молекулы стремятся ориентироваться так, чтобы направление большего значения ϵ совпадало с направлением приложенного поля. Значит, исходная планар-

ная структура нематика в электрическом поле может перейти в гомеотропную или наоборот. Эффект изменения структуры нематика во внешнем электрическом поле называют **переходом Фредерикса**. Общим свойством переориентации нематика под действием внешнего поля является ее пороговый характер. Критическое значение электрического поля, при котором совершается переход Фредерикса, вычисляется по следующей формуле:

$$E_c = \frac{\pi}{d} \left(\frac{4\pi K}{\Delta\varepsilon} \right)^{1/2},$$

где d – толщина слоя жидкого кристалла, K – его коэффициент упругости, $\Delta\varepsilon = \varepsilon_{\parallel} - \varepsilon_{\perp}$ анизотропия диэлектрической проницаемости жидкого кристалла. Здесь ε_{\parallel} – диэлектрическая проницаемость жидкого кристалла, когда поляризация света параллельна директору, ε_{\perp} – диэлектрическая проницаемость жидкого кристалла, когда поляризация света перпендикулярна директору. Из этого выражения следует, что критическое напряжение, прикладываемое к слою нематика для его переориентации, не зависит от толщины d :

$$V_c = \pi \left(\frac{4\pi K}{\Delta\varepsilon} \right)^{1/2}. \quad (7.1)$$

Подставив в выражение (7.1) типичные значения: $K = 10^{-6}$ дин, $\Delta\varepsilon \approx 3$, получим, что $V_c \approx 2$ В. Выше этого значения напряжения на ячейке изменения структуры быстро достигают насыщения.

Ориентационное действие электрического поля величиной E проявляется на протяжении **электрической когерентной длины**, отсчитываемой от опорной поверхности,

$$\lambda_k = \frac{1}{E} \left(\frac{4\pi K}{\Delta\varepsilon} \right)^{1/2}. \quad (7.2)$$

Подставив в это выражение типичные значения: $K = 10^{-6}$ дин, $\Delta\varepsilon \approx 3$ и $E = 10^3$ В/см, получим, что $\lambda_k \approx 10$ мкм. Следовательно, в практических приложениях

пленки нематика должны иметь толщину значительно больше двух электрических когерентных длин, что соответствует десяткам микрон.

Если увеличивать напряжение на ячейке выше порога Фредерикса, то вблизи десятка вольт структура нематика претерпевает следующее скачкообразное изменение: вдоль поверхности ячейки образуются круговые потоки молекул нематика, что формирует структуру

типа упорядоченного слоя вращающихся цилиндров. Это так называемые **домены Капустина-Вильямса**. Их возникновение обусловлено проявлением ионной проводимости жидкого кристалла в электрическом поле: ионы движутся к противоположно заряженной опорной поверхности, где и меняют заряд. Затем ионы движутся

в обратном направлении и все повторяется на другой опорной поверхности. И так много раз. Из-за изменения ориентации молекул в пределах домена он выступает как неоднородность диэлектрической проницаемости ϵ . Рисунок 57 поясняет возникновение доменов (окружности в центре рисунка с указанным направлением вращения ионов молекул в домене), изменения показателя преломления из-за изменения в пространстве ориентации директора в домене (синусоидальная линия в центре рисунка), а также причину дифракции света при прохождении им слоя нематика с доменами.

Дальнейшее повышение напряжение на ячейке приводит к сначала к увеличению их числа в направлении действия электрического поля, а затем к разрушению доменов и появлению хаотических турбулентных потоков в слое нематика. В результате такого состояния структуры жидкого кристалла он рассеивает падающий на него свет равномерно во все стороны. Поэтому данное состояние нематика **называют динамическим рассеянием света**, при котором рассеяние падающего света происходит во всех направлениях, что облегчает визуальное наблюдение измененного состояния жидкого кристалла.

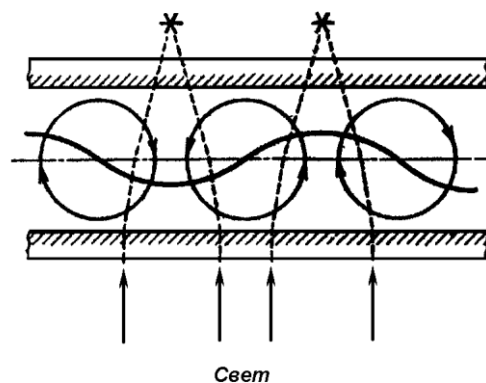


Рис. 57

Перечисленные структурные состояния нематика в электрическом поле могут быть использованы в электронике: для отображения и хранения информации, преобразования световых потоков и т.д.

Холестерики. Следует заметить, что шаг холестерической спирали чувствителен к электрическому полю, температуре, химическим примесям на поверхности холестерика, радиационному излучению и т.д.: при приложении, например, электрического поля шаг спирали увеличивается с ростом величины поля. При приближении к некоторому пороговому напряжению (несколько вольт) шаг холестерической спирали начинает резко увеличиваться, а при пороговом напряжении

$$E_p = \left(\frac{\pi^2}{2\rho_0} \right) \cdot \left(\frac{4\pi K}{\Delta\varepsilon} \right)^{1/2} \quad (7.3)$$

спираль вообще распрямляется. В результате этого холестерик переходит в нематик. Здесь ρ_0 – шаг холестерической спирали в отсутствие воздействия, который обычно выбирают в десятки раз меньше толщины слоя жидкого кристалла. Поэтому значение E_p оказывается больше, чем поле, необходимое для перехода Фредерикса. Если продолжать увеличивать напряжение на пленке холестерика, то далее можно обнаружить те же структурные преобразования, о которых говорилось при обсуждении свойств нематика: домены Капустина-Вильямса, динамическое рассеяние.

Поведение холестерика в электрических полях сложнее нематика. Так, динамическое рассеяние в холестериках может обладать эффектом памяти: рассеивающее состояние холестерика сохраняется в течение длительного времени после снятия поля. Время памяти, т.е. сохранения рассеивающего состояния, зависит от свойств конкретного холестерика и может колебаться от минут до нескольких лет. Наложение переменного электрического поля возвращает холестерик в исходное состояние.

7.5. Оптические свойства жидких кристаллов

Для понимания принципов работы оптических жидкокристаллических ячеек необходимо вспомнить раздел оптики, в котором рассматриваются особенности прохождения через оптически активный кристалл линейно поляризованного излучения с заданной поляризацией и направлением распространения излучения относительно оптической оси кристалла.

Нематики. Нематик является оптически одноосным кристаллом, оптическая ось которого совпадает с направлением директора. Поэтому оптические характеристики планарной и гомеотропной структур различны. В ячейке с гомеотропной структурой свет распространяется вдоль оптической оси (вдоль директора), поэтому, как во всяком одноосном кристалле, в этом случае оптические характеристики ячейки не зависят от поляризации света: свет проходит жидкокристаллическую оптическую ячейку без двулучепреломления, а значит без изменения фазы. При падении же света на ячейку с планарной структурой прохождение света сквозь нее зависит от поляризации падающего света: без изменения поляризации в ней распространяется только свет, поляризованный ортогонально директору. В случае поляризации по директору свет расщепляется на обыкновенную волну и необыкновенную, которые имеют разные скорости распространения и взаимно перпендикулярные поляризации. В итоге на выходе из оптической ячейки эти волны складываются в одну, вектор поляризации которой оказывается повернут на некоторый угол относительно поляризации на входе. Подбирая толщину ЖК или значения его диэлектрических проницаемостей ε_{\parallel} и ε_{\perp} , можно добиться того, что на выходе поляризация будет перпендикулярна к поляризации на входе в планарную оптическую ячейку. Используя поляроидные пленки, нанесенные на опорные поверхности, можно сделать гомеотропную структуру непрозрачной для света, скрестив верхний и нижний поляроиды (рис. 55, б). Планарная структура при таком же положении поляроидов пропускает свет, если ориентация поляроида не совпадает с направлением директора. Отсюда следует принцип, на основе которого с помо-

шью жидкокристаллической ячейки можно отображать информацию – путем прикладываемого напряжения переводить планарную текстуру в гомеотропную и обратно. В планарной структуре направление распространения света перпендикулярно оптической оси. Поэтому показатель преломления света различен для света, плоскость поляризации которого перпендикулярна направлению директора в слое, и света, поляризация которого содержит это направление. Знак оптической анизотропии может быть как положительным, так и отрицательным. В нематике с $\Delta\epsilon < 0$ фазовая скорость света, поляризованного перпендикулярно направлению директора, оказывается меньше фазовой скорости света, поляризованного в направлении директора. Переход Фредерикса, в котором при отсутствии поля нематик с $\Delta\epsilon > 0$, образует планарную текстуру, а при наложении поля переходит в гомеотропную, принято называть **S-эффектом**.

В **B-эффекте** исходная текстура гомеотропная, а поле переводит ее в планарную для чего требуется нематик с отрицательной диэлектрической анизотропией $\Delta\epsilon$. Для **T-эффекта** исходная планарная структура берется закрученной (ориентация молекул на одной поверхности составляет угол 90° относительно ориентации на другой). Приложенное поле, как и в случае S-эффекта, изменяет планарную ориентацию молекул нематика с $\Delta\epsilon > 0$ на гомеотропную. Работа ячейки, использующей T-эффект, основывается на свойстве закрученной твист-структуры вращать плоскость поляризации света.

Конструкция оптической ячейки. В качестве примера рассмотрим конструкцию планарной ячейки, предназначенной для отображения информации на твист-эффекте, как наиболее используемом в оптоэлектронике. Оптическая ячейка образована двумя стеклянными пластинками, на внутренние поверхности которых напылены металлические электроды, имеющие высокий коэффициент пропускания света. На внешние поверхности нанесены поляроидные пленки, пропускающие свет только определенной поляризации. Эти поляроидные пленки могут пропускать как свет одной поляризации, так и свет взаимно перпендикулярных поляризаций. Верхний поляроид пропускает поляризацию,

параллельную директору жидкого кристалла, который выбирают с положительной оптической анизотропией.

Исходная ориентация нематика закрученная, то есть мы имеем так называемую твист-ориентацию молекул (см. рис. 58, а). Свет падает на верхний поляризатор и становится плоскополяризованным в соответствии с его поляризацией. При отсутствии электрического поля (то есть в выключенном состоянии) свет, “следуя” твист-ориентации нематика, меняет свое направление в соответствии с оптической осью нематика и на выходе будет иметь то же

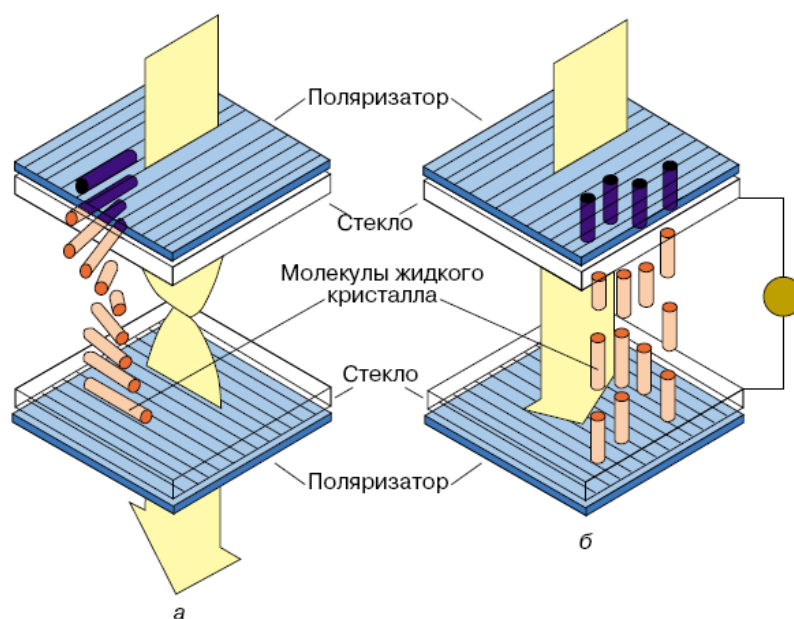


Рис.58

направление поляризации, что и нижний поляризатор (см. рис. 58, а). При включении электрического поля для нематического жидкого кристалла с положительной диэлектрической анизотропией ($\Delta\epsilon > 0$) произойдет переход от закрученной твист-ориентации к гомеотропной ориентации молекул, то есть длинные оси молекул повернутся в направлении, перпендикулярном к электродам, и спиральная структура разрушится (рис. 58, б). Теперь свет, не изменив направления исходной поляризации, совпадающей с поляризацией верхнего поляризатора, будет иметь направление поляризации, противоположное нижнему поляризатору, а они, как видно на рис. 58, б, находятся в скрещенном положении. В этом случае ячейка — «непрозрачна». Используя это свойство ячейки, можно создать электронный затвор оптического излучения.

Эффект «гость-хозяин». Электрооптические эффекты в ячейке могут проявляться не только в переходах из непрозрачного состояния в прозрачное или наоборот, но также в изменении окраски ячейки – это так называемый эффект «гость – хозяин». Сущность данного эффекта заключается в следующем. Если в нематик (хозяин) в небольшом количестве добавить молекулы дихроичного красителя (гость), которые ориентируются нематиком так же, как и собственные молекулы, то при переходе Фредерикса вместе с переориентацией молекул нематика происходит переориентация молекул красителя. В результате этого изменяется окраска жидкокристаллической ячейки из-за изменения ориентации молекул дихроика. Это происходит потому, что дихроичные молекулы красителя имеют линию поглощения света на строго определенной длине волны, причем величина коэффициента поглощения на этой волне зависит от взаимной ориентации поляризации падающего на него света и длинной оси молекул дихроика. Так что свет с поляризацией перпендикулярной оси дихроичной молекулы ею не поглощается, а с поляризацией параллельной оси дихроика – поглощается. В последнем случае краситель приобретает цвет, определяемый линией поглощения дихроика: это может быть и красный цвет ($\lambda=0,64$ мкм), и зеленый ($\lambda=0,55$ мкм) и т.д. В соответствующий линии поглощения цвет окрашивается оптическая ячейка.

Поскольку молекул красителя в нематик добавляют мало, все диэлектрические характеристики жидкого кристалла сохраняются. Можно ввести два красителя с различными частотами дихроичных линий поглощения и тем самым добиться двойного изменения окраски ячейки при переключении поля: один цвет до действия поля и другой – после перехода Фредерикса.

Смектики. Оптические свойства смектиков находятся в прямой зависимости от их строения. Смектик *A*-типа, как и нематик, является одноосной средой в оптическом отношении – в нем существует одно направление распространения света, совпадающее с директором, для которого показатель преломления не зависит от поляризации света.

Смектики С-типа являются оптически двухосными, т.к. ориентационные движения молекул в смектиках отличаются от таковых в нематике. Если в нематике молекулы могут свободно вращаться вокруг своих длинных осей, то в смектиках в общем случае вращательное движение заторможено, т.е. хотя молекула и вращается, но с наибольшей вероятностью она принимает определенное ориентационное положение. Поскольку свет «видит» усредненное положение молекул, то в случае свободного вращения усредненная молекула осесимметрична, а при заторможенном вращении она выглядит как часть плоскости (блиноподобна). Из теории анизотропных сред известно, что ансамблю осесимметричных молекул соответствует оптически одноосная среда, а системе молекул, у которых отсутствует ось вращения, соответствует двухосная среда.

В природе встречаются смектики, обладающие большим электрическим дипольным моментом. Происхождение дипольного момента связано с тем, что образующие смектик молекулы обладают собственным электрическим моментом: центры «тяжести» (говорят – центроиды) положительного и отрицательного зарядов в отдельной молекуле сдвинуты в пространстве относительно друг друга. Направление дипольного момента изменяется от слоя к слою, образуя спиральную структуру. Если эту спираль раскрутить приложением электрического поля, перпендикулярного оптической оси, то можно обнаружить дипольный момент. При этом резко изменяются оптические свойства смектика, что можно использовать в приборостроении.

Холестерики. Этот тип жидких кристаллов характеризуется спиральным поведением директора. Но с изменением направления директора изменяется диэлектрическая проницаемость. Следовательно, по толщине кристалла значение диэлектрической проницаемости изменяется от ε_{\parallel} – диэлектрическая проницаемость жидкого кристалла, когда поляризация света параллельна директору, до ε_{\perp} – диэлектрическая проницаемость жидкого кристалла, когда поляризация света перпендикулярна директору. Диэлектрические свойства холестерика периодичны по координате с периодом $d = \rho_0/2$. В холестерике периодом служит не шаг спирали ρ_0 , на котором директор совершает оборот на 2π , а

$p_0/2$, так как ориентации директора \vec{n} и $-\vec{n}$ эквивалентны, и период диэлектрических свойств холестерика

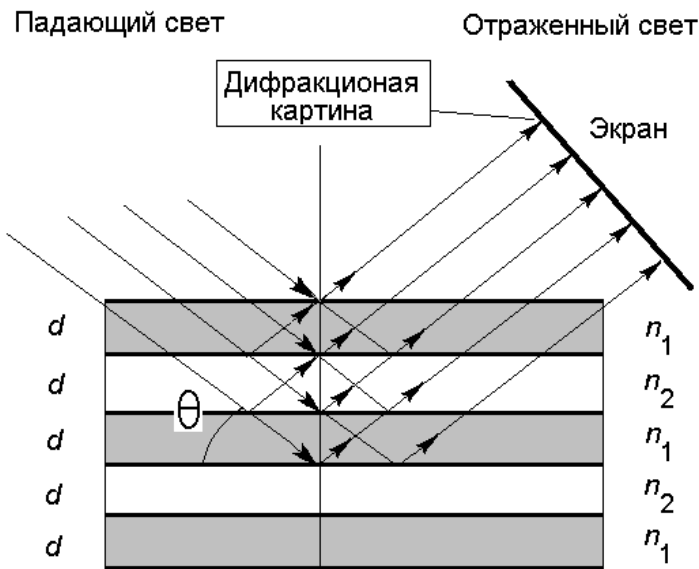


Рис. 59

оказывается вдвое меньше, чем расстояние, на котором директор совершает полный оборот. По этой причине холестерик можно представить в виде чередующихся слоев двух типов с различными значениями диэлектрической проницаемости: ϵ_{\parallel} и ϵ_{\perp} .

Поскольку шаг спирали выбирается в диапазоне видимого света, то свет будет испытывать дифракцию на периодической структуре холестерика. Если на такую структуру под углом падения θ (это угол между поверхностью кристалла и направлением падения света) падает белый свет (т.е. излучение с широким диапазоном значений длин волн), то в отраженном свете будет видна дифракционная картина, полученная от оптического излучения только одной длины волны λ из всего диапазона длин волн падающего излучения. Внешне это будет выглядеть как окрас слоя холестерика в цвет длиной λ (см. рис. 59). Это явление называется **дифракцией Вульфа-Брэгга**. Условие наблюдения этого вида дифракции таково:

$$\sin\theta = \frac{m\lambda}{2d},$$

где $d = p/2$ - толщина слоя неоднородности холестерика (взаимнообратная ориентация директора в холестерике на шаге спирали идентична по оптическим свойствам), p - шаг холестерической спирали, m - целое число, равное порядку дифракционного максимума.

Как указывалось ранее, шаг спирали холестерика чувствителен к внешним воздействиям. Его легко изменить, нагревая холестерик, помещая в элек-

трическое или магнитное поле и т.д. Отсюда следует один из принципов использования холестерических кристаллов в оптоэлектронике: плавное изменение цвета оптической ячейки в широких пределах при действии электрического поля, теплового или ионизирующего излучения.

7.6. Применение жидких кристаллов в оптоэлектронике

В виду разной изученности жидких кристаллов в современной оптоэлектронике используются преимущественно нематики и частично холестерики. Смектики не проявляют выраженного перехода Фредерикса, и потому в настоящее время их использование в электронике ограничено.

Жидкие кристаллы в оптоэлектронике применяют главным образом для создания модуляторов, построения дефлекторов оптического излучения, интегрально-оптических элементов различного назначения, систем обработки оптической информации – оптических транспарантов, управляемых электрическим полем или светом, устройств отображения информации. Использование нематиков базируется преимущественно на использовании эффекта Фредерикса, требующем импульсов напряжений малой амплитуды (до 5 Вольт). Это обстоятельство дополняет хорошую интегрируемость жидкокристаллических оптических элементов с микропроцессорной техникой и позволяет создавать программно управляемые экономные, малогабаритные и надежные оптоэлектронные элементы и устройства.

1. Жидкокристаллические модуляторы представляют собой оптическую планарную или гомеотропную ячейку, которую размещают между скрещенными или параллельными поляризаторами. Падающее оптическое излучение имеет линейную поляризацию. В зависимости от конструкции модулятора и используемого электрооптического эффекта при наложении электрического поля, создающего один из перечисленных выше структурных эффектов, его пропускная способность изменяется от нуля до 100% в соответствии с законом изменения

прикладываемого напряжения. Предельная частота модуляции такого модулятора обычно составляет несколько килогерц. Поэтому такие модуляторы чаще используются как оптические затворы. Однако скоро ЖК модуляторы будут работоспособны до сотен килогерц и выше.

2. Дефлекторы оптического излучения могут быть созданы на основе многочисленных электрооптических эффектов в жидких кристаллах: переходе Фредерикса, доменах Капустина-Вильямса, использовании нарушения полного внутреннего отражения и т.д. Назначение дефлектора – отклонение оптического луча в нужном направлении. Наиболее просто этого достичь, если в жидкокристаллической ячейке выбором треугольной топологии полевых электродов сформировать клиновидную призму, изменяющую свой относительный показатель преломления в электрическом поле от нуля до величины, определяемой анизотропией диэлектрической проницаемости.

3. Интегрально-оптические элементы на основе жидких кристаллов представляют собой элементы, встроенные в оптическую схему оптоэлектронного устройства. Это активные линзы, волноводные модуляторы, волноводные дефлекторы и т.д. Их применение позволяет программно управлять оптической схемой оптоэлектронного устройства, изменять его функции. Подавая на эти элементы в соответствии с алгоритмом работы устройства напряжения выше порога Фредерикса, можно ЖК призмой изменять направление распространения светового луча; активируя ЖК линзу, включать-выключать фокусировку; оптической ячейкой осуществлять модуляцию света цифровым сигналом. Однако ввиду значительных оптических потерь в ЖК интегрально-оптические элементы применяют в волноводных линиях.

4. Оптоэлектронные транспаранты. Для современных систем обработки информации большое значение имеют оптоэлектронные транспаранты, управляемые полем или светом. Транспаранты представляют собой пространственные модуляторы или, точнее, пространственные световые ключи, управляющие по заданному временному закону одним или несколькими параметрами светового потока. Жидкие кристаллы являются удобными рабочими веще-

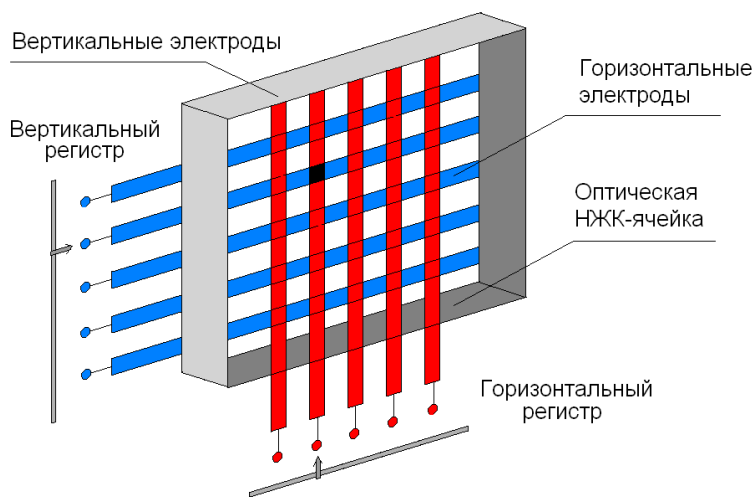


Рис. 60

представляет собой матрицу, составленную из большого числа независимых оптических ячеек. На матрицу нанесены прозрачные металлические электроды в виде полосок, направленных взаимно перпендикулярно на противоположных опорных стеклах (см. рис. 60). При подаче напряжения на определенную комбинацию вертикальных и горизонтальных полосок выбранные таким способом ячейки изменяют свои оптические свойства. На представленном рисунке переход Фредерикса происходит в ячейке с координатами (2, 4). Если транспарант работает на просвет, то обе опорные поверхности делают прозрачными. Если же транспарант работает на отражение, то одна опорная поверхность делается зеркальной.

Транспаранты, управляемые светом, отличаются от описанных выше тем, что в них жидкокристаллическая ячейка включается последовательно с высокоомным фоторезистором. Прикладываемое к транспаранту постоянное напряжение берется выше напряжения Фредерикса. В отсутствии освещения из-за большого падения напряжения на фоторе-

ствами для транспарантов ввиду малых управляющих напряжений и потребляемой мощности, хорошего контраста и возможности изготовления транспарантов практически любых размеров.

Так, жидкокристаллический транспарант, управляемый электрическим полем, представляет собой матрицу, составленную из большого числа независимых оптических ячеек. На матрицу нанесены прозрачные металлические электроды в виде полосок, направленных взаимно перпендикулярно на противоположных опорных стеклах (см. рис. 60). При подаче напряжения на определенную комбинацию вертикальных и горизонтальных полосок выбранные таким способом ячейки изменяют свои оптические свойства. На представленном рисунке переход Фредерикса происходит в ячейке с координатами (2, 4). Если транспарант работает на просвет, то обе опорные поверхности делают прозрачными. Если же транспарант работает на отражение, то одна опорная поверхность делается зеркальной.

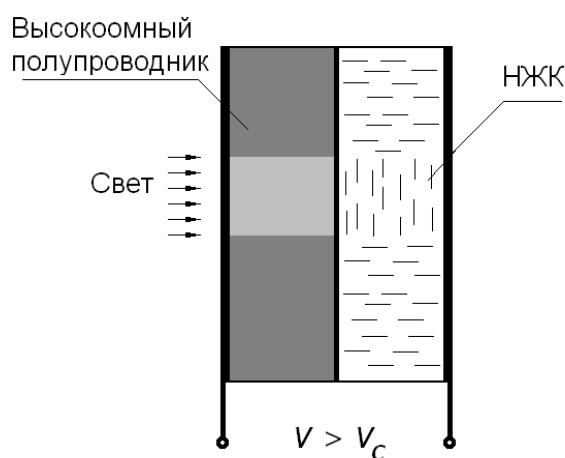


Рис.61

зисторе напряжения на оптической ячейке недостаточно для структурных изменений в ней. Освещение фотослоя повышает его проводимость в местах освещения, и тогда в освещенных точках транспаранта приложенное напряжение перераспределяется в пользу ЖК слоя. Это вызывает переход Фредерикса и изменение структуры ЖК, что обнаруживается в проходящем или отраженном свете. Таким образом, изображение на ЖК транспаранте в точности повторяет световую картинку, спроецированную на его заднюю стенку. Если проецируемое на транспарант изображение инфракрасное, то такой транспарант позволяет сделать его видимым.

Жидкокристаллические транспаранты используются для формирования двумерных массивов информации, выполнения логических операций, преобразования яркости излучения, перестраиваемой фильтрации изображения и т.д.

5. Устройства отображения информации на жидких кристаллах характеризуются дешевизной, низким энергопотреблением, малыми управляющими напряжениями и хорошей технологической совместимостью с интегральными микросхемами. Они используют переход Фредерикса в твист-структуре, а также эффект динамического рассеяния света в S-эффекте только в черно-белом цвете. Индикаторы третьего поколения полноцветные, что достигается применением эффекта «гость-хозяин» в нематике либо использованием холестерика с электрически регулируемым шагом спирали.

9. ПЕРСПЕКТИВЫ РАЗВИТИЯ ОПТОЭЛЕКТРОНИКИ

В настоящее время достаточно четко просматриваются три перспективных направления развития оптоэлектроники на ближайшие годы как одного из направлений развития твердотельной электроники: использование для создания приборов и устройств электроники новых материалов; применение новых физических явлений и принципов; дальнейшая миниатюризация приборов за счет использования квантовых эффектов в твердом теле.

Новые материалы. Применение новых материалов для создания элементов оптоэлектроники способно существенным образом способствовать развитию этой области науки. Так, изготовление генераторов оптического излучения и фотоприемных элементов на основе гетеропереходов (p - n -переходы, полученные на основе полупроводниковых материалов с различной шириной запрещенной зоны), создать систему «приемник-передатчик» на оптимальной длине волны для заданного оптического тракта. Это происходит за счет получения узкой спектральной характеристики обоих элементов связи, а также за счет возможности управления спектральной характеристикой в широком диапазоне длин волн.

Широкие возможности в области совершенствования элементной базы электроники и создания новых приборов сулит также использование нового типа полупроводникового материала: сверхрешеток, получаемых из тонких 100-ангстремных слоев полупроводниковых материалов с различной шириной запрещенной зоны. Подбором свойств исходных полупроводников можно получать многослойную структуру, которая будет иметь желаемую ширину запрещенной зоны.

Новые физические явления и принципы – электронная и оптическая синергетика для создания сверхскоростных бистабильных элементов; высокотемпературная сверхпроводимость; голографические оптические методы записи информации в кристаллах; переходы металл-полупроводник; использование эффекта фотонного эха; использование свойств «голубой» фазы в жидких кристаллах; использование нелинейных свойств жидких кристаллов для преобразования оптического излучения; применение твердотельных автоволн для создания пространственно-временных модуляторов излучения.

Разработка новых технологий. На этапе становления электроники (первые лет 15) основными направлениями ее развития были формирование номенклатуры базовых элементов и их реализация в виде дискретных приборов. Дальнейший прогресс этой области науки связан с переносом центра внимания на создание функционально сложных и конструктивно законченных приборов,

устройств, систем. Именно в них наиболее полно проявляются преимущества биоэлектроники перед микроэлектроникой. Особенно глобальные изменения в оптоэлектронике будут связаны с использованием идей и технологии нового направления в микроэлектронике – наноэлектроники (электроники атомного уровня), которое принципиальным образом изменит элементную базу всей современной оптоэлектроники, конструкторские подходы к созданию электронных устройств и систем.

9. ЛИТЕРАТУРА

1. Епифанов Ю.И. Физические основы микроэлектроники. – М.: Высшая школа, 1971. – 388 с.
2. Верещагин И.К., Косяченко Л.А., Кокин С.М. Введение в оптоэлектронику. – М.: Высшая школа, 1991. – 189 с.
3. Носов Ю.Р. Оптоэлектроника. – М.: Радио и связь, 1989. – 350 с.
4. Берг А., Дин П. Светодиоды: Пер. с англ. – М., 1973. – 88 с.
5. Мосс Т., Баррел Г., Эллис Б. Полупроводниковая оптоэлектроника. – М.: Мир, 1976.– 430 с.
6. Панков Ж. Оптические переходы в полупроводниках. – М.: Мир, 1973. – 393 с.
7. Ван дер Зил. Шумы. Источники, описание, измерения: Перевод с англ. – М.: Сов. радио, 1973. – 226 с.
8. Пикин С.А., Блинов Л.М. Жидкие кристаллы.– М.: Наука, 1982. -207 с.