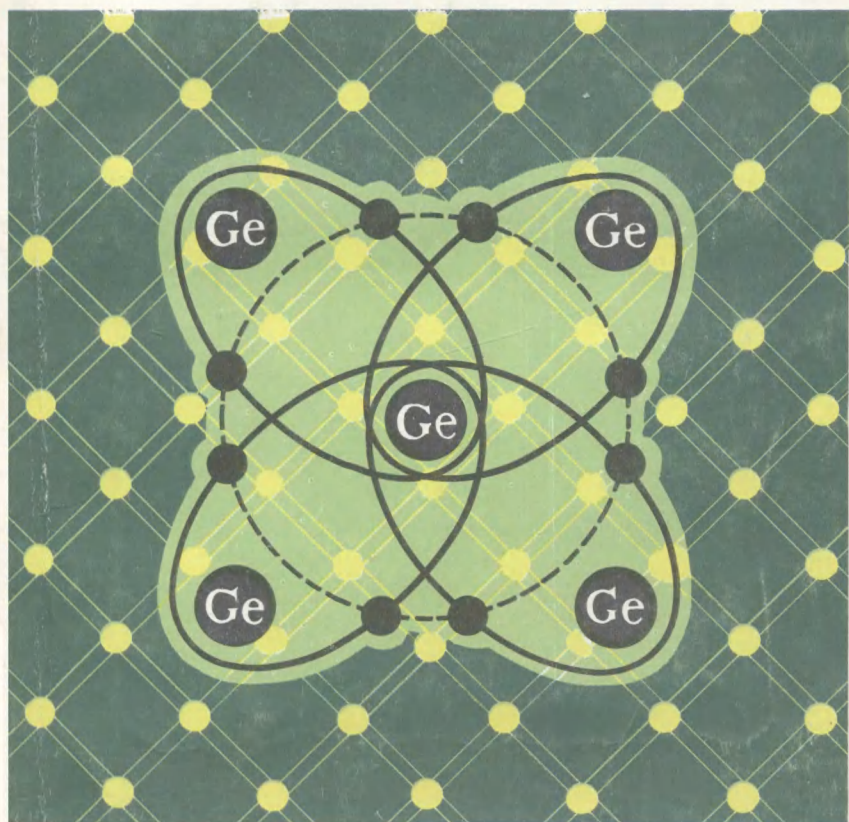


МР
И
знаний

А. М. ПОЛЯКОВ

Разгаданный полупроводник



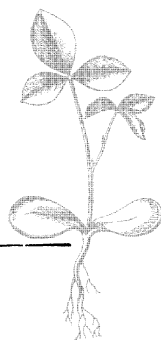
МИР ЗНАНИЙ

А.М. ПОЛЯКОВ

Разгаданный полупроводник

*Книга для внеклассного чтения
VIII—X классы*

Москва «Просвещение» 1981



Scan AAW

ББК 22.379
П54

Рецензент:
канд. физ.-мат. наук *О. Ф. Кабардин*

Поляков А. М.

П54 Разгаданный полупроводник: Кн. для внеклас.
чтения. VIII—X кл.— М.: Просвещение, 1981.—
160 с., ил.— (Мир знаний).

В книге на основе квантовых представлений и зонной теории дано доступное и вместе с тем достаточно глубокое изложение теории полупроводников и действия основных полупроводниковых приборов (диодов, триодов, лазеров).

Книга предназначена для учащихся старших классов средней школы, интересующихся физикой.

П $\frac{60601-530}{103(03)-81}$ 236—81 4306021100

ББК 22.379
531.9

© Издательство «Просвещение», 1981 г.

Глава 1.

О ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ

В ту пору, когда исследователи впервые столкнулись с полупроводниками, существовало четкое разделение всех твердых тел на две большие, принципиально различные по своим свойствам группы — на проводники (к которым относились все металлы) и изоляторы (или диэлектрики). Новые материалы нельзя было отнести ни к одной из этих групп: с одной стороны, они проводили электрический ток, хотя и гораздо хуже, чем металлические проводники, а с другой — проводимость обнаруживалась у них далеко не во всех случаях. Тем не менее само наличие проводимости у этих новых веществ предопределило их название — *полупроводники* (плохие проводники).

В дальнейшем было выявлено, что проводимость полупроводников отличается от проводимости металлов как по своей природе, так и по характеру влияния на нее различных внешних факторов. Так, влияние температуры на проводимость металлических проводников и полупроводников оказалось диаметрально противоположным: если у металлов повышение температуры вызывает постепенное уменьшение проводимости, то нагрев полупроводников приводит к резкому возрастанию их проводимости. Введение примесей влияет на проводимость металлических проводников и полупроводников также по-разному: в металлах наличие примеси приводит, как правило, к ухудшению их проводящих свойств, в то время как в полупроводниках введение даже ничтожного количества определенных примесей является причиной увеличения проводимости в десятки и сотни тысяч раз. Наконец, если направить на проводник пучок света или

поток каких-либо частиц, то это практически не скажется на его проводимости; в полупроводнике же воздействие излучения или бомбардировка частицами вызывает очень сильное увеличение проводимости.

Интересно отметить, что все эти свойства полупроводников в значительной степени характерны и для диэлектриков, поэтому было бы гораздо правильнее называть полупроводники полуизоляторами или полудиэлектриками.

Чтобы понять поведение полупроводников в различных условиях, объяснять их свойства, предсказывать новые эффекты, надо подробно рассмотреть особенности структуры этих веществ. Поэтому мы начнем рассказ с того, как устроены атомы вещества.

§ 1. КАК УСТРОЕНЫ АТОМЫ? АТОМ ВОДОРОДА

Из курса физики вы знаете, что атом любого вещества состоит из ядра и обращающихся вокруг него электронов. Такую модель атома предложил выдающийся английский физик Э. Резерфорд. Основываясь на этой модели, один из основоположников квантовой механики датский физик Н. Бор в 1913 году произвел первые правильные расчеты атома водорода, достаточно хорошо совпавшие с экспериментальными данными. Теория атома водорода, предложенная Бором, сыграла чрезвычайно важную роль в развитии квантовой механики, хотя в дальнейшем и претерпела существенные изменения.

Атом водорода. Постулаты Бора. Согласно модели Резерфорда—Бора атом водорода состоит из однократно заряженного положительного ядра и одного электрона, обращающегося вокруг него. В первом приближении можно предположить, что движение электрона происходит по траектории, представляющей собой окружность, в центре которой находится неподвижное ядро. В соответствии с требованиями классической электродинамики всякое ускоренное движение заряженного тела (в том числе и электрона) должно сопровождаться излучением электромагнитных волн. В рассматриваемой модели атома электрон движется с колоссальным центростремительным ускорением, и поэтому он должен был бы непрерывно испускать свет. При этом энергия его должна

была бы уменьшаться, а сам электрон должен был бы все ближе и ближе смещаться к ядру. Закончилось бы это тем, что электрон объединился бы с ядром («упал» бы на ядро). Однако ничего подобного не происходит и атомы в невозбужденном состоянии не испускают света. Для объяснения этого факта Бор предложил два основных постулата.

Согласно первому постулату Бора электрон может находиться только на таких орбитах, для которых момент количества движения электрона (то есть произведение количества движения электрона mv на радиус орбиты r) кратен значению $\frac{h}{2\pi}$ (где h — постоянная

Планка)¹. Пока электрон находится на одной из таких орбит, излучения энергии не происходит. Каждой разрешенной орбите электрона соответствует определенная энергия, или определенное *энергетическое состояние атома*, называемое *стационарным*. Находясь в стационарном состоянии, атом не излучает света. Математически первый постулат Бора можно записать так:

$$mvr = n \frac{h}{2\pi},$$

где $n=1, 2, 3, \dots$ — некоторое целое число, называемое *главным квантовым числом*.

Второй постулат Бора содержит утверждение, что поглощение или испускание света атомом происходит только при переходах атома из одного стационарного состояния в другое. При этом энергия поглощается или испускается определенными порциями, квантами, значение которых $h\nu$ определяется разностью энергий, соответствующих начальному и конечному стационарным состояниям атома:

$$h\nu = W_m - W_n,$$

где W_m — энергия начального состояния атома, W_n —

¹ Постоянная Планка является универсальной физической константой и имеет смысл произведения энергии на время, называемого в механике *действием*. Благодаря тому что величина h представляется как бы элементарным количеством действия, постоянная Планка называется *квантом* (порцией) *действия*. Введение в физику понятия кванта действия было началом создания важнейшей физической теории XX века — квантовой теории. Квант действия очень мал: $h = 6,62 \cdot 10^{-34}$ Дж·с.

энергия его конечного состояния, ν — частота света, испущенного или поглощенного атомом. Если $W_m > W_n$, то энергия излучается атомом, если же $W_m < W_n$, то поглощается.

Позже кванты света получили название *фотонов*.

Таким образом, по теории Бора электрон в атоме не может изменять свою траекторию постепенно (непрерывно), а может лишь «перепрыгивать» с одной стационарной орбиты на другую. При переходе со стационарной орбиты, более удаленной от ядра, на стационарную орбиту, расположенную ближе к ядру, как раз и происходит испускание света.

Радиусы орбит и энергетические уровни атомов. Радиусы разрешенных электронных орбит можно найти, используя закон Кулона, соотношения классической механики и первый постулат Бора. Их значения определяются выражением

$$r = n^2 \frac{h^2}{4\pi^2 m e^2}.$$

Самой близкой к ядру разрешенной орбите соответствует $n=1$. Используя полученные экспериментально значения величин m , e и h , находим для ее радиуса значение

$$r_1 = 0,53 \cdot 10^{-8} \text{ см.}$$

Эта величина как раз и принимается за радиус атома водорода. Любая другая орбита с квантовым числом n имеет радиус

$$r_n = n^2 r_1.$$

Таким образом, радиусы последовательно расположенных электронных орбит возрастают как квадрат числа n (рис. 1).

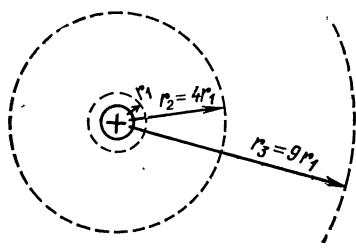


Рис.

Значение полной энергии атома, соответствующей нахождению электрона на n -ой орбите, определяется формулой

$$W_n = - \frac{2\pi^2 m e^4}{n^2 h^2}.$$

Эти значения энергий называются *энергетическими уровнями* атома. Если по вертикальной оси откладывать возможные значения энергии атома, то можно получить так называемый *энергетический спектр* разрешенных состояний атома (рис. 2).

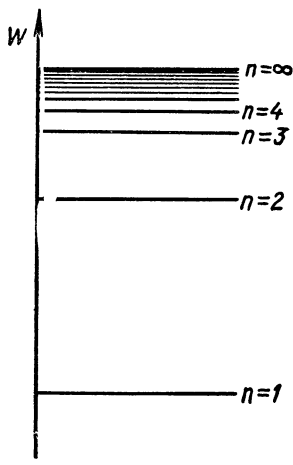


Рис. 2

Расстояние между последовательно расположенными энергетическими уровнями быстро уменьшается. Это можно легко объяснить: увеличение энергии атома (за счет поглощения атомом энергии извне) сопровождается переходом электрона на все более удаленные орбиты, где взаимодействие между ядром и электроном становится более слабым. По этой причине переход между соседними удаленными орбитами связан с очень малым изменением энергии. Энергетические уровни при этом располагаются настолько близко, что спектр становится практически непрерывным. В верхней части непрерывный спектр заканчивается уровнем ионизации атома ($n = \infty$), соответствующим полному отделению электрона от ядра (электрон становится свободным).

Знак «—» в выражении для полной энергии атома указывает на то, что энергия атома тем меньше, чем ближе к ядру находится электрон. Для того чтобы удалить электрон от ядра, необходимо затратить определенную энергию, то есть сообщить атому некоторую энергию извне. Энергия атома принимается равной нулю при $n = \infty$, т. е. в случае, когда атом ионизирован. Именно поэтому значениям $n \neq \infty$ соответствуют отрицательные значения энергии. Уровню с $n = 1$ соответствуют минимальная энергия атома и минимальный радиус разрешенной орбиты электрона. Этот уровень называется *основным* или *невозбужденным*. Уровни с $n = 2, 3, 4, \dots$ называются *уровнями возбуждения*.

Квантовые числа. В теории Бора предполагалось, что электронные орбиты имеют вид окружностей. Эта тео-

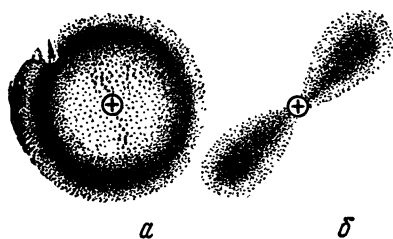


Рис. 3

рия дала достаточно хорошие результаты только при рассмотрении самого простого атома — атома водорода. Но уже при расчете атома гелия она не смогла дать количественно правильные результаты. Определенным шагом вперед была планетарная модель атома,

предполагавшая движение электронов подобно планетам солнечной системы по эллиптическим орбитам, в одном из фокусов которых располагалось ядро. Однако и эта модель быстро исчерпала себя, не дав ответа на многие вопросы.

Это связано с принципиальной невозможностью определения характера движения электрона в атоме. В доступном нашему наблюдению макром мире нет аналогов этого движения. Мы не можем не только проследить путь движения электрона, но и даже определить точно его местонахождение в какой-либо определенный момент времени. Само понятие орбиты, или траектории движения электрона в атоме, лишено физического смысла. Никакой определенной последовательности появления электрона в различных точках пространства установить нельзя, он оказывается как бы «размазанным» в некоторой области, называемой обычно *электронным облаком*. Облако это, например, для невозбужденного атома водорода имеет форму шара, но плотность его не одинакова. Вероятность обнаружения электрона будет наибольшей вблизи сферы с радиусом r_1 , соответствующим радиусу первой боровской орбиты. В дальнейшем под орбитой электрона в атоме мы будем понимать геометрическое место точек, которые характеризуются наибольшей вероятностью обнаружения электрона, или, другими словами, область пространства с наибольшей плотностью электронного облака.

Всегда сферическим электронное облако будет лишь для случая невозбужденного состояния атома водорода, когда главное квантовое число $n=1$ (рис. 3, *a*). Если же $n=2$, то, помимо сферического облака, размеры которого будут теперь в четыре раза больше, электрон может со-

здать облако в виде своеобразной гантели (рис. 3, б). С появлением несферичности области преимущественной локализации электрона (электронного облака) связано введение второго квантового числа l , называемого *орбитальным квантовым числом*. Каждому значению главного квантового числа n соответствуют положительные целочисленные значения квантового числа l от нуля до $(n-1)$:

$$l = 0, 1, 2, \dots, (n-1).$$

Так, если $n=1$, то l имеет единственное значение, равное нулю. Если же $n=3$, то l может принимать значения 0, 1, 2. При $n=1$ имеется только сферическая орбита, поэтому и $l=0$. Когда $n=2$, возможны как сферическая, так и гантелеобразные орбиты, поэтому и l может быть разным либо нулю, либо единице.

Если $n=3$, то $l=0, 1, 2$. Электронное облако, соответствующее значению $l=2$, приобретает уже довольно сложный характер. Для нас, однако, важна не форма электронного облака, а то, какая ему соответствует энергия атома.

Энергия атома водорода определяется только значением главного квантового числа n и не зависит от значения орбитального числа l . Иначе говоря, если $n=3$, то атом будет иметь определенную энергию W_3 независимо от того, на какой из возможных орбит, соответствующих данному значению n и различным возможным значениям l , находится электрон. Это означает, что при возвращении с уровня возбуждения на основной уровень атом будет испускать фотоны, энергия которых не зависит от значения l .

Рассматривая пространственную модель атома, необходимо иметь в виду, что электронные облака в нем имеют строго определенную ориентацию. Положение электронного облака в пространстве относительно выбранного каким-либо образом направления задается *магнитным квантовым числом* m , которое может принимать целочисленные значения от $-l$ до $+l$, включая 0. При данной форме (данном значении l) электронное облако может иметь несколько различных ориентаций в пространстве. При $l=1$ их будет три, соответствующих значениям магнитного квантового числа m , равным -1 , 0 и $+1$. Если $l=2$, то различных ориентаций электронно-

го облака будет 5, соответствующих значениям $m = -2, -1, 0, +1$ и $+2$. Естественно, что если уж форма электронного облака в свободном атоме водорода не влияет на энергию атома, то тем более не влияет на энергию атома ориентация этого облака в пространстве.

Наконец, при более детальном рассмотрении экспериментальных данных выяснилось, что сами электроны могут находиться на орбитах в двух возможных состояниях, определяемых направлением так называемого *спина* электрона.

Но что такое спин электрона?

В 1925 году английские физики Дж. Уленбек и С. Гоудсмит для объяснения тонкой структуры линий в оптических спектрах некоторых элементов предложили гипотезу, согласно которой каждый электрон вращается вокруг своей собственной оси подобно волчку или веретену. При таком вращении электрон приобретает некоторый момент импульса, который и получил название *спина* (в переводе с английского *спин* означает вращение, веретено). Поскольку вращение может происходить по часовой стрелке или против, то и спин (иначе говоря, вектор момента импульса) может иметь два направления. В единицах $\frac{h}{2\pi}$ спин равен $1/2$, а благодаря различным направлениям имеет знак «+» или «-». Таким образом, ориентация электрона на орбите определяется *спиновым квантовым числом* σ , равным $\pm 1/2$. Отметим, что и ориентация спина, как и ориентация орбиты электрона, не влияет на энергию атома водорода, находящегося в свободном состоянии.

Более поздние исследования и расчеты показали, что объяснить спин электрона простым вращением его вокруг оси нельзя. При подсчете угловой скорости вращения электрона для объяснения экспериментальных данных выяснилось, что линейная скорость точек, лежащих на экваторе электрона (в предположении, что электрон имеет шарообразную форму), должна быть больше скорости света, чего не может быть. Спин является некоторой неотъемлемой характеристикой электрона, такой, например, как его масса или заряд.

Квантовые числа — адрес электрона в атоме. Итак, мы выяснили, что для описания движения электрона в атоме, или, как говорят физики, для определения

состояния электрона в атоме, необходимо задать набор из четырех квантовых чисел: n , l , m и σ .

Главное квантовое число n определяет, грубо говоря, размеры электронной орбиты. Чем больше n , тем большее пространство охватывает соответствующее электронное облако. Задаваясь значением n , мы тем самым определяем номер электронной оболочки атома. Само число n может принимать любые целочисленные значения от 1 до ∞ :

$$n = 1, 2, 3, \dots$$

Орбитальное квантовое число l определяет форму электронного облака. Из всей совокупности орбит, относящихся к одному и тому же значению n , орбитальное число l выделяет орбиты, имеющие одинаковую форму. Каждому значению l соответствует своя подоболочка. Число подоболочек равно n , так как l может принимать значения от 0 до $(n-1)$:

$$l = 0, 1, 2, \dots (n-1).$$

Магнитное квантовое число m определяет пространственную ориентацию орбиты в группе орбит, имеющих одинаковую форму, то есть относящихся к одной подоболочке. В каждой подоболочке насчитывается $(2l+1)$ различно ориентированных орбит, поскольку m может принимать значения от 0 до $\pm l$:

$$m = -l, -(l-1), \dots, -1, 0, +1, \dots, +(l-1), +l.$$

Наконец, спиновое квантовое число σ определяет ориентацию спина электрона на заданной орбите. Значений у σ всего два:

$$\sigma = \pm \frac{1}{2}.$$

Рассматривая атом водорода и оперируя понятиями «оболочка», «подоболочка», «орбита», мы говорили не столько о строении атома, сколько о возможностях, открывающихся перед единственным электроном, содержащимся в этом атоме. Электрон в атоме водорода может переходить с оболочки на оболочку и с орбиты на орбиту в пределах одной оболочки.

Гораздо сложнее оказывается картина распределения электронов и возможностей их переходов в многоэлектронных атомах,

§ 2. МНОГОЭЛЕКТРОННЫЕ АТОМЫ

Принцип запрета Паули. При рассмотрении вопроса о строении многоэлектронных атомов весьма важным является принцип запрета, сформулированный в 1925 году известным швейцарским физиком В. Паули. Согласно этому принципу в атоме в любом состоянии, описываемом набором четырех квантовых чисел (n, l, m, σ), не может находиться более одного электрона. Иначе говоря, на любой атомной стационарной орбите одновременно может находиться только один или два электрона, причем в последнем случае спины электронов должны иметь противоположные направления, то есть для одного электрона $\sigma = +1/2$, а для другого $\sigma = -1/2$.

Принимая во внимание принцип запрета Паули и зная количество стационарных орбит, характеризующихся различными квантовыми числами, можно точно определить возможное число электронов в каждой оболочке и подоболочке атома (см. табл. 1).

Как расселяются электроны по оболочкам. Первая оболочка с главным квантовым числом $n=1$ не делится на подоболочки, так как ей соответствует только одно значение l , равное нулю. А поскольку и $m=0$, то можно сделать вывод, что первая оболочка состоит всего из одной орбиты, на которой согласно принципу Паули могут находиться всего два электрона.

Вторая оболочка ($n=2$) состоит из двух подоболочек, так как l может быть равно 0 или 1. В атомной физике при описании подоболочек принято вместо численных значений l употреблять буквенные обозначения. Так, независимо от значения главного квантового числа n все подоболочки, характеризующиеся значением $l=0$, обозначаются буквой s , подоболочки с $l=1$ обозначаются буквой p , при $l=2$ вводится обозначение d и т. д. В связи с этим говорят, что вторая оболочка состоит из s - и p -подоболочек. При этом s -подоболочка ($l=0$) состоит из одной круговой орбиты и в ней могут расположиться всего два электрона, в то время как p -подоболочка состоит из трех орбит (m может быть равно $-1, 0$ и $+1$) и на ней могут находиться шесть электронов. Общее же число электронов, могущих расположиться на второй оболочке, равно восьми.

Аналогичным образом можно подсчитать возможное

Т а б л и ц а 1

Квантовые числа				Обозначение уровня подоболочки	Число орбит в подоболочке	Число электронов на уровне	Общее число электронов в оболочке
n	l	m	σ				
1	0	0	$\pm \frac{1}{2}$	1s	1	2	2
2	0	0	$\pm \frac{1}{2}$	2s	1	2	8
	1	-1, 0, +1	$\pm \frac{1}{2}$	2p	3	6	
3	0	0	$\pm \frac{1}{2}$	3s	1	2	18
	1	-1, 0, +1	$\pm \frac{1}{2}$	3p	3	6	
	2	-2, -1, 0, +1, +2	$\pm \frac{1}{2}$	3d	5	10	

число электронов на любой оболочке и подоболочке. Например, на 3d-подоболочке ($n=3$, $l=2$) может быть 10 электронов: по два электрона на каждой из пяти орбит, характеризующихся различными значениями квантового числа m . Максимальное число электронов, могущих разместиться в той или иной подоболочке, равно $2(2l+1)$. В спектроскопии различным оболочкам обычно присваивают буквенные обозначения — термы: первая оболочка обозначается буквой K , вторая — L , третья — M и т. д.

В то время как в атоме водорода единственный электрон находится в центрально-симметричном поле атомного ядра и энергия электрона определяется только значением главного квантового числа n и не зависит от значений остальных квантовых чисел, в многоэлектронных атомах каждый электрон находится в поле, создаваемом как ядром, так и остальными электронами. Поэтому энергия электрона в многоэлектронных атомах оказывается зависящей не только от главного квантового

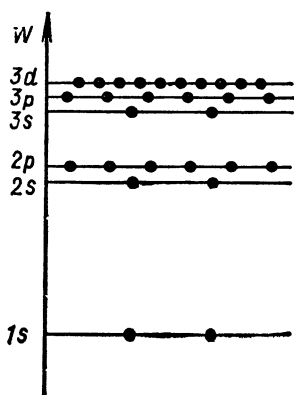


Рис. 4

числа n , но и от орбитального числа l . При этом она остается независимой от значений m и σ .

Такая специфика много-электронных атомов приводит к существенным отличиям их энергетического спектра от спектра атома водорода. На рисунке 4 приведена часть спектра многоэлектронного атома (энергетические уровни трех первых оболочек атома). Кружочки на уровнях обозначают максимальное число электронов, могущих располо-

житься в соответствующей подоболочке.

Известно, что всякая система, будучи представлена самой себе, стремится перейти в состояние с наименьшей энергией. Не является в этом смысле исключением и атом. При заполнении оболочек атома электроны стремятся занять самые нижние уровни, и если бы не ограничения, выражаемые принципом запрета Паули, то они все расположились бы на первом уровне. В атоме водорода единственный электрон занимает самую нижнюю орбиту, принадлежащую уровню $1s$. В атоме гелия на этой же орбите располагается и второй электрон, и первая оболочка атомной структуры оказывается заполненной. Отметим, что гелий является инертным газом, и его высокая устойчивость обусловлена именно заполненностью внешней оболочки.

В атоме лития имеются всего три электрона. Два из них располагаются в первой оболочке, а третий — во второй оболочке с $n=2$ (его размещение на первой оболочке запрещено принципом Паули). Литий является щелочным металлом с валентностью, равной 1. Это означает, что электрон, находящийся на второй оболочке, слабо связан с атомным остатком и легко отделяется от него. Об этом можно судить и по значению потенциала ионизации, который для лития равен всего 5,37 В, тогда как для гелия он равен 24,45 В.

По мере увеличения числа электронов в атомах происходит заполнение лежащих выше подоболочек и обо-

лочек. Так, начиная с бора, имеющего 5 электронов, происходит заполнение $2p$ -подоболочки, завершающееся в атоме неона, имеющего полностью заполненную вторую оболочку и характеризующегося благодаря этому высокой стабильностью инертного газа. Одиннадцатый электрон в атоме натрия начинает заселение третьей оболочки (подоболочки $3s$) и т. д.

§ 3. ВЫРОЖДЕНИЕ ЭНЕРГЕТИЧЕСКИХ УРОВНЕЙ В СВОБОДНЫХ АТОМАХ. СНЯТИЕ ВЫРОЖДЕНИЯ ПОД ВЛИЯНИЕМ ВНЕШНЕГО ВОЗДЕЙСТВИЯ

Вырожденные состояния. Мы уже отмечали, что в многоэлектронных атомах энергия электронов определяется только значениями квантовых чисел n и l и не зависит от значений чисел m и σ . Это наглядно видно и из энергетического спектра, приведенного на рисунке 4. В самом деле, например, все 6 электронов, находящихся на $3p$ -подоболочке, хотя и характеризуются различными m и σ , имеют одну и ту же энергию. Состояния, описываемые различными наборами квантовых чисел, но имеющие одну и ту же энергию, называются *вырожденными*. Соответственно и уровни, к которым относятся эти состояния, также называются вырожденными. Однако вырожденность уровней сохраняется до тех пор, пока атомы находятся в свободном состоянии. Если же атомы поместить в достаточно сильное магнитное или электрическое поле, то вырожденность уровней частично или полностью снимается. Выясним, как это происходит на примере снятия вырождения по квантовому числу m .

Внешнее поле снимает вырождение. Различные значения числа m отражают различную ориентацию в пространстве одинаковых электронных орбит. В отсутствие внешнего поля различная ориентация орбит действительно не влияет на энергию электронов. Однако если поместить атом во внешнее поле, то оно по-разному будет влиять на электроны, находящиеся на орбитах, ориентированных различным образом по отношению к его направлению. Благодаря этому изменения энергии электронов, находящихся на одинаковых по форме, но различных по ориентации орбитах, будут различными как по модулю, так и по знаку: энергия одних электронов несколько возрастет, а других — уменьшится. При этом

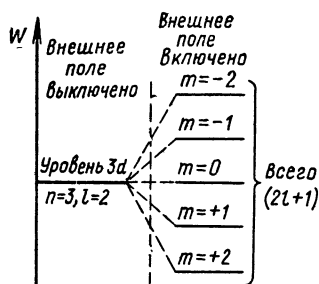


Рис. 5

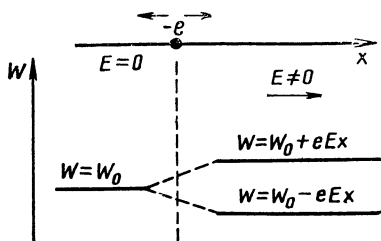


Рис. 6

и уровни в спектре, отражающие энергию различных электронов, изменят свое расположение. Более того, вместо одного энергетического уровня, к которому относились все электроны, находящиеся на одинаковых орбитах, в спектре появятся несколько подуровней, число которых будет равно числу различным образом ориентированных одинаковых орбит, то есть числу возможных значений квантового числа m . Пример воздействия внешнего электрического поля на уровень $3d$, для которого $n=3$ и $l=2$, приведен на рисунке 5. Мы видим, что происходит одновременно и расщепление уровня на подуровни, и смещение подуровней.

Такой процесс внесения различия в ранее неразличимые с энергетической точки зрения вырожденные уровни называется *снятием вырождения*. Проиллюстрируем снятие вырождения еще одним примером.

Рассмотрим электрон с некоторой энергией W_0 , находящийся в одномерном пространстве, характеризуемом координатой x (рис. 6). В отсутствие внешнего поля состояние такого электрона описывается одним энергетическим уровнем W_0 , независимо от направления его движения. Иначе говоря, в отсутствие внешнего поля имеется двукратное вырождение энергетического уровня W_0 . При включении внешнего электрического поля, направленного, например, вдоль оси x , энергия электрона становится зависящей от направления движения. При движении в направлении оси x электрон будет тормозиться внешним полем и энергия его будет равна $W_0 - eEx$ (где x — пройденное расстояние),

При движении в противоположном направлении энергия электрона становится равной $W_0 + eEx$. Соответственно и на энергетическом спектре появление двух состояний электрона отражается расщеплением вырожденного уровня W_0 на два невырожденных: $W_0 - eEx$ и $W_0 + eEx$, т. е. происходит снятие вырождения под действием внешнего электрического поля. Снятие вырождения может происходить и при воздействии на вырожденную систему внешнего магнитного поля.

§ 4. ОБРАЗОВАНИЕ ЭНЕРГЕТИЧЕСКИХ ЗОН В КРИСТАЛЛАХ

Расщепление энергетических уровней в кристалле. Произведем мысленно следующий эксперимент. Возьмем N атомов некоторого вещества и расположим их в пространстве на достаточно большом расстоянии друг от друга, но так, чтобы их размещение воспроизводило структуру кристалла этого вещества. Пока расстояния между атомами велики, можно пренебречь их взаимодействием и считать их свободными. В каждом из них имеются вырожденные уровни с кратностью вырождения, равной числу различным образом ориентированных одинаковых орбит в соответствующих подболочках. Начнем теперь сближать атомы, сохраняя их прежнее взаимное расположение. По мере сближения атомы начинают испытывать воздействие от приближающихся соседей, аналогичное тому, которое они испытывали бы при помещении во внешнее электрическое поле. Чем меньше расстояния между атомами, тем более сильным оказывается взаимодействие между ними. Благодаря этому взаимодействию происходит снятие вырождения с энергетических уровней, характеризующих свободные атомы: каждый вырожденный уровень расщепляется на $(2l + 1)$ невырожденных подуровней. В кристалле все атомы находятся, вообще говоря, в одинаковых условиях (за исключением тех, которые образуют внешнюю границу кристалла). Поэтому может показаться, что в общий энергетический спектр, характеризующий кристалл в целом, каждый из атомов должен внести одинаковый набор расщепленных невырожденных подуровней: по одному подуровню $1s$, по три подуровня $2p$, по пять подуровней $3d$ и т. д. На каждом подуровне могут расположиться 2 электрона с противоположными спинами. И хотя

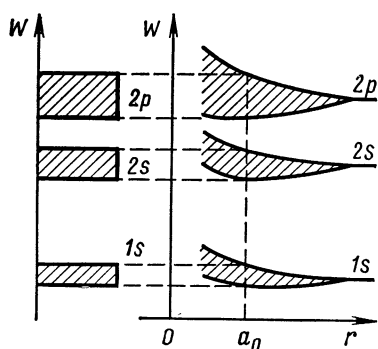


Рис. 7

такое расщепление в действительности происходит, соответствующие подуровни, получившиеся из одинаковых атомных уровней, оказываются отличными друг от друга по энергии: одни из них в энергетическом спектре кристалла располагаются несколько выше исходных уровней отдельных атомов, другие — несколько ниже. Объясняется это различие проявлением принципа Паули, обобщенным на весь кристалл как единый коллектив. Согласно этому принципу в кристалле не может быть двух одинаковых по энергии невырожденных подуровней. Поэтому каждый энергетический уровень при образовании кристалла расщепляется в энергетическую зону, состоящую из $N(2l+1)$, отличающихся по энергии невырожденных подуровней. Так, уровень $1s$ расщепляется в зону $1s$, состоящую из N подуровней, на которых могут расположиться $2N$ электронов, уровень $2p$ размывается в зону $2p$, состоящую из $3N$ подуровней и способную вместить $6N$ электронов и т. д.

Схематически процесс образования энергетических зон кристалла из дискретных энергетических уровней отдельных атомов показан на рисунке 7. Чем меньше расстояние r , тем большему воздействию подвергаются атомы со стороны соседей и тем больше «размытие» уровней. Энергетический спектр кристалла определяется размытостью уровней, соответствующей межатомному расстоянию a_0 , характерному для данного кристалла.

Степень размытия уровней зависит от глубины их расположения в атоме. Внутренние электроны характеризуются сильной связью с ядром и сильной экранировкой от внешнего воздействия наружными электронными оболочками, поэтому соответствующие им уровни испытывают слабое размытие. Электроны, находящиеся на внешних оболочках, естественно, подвергаются наибольшему воздействию со стороны поля кристаллической ре-

шетки, и соответствующие им уровни претерпевают наибольшее размытие. Следует отметить, что размытие уровней в зоны происходит независимо от того, имеются на соответствующих атомных уровнях электроны или они пусты. В последнем случае размытие уровней отражает расширение диапазона возможных энергий, которые может приобрести электрон в кристалле.

Разрешенные и запрещенные зоны. Из приведенного выше рассмотрения следует, что каждому разрешенному атомному энергетическому уровню в кристалле соответствует целая зона разрешенных значений энергии — *разрешенная зона*. Разрешенные зоны чередуются с зонами запрещенных энергий, или *запрещенными зонами*. В чистом кристалле электроны не могут обладать энергией, лежащей в диапазоне запрещенных зон. Чем выше по шкале энергии расположен разрешенный атомный уровень, тем более размытой оказывается соответствующая зона. Ширина же запрещенных зон по мере увеличения энергии становится все более узкой.

Расстояние между подуровнями в разрешенной зоне очень мало. В реальных кристаллах с размерами в пределах от 1 до 100 см³ расстояние между подуровнями по порядку величины лежит в пределах от 10^{-22} до 10^{-24} эВ. Это различие в энергии настолько ничтожно, что обычно зоны рассматриваются как непрерывные. Тем не менее сам факт дискретности подуровней в зонах и то обстоятельство, что число подуровней в зоне всегда конечно, играют принципиальную роль во всей физике кристаллических сред, так как именно в зависимости от степени заполнения зон электронами все твердые тела делятся на проводники, полупроводники и диэлектрики.

§ 5. ЗАПОЛНЕНИЕ ЭНЕРГЕТИЧЕСКИХ ЗОН ЭЛЕКТРОНАМИ

Заполненные уровни создают заполненные зоны, а пустые уровни — пустые зоны. Поскольку энергетические зоны в твердых телах образуются из уровней отдельных атомов, то вполне очевидно, что и заполнение их электронами будет определяться в первую очередь степенью заполненности электронами соответствующих атомных уровней.

Возьмем для примера кристалл лития. В свободном состоянии атомы лития имеют по три электрона. Два из

них расположены на $1s$ -оболочке, полностью заполняя ее. Третий электрон принадлежит $2s$ -подоболочке, заполняя ее лишь наполовину. Соответственно при образовании кристалла зона $1s$ оказывается полностью заполненной, а зона $2s$ — заполненной только наполовину; зоны $2p$, $3s$, $3p$ и т. д. в невозбужденном кристалле лития оказываются пустыми, поскольку были пустыми уровни, из которых эти зоны образовались.

Аналогичная картина наблюдается у всех щелочных металлов. Так, при образовании кристалла натрия зоны $1s$, $2s$ и $2p$ оказываются полностью заполненными, так как в атомах натрия соответствующие уровни полностью укомплектованы электронами (на уровне $1s$ — два электрона, на уровне $2s$ — два электрона, а на уровне $2p$ — шесть электронов). Одиннадцатый электрон в атоме натрия заполняет уровень $3s$ только наполовину, благодаря этому и зона $3s$ заполняется электронами наполовину.

При образовании кристаллов из атомов с полностью укомплектованными уровнями создающиеся зоны в общем случае также оказываются целиком заполненными. Так, если бы мы создавали кристалл из атомов неона, то в энергетическом спектре такого кристалла зоны $1s$, $2s$, $2p$ были бы целиком заполненными (в каждом атоме неона имеется по 10 электронов, которые и заполняют соответствующие энергетические уровни). Остальные зоны, расположенные выше (зоны $3s$, $3p$ и т. д.), оказались бы пустыми.

Перекрывание энергетических зон в кристалле. В некоторых случаях вопрос о заполнении энергетических зон электронами оказывается более сложным. Это относится к кристаллам щелочноземельных элементов и к кристаллам, имеющим решетку типа алмаза, среди которых наибольший интерес представляют кристаллы типичных полупроводников — кремния и германия.

На первый взгляд кристаллы щелочноземельных элементов должны иметь в энергетическом спектре только заполненные и пустые зоны. В самом деле, атомы, например, бериллия, имеющие по четыре электрона, характеризуются двумя заполненными уровнями $1s$ и $2s$. Атомы магния со своими 12 электронами также имеют заполненные уровни $1s$, $2s$, $2p$ и $3s$. Однако в кристаллах щелочноземельных элементов верхние энергетические зоны, в создании которых принимают участие заполнен-

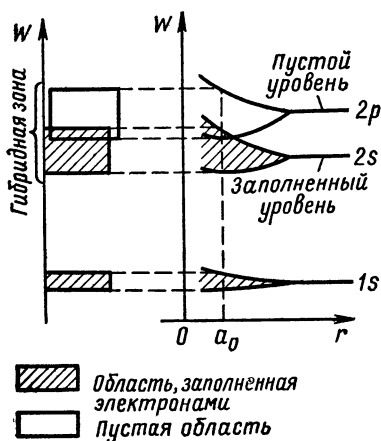


Рис. 8

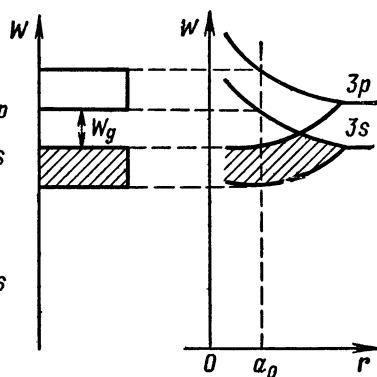


Рис. 9

ные атомные уровни, в действительности оказываются заполненными лишь частично. Объясняется это тем, что в процессе образования кристаллов этих элементов энергетические зоны, соответствующие верхним уровням, размываются настолько сильно, что происходит их перекрытие. В результате такого перекрытия образуются *гибридные зоны*, в состав которых входят как заполненные, так и пустые уровни. Например, в кристалле бериллия в создании гибридной зоны принимают участие заполненные уровни $2s$ и пустые уровни $2p$ (рис. 8), а в кристалле магния — заполненные уровни $3s$ и пустые уровни $3p$. Благодаря эффекту перекрытия верхние энергетические зоны в кристаллах щелочноземельных элементов и оказываются заполненными лишь частично.

К совершенно противоположному результату приводит перекрытие зон в кристаллах полупроводников, имеющих решетку типа алмаза. В атомах кремния, например, на уровне $3p$ (в подболочке $3p$) находятся всего два электрона, в то время как на этом уровне могут расположиться шесть электронов. Вполне естественно ожидать, что при образовании кристалла кремния верхняя энергетическая зона (зона $3p$) окажется заполненной лишь частично, в то время как предшествующая ей зона (зона $3s$) будет заполнена целиком (она образуется из полностью заполненного уровня $3s$). Однако в дейст-

вительности в процессе образования кристалла кремния эффект перекрытия приводит не только к возникновению гибридной зоны, составленной из подуровней $3s$ и $3p$, но и к дальнейшему расщеплению создавшейся гибридной зоны на две подзоны, разделенные между собой промежуток запрещенных энергий W_g (рис. 9). Всего в гибридной зоне $3s+3p$ должно быть по 8 вакантных электронных мест на каждый атом (2 места от подоболочки $3s$ и 6 мест от подоболочки $3p$). При расщеплении же гибридной зоны в каждой из подзон оказывается по 4 места на атом. Стремясь занять нижние энергетические уровни, электроны третьих оболочек атомов кремния (а их всего четыре — два на подоболочке $3s$ и два на подоболочке $3p$) как раз и заполняют нижнюю подзону. Верхняя подзона при этом остается пустой.

§ 6. ДЕЛЕНИЕ ТВЕРДЫХ ТЕЛ НА ПРОВОДНИКИ, ПОЛУПРОВОДНИКИ И ДИЭЛЕКТРИКИ

Физические свойства твердых тел, и в первую очередь их электрические свойства, определяются не тем, как образовались зоны, а тем, как они заполнены. С этой точки зрения все кристаллические тела можно разделить на две принципиально различные группы.

Проводники. В первую группу входят тела, в энергетическом спектре которых над целиком заполненными зонами располагается зона, заполненная частично (рис. 10, а). Как мы видели, частичное заполнение зон наблюдается у щелочных металлов, у которых верхняя зона образуется из незаполненных атомных уровней, а также у кристаллов щелочноземельных элементов, верхняя зона которых благодаря перекрытию заполненных и пустых зон является гибридной. Все тела, входящие в первую группу, являются **проводниками**.

Полупроводники и диэлектрики. Во вторую группу объединяются тела, у которых над целиком заполненными зонами располагаются совершенно пустые зоны (рис. 10, б, в). В эту группу входят и кристаллы, имеющие структуру алмаза, такие, как кремний, германий, серое олово, собственно алмаз и др. К этой группе относятся и многие химические соединения — окислы металлов, карбиды, нитриды металлов, корунд (Al_2O_3) и др. Вторая группа твердых тел объединяет **полупроводники**

и диэлектрики. Самая верхняя заполненная зона в группе этих кристаллов называется *валентной зоной*, а находящаяся над ней первая пустая зона — *зоной проводимости*. Самый верхний уровень валентной зоны называется *потолком* валентной зоны и обозначается W_v (индекс v происходит от английского слова *valency* — *валентность*). Самый нижний уровень зоны проводимости называется *дном* зоны проводимости и обозначается W_c (индекс c происходит от английского слова *conductivity* — *проводимость*).

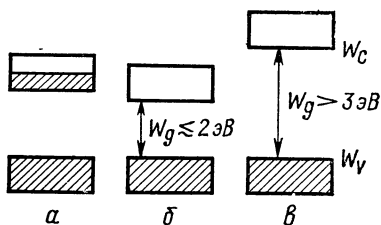


Рис. 10

Принципиальной разницы между полупроводниками и диэлектриками нет. Деление их в пределах второй группы довольно условно и определяется шириной зоны запрещенных энергий W_g , отделяющей целиком заполненную зону от пустой. Тела, имеющие ширину запрещенной зоны $W_g \leq 2$ эВ, относят в подгруппу полупроводников. Типичными представителями их являются германий ($W_g \approx 0,7$ эВ), кремний ($W_g \approx 1,2$ эВ), арсенид галлия GaAs ($W_g \approx 1,5$ эВ), антимонид индия InSb ($W_g \approx 0,2$ эВ).

Тела, у которых $W_g > 3$ эВ, относят к диэлектрикам. Широко известными диэлектриками являются корунд ($W_g \approx 7$ эВ), алмаз ($W_g > 5$ эВ), нитрид бора ($W_g \approx 4,5$ эВ) и др.

Условность деления твердых тел в пределах второй группы на диэлектрики и полупроводники подчеркивается тем, что многие общепризнанные диэлектрики в настоящее время начинают использоваться в технике как полупроводники. Так, уже сегодня в полупроводниковых устройствах применяется карбид кремния с шириной запрещенной зоны около 3 эВ. И даже такой классический представитель группы диэлектриков, как алмаз, в последнее время изучается в плане возможного использования его в полупроводниковой технике.

Степень заполнения зон электронами и проводимость кристаллов. Рассмотрим свойства кристалла с частично заполненной верхней зоной при абсолютном нуле темпе-

ратуры ($T=0$). В этих условиях в отсутствие внешнего электрического поля все электроны займут самые нижние уровни в зоне, располагаясь на них попарно в соответствии с принципом Паули.

Создадим теперь в кристалле внешнее электрическое поле с напряженностью E . Поле это будет действовать на каждый электрон с силой $F = -eE$. Приобретение ускорения электроном означает увеличение его энергии, что должно сопровождаться переходом электрона на более высокие уровни. Поскольку в частично заполненной зоне имеется много свободных энергетических уровней, то такие переходы оказываются вполне возможными. А так как энергетическое расстояние между уровнями очень мало, то даже ничтожные электрические поля вызывают переход электронов на более высоко расположенные уровни. Таким образом, в твердых телах с частично заполненной зоной внешнее электрическое поле увеличивает скорость движения электронов в направлении действия силы со стороны этого поля, что и означает появление электрического тока. Именно такие тела мы называем проводниками.

В отличие от проводников тела, характеризующиеся наличием только заполненных или пустых зон, не способны проводить электрический ток. Внешнее поле в таких твердых телах не может создать направленное движение электронов, так как приобретение электроном дополнительной энергии за счет действия поля означало бы переход его на более высокий уровень; между тем уровни валентной зоны все заняты. В пустой же зоне проводимости, хотя и много свободных мест, нет электронов, а практически используемые электрические поля оказываются не способными сообщить электронам такую энергию, чтобы перевести их из валентной зоны в зону проводимости (здесь исключаются из рассмотрения электрические поля, способные вызвать пробой диэлектрика). В силу всех этих причин даже в полупроводниках внешнее поле не может вызвать появление электрического тока. Вот почему при температуре, соответствующей абсолютному нулю, полупроводник в отношении электропроводности ничем не отличается от диэлектрика.

ЭЛЕКТРОПРОВОДНОСТЬ ТВЕРДЫХ ТЕЛ

§ 7. СИЛЫ СВЯЗИ В КРИСТАЛЛИЧЕСКОЙ РЕШЕТКЕ

Кристалл — система атомов, находящаяся в состоянии устойчивого равновесия. Что определяет образование из отдельных атомов строго упорядоченной кристаллической решетки? Почему в процессе образования кристалла атомы не могут неограниченно сближаться, чем определяется прочность кристалла?

В поисках ответа на эти вопросы мы должны предположить, что существуют силы притяжения $F_{\text{пр}}$ и силы отталкивания $F_{\text{от}}$ между атомами, приходящие в равновесие при образовании кристаллической структуры. Независимо от природы этих сил характер их зависимости от расстояния между атомами оказывается одинаковым (рис. 11, а). На расстоянии $r > a_0$ преобладают силы притяжения; при $r < a_0$ большими оказываются силы отталкивания. На некотором расстоянии $r = a_0$, вполне определенном для данного кристалла, силы притяжения и отталкивания уравниваются друг друга, и результирующая сила F_r (ход которой отражает кривая 3) становится равной нулю. Энергия взаимодействия между частицами при этом достигает наименьшего значения $W_{\text{кр}}$ (рис. 11, б). Поскольку энергия взаимодействия при $r = a_0$ минимальна, то в отсутствие внешнего возбуждения атомы не могут уйти из этого положения: как удаление атомов друг от друга, так и дальнейшее их сближение приводит к увеличению энергии взаимодействия. Такая ситуация означает, что при $r = a_0$ рассматриваемая система атомов находится в состоянии устойчивого равновесия. Это состояние как раз и соответствует образованию твердого тела со строго определенной внутренней структурой — кристалла.

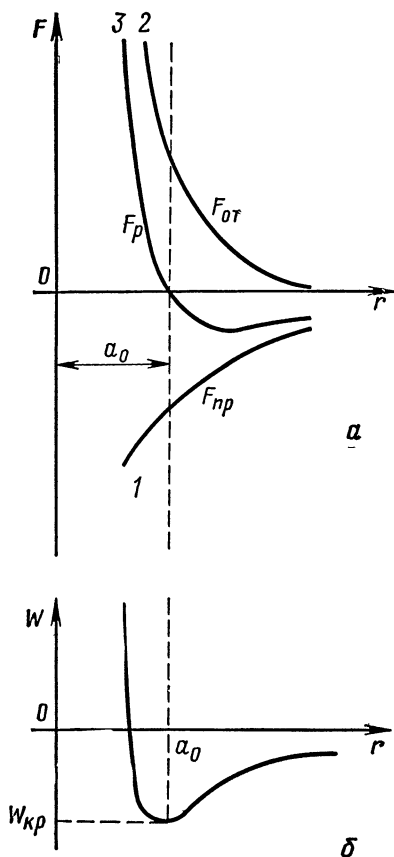


Рис. 11

Силы отталкивания и силы притяжения. Силы отталкивания, как видно из хода кривой 2 на рисунке 11, очень быстро возрастают с уменьшением расстояния r между атомами. Преодоление этих сил связано с затратой больших энергий. Например, при сближении протона с атомом водорода уменьшение расстояния от $r=2a$ до

$$r = \frac{a}{2}$$

(где a — радиус первой боровской орбиты) сопровождается возрастанием энергии отталкивания в 300 раз. Отталкивание легких атомов, ядра которых слабо экранированы электронными оболочками, обусловлено в первую очередь взаимодействием ядер. В случае же сближения многоэлектронных атомов отталкивание объясняется взаимодействием внутренних заполненных

электронных оболочек. При этом отталкивание происходит не только потому, что электронные оболочки имеют одинаковый знак заряда, но и потому, что при сближении атомов происходит перестройка электронных оболочек. На очень малых расстояниях должно происходить перекрытие электронных оболочек и возникновение орбит, общих для обоих атомов. Однако поскольку на первичных заполненных орбитах все места заняты и появление избыточных электронов на них согласно принципу Паули невозможно, то определенная часть электронов должна

перейти на более высокие оболочки. Такой переход связан с повышением общей энергии системы, чем и объясняется проявление сил отталкивания.

Как видно, природа сил отталкивания является в определенной степени общей для всех атомов и не зависит от строения внешних, незаполненных оболочек. В противоположность этому природа сил притяжения между атомами оказывается гораздо более разнообразной и определяется она как раз строением и степенью заполненности внешних электронных оболочек. В соответствии с природой сил притяжения получили название и силы связи в целом. Наиболее важными при рассмотрении строения кристаллов являются ионная, ковалентная и металлическая связи, известные из школьного курса химии. Мы остановимся на рассмотрении только ковалентной связи, определяющей основные свойства полупроводниковых кристаллов.

Ковалентная связь играет основную роль при образовании молекул или кристаллов из одинаковых или однородных атомов. Естественно, что при взаимодействии одинаковых атомов не происходит передача электронов от одного атома к другому и образование ионов. Однако процесс перераспределения электронов и здесь играет основную роль. Но в этом случае перераспределение завершается не переходом какого-либо электрона от одного атома к другому, а о б о б щ е с т в л е н и е м некоторых электронов: определенные электроны оказываются принадлежащими одновременно нескольким атомам.

Рассмотрим образование ковалентной связи в молекуле водорода H_2 . Пока оба атома водорода находятся на достаточно большом расстоянии друг от друга, каждый из них «владеет» своим электроном, и вероятность обнаружить в пределах данного атома «чужой» электрон ничтожно мала. Так, при расстоянии между атомами $r=5$ нм каждый из электронов может появиться в соседнем атоме один раз в 10^{12} лет. По мере сближения атомов вероятность появления «чужого» электрона резко возрастает. При $r=0,2$ нм частота перехода достигает уже $10^{14}c^{-1}$, а при дальнейшем сближении частота обмена электронами между атомами становится настолько большой, что «свой» и «чужой» электроны появляются одинаково часто у обоих ядер. В этом случае говорят, что электроны обобществлены. Связь, основанная на

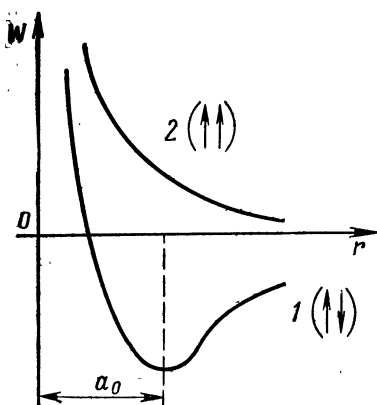


Рис. 12

«совместном владении» атомами двумя электронами, и называется *ковалентной*.

Создание ковалентной связи, естественно, должно быть выгодно и с энергетической точки зрения. На рисунке 12 кривая 1 отражает ход суммарной энергии двух атомов водорода по мере их сближения (за нулевой уровень принята суммарная энергия двух бесконечно удаленных атомов водорода, не взаимодействующих друг с другом).

Мы видим, что по мере сближения атомов их суммарная энергия уменьшается, достигая минимума при $r = a_0$, соответствующего расстоянию между атомами в молекуле водорода.

Не всякие два атома водорода могут создать молекулу. Описанный процесс образования молекулы водорода возможен только в том случае, если электроны в объединяющихся атомах имеют противоположно направленные спины. Лишь при различной ориентации спинов электроны могут расположиться на одной электронной орбите, являющейся общей орбитой для объединенных атомов. Если же в сближающихся атомах водорода электроны имеют одинаково направленные спины, то есть находятся в одном и том же состоянии, определяемом одинаковым набором всех четырех квантовых чисел (n , l , m и σ), то согласно принципу Паули они не могут разместиться на одной и той же орбите. Такие атомы по мере сближения будут не притягиваться, а отталкиваться, а их суммарная энергия с уменьшением r будет не уменьшаться, а возрастать (кривая 2 на рисунке 12). Ясно, что в этом случае образования молекулы не происходит.

Важнейшей особенностью ковалентных связей является их *насыщаемость*. Свойство это также связано с проявлением *принципа* Паули и заключается в невозможности участия в создании ковалентной связи треть-

его электрона. Объединившись в одну ковалентную связь, электроны уже не могут создать другой связи и в то же время «не разрешают» другим электронам проникнуть на созданную объединенную орбиту. Именно поэтому, например, молекула водорода состоит из двух атомов и не может быть создана молекула H_3 .

Полупроводники — типичные представители ковалентных кристаллов. Наиболее характерные черты ковалентных связей просматриваются в кристаллах, имеющих структуру алмаза. Типичными представителями этой группы являются кристаллы полупроводников, среди которых наиболее известны кремний и германий. Атомы этих элементов имеют на внешней оболочке по четыре валентных электрона, каждый из которых создает по одной ковалентной связи с четырьмя близлежащими соседями (рис. 13). Образую ковалентную связь с соседом, атом отдает в «частичное пользование» один из своих валентных электронов и одновременно приобретает электрон соседа на тех же основаниях. В результате каждый из атомов, образующих кристалл, как бы достраивает свою внешнюю оболочку до полной заселенности (по 8 электронов), создавая устойчивую структуру, аналогичную структуре атомов инертных газов (на рисунке 13 эти 8 электронов условно расположены на круговой орбите, обозначенной пунктиром). Благодаря неразличимости электронов и возможности обмена электронами между атомами все валентные электроны в одинаковой степени принадлежат всем атомам кристалла. Кристалл полупроводника предстает как одна гигантская молекула, в которой атомы соединены между собой ковалентными связями. На рисунках такие кристаллы условно

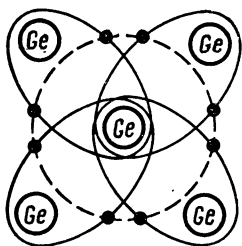


Рис. 13

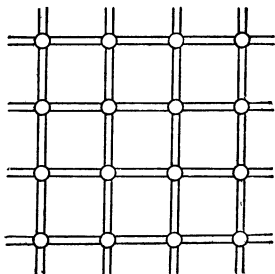


Рис. 14

отображаются плоскостной фигурой (рис. 14), где каждая двойная линия между атомами обозначает ковалентную связь, образованную двумя электронами.

§ 8. ЭЛЕКТРИЧЕСКАЯ ПРОВОДИМОСТЬ МЕТАЛЛОВ

Наиболее полно и точно явление электрической проводимости металлов описывает квантовая теория твердых тел. Однако для выяснения наиболее общих вопросов можно ограничиться и рассмотрением на основе классической электронной теории. Согласно этой теории совокупность электронов в кристалле можно с известным приближением уподобить идеальному газу, считая движение электронов подчиняющимся законам классической механики. При этом взаимодействие электронов между собой не рассматривается вообще, а взаимодействие электронов с ионами кристаллической решетки сводится к обычным упругим столкновениям.

Металлы содержат огромное число свободных электронов, перемещающихся в межузельном пространстве кристалла. В 1 см^3 насчитывается около 10^{23} атомов. Следовательно, при валентности металла Z концентрация n свободных электронов, называемых еще электронами проводимости, равна $10^{23} Z \text{ см}^{-3}$. Все они находятся в беспорядочном хаотическом тепловом движении, перемещаясь в пространстве кристалла с колоссальной скоростью, среднее значение которой около 10^8 см/с . В силу хаотичности теплового движения число электронов, двигающихся в каком-либо направлении, в среднем всегда равно числу электронов, двигающихся во встречном направлении, благодаря чему в отсутствие внешнего электрического поля заряд, переносимый электронами через любое сечение кристалла, равен нулю. Под действием электрического поля каждый электрон приобретает дополнительную скорость, благодаря чему весь коллектив электронов в металле начинает перемещаться в направлении, противоположном направлению напряженности приложенного поля. Появление направленного движения электронов и определяет возникновение в проводнике электрического тока.

На каждый электрон электрическое поле напряженностью E действует с силой $F = eE$. Под действием этой силы электрон приобретает ускорение

$$a = \frac{F}{m} = \frac{eE}{m},$$

где e — заряд электрона, а m — его масса.

Согласно законам классической механики в свободном пространстве скорость электронов возрастала бы неограниченно; то же наблюдалось бы при их движении в строго периодическом поле (например, в идеальном кристалле с покоящимися в узлах атомами).

В действительности же благодаря нарушениям периодичности в потенциальном поле решетки направленное перемещение электронов в кристалле оказывается совсем незначительным. Нарушения эти в первую очередь связаны с тепловыми колебаниями атомов (в случае металлов — атомных остатков) в узлах кристаллической решетки (при этом амплитуда колебаний тем больше, чем выше температура кристалла). Кроме того, в кристалле всегда имеются различные дефекты, обусловленные наличием атомов примесей, пустых мест в узлах, атомов в междоузлиях, дислокаций. Влияют также границы блоков кристаллов, трещины, полости и т. д.

В этих условиях электроны все время испытывают столкновения и растрачивают приобретенную в электрическом поле энергию. Поэтому в действительности скорость электронов под действием силы внешнего поля увеличивается только на участке между двумя столкновениями. Средняя длина этого участка называется *длиной свободного пробега электрона* и обозначается через λ .

Итак, ускоряясь на длине свободного пробега, электрон приобретает дополнительную скорость направленного движения

$$\Delta v = a\tau,$$

где τ — время свободного пробега, или среднее время между двумя последовательными соударениями электрона с дефектами. Зная длину свободного пробега λ , можно время свободного пробега τ вычислить по формуле

$$\tau = \frac{\lambda}{v_0 + \Delta v},$$

где v_0 — скорость хаотического теплового движения электрона. Длина свободного пробега электрона λ обыч-

но очень мала и не превышает 10^{-5} см. Поэтому малыми оказываются и время свободного пробега τ , и сама добавка скорости Δv . Так как $\Delta v \ll v_0$, то

$$\tau \approx \frac{\lambda}{v_0}.$$

Принимая, что при столкновении с дефектом электрон практически полностью теряет скорость направленного движения, можно среднюю скорость направленного движения, называемую *скоростью дрейфа*, выразить так:

$$\bar{v} = \frac{\Delta v}{2} = \frac{eE}{2m} \cdot \tau = \frac{e\lambda}{2mv_0} E = uE.$$

Коэффициент пропорциональности

$$u = \frac{e}{2m} \frac{\lambda}{v_0}$$

между средней скоростью дрейфа \bar{v} и напряженностью поля E называется *подвижностью электронов*.

Название этой величины точно отражает ее физический смысл: подвижность — это скорость дрейфа, которую приобретают электроны в электрическом поле единичной напряженности. Более строгий расчет, учитывающий тот факт, что и при хаотическом тепловом движении электроны перемещаются не с постоянной скоростью v_0 , а имеют различные скорости, приводит к вдвое большему значению для подвижности электронов:

$$u = \frac{e}{m} \frac{\lambda}{v_0}.$$

Соответственно и для скорости дрейфа более точным является выражение

$$\bar{v} = \frac{e\lambda}{mv_0} E.$$

Найдем теперь выражение для плотности тока в металлах. Так как под действием внешнего электрического поля электроны приобретают дополнительную скорость дрейфа \bar{v} , то за единицу времени через любую площадку, перпендикулярную напряженности поля, пройдут все электроны, отстоящие от этой площадки на расстоянии, не превосходящем \bar{v} . Через площадку площадью S за то же время пройдут все электроны, заключенные в объеме

параллелепипеда длиной \bar{v} (рис. 15). Если концентрация свободных электронов в металле n , то число их в объеме этого параллелепипеда будет равно $n\bar{v}S$. Плотность тока, определяемая зарядом, перенесенным этими электронами через единичную площадь, выразится так:

$$j = \frac{en\bar{v}S}{S} = \frac{ne^2\lambda}{mv_0}E.$$

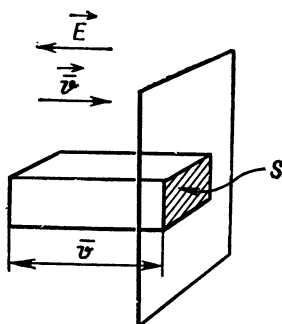


Рис. 15

Отношение плотности тока к напряженности поля, вызвавшего появление этого тока, называется *удельной электрической проводимостью* (обозначается через σ). Очевидно, что

$$\sigma = \frac{j}{E} = \frac{ne^2\lambda}{mv_0} = ne\mu.$$

Величина ρ , обратная удельной электрической проводимости, называется *удельным электрическим сопротивлением*:

$$\rho = \frac{1}{\sigma}.$$

Обратим внимание на то, что возникновение электрического тока в проводнике четко связано с появлением дрейфа электронов. Скорость дрейфа оказывается очень небольшой — в реальных электрических полях ее значение обычно не превышает скорости пешехода. В то же время ток распространяется по проводам практически мгновенно и появляется во всех участках замкнутой цепи почти одновременно. Объясняется это колоссальной скоростью распространения самого электрического поля. При включении источника напряжения электрическое поле приходит в самые удаленные места рассматриваемой цепи со скоростью света и вызывает появление дрейфового перемещения сразу у всего коллектива электронов.

§ 9. ЭЛЕКТРИЧЕСКАЯ ПРОВОДИМОСТЬ ПОЛУПРОВОДНИКОВ

Как и в металлах, появление электрического тока в полупроводниках связано с возникновением дрейфа носителей заряда. Но если в металлах наличие свободных электронов в кристалле обусловлено самой природой металлической связи, то появление носителей заряда в полупроводниках определяется многими факторами, среди которых наиболее важными являются чистота полупроводника и его температура.

Полупроводники делятся на чистые, или, как их называют, *собственные*, и *примесные*, или *легированные*. Примесные полупроводники в свою очередь, в зависимости от типа вводимой примеси, делятся на *донорные*, или электронные, и *акцепторные*, или дырочные. Рассмотрим каждую из этих групп полупроводников отдельно.

Собственные полупроводники

Собственными называются полупроводники высокой степени очистки. В этом случае свойства всего кристалла определяются только свойствами собственных атомов полупроводникового элемента.

Электронная проводимость. При температуре, близкой к абсолютному нулю, все атомы кристалла связаны между собой ковалентными связями, в создании которых заняты все валентные электроны. И хотя, как мы уже отмечали ранее, все валентные электроны в одинаковой степени принадлежат всем атомам кристалла и могут переходить от одного атома к другому, тем не менее электрической проводимостью кристалл в таких условиях не обладает. Всякий переход электрона от атома к атому сопровождается встречным переходом; при этом прямой и встречный переходы происходят одновременно и приложенное электрическое поле не может создать направленного перемещения зарядов. Свободных же электронов в условиях сверхнизких температур нет.

С позиции зонной теории такая ситуация соответствует тому, что валентная зона целиком заполнена, а зона проводимости совершенно пуста.

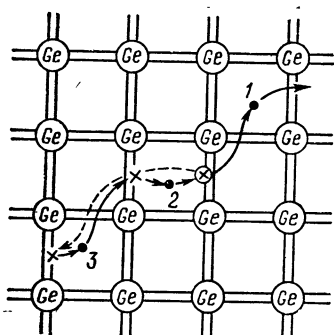


Рис. 16

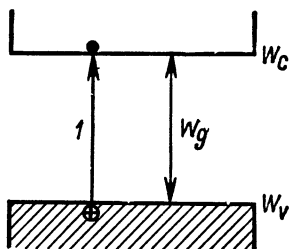


Рис. 17

При повышении температуры тепловые колебания кристаллической решетки сообщают электронам дополнительную энергию. В определенных условиях энергия электрона оказывается больше энергии ковалентной связи, и он, разрывая эту связь, переходит в межузельное пространство кристалла, становясь «свободным». Такой электрон может свободно перемещаться в межузельном пространстве кристалла независимо от перемещения других электронов (электрон 1 на рисунке 16).

На диаграмме энергетических уровней «освобождение» электрона означает переход электрона из валентной зоны в зону проводимости (рис. 17). Энергия разрыва ковалентной связи в кристалле как раз и равна ширине запрещенной зоны W_g , то есть энергии, которую должен иметь электрон, чтобы из валентного он мог стать электроном проводимости. Понятно, что, чем уже запрещенная зона кристалла, тем при более низкой температуре начнут появляться свободные электроны. Иначе говоря, при одной и той же температуре кристаллы с более узкой запрещенной зоной будут обладать большей про-

Таблица 2

Вещество	W_g , эВ	n , см ⁻³
Антимонид индия	0,2	10^{16}
Германий	0,7	10^{13}
Алмаз	5,0	10^3

водимостью из-за большей концентрации электронов n в зоне проводимости. Соответствующие данные для случая комнатной температуры приведены в таблице 2.

Если, например, алмаз нагреть до температуры 600 К, то концентрация свободных электронов в нем увеличивается настолько, что становится сравнимой с концентрацией электронов проводимости в германии, находящемся при комнатной температуре. Это еще раз убеждает в том, что деление твердых тел на диэлектрики и полупроводники носит условный характер.

Дырочная проводимость. Возникновение большого числа свободных электронов по мере повышения температуры оказывается лишь одной из причин проводимости собственного полупроводника. Другая причина связана с изменением структуры валентных связей в кристалле, вызванным переходом валентных электронов в межузельное пространство. Каждый электрон, уходящий в межузельное пространство и превращающийся в электрон проводимости, оставляет в системе валентных связей в кристалле вакантное место — «дырку» (на рисунке 16 дырка изображена светлым кружком; знаком «X» отмечен разрыв связи, обусловленный уходом электрона). На освободившееся место может перейти валентный электрон от любого соседнего атома. В свою очередь на место, освободившееся после ухода этого электрона, может перейти электрон от следующего атома и т. д. Такие переходы электронов на вакантные места уже не связаны с обязательными встречными переходами (как это было в полностью укомплектованной системе валентных связей кристалла), и поэтому возникает возможность направленного переноса заряда по кристаллу. В отсутствие внешнего поля подобные переходы одинаково возможны во всех направлениях, благодаря чему суммарный заряд, переносимый через любую площадку в кристалле, равен нулю. Однако при включении внешнего электрического поля такие переходы приобретают направленный характер: электроны в системе валентных связей перемещаются в том же направлении, что и свободные электроны проводимости. Перемещение электронов в цепочке таких переходов происходит последовательно, как бы по очереди — каждый электрон переходит на место, освободившееся от его предшественника. Если же рассмотреть результат такого последовательного про-

цесса, то его можно представить как перемещение во встречном направлении самого вакантного места.

Рассмотрим для наглядности цепочку из шашек, в которой имеется одно вакантное место (рис. 18, а). Последова-

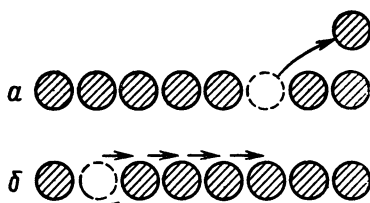


Рис. 18

тельное перемещение слева направо четырех шашек (рис. 18, б) можно рассматривать как перемещение на четыре места во встречном направлении самого вакантного места. Нечто подобное происходит и в полупроводнике. Последовательный переход электронов 2 и 3 (рис. 16) на место, освободившееся от ухода электрона 1, равносильно переходу в противоположном направлении самой вакансии, путь которой указан штриховой линией.

В физике полупроводников подобные вакансии и получили название дырок. Каждой дырке приписывается положительный заряд $+e$, численно равный заряду электрона. Такой подход позволяет вместо описания последовательных переходов цепочки электронов (каждого в соседний атом) рассматривать ряд переходов одной дырки, что существенно упрощает расчеты.

С позиций зонной теории «дырочная» проводимость в собственном полупроводнике также легко объяснима. Уход электронов в зону проводимости (см. рис. 17) сопровождается образованием вакантных мест — дырок в ранее полностью заполненной валентной зоне. Поэтому электроны, оставшиеся в валентной зоне, получают возможность переходить на более высокие освободившиеся энергетические уровни. А это значит, что они могут получить ускорение (за счет действия внешнего электрического поля) и тем самым принять участие в направленном переносе заряда — создании электрического тока.

Число дырок равно числу свободных электронов. В собственном полупроводнике мы рассматриваем два типа носителей заряда: электроны (носители отрицательного заряда) и дырки (носители положительного заряда). Число дырок всегда равно числу электронов, так как появление электрона в зоне проводимости однознач-

но связано с появлением дырки в валентной зоне. Поэтому в создании проводимости собственных полупроводников дырки и электроны выступают на равных правах. Разница заключается лишь в том, что электронная проводимость обусловлена перемещением свободных электронов в межузельном пространстве кристалла (то есть перемещением электронов, перешедших в зону проводимости), в то время как дырочная проводимость связана с переходом электронов от атома к атому в системе ковалентных связей кристалла (то есть переходами электронов, оставшихся в валентной зоне).

В соответствии с наличием в собственных полупроводниках двух типов носителей зарядов и само выражение для удельной электрической проводимости собственных полупроводников представляется в виде двух слагаемых:

$$\sigma = en_i u_n + ep_i u_p ,$$

где n_i — концентрация электронов в собственном полупроводнике, p_i — концентрация дырок в собственном полупроводнике, u_n и u_p — соответственно подвижности электронов и дырок¹.

Несмотря на внешнее равноправие электронов и дырок и равенство их концентраций ($n_i = p_i$), вклад электронной проводимости в проводимость собственного полупроводника, как правило, больше вклада дырочной проводимости. Объясняется это большей подвижностью электронов по сравнению с дырками. Так, в германии подвижность электронов почти в два раза больше подвижности дырок, а, например, в антимониде индия InSb отношение подвижности электронов к подвижности дырок достигает 80.

Забегая немного вперед, отметим, что появление проводящих свойств в полупроводнике может быть обусловлено не только повышением температуры, но и другими внешними воздействиями, например облучением светом или бомбардировкой быстрыми электронами. Важно

¹ В полупроводниковой технике для различных обозначений используются буквы английского алфавита. Буква n используется для обозначения самих электронов, их концентрации и различных электронных характеристик (negativ — отрицательный), буква p — для аналогичных обозначений, когда речь идет о дырках (positiv — положительный), а индекс i указывает на то, что соответствующая величина относится к собственному полупроводнику (intrinsic — присущий чему-либо, собственный).

лишь, чтобы внешнее воздействие вызывало переход электронов из валентной зоны в зону проводимости или, как говорят, чтобы были созданы условия для генерации свободных носителей заряда в объеме полупроводника.

Собственная проводимость со строгим равенством концентраций носителей различных знаков ($n_i = p_i$) может быть реализована только в сверхчистых идеальных кристаллах полупроводника. В реальных условиях мы всегда имеем дело с кристаллами, в той или иной степени загрязненными различными примесями. Более того, именно примесные полупроводники и представляют наибольший интерес в полупроводниковой технике.

Примесные полупроводники

Донорные примеси. Наличие примесных атомов в объеме собственного полупроводника приводит к некоторому изменению энергетического спектра кристалла. Если в собственном полупроводнике валентные электроны могут иметь энергию только в области разрешенных зон (в пределах валентной зоны или зоны проводимости), а в зоне запрещенных энергий им «находиться» запрещено, то электроны атомов некоторых определенных примесей могут иметь энергию, лежащую в пределах запрещенной зоны. Таким образом в энергетическом спектре появляются дополнительные разрешенные примесные уровни в запрещенной области между потолком W_v валентной зоны и дном W_c зоны проводимости.

Рассмотрим сначала возникновение примесных уровней на примере так называемого *донорного полупроводника*, получающегося при введении, например, в кристалл четырехвалентного германия примеси в виде атомов пятивалентного мышьяка (рис. 19). Четыре электрона из пяти валентных электронов атома мышьяка участвуют в образовании ковалентных связей с четырьмя ближайшими соседними атомами германия и тем самым принимают участие в создании кристаллической решетки. Эти электроны находятся в аналогичных условиях с валентными электронами атомов основного материала — германия, благодаря чему они практически имеют такие же значения энергии, что и электроны атомов германия, и располагаются в энергетическом спектре в пределах валентной зоны. Поэтому эти электроны атомов мышьяка

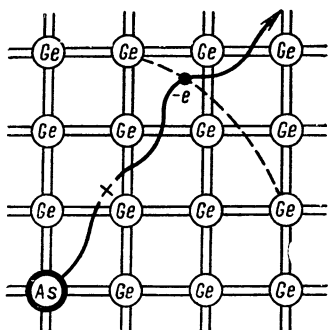


Рис. 19

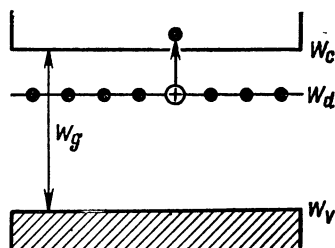


Рис. 20

изменения в энергетический спектр германия не вносят. Пятый же электрон в образовании ковалентных связей не участвует. Но, оставаясь принадлежащим атому мышьяка, он продолжает движение в поле атомного остатка. Однако взаимодействие между этим электроном и атомным остатком оказывается значительно ослабленным, как становятся ослабленными силы кулоновского взаимодействия между двумя электрическими зарядами, помещенными в диэлектрик. Диэлектрическая проницаемость германия $\epsilon = 16$, следовательно, сила взаимодействия между атомным остатком мышьяка и пятым валентным электроном оказывается ослабленной в 16 раз, а энергия их связи — уменьшенной почти в 250 раз. Благодаря этому радиус орбиты пятого электрона увеличивается в 16 раз, и для того чтобы оторваться ему от атома и превратиться в электрон проводимости, оказывается достаточной энергия всего 0,01 эВ.

На языке зонной теории создавшаяся ситуация как раз и означает появление в энергетическом спектре кристалла дополнительного разрешенного уровня, соответствующего энергии пятого валентного электрона мышьяка. Располагается он вблизи дна зоны проводимости (рис. 20) и отстоит от нее на расстоянии $W_d \approx 0,01$ эВ¹.

При низких температурах, близких к абсолютному нулю, все пятые электроны примесных атомов мышьяка

¹ Индекс d происходит от английского слова *donor* — жертвующий, отдающий. Соответственно и сами примеси, и энергетические уровни, образующиеся при введении этих примесей, называются *донорными*.

остаются связанными со своими атомными остатками, иначе говоря, находятся на своем донорном уровне. Зона проводимости при этом остается пустой, и донорный полупроводник, как и собственный полупроводник, при $T=0$ ничем не отличается от типичного диэлектрика. Однако уже при незначительном повышении температуры, когда энергия тепловых колебаний решетки становится сравнимой с энергией связи $W_d \approx 0,01$ эВ, происходит отрыв пятых электронов от атомов мышьяка и переход их в зону проводимости. Донорный полупроводник приобретает проводящие свойства за счет появления свободных электронов в межузловом пространстве кристаллической решетки.

Обратим особое внимание на то, что положительные заряды, остающиеся после ухода электронов с донорных уровней, принципиально отличаются от дырок в собственном полупроводнике. Ушедшие электроны атомов примеси не участвовали в создании ковалентных связей кристалла и не принадлежали валентной зоне, поэтому оставшиеся положительные заряды представляют собой положительно заряженные ионы донорной примеси (в рассматриваемом случае мышьяка), зафиксированные в кристаллической решетке и не принимающие никакого участия в электропроводности кристалла.

Так как электронная проводимость является основной проводимостью в кристаллах с донорной примесью, то полупроводники с донорными примесями называют еще *электронными полупроводниками* или полупроводниками *n*-типа.

Электронная проводимость является преимущественной проводимостью в донорных полупроводниках при достаточно низких температурах. В области повышенных температур, например при комнатной температуре, в зоне проводимости, помимо электронов, перешедших с донорного уровня, появляются электроны, перешедшие из валентной зоны благодаря разрыву валентных связей. Такие переходы, как мы знаем, сопровождаются появлением в валентной зоне дырок и образованием дырочной проводимости. Тем не менее электронная проводимость остается во много раз превосходящей дырочную проводимость.

Например, если в кристалле германия на 10 миллионов атомов германия приходится всего один атом мышья-

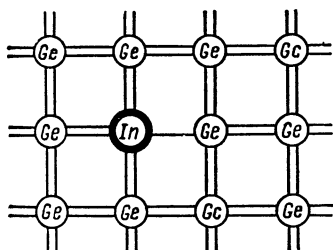


Рис. 21

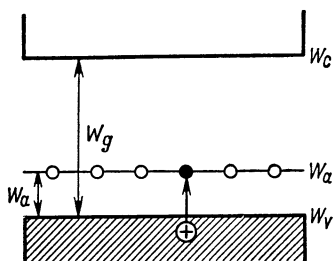


Рис. 22

яка, то при комнатной температуре концентрация электронов проводимости оказывается выше концентрации дырок почти в 2000 раз.

Носители заряда, концентрация которых в рассматриваемом полупроводнике преобладает, называются *основными носителями*; носители же заряда противоположного знака называются *неосновными*. В донорном полупроводнике основными носителями, естественно, являются электроны, а неосновными — дырки.

Акцепторные полупроводники. Рассмотрим теперь случай, когда в кристалле германия вместо пятивалентного атома мышьяка имеется примесный трехвалентный атом индия (рис. 21). Для создания ковалентных связей с четырьмя ближайшими соседями атому индия не хватает одного электрона, то есть в кристаллической решетке германия одна двойная связь оказывается неупакованной. В принципе создание полноценной ковалентной связи с четвертым соседом может быть обеспечено переходом к атому индия электрона от другого атома германия, но для этого электрону необходима некоторая дополнительная энергия. Поэтому при температуре, близкой $T=0$, когда такой дополнительной энергии электрону получить неоткуда, валентные электроны германия остаются при своих атомах, находясь в валентной зоне, а примесные атомы индия так и остаются нейтральными атомами с неупакованными четвертыми связями. Однако само наличие атомов индия в кристалле создает принципиальную возможность для перехода электронов, получивших некоторую дополнительную энергию W_a , на более высокие энергетические уровни, обусловленные созданием дополнительных связей ато-

мами индия (рис. 22). Ясно, что при $T=0$ рассматриваемый полупроводник электропроводностью не обладает, так как нет свободных носителей заряда (ни электронов в зоне проводимости, ни дырок в валентной зоне).

При повышении температуры, когда электроны за счет тепловых колебаний решетки

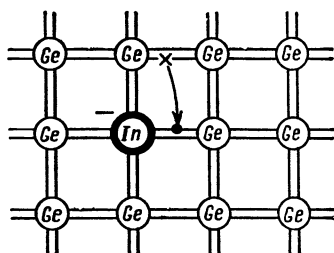


Рис. 23

получают дополнительную энергию порядка W_a (в рассматриваемом случае $W_a \approx 0,01$ эВ, появляется возможность для перехода их от атомов германия к атомам индия. На месте ушедшего электрона при этом остается вакантное место — дырка. Естественно, что возможен и обратный переход, то есть возвращение электрона к атому германия. Однако если за время пребывания электрона в атоме индия его вакантное место окажется уже занятым другим валентным электроном, то он так и останется в атоме индия, который при этом превратится в отрицательно заряженный ион, связанный с кристаллической решеткой и в силу этого лишенный подвижности. Вакантное место в системе валентных связей, образовавшееся после ухода электрона (рис. 23), превращается, таким образом, в свободную дырку. Образование дырок в валентной зоне (см. рис. 22) и означает появление в кристалле дырочной проводимости. Благодаря такому типу проводимости и сами полупроводники получили название *дырочных полупроводников* или полупроводников *p-типа*. Примеси, вводимые в полупроводник для захвата электронов из валентной зоны, получили название *акцепторов*¹, из-за чего энергетические уровни этих примесей называются *акцепторными уровнями*, а сами полупроводники с такими примесями — *акцепторными полупроводниками*.

¹ От английского acceptor — принимающий, захватывающий.

§ 10. ВЛИЯНИЕ ТЕМПЕРАТУРЫ НА КОНЦЕНТРАЦИЮ НОСИТЕЛЕЙ ЗАРЯДА В ПОЛУПРОВОДНИКАХ

Рассматривая характерные особенности полупроводников с различными типами проводимости, мы все время подчеркивали принципиальную зависимость самого появления проводящих свойств от температуры. Только при температурах, существенно отличных от абсолютного нуля, полупроводники становятся способными проводить электрический ток благодаря появлению свободных носителей заряда.

Рассмотрим теперь, каково различие влияния роста температуры на электрические свойства собственных и примесных полупроводников.

Влияние роста температуры на проводимость собственных полупроводников. Специфика собственных полупроводников в первую очередь состоит в том, что носители зарядов обоих знаков — и электроны, и дырки — появляются одновременно и в одном и том же количестве. После того как энергия тепловых колебаний решетки оказывается достаточной для перевода электронов из валентной зоны в зону проводимости, дальнейшее увеличение температуры сопровождается исключительно быстрым ростом концентрации свободных носителей заряда. В собственном германии, например, увеличение температуры от 100 до 600 К повышает концентрацию носителей на 17 порядков (в 10^{17} раз). Для того чтобы графически отразить эту зависимость, обычно пользуются полулогарифмическим масштабом, откладывая по оси абсцисс величину $1/T$, обратную температуре, а по оси ординат — значение $\lg n_i$ логарифма от концентрации (рис. 24). В таком масштабе зависимость $n_i = f(1/T)$ выражается прямой, пересекающей с осью абсцисс в некоторой точке $1/T$. Нарастание концентрации носителей по мере роста температуры происходит из-за того, что реализуются переходы в зону проводимости электронов все с более и более глубинных уровней валентной зоны.

Если бы мы продолжили прямую $n_i = f(1/T)$ до пересечения с осью ординат (что соответствовало бы значению $T = \infty$), то получили бы значение $\lg n_{i0}$ (где n_{i0} — концентрация валентных электронов рассматриваемого полупроводника). Получить соответствующее состояние экспериментально не представляется возможным, так

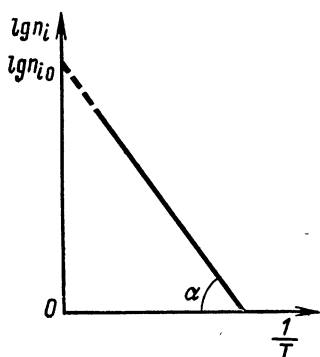


Рис. 24

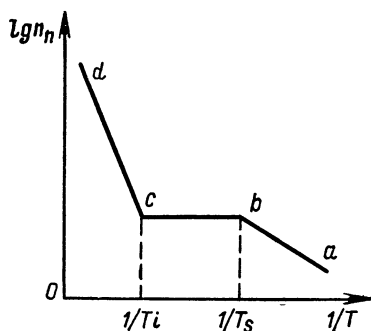


Рис. 25

как, прежде чем все электроны, создающие ковалентные связи, перейдут в зону проводимости, произойдет разрушение кристаллической решетки и полупроводник перейдет в расплавленное состояние.

Поскольку в собственном полупроводнике при любой температуре концентрации электронов и дырок одинаковы, то зависимость $\lg p_i = f(1/T)$ выразится точно такой же прямой, как и приведенная на рисунке 24.

В силу равенства концентраций $n_i = p_i$ можно записать:

$$n_i p_i = n_i^2.$$

А поскольку в каждом собственном полупроводнике концентрация носителей зависит только от температуры, то очевидно, что при всякой фиксированной температуре произведение $n_i p_i$ оказывается постоянной величиной (*закон действующих масс*):

$$n_i p_i = n_i^2 = \text{const.}$$

При температуре 300 К (обычно именно эта температура принимается за комнатную) значение n_i^2 для кремния равно $4,2 \cdot 10^{21} \text{ см}^{-6}$, а для германия — $6,25 \cdot 10^{26} \text{ см}^{-6}$. Интересно отметить, что закон действующих масс оказывается справедливым и для примесных полупроводников.

Примесные полупроводники. Так как характер температурной зависимости концентраций свободных носителей в донорных и акцепторных полупроводниках совер-

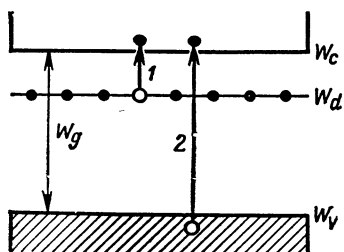


Рис. 26

шенно одинаков, то мы ограничимся рассмотрением только полупроводника n -типа.

Типичная кривая зависимости $\lg n_n = f(1/T)$ для примесных полупроводников приведена на рисунке 25. В области очень низких температур, близких к абсолютному нулю, свободных носителей

в зоне проводимости донорного полупроводника, как и в случае собственного полупроводника, нет. Однако по мере повышения температуры электроны проводимости в донорном полупроводнике появляются намного раньше, чем в собственном. Это и понятно, ведь для отрыва пятого валентного электрона от донорного атома, то есть для перевода электрона с донорного уровня W_d в зону проводимости, необходима энергия, почти в 100 раз меньшая, чем для перехода электрона из валентной зоны в зону проводимости. Вспомним, что для кремния, например, $W_d \approx 0,01$ эВ, в то время как $W_g \approx 1,2$ эВ. Поэтому сначала в зоне проводимости появятся электроны в результате переходов, обозначенных на рисунке 26 цифрой 1. Конечно, разницы между этими электронами и электронами, переходящими в зону проводимости из валентной зоны, нет, но тем не менее иногда их называют примесными электронами, подчеркивая их происхождение.

С ростом температуры число примесных электронов, перешедших в зону проводимости, довольно быстро увеличивается, а число электронов, оставшихся на донорном уровне, уменьшается — происходит истощение примесного уровня. При некоторой температуре T_s (температуре истощения) все электроны с донорного уровня оказываются переведенными в зону проводимости. Концентрация электронов проводимости в этом случае становится практически равной концентрации донорной примеси N_d в полупроводнике:

$$n_n \approx N_d.$$

Истощение примесных уровней в большинстве полупроводников происходит в области достаточно низких

температур. Например, в германии с концентрацией примесных атомов мышьяка $N_d \approx 10^{16} \text{ см}^{-3}$ температура истощения донорного уровня составляет около 30 К. Естественно, что процесс истощения зависит от значения энергии активации W_d примесных центров и от их концентрации N_d , поэтому и температура истощения T_s тем выше, чем больше энергия активации и чем выше концентрация примеси.

При $T > T_s$ число свободных электронов в довольно значительном интервале температур не изменяется (участок bc на рисунке 25). Объясняется это тем, что примесные уровни уже полностью истощены, а для возбуждения валентных электронов энергия тепловых колебаний решетки оказывается еще недостаточной. В этом интервале температур можно для определения концентрации свободных носителей пользоваться соотношением $n_n \approx N_d$.

В области достаточно высоких температур все более интенсивным становится процесс перехода электронов в зону проводимости из валентной зоны (на рисунке 26 они обозначены цифрой 2). При некоторой температуре T_i число таких переходов становится настолько большим, что концентрация собственных электронов, то есть электронов, пришедших из валентной зоны, становится сравнимой с концентрацией примесных электронов. Температура T_i называется *температурой перехода к собственной проводимости*. Дальнейшее повышение температуры приводит к столь быстрому росту концентрации собственных носителей (участок cd на рисунке 25), что примесными электронами уже можно пренебречь и считать, что $n_n \approx n_i$. В этой области температур примесный донорный полупроводник как по характеру проводимости (она теперь становится смешанной — и электронной и дырочной, так как уход электронов из валентной зоны сопровождается одновременным образованием дырок), так и по концентрации свободных носителей заряда становится практически ничем не отличимым от собственного, беспримесного полупроводника.

Для большинства примесных полупроводников температура перехода к собственной проводимости существенно превышает комнатную температуру. Так, для германия n -типа с концентрацией примеси $N_d \approx 10^{16} \text{ см}^{-3}$ температура T_i приблизительно равна 450 К. Значение

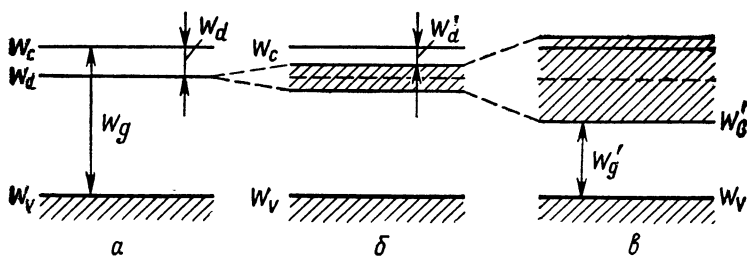


Рис. 27

T_i зависит от концентрации примеси N_d и ширины запрещенной зоны W_g полупроводника: чем больше концентрация примеси и чем шире запрещенная зона, тем выше температура перехода к собственной проводимости.

Вырожденные полупроводники. Особый интерес представляют полупроводники с очень большой концентрацией примеси $\sim 10^{18}—10^{19} \text{ см}^{-3}$. Такие полупроводники называются *вырожденными полупроводниками* или *полуметаллами*. Вырожденные полупроводники занимают очень важное место в полупроводниковой технике.

Введение в собственный полупроводник донорной примеси, как мы уже отмечали ранее, приводит к появлению в энергетическом спектре полупроводника (точнее, в его запрещенной зоне) некоторого дискретного донорного уровня, отстоящего от дна зоны проводимости на весьма небольшом энергетическом расстоянии W_d (рис. 27, *a*). Этот уровень остается достаточно узким, и при не слишком больших концентрациях примеси (пока примесные атомы находятся на столь большом расстоянии друг от друга, что можно пренебречь их взаимодействием) его можно представить в виде линии. По мере увеличения концентрации примесные атомы располагаются все ближе друг к другу, и при этом орбиты их пятых валентных электронов, не участвующих в создании решетки, начинают перекрываться. Для этих электронов появляется возможность беспрепятственного перехода от одного примесного атома к другому. Иначе говоря, пятые валентные электроны перестают принадлежать отдельным конкретным примесным атомам и становятся обобщенными. Это означает, что они могут перемещать-

ся от одного примесного атома к другому в пределах всего кристалла.

На языке зонной теории увеличение концентрации примеси, сопровождающееся появлением взаимного воздействия примесных атомов друг на друга, приводит к снятию вырождения в этих атомах, к расщеплению примесного донорного уровня на подуровни и к образованию примесной зоны (рис. 27, б). По существу в системе примесных атомов происходит тот же процесс образования энергетической зоны из отдельного вырожденного уровня, какой наблюдается при образовании зон из энергетических уровней изолированных атомов, когда они сближаются, образуя кристалл (см. рис. 7).

Чем выше концентрация примеси, тем ближе располагаются примесные атомы друг к другу, тем сильнее оказывается их взаимное воздействие друг на друга и тем больше размывается примесный уровень. В конце концов примесная зона расширяется настолько, что происходит перекрытие ее с зоной проводимости (рис. 27, в), благодаря чему образуется своеобразная гибридная зона, заполненная электронами лишь частично. Примесным электронам, ранее находившимся только в примесной зоне, теперь ничто не препятствует переходить в зону проводимости. Наличие более высоко расположенных свободных уровней в зоне проводимости определяет возможность для примесных электронов приобретать дополнительную энергию и, следовательно, ускорять свое движение во внешнем электрическом поле. А это, как мы уже видели, как раз и означает появление у кристалла проводящих свойств.

Подчеркнем здесь, что наличие проводящих свойств у вырожденных полупроводников связано только с высокой концентрацией донорной примеси, приводящей к перекрытию примесной зоны с зоной проводимости, и не связано с температурой кристалла. Высокая концентрация электронов проводимости в вырожденных полупроводниках имеет место даже при абсолютном нуле температуры ($T=0$). По этой причине их иногда и называют полуметаллами. Такая независимость концентрации электронов проводимости от температуры сохраняется в вырожденных полупроводниках вплоть до температуры T_i . При более высоких температурах на первый план выступают переходы электронов в зону

проводимости из валентной зоны, вызывающие появление в полупроводнике собственной проводимости.

По мере увеличения концентрации примеси размытие примесных уровней в зону на первых порах сопровождается уменьшением энергии активации примесных электронов: величина W'_d (см. рис. 27, б) меньше W_d (см. рис. 27, а). Это приводит к появлению примесных электронов проводимости при более низких температурах. Если же концентрация примеси настолько велика, что происходит перекрытие примесной зоны с зоной проводимости, то наблюдается уменьшение ширины запрещенной зоны: W'_g (см. рис. 27, в) меньше, чем W_g (см. рис. 27, а). Например, в германии при концентрации примеси порядка 10^{19} см^{-3} ширина запрещенной зоны уменьшается с 0,7 эВ до 0,5 эВ.

Зависимость $\lg p_p = f(1/T)$ для полупроводников p -типа описывается графиком, не отличающимся от приведенного на рисунке 25. Участки ab и bc в этом случае соответствуют дырочной примесной проводимости, а участок cd — собственной проводимости. Нарастание концентрации дырок по мере роста температуры на участке ab обусловлено захватом электронов из валентной зоны атомами акцепторной примеси и появлением свободных дырок, способных переносить заряд. После заполнения акцепторного уровня электронами, то есть когда все примесные атомы доукомплектуют свои связи в кристаллической решетке, наступает режим насыщения с характерной независимостью концентрации примесных дырок от температуры (участок bc). При более высоких температурах (участок cd) наблюдается резкое возрастание концентрации собственных носителей, обусловленное переходом электронов из валентной зоны в зону проводимости.

Необходимо отметить, что во всех примесных полупроводниках практически при любой температуре, помимо основных носителей, всегда имеются и неосновные носители. В области низких температур число неосновных носителей, конечно, мало. Концентрации основных и неосновных носителей взаимосвязаны законом действующих масс, согласно которому произведение концентраций основных и неосновных носителей в любом полупроводнике равно квадрату концентрации собственных носителей в соответствующем чистом полупро-

воднике при этой же температуре: $n_n p_n = n_i^2$, $p_p n_p = n_i^2$, где n_n и p_n — концентрации электронов и дырок в электронном полупроводнике, а p_p и n_p — концентрации дырок и электронов в дырочном полупроводнике. Из закона действующих масс следует, что введение в полупроводник какой-либо активной примеси вместе с повышением концентрации основных носителей одновременно приводит к снижению концентрации неосновных носителей так, что произведение концентраций этих носителей остается неизменным. Например, если в чистый германий, имеющий при комнатной температуре концентрацию собственных носителей $n_i \approx 10^{13} \text{ см}^{-3}$, ввести донорную примесь в таком количестве, что число свободных электронов увеличится в 1000 раз и концентрация основных носителей при этом станет равной $n_n \approx 10^{16} \text{ см}^{-3}$, то число неосновных носителей (дырок) в этом случае уменьшится в 1000 раз и концентрация их станет равной $p_n \approx 10^{10} \text{ см}^{-3}$. Таким образом, концентрация неосновных носителей окажется в миллион раз меньше концентрации основных носителей. Такое уменьшение концентрации неосновных носителей объясняется следующим образом.

Появление большого числа электронов проводимости приводит к тому, что нижние энергетические уровни в зоне проводимости оказываются занятыми. Поэтому электроны из валентной зоны могут переходить в зону проводимости только на более высокие, еще не заполненные уровни. Для этих переходов необходима большая энергия, чем в случае свободной зоны проводимости. Вследствие этого вероятность переходов электронов из валентной зоны в зону проводимости и, следовательно, число дырок, образующихся в валентной зоне, уменьшаются.

§ 11. ЗАВИСИМОСТЬ ЭЛЕКТРИЧЕСКОЙ ПРОВОДИМОСТИ ПОЛУПРОВОДНИКОВ ОТ ТЕМПЕРАТУРЫ

Как мы уже видели, удельная проводимость выражается формулой

$$\sigma = en\mu,$$

где n — концентрация носителей заряда, определяющих

проводящие свойства данного тела, а μ — подвижность этих носителей. Носителями заряда могут быть как электроны, так и дырки. Интересно отметить, что, хотя, как известно, у большинства металлов свободными носителями заряда являются электроны, у некоторых металлов роль свободных носителей заряда выполняют дырки. Типичными представителями металлов с дырочной проводимостью являются цинк, бериллий и некоторые другие.

Для выяснения зависимости проводимости от температуры необходимо знать температурную зависимость концентрации свободных носителей и их подвижности. В металлах концентрация свободных носителей заряда не зависит от температуры. Поэтому изменение проводимости металлов в зависимости от температуры полностью определяется температурной зависимостью подвижности носителей. В полупроводниках, напротив, концентрация носителей резко зависит от температуры, а температурные изменения подвижности практически оказываются незаметными. Однако в тех областях температур, где концентрация носителей оказывается постоянной (область истощения и область насыщения примесей), ход температурной зависимости проводимости полностью определяется температурным изменением подвижности носителей.

Значение же самой подвижности определяется процессами рассеяния носителей на различных дефектах кристаллической решетки, то есть изменением скорости направленного движения носителей при их взаимодействии с различными дефектами. Наиболее существенным является взаимодействие носителей с ионизированными атомами различных примесей и с тепловыми колебаниями решетки кристалла. В различных областях температуры процессы рассеяния, обусловленные этими взаимодействиями, сказываются по-разному.

В области низких температур, когда тепловые колебания атомов столь малы, что ими можно пренебречь, основное значение имеет рассеяние на ионизированных атомах примеси. В области же высоких температур, когда в процессе тепловых колебаний атомы решетки значительно смещаются от положения устойчивого равновесия в кристалле, на первый план выступает тепловое рассеяние.

Рассеяние на ионизированных атомах примеси. В примесных полупроводниках концентрация примесных атомов во много раз превосходит концентрацию примесей в металлах. Даже при достаточно низкой температуре большая часть примесных атомов находится

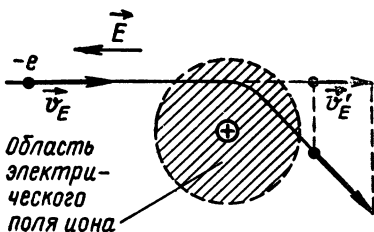


Рис. 28

в ионизированном состоянии, что представляется вполне естественным, поскольку само происхождение проводимости полупроводников связано в первую очередь с ионизацией примесей. Рассеяние носителей на ионах примеси оказывается гораздо более сильным, чем рассеяние на нейтральных атомах. Объясняется это тем, что если рассеяние носителя на нейтральном атоме происходит при непосредственном столкновении, то для рассеяния на ионизированном атоме достаточно носителю попасть в область электрического поля, создаваемого ионом (рис. 28). Когда электрон пролетает сквозь область электрического поля, создаваемого положительным ионом, траектория его полета претерпевает изменение, как показано на рисунке; при этом скорость его направленного движения v_E , приобретенная благодаря воздействию внешнего поля, уменьшится до v'_E . Если электрон проходит достаточно близко около иона, то после рассеяния направление движения электрона может оказаться вообще противоположным направлению действия внешнего электрического поля.

Рассматривая задачу о рассеянии заряженных частиц на заряженных центрах, выдающийся английский физик Э. Резерфорд пришел к выводу, что длина свободного пробега частиц пропорциональна четвертой степени их скорости:

$$\lambda \sim v_0^4.$$

Применение этой зависимости к рассеянию носителей в полупроводниках привело к очень интересному и, на первый взгляд, неожиданному результату: *подвижность носителей в области низких температур должна с повышением температуры расти*. В самом деле, подвижность

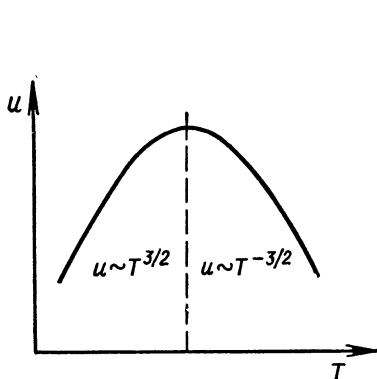


Рис. 29

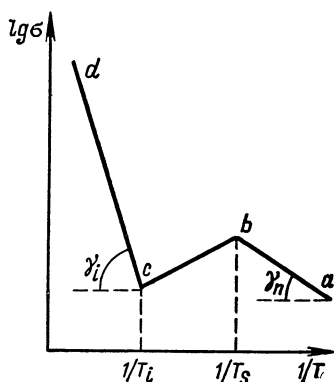


Рис. 30

носителей оказывается пропорциональной кубу скорости их движения:

$$u \sim \frac{\lambda}{v_0} \sim \frac{v_0^4}{v_0} = v_0^3.$$

В то же время средняя кинетическая энергия носителей заряда в полупроводниках пропорциональна температуре $W_k \sim kT$, а, значит, средняя тепловая скорость пропорциональна корню квадратному из T (ведь $W_k = \frac{mv^2}{2}$).

Следовательно, подвижность носителей находится в следующей зависимости от температуры:

$$u \sim v_0^3 \sim (\sqrt{T})^3 = T^{3/2}.$$

В области низких температур, когда рассеяние на ионизированных примесях играет основную роль и когда тепловыми колебаниями атомов решетки можно пренебречь, подвижность носителей растет по мере повышения температуры пропорционально $T^{3/2}$ (левая ветвь кривой $u(T)$ на рисунке 29). Качественно такая зависимость вполне объяснима: чем больше тепловая скорость носителей, тем меньше времени находятся они в поле ионизированного атома и тем меньше искажение их траектории. Благодаря этому возрастает длина свободного пробега носителей и увеличивается их подвижность.

Рассеяние на тепловых колебаниях. С повышением температуры средняя скорость теплового движения но-

сителей возрастает настолько, что вероятность их рассеяния на ионизированных примесях становится очень малой. Вместе с этим амплитуда тепловых колебаний атомов решетки при этом возрастает, так что на первый план выступает рассеяние носителей на тепловых колебаниях. Благодаря росту рассеяния на тепловых колебаниях уменьшается по мере нагрева полупроводника длина свободного пробега носителей и, следовательно, их подвижность.

Конкретный ход зависимости $\mu(T)$ в области высоких температур для различных полупроводников неодинаков. Он определяется природой полупроводника, шириной запрещенной зоны, концентрацией примеси и некоторыми другими факторами. Однако для типичных ковалентных полупроводников, в частности для германия и кремния, при не слишком больших концентрациях примеси зависимость $\mu(T)$ имеет вид:

$$\mu \sim T^{-3/2}$$

(см. правую ветвь кривой на рисунке 29).

Итак, подвижность носителей в полупроводниках в области низких температур растет прямо пропорционально $T^{3/2}$, а в области высоких температур падает обратно пропорционально $T^{3/2}$.

Зависимость проводимости полупроводника от температуры. Зная ход температурной зависимости подвижности и концентрации носителей в полупроводниках, можно установить характер температурной зависимости проводимости полупроводников. Схематически зависимость $\lg \sigma = f(1/T)$ показана на рисунке 30. Ход этой кривой очень близок к ходу кривой зависимости $\lg n = f(1/T)$, приведенной на рисунке 25. Поскольку зависимость концентрации носителей от температуры гораздо сильнее температурной зависимости их подвижности, то в областях примесной проводимости (участок ab) и собственной проводимости (участок cd) зависимость удельной проводимости $\sigma(T)$ практически полностью определяется ходом зависимости концентрации носителей от температуры. Углы наклона этих участков графика зависят соответственно от энергии ионизации атомов донорной примеси и от ширины запрещенной зоны полупроводника. Тангенс угла наклона γ_n пропорционален энергии отрыва пятого валентного электрона атома донорной примеси.

Поэтому, получив экспериментально график изменения проводимости полупроводника при нагреве на примесном участке ab , можно определить значение энергии активации донорного уровня, то есть энергетическое расстояние донорного уровня W_d от дна зоны проводимости (см. рис. 20). Тангенс угла наклона γ_i пропорционален энергии перехода электрона из валентной зоны в зону проводимости, то есть энергии создания собственных носителей в полупроводнике. Таким образом, получив экспериментально ход зависимости проводимости от температуры на собственном участке cd , можно определить ширину запрещенной зоны W_g (см. рис. 17). Величины W_d и W_g являются важнейшими характеристиками полупроводника.

Основное различие между зависимостями $\sigma(T)$ и $n(T)$ наблюдается на участке bc , расположенном между температурой истощения примесей T_s и температурой перехода к собственной проводимости T_i . Этот участок соответствует ионизированному состоянию всех примесных атомов, а для создания собственной проводимости энергия тепловых колебаний оказывается еще недостаточной. Поэтому концентрация носителей, будучи практически равной концентрации примесных атомов, не изменяется при увеличении температуры. Ход температурной зависимости проводимости на этом участке определяется ходом зависимости от температуры подвижности носителей. В большинстве случаев при умеренной концентрации примеси основным механизмом рассеяния носителей в этой области температур является рассеяние на тепловых колебаниях решетки. Этот механизм обуславливает уменьшение подвижности носителей и, следовательно, проводимости полупроводников с ростом температуры на участке bc .

В вырожденных полупроводниках благодаря большой концентрации примесей, обуславливающей перекрытие электрических полей ионов, рассеяние носителей на ионизированных атомах примеси сохраняет основное значение вплоть до высоких температур. А для этого механизма рассеяния как раз и характерно увеличение подвижности носителей с ростом температуры.

Глава 3.

НЕРАВНОВЕСНЫЕ ПРОЦЕССЫ В ПОЛУПРОВОДНИКАХ

§ 12. ГЕНЕРАЦИЯ И РЕКОМБИНАЦИЯ НЕРАВНОВЕСНЫХ НОСИТЕЛЕЙ ЗАРЯДА

Генерация свободных носителей. Мы видели, что при температуре, отличной от абсолютного нуля, любой полупроводник характеризуется некоторой концентрацией свободных носителей. Чем выше температура, тем, вообще говоря, выше концентрация свободных носителей. Появление свободных носителей объясняется переходом электронов из валентной зоны или с донорных уровней в зону проводимости. Этот процесс называется *генерацией свободных носителей* (рис. 31).

Рекомбинация свободных носителей. Однако если бы процесс генерации был единственным в полупроводнике, то к уже перешедшим в зону проводимости электронам присоединялись бы все новые и новые и через некоторое время в зоне проводимости оказались бы все валентные электроны. Экспериментально такая картина не наблюдается. Объясняется это тем, что наряду с про-

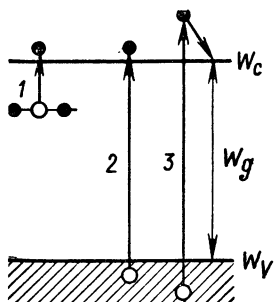


Рис. 31

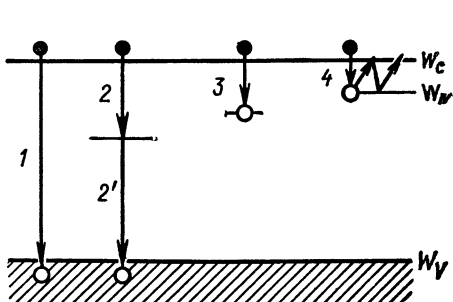


Рис. 32

цессом генерации одновременно происходит процесс *рекомбинации свободных носителей*. По истечении некоторого промежутка времени электроны из зоны проводимости возвращаются в валентную зону или на пустые места какого-либо примесного уровня (рис. 32). Иначе говоря, свободный электрон, пропутешествовав в межатомном пространстве кристалла, рано или поздно встречается с вакантным местом и занимает его. При этом если это вакантное место — положительно заряженный ион примеси, то происходит исчезновение одного электрона проводимости, а если вакантным местом является разорванная связь в кристаллической решетке — дырка, то происходит исчезновение сразу двух свободных носителей: и электрона, и дырки.

Равновесные носители. Концентрация свободных носителей в полупроводнике определяется одновременно протекающими встречными процессами — генерацией и рекомбинацией. Тепловые процессы в твердых телах всегда инерционны, то есть протекают достаточно медленно, поэтому практически при любой температуре процессы тепловой генерации и рекомбинации успевают прийти в равновесие. Благодаря этому и концентрация носителей, определяемая только тепловыми процессами, называется *равновесной*, а сами носители — *равновесными*. Поскольку для равновесных носителей процессы генерации и рекомбинации всегда оказываются уравновешенными (взамен каждой рекомбинирующей пары вследствие термического возбуждения тут же создается новая пара носителей заряда), то обычно сами процессы генерации и рекомбинации в условиях теплового равновесия не принимаются во внимание.

Неравновесные носители. Наряду с тепловым возбуждением могут действовать и другие факторы, создающие свободные носители в полупроводниках, например облучение светом, бомбардировка различными частицами, вызывающими ионизацию, и т. д. Свободные носители могут также проникать через контакт из другого тела. Свободные носители, появляющиеся в полупроводнике, благодаря этим факторам являются *избыточными* по отношению к равновесным и называются *неравновесными носителями*.

Механизм возникновения неравновесных носителей принципиально отличается от механизма образования

равновесных носителей. При тепловой генерации подводимое извне тепло определяет увеличение энергии тепловых колебаний атомов в решетке кристалла, и когда энергии этих колебаний оказывается достаточно для разрыва связей валентных электронов, происходит ионизация атомов, сопровождающаяся образованием равновесных носителей. Таким образом, подводимое тепло создает свободные носители, действуя через посредника — кристаллическую решетку полупроводника. Благодаря этому концентрация равновесных носителей оказывается одинаковой во всем объеме полупроводника.

При создании же неравновесных носителей энергия внешнего источника возбуждения (энергия фотонов света, энергия бомбардирующих электронов или других частиц) передается непосредственно валентным электронам, в то время как энергия кристаллической решетки сохраняется практически неизменной. Поэтому, во-первых, неравновесные носители при их возникновении не находятся в тепловом равновесии с кристаллом и, во-вторых, распределение их плотности по объему может быть очень неравномерным — концентрация неравновесных носителей обычно уменьшается по мере удаления от поверхности кристалла или от освещенного места.

Непосредственная передача энергии носителю заряда часто приводит к тому, что электрон, например, приобретает энергию большую, чем необходимо для возможности преодоления им запрещенной зоны. В этом случае избыточные электроны забрасываются в зону проводимости на высокие энергетические уровни (переход 3 на рисунке 31), причем кинетическая энергия таких неравновесных носителей оказывается значительно больше средней энергии $\frac{3}{2}kT$, которой обладают располагающиеся у дна зоны проводимости тепловые равновесные электроны. В дальнейшем эта избыточная кинетическая энергия быстро рассеивается в процессе столкновений неравновесных электронов с дефектами кристаллической решетки. За время 10^{-10} с быстрые электроны претерпевают около тысячи актов рассеяния, в результате чего их кинетическая энергия становится одинаковой с энергией тепловых электронов. Можно считать, что кинетическая энергия неравновесных носителей практически сразу

после их возникновения становится сравнимой с энергией равновесных носителей, благодаря чему их подвижность имеет то же значение, что и подвижность равновесных носителей. Поэтому изменение проводимости при образовании неравновесных носителей обусловлено лишь изменением суммарной концентрации носителей; мгновенные отличия энергии и подвижности неравновесных носителей, имеющиеся непосредственно после их создания, не играют роли.

Условие электронейтральности. Обозначив через Δn и Δp концентрации избыточных электронов и дырок, можно записать полные концентрации носителей так:

$$\begin{aligned} n &= n_0 + \Delta n, \\ p &= p_0 + \Delta p, \end{aligned}$$

где n_0 и p_0 — соответственно равновесные концентрации электронов и дырок. Если при внешнем воздействии на полупроводник создавались только собственные носители заряда (возбуждались переходы электронов из валентной зоны в зону проводимости), а сам полупроводник не содержит какого-либо объемного заряда, то концентрация неравновесных электронов должна быть равна концентрации избыточных дырок:

$$\Delta n = \Delta p.$$

Это — *условие электронейтральности*.

Так как подвижность избыточных носителей, как мы уже отмечали, практически не отличается от подвижности равновесных носителей, то для дополнительной электропроводности $\Delta\sigma$, возникающей в полупроводнике под воздействием внешнего возбуждающего источника, можно записать выражение

$$\Delta\sigma = e(u_n \cdot \Delta n + u_p \cdot \Delta p),$$

а для полной электропроводности — выражение

$$\sigma = e[u_n (n_0 + \Delta n) + u_p (p_0 + \Delta p)].$$

После выключения возбуждающего источника полупроводник постепенно возвращается в состояние равновесия. При этом концентрация свободных носителей в нем вновь становится равновесной. Восстановление равновесия происходит благодаря процессу рекомбина-

ции. Однако до акта рекомбинации (до встречи электрона с дыркой) каждый носитель некоторое время находится в свободном состоянии. Это время зависит от многих причин и колеблется в широких пределах — от 10^2 до 10^{-9} с. Поэтому для характеристики коллектива неравновесных носителей в целом вводят так называемое *среднее время жизни*: для неравновесных электронов τ_n и для дырок τ_p . Практически среднее время жизни, например, электрона определяется как средний отрезок времени между моментом генерации избыточного электрона и моментом захвата его дыркой.

Скорость рекомбинации. С понятием среднего времени жизни связано понятие скорости рекомбинации (скорости изменения концентрации избыточных носителей). За скорость рекомбинации принимают число рекомбинирующих пар носителей в единицу времени $\Delta n/\tau_n$ (где Δn — концентрация неравновесных носителей в данный момент времени t). Понятно, что скорость рекомбинации тем больше, чем больше концентрация неравновесных носителей. Со временем скорость рекомбинации уменьшается, так как благодаря процессу рекомбинации число избыточных носителей падает.

Простой расчет показывает, что убывание концентрации неравновесных носителей со временем происходит по экспоненциальному закону:

$$\Delta n = \Delta n_0 e^{-\frac{t}{\tau_n}},$$

где Δn_0 — начальная концентрация избыточных носителей в момент времени $t=0$ (момент выключения возбуждающего источника), $e \approx 2,7$ — основание натурального логарифма.

Из этой формулы следует, что по истечении времени τ_n после выключения возбуждающего источника концентрация неравновесных носителей уменьшится в e раз. Действительно, при $t=\tau_n$ концентрация Δn становится равной $\Delta n_0/e$. График зависимости Δn от t приведен на рисунке 33. По этому графику можно практически определить величину τ_n как отрезок, отсекаемый на оси времени касательной к кривой в точке $t=0$.

Понятие сечения захвата. Процесс рекомбинации предполагает захват электрона дыркой. Способность дырки совершать захват электрона характеризуется так

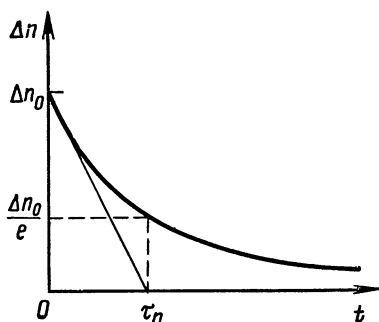


Рис. 33

называемым сечением захвата. Проведем через дырку плоскость, перпендикулярную направлению движения электрона, приближающегося к дырке. Если электрон, пересекая эту плоскость, пролетает мимо дырки на большом расстоянии, то он не будет захвачен, а лишь в большей или меньшей мере изменит направление своего движения. Но если

электрон пересекает эту плоскость на малом расстоянии от дырки, т. е. внутри круга достаточно малого радиуса, то он оказывается захваченным ею. Площадь этого круга и называется *эффективным сечением захвата* или просто сечением захвата. Сечение захвата сильно зависит от скорости электрона относительно дырки и природы этой дырки.

Пути рекомбинации. Различают два случая рекомбинации носителей заряда. В одном случае электрон совершает прямой переход из зоны проводимости в валентную зону (переход 1 на рисунке 32). Это так называемая *межзонная рекомбинация*. В другом случае переход электрона из зоны проводимости в валентную зону происходит в два этапа: сначала электрон из зоны проводимости переходит на некоторый промежуточный уровень (переход 2), созданный в кристалле каким-либо примесным атомом, а потом он с этого уровня переходит на свободное место в валентную зону (переход 2'). Такой процесс называется *рекомбинацией через примесные центры*. В обоих случаях переход электрона из зоны проводимости в валентную зону сопровождается выделением энергии, приблизительно равной ширине запрещенной зоны W_g (как правило, в рекомбинации участвуют электроны, находящиеся у дна зоны проводимости, и дырки, расположенные у потолка валентной зоны).

Освобождающаяся при рекомбинации энергия может выделяться в виде кванта излучения либо перейти в энергию кристаллической решетки. В соответствии с этим

говорят об *излучательной* и *безызлучательной рекомбинации*.

Излучательная рекомбинация. При излучательной рекомбинации энергия электрона, объединяющегося с дыркой, преобразуется в энергию фотона $h\nu$. В этом случае акт рекомбинации подобен возвращению возбужденного атома в основное состояние (электрон с возбужденной орбиты переходит на основную орбиту). Энергия фотона, испускаемого при межзонной рекомбинации, согласно закону сохранения энергии равна

$$h\nu = W_c - W_v = W_g,$$

где W_c — энергия электрона до рекомбинации (у дна зоны проводимости), а W_v — энергия электрона после рекомбинации (у потолка валентной зоны).

Излучательная рекомбинация в большинстве случаев связана с прямыми межзональными переходами электронов. Однако такие переходы в полупроводниках имеют место лишь в хорошо очищенных кристаллах с узкой запрещенной зоной, например в кристаллах антимонида индия (InSb), у которых ширина запрещенной зоны W_g равна всего 0,18 эВ. Время жизни электрона в этих кристаллах по отношению к излучательной рекомбинации мало (порядка 10^{-7} с). Поэтому практически все электроны рекомбинируют с испусканием фотона.

По мере увеличения ширины запрещенной зоны вероятность излучательной рекомбинации становится все меньше. Например, в собственном германии, для которого при комнатной температуре $W_g = 0,72$ эВ, время жизни электрона по отношению к излучательной рекомбинации составляет 1 с, в то время как по отношению к безызлучательной рекомбинации время жизни по порядку величины равно 10^{-3} — 10^{-4} с. Это означает, что на несколько тысяч безызлучательных рекомбинационных переходов приходится всего один акт рекомбинации с испусканием фотона.

Безызлучательная рекомбинация. Безызлучательная рекомбинация, характерная для полупроводников с достаточно широкой запрещенной зоной, протекает, как правило, с участием примесных центров. Объясняется это тем, что высвобождающаяся при переходе электрона из зоны проводимости в валентную зону довольно большая

энергия не может быть передана кристаллической решетке за один раз. Вероятность такого акта столь же мала, как, например, вероятность одновременного столкновения в одной точке сразу десяти частиц. В случае же рекомбинации через примесные уровни энергия электрона передается кристаллической решетке в два этапа. При этом чем ближе к середине запрещенной зоны находится энергетический уровень примесного атома, тем более вероятной оказывается рекомбинация через него, потому что энергия в этом случае выделяется в виде двух приблизительно одинаковых порций, каждая из которых равна приблизительно половине W_g . На языке зонной теории это означает, что на первом этапе примесный центр захватывает электрон, а на втором — дырку. В принципе этапы этого процесса могут происходить и в обратном порядке. Их последовательность определяется вероятностью захвата примесным центром электрона и дырки. Если энергетический уровень примесного центра расположен ближе к дну зоны проводимости, то вероятность захвата электрона будет выше вероятности захвата дырки, так как при захвате электрона будет выделяться меньшая энергия (вообще, чем меньше выделяемая энергия, тем процесс более вероятен). В этом случае электрону после его захвата примесным центром приходится некоторое время «ждать» дырку. Дело в том, что хотя дырок может быть и много, не с каждой из них может прорекомбинировать электрон. Рекомбинация возможна лишь в тех случаях, когда электрон имеет возможность отдать выделяемую энергию кристаллической решетке.

Центры захвата и центры прилипания. Может случиться так, что процесс рекомбинации избыточных носителей уже закончился, а электрон, находясь на примесном центре, так и не дождался нужной дырки. Обычно такая ситуация бывает, когда уровень примесного центра находится достаточно далеко от середины запрещенной зоны (переход 3 на рисунке 32). В этом случае примесный центр называется *центром захвата* электрона.

Если уровень примесного центра расположен недалеко от дна зоны проводимости, так что энергетическое расстояние $W_c - W_{np}$ оказывается сравнимым с энергией тепловых колебаний решетки, то электрон, попавший в такой центр, может быть снова заброшен в зону прово-

димости (случай 4 на рисунке 32). Этот процесс может повториться многократно, прежде чем электрон «свалится» в валентную зону. Такие примесные центры называются *центрами прилипания*. Наличие центров прилипания может существенно увеличить среднее время жизни неравновесных носителей.

В высокочастотных полупроводниковых приборах необходимо, чтобы избыточные носители быстро исчезали после выключения возбуждающего сигнала. Именно поэтому в таких приборах, создаваемых на основе монокристаллов кремния или германия, для получения большой скорости рекомбинации используются примеси золота, никеля, меди и некоторых других элементов, которые создают рекомбинационные уровни вблизи середины запрещенной зоны.

Поверхностная рекомбинация. Исчезновение свободных носителей может происходить не только в результате объемной рекомбинации, но и вследствие *поверхностной рекомбинации*, которая в большинстве случаев протекает гораздо более интенсивно, чем объемная. Объясняется это тем, что на поверхности полупроводника всегда имеется большое количество адсорбированных примесных атомов и различных дефектов, являющихся эффективными центрами рекомбинации. Особенно значительна роль поверхностной рекомбинации в случае тонких образцов, когда объем мал, а поверхность относительно велика.

Поскольку рекомбинация на поверхности протекает всегда более интенсивно, чем в объеме, то вблизи поверхности число свободных носителей оказывается меньшим, чем в глубине полупроводника. Образующийся перепад концентрации вызывает поток свободных носителей изнутри полупроводника к поверхности. Чем больше разность концентраций (то есть чем интенсивнее протекает рекомбинация носителей на поверхности), тем с большей скоростью движутся носители к границе полупроводника. Для уменьшения интенсивности процесса поверхностной рекомбинации образцы полупроводника подвергаются специальной обработке (химическому травлению, тщательной отмывке и т. п.), обеспечивающей удаление с поверхности рекомбинационных центров. При хорошей обработке поверхности удастся снизить скорость процесса поверхностной рекомбинации в сотни раз.

§ 13. ДИФФУЗИОННЫЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

Каким бы образом ни создавались неравновесные носители — введением избыточных носителей из контактирующего с полупроводником металла, генерацией их при облучении светом или при бомбардировке быстрыми частицами, — всегда в полупроводнике создается неравномерное распределение этих носителей. Вполне понятно, что у той поверхности полупроводника, где создаются неравновесные носители, их концентрация гораздо выше, чем в глубине полупроводника. Разность концентраций приводит к диффузии и неравновесных носителей из области высокой концентрации в область пониженной концентрации.

Диффузионный ток. Диффузионный поток неравновесных носителей электрических зарядов представляет собой электрический ток, который получил название *диффузионного*. Диффузионный ток может быть как электронным, так и дырочным. Рассмотрим некоторые характерные особенности и закономерности диффузии неравновесных электронов.

Пусть в полупроводнике концентрация электронов по мере удаления от внешней границы (которой на рисунке 34 соответствует координата $x=0$) в глубину падает в соответствии с приведенной на этом рисунке кривой. Рассечем мысленно полупроводник в точке $x=x_0$ плоскостью, перпендикулярной оси X , и выделим из объема два соседних слоя толщиной Δx . Все электроны, имевшиеся в слое 1 в некоторый начальный момент, в результате беспорядочного теплового движения через некоторое время покинут его. Поскольку вероятности движения электронов вправо и влево одинаковы, то можно считать, что половина электронов из слоя 1 перейдет через плоскость $x=x_0$. В то же самое время половина электронов из слоя 2, имеющего такую же тол-

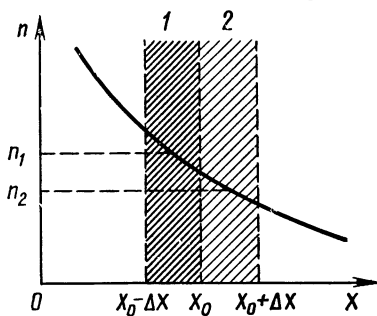


Рис. 34

тате беспорядочного теплового движения через некоторое время покинут его. Поскольку вероятности движения электронов вправо и влево одинаковы, то можно считать, что половина электронов из слоя 1 перейдет через плоскость $x=x_0$. В то же самое время половина электронов из слоя 2, имеющего такую же тол-

щину Δx , пересечет эту плоскость в обратном направлении. Поскольку средняя концентрация электронов n_1 в первом слое больше средней концентрации электронов n_2 в слое 2, то число электронов, пересекающих границу раздела между слоями слева направо, будет больше числа электронов, идущих навстречу. Разность между этими потоками как раз и составит тот результирующий диффузионный поток, наличие которого и определяет появление диффузионного тока.

Диффузионный поток будет тем больше, чем больше различие между концентрациями электронов в контактирующих слоях. Разница же между n_1 и n_2 определяется значением изменения концентрации электронов на единице длины в направлении, перпендикулярном границе раздела между слоями ($\Delta n / \Delta x$). Плотность i_n электронного диффузионного тока можно выразить формулой

$$i_n = eD_n \frac{\Delta n}{\Delta x},$$

где D_n — так называемый *коэффициент диффузии электронов*.

В случае акцепторного полупроводника плотность тока, обусловленного движением дырок, выражается аналогичной формулой:

$$i_p = -eD_p \frac{\Delta p}{\Delta x},$$

где D_p — *коэффициент диффузии дырок*.

Соотношение Эйнштейна. Коэффициент диффузии D зависит от природы и структуры вещества. Он пропорционален подвижности u носителей и абсолютной температуре T кристалла:

$$D = \frac{kT}{e} u.$$

Это — *соотношение Эйнштейна*.

Поскольку подвижность электронов больше подвижности дырок, то и коэффициент диффузии для электронов всегда больше, чем для дырок. Например, в германии коэффициент диффузии электронов при комнатной температуре в два раза больше коэффициента диффузии дырок, а в кремнии — в три раза.

Дырки преследуют электроны. Интересно отметить, что в однородном полупроводнике диффузия носителей какого-либо одного знака практически не приводит к нарушению условия электронейтральности: при замере концентраций избыточных носителей в любой точке такого полупроводника всегда оказывается, что $\Delta n = \Delta p$. Объясняется это тем, что диффузия носителей одного знака сопровождается одновременным параллельным переносом носителей противоположного знака. Поясним это конкретным примером.

Пусть из какого-либо объема полупроводника в результате диффузии ушла часть электронов. В первый момент такой уход, естественно, нарушает условие электронейтральности: Δp становится больше Δn и в рассматриваемой области возникает объемный заряд, равный $e(\Delta p - \Delta n)$. Появление объемного заряда сопровождается возникновением электрического поля, направленного в сторону перемещения ушедших электронов. Напряженность этого поля может быть очень большой: в обычных полупроводниках она достигает значений 10^5 В/см и более. Такое сильное поле вызывает интенсивный поток дырок, которые, уходя из рассматриваемого объема, восстанавливают нарушенное равновесие и тем самым способствуют исчезновению возникшего объемного заряда. Кроме того, дырки, догоняя ушедшие электроны, не дают возникнуть нарушению электронейтральности в новых областях. При внезапном нарушении электронейтральности восстановление ее происходит, конечно, не мгновенно, а по истечении некоторого времени τ_0 , называемого *временем диэлектрической релаксации*. Однако время это настолько мало (в обычных условиях — порядка 10^{-12} с), что практически диффузионный поток носителей одного знака сопровождается параллельным переносом носителей противоположного знака, благодаря чему в любом объеме однородного полупроводника всегда сохраняется условие электронейтральности.

Диффузия и рекомбинация. Процессы диффузии и рекомбинации неравновесных носителей тесно связаны друг с другом. Проиллюстрируем это на примере распространения неравновесных дырок в электронном полупроводнике. Предположим, что на одной из поверхностей полупроводника n -типа имеется источник дырок,

создающий на этой грани (при $x=0$) некоторую избыточную концентрацию Δp_0 неравновесных дырок. Поскольку внутри полупроводника концентрация дырок меньше, чем у поверхности, то возникает диффузия дырок в глубь полупроводника, то есть появляется диффузионный ток. Если бы внутри полупроводника не происходила рекомбинация носителей, то через некоторое время избыточные дырки достигли бы противоположной грани образца и на всем протяжении полупроводника установилась бы постоянная концентрация Δp_0 избыточных носителей. Однако в действительности такой картины не наблюдается, так как, диффундируя от поверхности, неравновесные дырки рекомбинируют с электронами, движущимися к поверхностному слою изнутри полупроводника, и их концентрация уменьшается от поверхности в глубь образца.

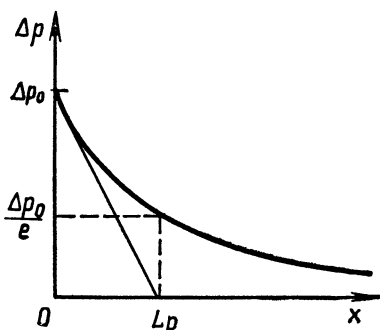


Рис. 35

Уменьшение концентрации избыточных дырок по мере удаления от границы полупроводника, на которой создаются избыточные носители, выражается экспоненциальным законом (рис. 35):

$$\Delta p = \Delta p_0 e^{-\frac{x}{L_p}}.$$

Параметр L (для дырок L_p) называется *диффузионной длиной носителей*. Величина L_p равна длине, на которой концентрация избыточных дырок уменьшается в e раз.

Связь между процессами диффузии и рекомбинации устанавливается соотношением

$$L_p = \sqrt{D_p \tau_p},$$

где τ (для дырок τ_p) — среднее время жизни неравновесных носителей (напомним, что среднее время жизни неравновесных носителей отражает средний промежуток времени между моментом их возникновения и моментом рекомбинации). Диффузионная длина L — это среднее

расстояние, которое проходят неравновесные носители в процессе диффузии за среднее время жизни, то есть средний путь от места рождения до места рекомбинации.

§ 14. ФОТОПРОВОДИМОСТЬ И ПОГЛОЩЕНИЕ СВЕТА

Одним из примеров протекания неравновесных процессов в полупроводниках является фотопроводимость, заключающаяся в появлении или изменении проводящих свойств полупроводника под действием какого-либо излучения (инфракрасного, видимого или ультрафиолетового). Как правило, облучение полупроводника светом сопровождается увеличением его электропроводности. Увеличение проводимости $\sigma = n e u$ объясняется ростом концентрации n свободных носителей (подвижность u неравновесных носителей практически не отличается от подвижности равновесных).

Фотоактивные процессы. Образование избыточных подвижных носителей при воздействии света возможно по следующим трем основным причинам (рис. 36):

1) кванты света, взаимодействуя с электронами, находящимися на примесных донорных уровнях, и отдавая им свою энергию, переводят их в зону проводимости, увеличивая тем самым концентрацию электронов проводимости (случай 1);

2) кванты света возбуждают электроны, находящиеся в валентной зоне, и переводят их на акцепторные уровни, создавая тем самым свободные дырки в валентной зоне и увеличивая дырочную проводимость полупроводника (случай 2);

3) кванты света переводят электроны из валентной зоны непосредственно в зону проводимости, создавая тем самым одновременно и подвижные дырки, и свободные электроны (случай 3).

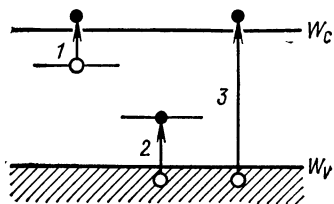


Рис. 36

Если в первых двух случаях каждое эффективное соударение кванта света с электроном сопровождается созданием одного свободного носителя (электрона или дырки), то в последнем случае каждый акт взаимо-

действия приводит к образованию сразу двух подвижных носителей заряда.

Характер нарастания и спада фотопроводимости. Рассмотрим процесс изменения проводимости под действием света на примере создания электронной фотопроводимости при возбуждении электронов с примесных уровней. Если предположить, что после включения источника света в единице объема полупроводника каждую секунду создается g свободных электронов, то в первое время нарастание числа электронов проводимости окажется очень быстрым, а затем темп этого нарастания будет постепенно уменьшаться. Объясняется это развивающимся процессом рекомбинации (скорость рекомбинации пропорциональна числу избыточных носителей, а значит, чем больше концентрация свободных электронов и чем больше число освободившихся вакантных мест на донорных уровнях, тем интенсивнее процесс рекомбинации). В конце концов процессы генерации электронов и их рекомбинации (точнее, возврат на пустые донорные уровни) уравниваются, и в полупроводнике установится некоторая стационарная концентрация $\Delta n_{\text{ст}}$ избыточных фотоэлектронов:

$$\Delta n_{\text{ст}} = g\tau_n,$$

где τ_n — время жизни фотоэлектронов:

Закон нарастания концентрации генерируемых светом электронов выражается формулой

$$\Delta n = \Delta n_{\text{ст}} \left(1 - e^{-\frac{t}{\tau_n}} \right).$$

С течением времени концентрация Δn асимптотически стремится к $\Delta n_{\text{ст}}$, как показано на рисунке 37, б (на рисунке 37, а показан ход изменения интенсивности I света, падающего на поверхность полупроводника).

При выключении освещения генерация электронов прекращается и концентрация Δn за счет процесса рекомбинации убывает до нуля по экспоненциальному закону

$$\Delta n = \Delta n_{\text{ст}} e^{-\frac{t}{\tau_n}},$$

и полупроводник возвращается в исходное состояние, характеризваемое равновесной концентрацией n_0 электро-

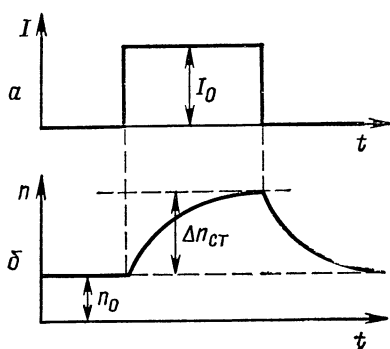


Рис. 37

нов. Скорость нарастания, как и скорость спада концентрации фотоэлектронов полностью определяется значением τ_n , которое зависит от особенностей конкретного образца.

Поскольку при освещении полупроводника изменяется только концентрация фотоэлектронов, а их подвижность остается неизменной, то

фотопроводимость полупроводника выразится так:

$$\Delta\sigma = eu_n \Delta n.$$

Характер изменения фотопроводимости со временем практически полностью повторяет временную зависимость концентрации фотоэлектронов, приведенную на рисунке 37. Значение же стационарной фотопроводимости определяется формулой

$$\Delta\sigma_{ст} = eu_n \Delta n_{ст} = eu_n g\tau_n,$$

из которой видно, что абсолютная величина фотопроводимости прямо пропорциональна времени жизни фотоэлектронов.

Квантовый выход. Интенсивность света, возбуждающего полупроводник, обычно выражают числом фотонов, падающих на образец. При освещении полупроводника часть фотонов отражается от его поверхности, другая часть может пройти сквозь полупроводник, а остальные фотоны поглощаются в объеме образца. Отношение числа фотоэлектронов g , освобожденных при поглощении света, к общему числу N поглощенных фотонов называется *квантовым выходом*:

$$\eta = \frac{g}{N}.$$

Если бы каждый акт поглощения фотона сопровождался появлением свободного электрона, то квантовый выход был бы равен единице. Однако, кроме фотоактивных, существует множество процессов поглощения фото-

нов, не сопровождающихся появлением фотоэлектронов. Поэтому обычно квантовый выход меньше единицы. В некоторых случаях квантовый выход оказывается большим единицы. Это возможно, если энергия одного фотона достаточна для создания двух или большего числа фотоэлектронов (облучение полупроводника ультрафиолетовым светом или рентгеновскими лучами). В собственном полупроводнике случай $\eta > 1$ реализуется следующим образом.

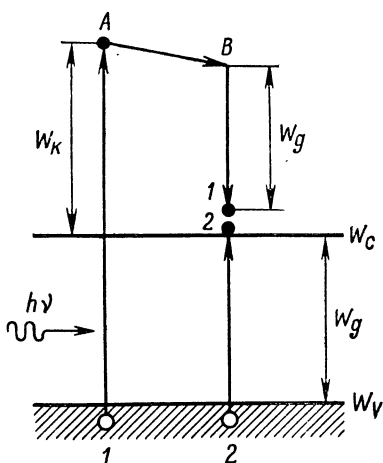


Рис. 38

Если энергия поглощенного фотона более чем вдвое превышает ширину запрещенной зоны, то получивший эту энергию электрон, перейдя в зону проводимости, будет обладать большой кинетической энергией, значение которой определяется расстоянием между энергетическим уровнем, на который перешел электрон, и дном зоны проводимости (рис. 38). Часть приобретенной энергии электрон, путешествуя в межузельном пространстве, может отдать кристаллической решетке, рассеиваясь на различных неоднородностях и примесях (участок AB). Но если на его пути встретится атом полупроводника (точка B на схеме), то он может отдать ему свою избыточную энергию (или часть ее), отрывая от него электрон и создавая еще одну пару свободных носителей заряда. Таким образом, в рассмотренном нами случае один фотон в результате двух последовательных актов ионизации создает в полупроводнике сразу четыре свободных носителя: два электрона и две дырки. Фотон с достаточно большой энергией может создать и большее число пар носителей.

Аналогично может происходить создание неравновесной проводимости и при бомбардировке полупроводника быстрыми заряженными частицами, например электронами: первичные электроны создают в объеме полупро-

водника быстрые вторичные электроны, которые в свою очередь, ионизируя атомы кристаллической решетки, обуславливают появление новых пар носителей. Такой процесс размножения свободных носителей продолжается до тех пор, пока энергия создаваемых при ионизации электронов проводимости достаточна для акта перевода электрона из валентной зоны в зону проводимости.

Квантовый выход может быть больше единицы также в том случае, когда первичный фотон (квант света), обладая при входе в полупроводник энергией, превосходящей не менее чем вдвое ширину запрещенной зоны ($h\nu_1 > 2W_g$), отдает при первой ионизации не всю свою энергию, а только часть ее, необходимую для перевода электрона из валентной зоны в зону проводимости (рис. 39). В результате этого процесса образуется новый фотон с энергией $h\nu_2$, большей ширины запрещенной зоны; такой фотон может создать еще одну пару носителей, столкнувшись с атомом полупроводника.

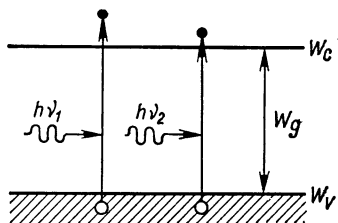


Рис. 39

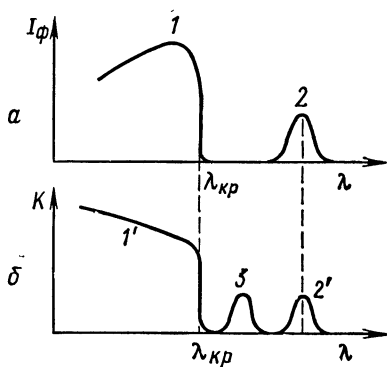


Рис. 40

Спектральная кривая фототока. Если полупроводниковый образец включить в электрическую цепь и осветить его, то в нем, помимо темнового тока (тока, текущего через образец в отсутствие освещения), возникнет дополнительный фототок I_{ϕ} , значение которого зависит от длины волны падающего света. При некоторых значениях длины волны λ фототок достигает больших величин, а в некоторых диапазонах изменения λ он оказывается равным нулю. Зависимость силы фототока от длины волны λ возбуждающего света называется *спектральной кривой фототока* (рис. 40, а).

Собственная фотопроводимость. Спектральная кривая, приведенная на рисунке 40, а, относится к примесному полупроводнику. На ней четко прослеживаются два максимума фототока, отражающие появление в полупроводнике под действием света собственных (участок 1) и примесных (участок 2) носителей заряда. Если полупроводник не имеет никаких примесей (чистый собственный полупроводник), то в спектре фототока будет наблюдаться только максимум 1. В этом случае отсутствие фотопроводимости при $\lambda > \lambda_{кр}$ объясняется тем, что в области больших длин волн энергии фотонов недостаточно для перевода электронов из валентной зоны в зону проводимости. Отсутствие фототока будет наблюдаться во всей длинноволновой области, где $h\nu < W_g$. При длине волны, которой соответствует энергия фотонов, сравнимая с шириной запрещенной зоны ($h\nu \approx W_g$), у электрона, находящегося вблизи потолка валентной зоны, появляется возможность за счет энергии поглощенного фотона перейти в зону проводимости. Так как такая возможность возникает у большого числа собственных электронов и при каждом акте взаимодействия фотона с атомом рождается пара носителей — электрон и дырка, то рост фототока происходит очень быстро.

Длина волны $\lambda_{кр}$, при которой появляется фотопроводимость, может быть определена из равенства

$$h\nu_{кр} = W_g,$$

определяющего положение так называемой *красной границы фотоэффекта*¹.

По мере уменьшения длины волны в области $\lambda < \lambda_{кр}$ фотопроводимость полупроводника падает. Объясняется это следующим образом.

При $h\nu \gg W_g$ поглощение света становится настолько сильным, что фактически весь падающий свет поглощается в очень тонком поверхностном слое полупроводника. Поскольку объем этого слоя невелик, то концентрация свободных носителей в нем (как электронов, так

¹ Здесь имеется в виду *внутренний фотоэффект*, когда электроны под действием света переводятся только в зону проводимости, в отличие от *внешнего фотоэффекта*, при котором поглощение фотона сопровождается выходом электрона за пределы полупроводника.

и дырок) становится очень большой, что обуславливает резкое возрастание скорости рекомбинации носителей и уменьшение времени их жизни. Росту скорости рекомбинации способствует и наличие в поверхностном слое большого числа различных примесей и дефектов, которые сильно снижают подвижность носителей. Таким образом, интенсивная рекомбинация и сильное рассеяние носителей на поверхностных дефектах в конце концов сводят на нет вклад фотопроводимости в общую проводимость образца. Чем лучше обработана поверхность кристалла, тем дальше в область коротких волн распространяется фоточувствительность образца.

Примесная фотопроводимость. Максимум 2 (см. рис. 40, а), соответствующий пику примесной фотопроводимости, в полупроводниках n -типа наблюдается при энергии фотонов, достаточной для возбуждения электронов, находящихся на донорных уровнях. В полупроводнике p -типа этот максимум соответствует энергии фотонов, достаточной для активации дырок, при которой электроны, переходя из валентной зоны на акцепторные уровни, оставляют в ней свободные дырки. Интересно отметить, что по обе стороны от максимума примесная фотопроводимость равна нулю. Если отсутствие фотопроводимости в области больших длин волн отражает тот факт, что энергии приходящих квантов света просто не хватает для ионизации примесных центров, то равенство нулю фототока в области более коротких волн указывает на резонансный характер взаимодействия фотонов с примесными атомами: когда энергия возбуждающих квантов света существенно отличается от энергии активации примесных центров, взаимодействия между ними не происходит.

Спектр поглощения. На рисунке 40, б приведена *спектральная кривая поглощения* света полупроводником (по оси ординат отложен коэффициент поглощения K). Спектр поглощения во многом похож на спектр фотопроводимости, но вместе с тем он имеет и свои особенности. Так, в полосе собственного поглощения (участок 1'), которая, как и полоса собственной фотопроводимости, обусловлена межзонными переходами электронов и начинается от красной границы $\lambda_{кр}$, по мере уменьшения длин волн света коэффициент поглощения не падает, а растет. Объясняется это тем, что причины, обуслови-

вающие падение фотопроводимости в области коротких длин волн, на поглощение света не влияют. Поэтому увеличение энергии возбуждающих фотонов сопровождается переводом в зону проводимости электронов со все более глубоких уровней валентной зоны.

Полоса примесного поглощения (участок с максимумом 2') практически точно соответствует полосе примесной фотопроводимости (участок 2) и обусловлена поглощением фотонов, вызывающим образование свободных примесных носителей при ионизации донорных или акцепторных центров.

Наряду с фотоактивными процессами поглощения света, сопровождающимися образованием избыточных свободных носителей заряда, в спектре поглощения могут быть участки, на которых поглощение квантов света не связано с образованием фотоносителей. Наиболее интересным с физической точки зрения является участок 3 *экситонного поглощения*.

Экситонное поглощение. Понятие об экситонах было введено еще в 1931 году известным советским физиком Я. И. Френкелем именно для того, чтобы объяснить, почему поглощение света, явно связанное с возбуждением электронов, не сопровождается изменением фотопроводимости. В дальнейшем представления Я. И. Френкеля были блестящим образом подтверждены экспериментально. Оказалось, что собственные атомы полупроводника могут поглощать энергию не только, когда $h\nu \geq W_g$, но и когда $h\nu < W_g$. Поскольку при этом энергии поглощаемого кванта света недостаточно для полной ионизации атома и перехода одного из валентных электронов в зону проводимости, то атом переходит в возбужденное состояние, при котором один из его валентных электронов располагается на возбужденной оболочке. Это возбужденное состояние атома и получило название *экситона* (от английского слова *excitation* — возбуждение). Благодаря тому что все собственные атомы в кристаллической решетке полупроводника эквивалентны, возбужденное состояние может передаваться в кристалле от атома к атому без изменения. При этом возбужденный атом возвращается в основное, нормальное состояние, а выделяющаяся энергия возбуждения передается соседнему атому, в котором электрон переходит на соответствующую возбужденную оболочку. Внешне такой про-

цесс воспринимается как перемещение экситона по кристаллу.

Экситон можно представить еще в виде пары электрон — дырка, связанных между собой кулоновскими силами притяжения. Ясно, что перемещение такой связанной пары отражает только распространение по кристаллу возбужденного состояния и не приводит к направленному переносу какого-либо заряда. Поэтому экситонное поглощение, имея свою собственную полосу в спектре поглощения (максимум 3 на рисунке 40, б), не влияет на ход фотопроводимости (на спектральной кривой фотопроводимости (см. рис. 40, а) соответствующего максимума нет).

Вообще говоря, экситонное поглощение может косвенным образом отразиться на фотопроводимости кристалла. В процессе перемещения по кристаллу экситон может быть разорван либо тепловыми колебаниями решетки, либо дополнительным квантом света с энергией, меньшей W_g (в отсутствие экситонного возбуждения такой квант вообще не поглотился бы в собственном полупроводнике). В результате разрыва экситона возникают два носителя заряда — электрон и дырка, которые вносят свой вклад в фотопроводимость. Экситон может встретить на пути и какой-либо примесный атом и отдать ему энергию, оторвав от него электрон и создав дополнительный свободный носитель заряда. Энергия рекомбинации экситона может также вызвать переход электрона из валентной зоны на акцепторный уровень. При этом создается свободная дырка, способная переносить заряд. Жизнь экситона может завершиться и при встрече с каким-либо неактивным примесным центром или с дефектом решетки. В этом случае энергия рекомбинации экситона передается тепловым колебаниям решетки.

§ 15. ЛЮМИНЕСЦЕНЦИЯ

Люминесценцией называют все виды «холодного» свечения. Различают около десятка различных видов люминесценции. Среди них есть, правда, и *термолюминесценция*, заключающаяся в свечении некоторых твердых тел (например, мрамора, алмаза) при слабом подогреве, но этот подогрев настолько мал, что природа наблюдаемого свечения не имеет никакой связи с тем повышением

температуры тела, которое необходимо для появления видимого теплового излучения.

Некоторые кристаллические тела светятся при трении друг о друга или при сильной деформации и разрушении. Такое свечение называется *триболюминесценцией*. Иногда свечение наблюдается при протекании различных химических процессов, и тогда оно называется *хемилюминесценцией*. Возникает свечение и при бомбардировке некоторых веществ электронами — *катодолюминесценция*. Разреженные газы при пропускании через них электрического тока излучают яркое свечение, цвет которого определяется природой светящегося газа; в этом случае говорят об *электролюминесценции*. Существует, кроме этого, *рентгенолюминесценция*, *радиолюминесценция* и др. Но наиболее распространенной является *фотолюминесценция* — свечение, обусловленное воздействием на вещество лучистой энергии — потока света.

Фотолюминесценция. Явление фотолюминесценции было открыто около 500 лет назад итальянским сапожником Винченцио Касчиорола, занимавшимся в свободное время алхимией. Но только в настоящее время после появления квантовой теории света и развития зонной теории оно получило достаточное объяснение. Различают два частных случая фотолюминесценции: *флюоресценцию* и *фосфоресценцию*.

Термин «флюоресценция» происходит от названия минерала флюорита (полевой шпат). Флюоресценция состоит в том, что под действием возбуждающего света некоторые вещества начинают сами излучать. Излучение это наблюдается одновременно с внешним облучением и исчезает практически сразу после прекращения облучения. Кроме флюорита, флюоресцируют цинковая обманка, ряд стекол и растворы многих химических соединений.

В отличие от флюоресценции фосфоресценция характеризуется определенной длительностью свечения вещества после выключения возбуждающего источника света. Со временем собственное свечение фосфоресцирующих веществ постепенно затухает. Некоторые вещества способны фосфоресцировать в продолжение нескольких месяцев, особенно при некотором подогреве. Фосфоресценция характерна для многих полупроводниковых соеди-

нений, в которые введены специальные примеси, называемые *активаторами*. Независимо от длительности послесвечения все вещества, обладающие способностью светиться под воздействием света, называются *люминофорами*. Следует отметить, что свечение люминофоров может вызываться воздействием невидимого излучения, особенно ультрафиолетового.

Явление фотолюминесценции связано с поглощением световой энергии и переводом атомов из нормального состояния в состояние возбуждения. Известно, что в возбужденном состоянии атом может находиться совсем незначительное время. Он довольно быстро должен возвратиться в стационарное состояние, выделив захваченную энергию. Однако поглощают свет практически все тела, а люминесцируют лишь некоторые.

Флюоресценция. Элементарная схема флюоресцирующего люминофора приведена на рисунке 41. Отличительной особенностью этой схемы является наличие активаторного уровня W_a . Возникает такой уровень при введении в кристалл специальных примесей — активаторов. По своим свойствам активаторные примеси отличаются от донорных и акцепторных примесей. Энергетические уровни, появляющиеся в кристалле при введении активаторных примесей, располагаются вблизи валентной зоны, то есть там, где обычно располагаются уровни атомов акцепторной примеси, способных захватывать электроны из валентной зоны. Однако по своему поведению атомы активаторов отличны от акцепторов — они не захватывают электроны. Находясь в невозбужденном состоянии, они, напротив, проявляют скорее свойства донорных атомов и могут отдавать свои валентные электроны в зону проводимости. Во всяком случае энергия, необходимая для отрыва валентного электрона от атома активатора, меньше, чем энергия отрыва валентного электрона от атома собственного полупроводника (на рисунке 41 это свойство отражено тем, что расстояние активаторного уровня W_a от зоны проводимости меньше W_g). Таким образом, активаторная примесь, создавая уровни в области, характерной для акцепторов, обладает в некоторой степени свойствами доноров. Образование активаторных уровней возможно и в процессе формирования кристаллов в естественных усло-

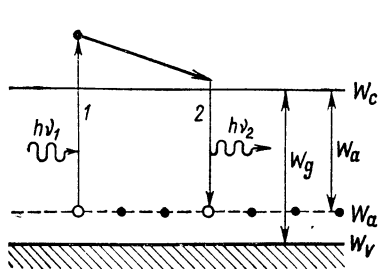


Рис. 41

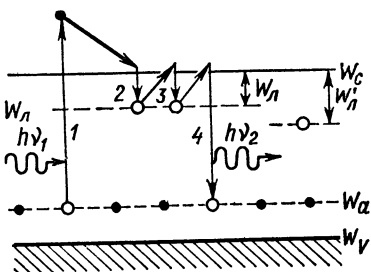


Рис. 42

виях. Поэтому ряд минералов обладает способностью к фотолюминесценции.

При освещении флюоресцирующего кристалла светом с энергией фотонов $h\nu_1 > W_a$ происходит поглощение фотонов атомами активатора, приводящее к их ионизации. Электроны, переходя с уровня W_a в зону проводимости (переход 1 на рисунке 41), превращаются в свободные электроны проводимости (появление фотолюминесценции всегда сопровождается возникновением фотопроводимости). Путешествуя в межузельном пространстве, электрон может встретиться с уже ионизированным атомом активатора и прорекомбинировать с ним. При возвращении электрона из зоны проводимости на уровень W_a (переход 2) излучается фотон с энергией $h\nu_2$, что и воспринимается как эффект флюоресценции. Перемещаясь в межузельном пространстве, электрон может растратить часть своей энергии, рассеиваясь на испытывающих тепловые колебания атомах решетки или на различных дефектах. Поэтому энергия излученного фотона не может превышать энергию поглощенного. Из это следует, что

$$\nu_2 \leq \nu_1, \text{ или } \lambda_2 \geq \lambda_1.$$

В этом суть установленного экспериментально английским физиком Д. Г. Стоксом правила, согласно которому длина волны света, испускаемого флюоресцирующим веществом, не может быть меньше длины волны возбуждающего света. Иначе говоря, если способное к флюоресценции вещество осветить, например, синим светом, то само оно может излучать свет зеленый, желтый или красный.

Время, которое проводит электрон в зоне проводимости, то есть время между моментом возбуждения электрона и моментом его рекомбинации, оказывается очень малым. В чистых полупроводниках, не содержащих, кроме атомов активатора, других примесей, это время может составить всего 10^{-9} с. Поэтому флюоресцирующее излучение наблюдается только непосредственно в процессе возбуждения кристалла первичным светом и исчезает практически мгновенно после выключения возбуждающего излучения.

Фосфоресценция. Для получения более или менее длительного свечения люминофора после окончания возбуждения в полупроводниковый кристалл необходимо, кроме активаторов, ввести примесные центры, выполняющие роль электронных ловушек. Энергетические уровни, создаваемые этими примесями, располагаются вблизи дна зоны проводимости (уровни W_d на рисунке 42), а сами центры по своим физическим свойствам аналогичны центрам прилипания и центрам захвата (см. § 12).

Начальная стадия возбуждения люминофора и создание в нем фотопроводимости происходит так же, как и в случае флюоресценции. Атомы активатора, поглощая фотоны с энергией $h\nu_1$, ионизируются и отдают свои валентные электроны в зону проводимости. Освободившийся электрон, встретившись с ионизированным атомом активатора, может прорекомбинировать с ним. При этом излучается фотон с энергией $h\nu_2$. Это механизм флюоресценции. Но возможны также встречи электронов с примесными ловушечными центрами (переход 2 на рисунке 42). В этом случае электроны теряют способность перемещаться по кристаллу и тем самым выключаются из участия в фотопроводимости. Кроме того, они не могут рекомбинировать с ионами активатора (переход с уровня W_d на уровень W_a запрещен). За счет тепловых колебаний решетки (если энергия теплового движения атомов кристалла сравнима с энергией активации ловушечного центра, то есть если $W_d \approx kT$) электрон может вернуться в зону проводимости (переход 3). После этого электрон может быть опять захвачен ловушкой или прорекомбинировать с повстречавшимся ионизированным атомом активатора; в последнем случае испускается фотон с энергией $h\nu_2$ (переход 4).

Из сказанного видно, что судьба электрона, попавшего в ловушку, не зависит от наличия первичного возбуждающего освещения. Поэтому как остаточная фотопроводимость, так и способность кристалла излучать вторичный свет (фосфоресценция) могут сохраняться достаточно длительное время и после выключения возбуждающего источника. Длительность послесвечения определяется глубиной залегания уровня $W_{\text{л}}$ и температурой кристалла. Если ловушечный уровень расположен недалеко от дна зоны проводимости и энергия активации ловушки при температуре кристалла T окажется сравнимой с энергией тепловых колебаний решетки ($W_{\text{л}} \approx kT$), то время послесвечения окажется небольшим. При более глубоком залегании ловушечного уровня (в случае если $W_{\text{л}} \gg kT$) послесвечение может оказаться достаточно длительным, так как для высвобождения электрона из ловушки в этом случае требуется флуктуация тепловых колебаний, энергии которой было бы достаточно для заброса электрона в зону проводимости. А это случается не так уж часто.

Стимулировать люминесценцию можно путем подогрева кристалла (термостимулированная люминесценция). Когда энергия тепловых колебаний становится сравнимой с $W_{\text{л}}$, происходит высвобождение электронов из ловушек и наблюдается люминесценция. Именно методом постепенного нагрева и определяют глубину залегания ловушечных уровней в запрещенной зоне кристалла. Всплески интенсивности люминесценции наблюдаются при тех температурах, которые удовлетворяют условию $W_{\text{л}} = kT$. Так, для кристалла, энергетическая схема которого приведена на рисунке 42, первый максимум послесвечения будет по мере нагревания наблюдаться при некоторой температуре T , удовлетворяющей условию $kT = W_{\text{л}}$, а второй максимум — при температуре T' , удовлетворяющей условию $kT' = W'_{\text{л}}$. Так как уровень $W'_{\text{л}}$ залегает более глубоко ($W'_{\text{л}} > W_{\text{л}}$) и для его возбуждения необходима большая энергия, то $T' > T$. Если уровень ловушек располагается достаточно далеко от дна зоны проводимости, то способность кристалла к люминесценции может сохраняться бесконечно долго. Так, некоторые минералы сохраняют в течение миллионов лет (со времени их образования до наших дней) способность люминесцировать при первом подогреве.

КОНТАКТНЫЕ ЯВЛЕНИЯ

§ 16. РАБОТА ВЫХОДА ЭЛЕКТРОНОВ ИЗ МЕТАЛЛА

Концентрация свободных электронов в металлах очень велика (порядка 10^{23} см⁻³). Находясь в постоянном хаотическом движении, они беспрерывно подходят к поверхности. Однако если металл не нагрет, то выхода электронов за пределы образца практически не наблюдается. Объясняется это следующим образом.

Внутри металла каждый свободный электрон при перемещении в межузельном пространстве взаимодействует со всеми окружающими его электронами, а также с положительно заряженными атомными остатками, создающими кристаллическую решетку. Благодаря равномерному распределению заряженных частиц в объеме кристалла равнодействующая всех сил, действующих на каждый электрон, оказывается в среднем практически равной нулю. При подходе же электрона к поверхности, а тем более при его выходе из кристалла равномерность распределения заряженных частиц вокруг него нарушается, и это приводит к возникновению сил, препятствующих выходу электрона за пределы кристалла. Можно рассмотреть две причины, затрудняющие выход электронов: образование у границы металла двойного электрического слоя и возникновение при выходе электрона так называемой силы зеркального отображения.

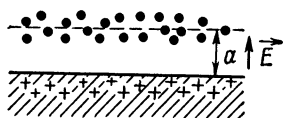


Рис. 43

Двойной электрический слой. Даже при абсолютном нуле температуры свободные электроны в металле обладают достаточно большой кинетической энергией. Однако притя-

жение к узлам кристаллической решетки, заряженным положительно, не дает электрону покинуть металл совсем. Благодаря этому металл оказывается окруженным электронным облаком (рис. 43). Образующийся у поверхности двойной электрический слой можно рассматривать как плоский конденсатор, одной обкладкой которого являются приповерхностные положительные ионы, оставшиеся нескомпенсированными после ухода электронов, а другой — представленное в виде тонкого слоя электронное облако. Очевидно, что напряженность поля внутри такого конденсатора может считаться постоянной. Если a — расстояние между указанными обкладками, то сила F_1 , препятствующая прохождению электрона, вычисляется по формуле

$$F_1 = \frac{e^2}{4a^2},$$

а энергия W_1 , которую необходимо затратить электрону на преодоление двойного слоя, по формуле

$$W_1 = F_1 a = \frac{e^2}{4a}.$$

Сила зеркального отображения. Преодолев двойной слой, электрон тем не менее еще не освобождается от воздействия со стороны металла. Дальнейшему удалению его препятствует *сила зеркального отображения*: на электрон, находящийся на расстоянии x от поверхности металла, действует со стороны этой поверхности сила, которая может быть определена как сила взаимодействия данного электрона с частицей, имеющей положительный заряд, равный модулю заряда электрона, и находящейся в глубине металлического тела на расстоянии x от поверхности (рис. 44). По закону Кулона сила зеркального отображения, действующая на электрон в вакууме, выражается формулой

$$F_2 = \frac{e^2}{4x^2}.$$

Если предположить, что сила F_1 при выходе за пределы двойного слоя непрерывно переходит в силу F_2 , то общий график силы, действующей на электрон по мере его удаления от поверхности металла, можно представить кривой, приведенной на рисунке 45. Энергия же,

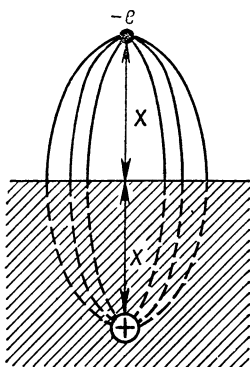


Рис. 44

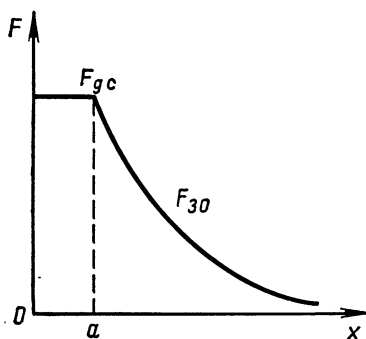


Рис. 45

необходимая для преодоления силы зеркального от-
ражения, равна

$$W_2 = \frac{e^2}{4a}.$$

Полная работа выхода. Электрон, покидающий ме-
талл, должен преодолеть своеобразный потенциальный
барьер, показанный на рисунке 46. Часть этого барьера
(на участке от 0 до a) создана силами двойного слоя,
а другая часть (на участке от a до ∞) — силами зер-
кального отражения. Общая же высота W_0 потен-
циального барьера равна той работе, которую должен
совершить электрон на пути от поверхности металла до
полного отделения от металла. Величина W_0 называется

*полной работой выхо-
да.* Экспериментально
определенные значения
полной работы выхода
для различных метал-
лов лежат в интервале
от 3 до 20 эВ.

Ход потенциального
барьера можно рас-
сматривать как кри-
вую, отражающую из-
менение потенциальной
энергии электрона по
мере его удаления от

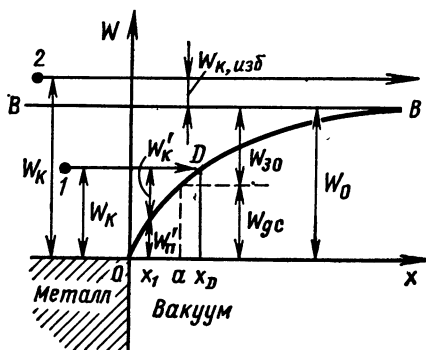


Рис. 46

поверхности металла. Действительно, электрон, находящийся внутри металла и обладающий некоторой кинетической энергией W_k (рассмотрим, например, электрон, обозначенный цифрой 1), может перемещаться в любом направлении без изменения энергии. Однако если он переходит границу металла — вакуум, то кинетическая энергия электрона в соответствии с законом сохранения энергии начинает превращаться в потенциальную. Так, если электрон отошел от границы на расстояние x_1 , то часть его кинетической энергии перейдет в потенциальную $W'_п$. Поскольку, дойдя до точки x_1 , электрон сохранил еще часть кинетической энергии W'_k , он может и дальше уходить от поверхности. Однако при удалении электрона на расстояние x_D уже вся его кинетическая энергия преобразуется в потенциальную (точка D на кривой потенциальной энергии электрона). В этой точке скорость электрона оказывается равной нулю, после чего он возвращается обратно.

Для того чтобы электрон имел возможность окончательно покинуть металл, его кинетическая энергия должна быть не меньше высоты потенциального барьера W_0 . Такой возможностью обладает, например, электрон, обозначенный на рисунке цифрой 2. Более того, поскольку начальная кинетическая энергия этого электрона превышает полную работу выхода, то, даже покинув металл, он сохранит некоторую скорость, соответствующую оставшейся неизрасходованной избыточной кинетической энергии:

$$W_{k, \text{ изб}} = \frac{mv^2}{2}.$$

Обычно за нулевой уровень отсчета энергии принимают не уровень неподвижного электрона в глубине металла, а уровень, соответствующий неподвижному электрону, находящемуся в вакууме достаточно далеко от поверхности металла и не испытывающему воздействия с его стороны. На рисунке 46 это уровень BB , к которому асимптотически приближается кривая потенциальной энергии электрона. В этом случае объем металла является для электрона потенциальной ямой, глубина которой равна W_0 . При таком выборе нулевого уровня электроны, находящиеся внутри металла и расположенные ниже уровня BB , имеют отрицательную

энергию. О полной же работе выхода тогда говорят, что это работа, которая должна быть совершена электроном, первоначально находившимся на дне потенциальной ямы, при его выходе из этой ямы.

§ 17. УРОВЕНЬ ФЕРМИ В МЕТАЛЛАХ И ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ФЕРМИ — ДИРАКА

Уровень Ферми. Несмотря на огромное количество свободных электронов в металле, располагаются они по энергетическим уровням потенциальной ямы в строгом порядке. Каждый из электронов занимает вакантное место на возможно более низком уровне. И это вполне естественно, так как всякая система, будучи предоставлена самой себе, то есть в отсутствие внешнего воздействия, всегда стремится перейти в состояние с наименьшей энергией. Распределение электронов по уровням подчинено принципу Паули, согласно которому никакие две частицы не могут находиться в совершенно одинаковых состояниях. В силу этого на каждом энергетическом уровне может расположиться не более двух электронов, да и то имеющих различные направления спинов. По мере укомплектования нижних уровней происходит заселение все более высоко расположенных уровней. Если в рассматриваемом образце металла имеется N свободных электронов, то в отсутствие теплового возбуждения, то есть при абсолютном нуле температуры ($T=0$), все свободные электроны разместятся попарно на $N/2$ нижних

уровнях (рис. 47). Самый высокий энергетический уровень потенциальной ямы металла, занятый электронами при $T=0$, называется *уровнем Ферми*¹ и обозначается бук-

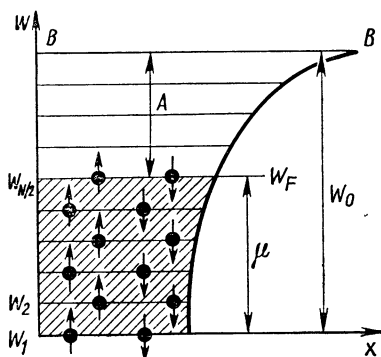


Рис. 47

¹ Свое название этот уровень получил в честь выдающегося итальянского физика Э. Ферми, разработавшего совместно с известным английским физиком П. Дираком теорию поведения коллективов частиц, ведущих себя как электроны в металле.

вой μ или W_F . Энергия электрона, находящегося на этом уровне, называется *энергией Ферми*. Все энергетические уровни, расположенные выше уровня Ферми, при $T=0$ оказываются абсолютно пустыми.

Вполне очевидно, что для выхода электронов, находящихся на уровне Ферми, за пределы металла должна быть совершена работа

$$A = W_0 - \mu.$$

Величина A , равная энергетическому расстоянию между уровнем удаленного электрона BB и уровнем Ферми, называется *термодинамической работой выхода* или просто *работой выхода*. Именно эта величина определяет поведение различных металлов при установлении контакта между ними или при создании контакта металл — полупроводник.

Функция распределения Ферми — Дирака. Характер распределения частиц по разным уровням или состояниям в тех или иных условиях определяется так называемой *функцией распределения*. В общем случае функция распределения описывает вероятность занятости того или иного уровня частицами. Если достоверно известно, что данный уровень заселен частицей, то говорят, что вероятность обнаружения частицы на этом уровне равна 1. Если же с полной достоверностью можно сказать, что на рассматриваемом уровне нет частиц, то говорят, что вероятность обнаружения частиц в рассматриваемом состоянии равна 0. Однако во многих случаях нельзя достоверно утверждать, что уровень заполнен или пуст. Тогда вероятность нахождения частицы на рассматриваемом уровне отлична от нуля, но меньше единицы. При этом чем больше вероятность обнаружить частицу на рассматриваемом уровне, тем ближе к единице оказывается значение функции распределения для соответствующего состояния.

Если по оси абсцисс откладывать значения энергии, соответствующей разным уровням, от дна потенциальной ямы до ее потолка, а по оси ординат — вероятность заполнения электронами соответствующих уровней, то мы получим график функции распределения

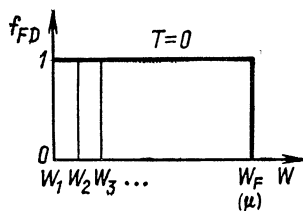


Рис. 48

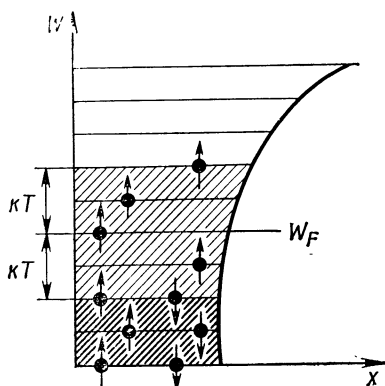


Рис. 49

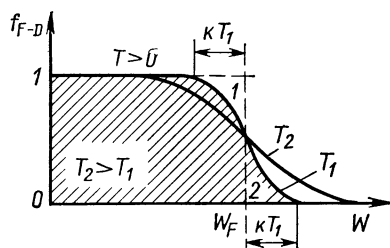


Рис. 50

Ферми — Дирака f_{F-D} . При $T=0$ он имеет вид, приведенный на рисунке 48. Часто этот график называют *ступенькой Ферми*. Из него видно, что при $T=0$ все уровни, вплоть до уровня Ферми, оказываются занятыми электронами. В точке $W=\mu$ функция распределения скачкообразно падает до нуля; это значит, что все уровни, расположенные выше уровня Ферми, пусты.

Влияние температуры.

При температурах, отличных от нуля, вид графика зависимости $f_{F-D}(W)$ отличается от приведенного на рисунке 48. Повышение температуры приводит к появлению теплового возбуждения электронов, которое они получа-

ют от тепловых колебаний кристаллической решетки. Благодаря этому возбуждению часть электронов, расположенных на наиболее высоких заполненных уровнях, переходит на пустые уровни, лежащие выше уровня Ферми (рис. 49). Вероятность обнаружения электронов на этих уровнях становится уже отличной от нуля. Одновременно с этим из-за ухода части электронов с некоторых уровней, расположенных непосредственно под уровнем Ферми, вероятность заполнения их окажется меньше единицы. Таким образом, повышение температуры приводит к некоторому «размытию» границы ступеньки Ферми: вместо скачкообразного изменения от 1 к 0 функция распределения совершает плавный переход. На рисунке 50 пунктиром показан вид графика функции распределения электронов по уровням при $T=0$, а сплошными линиями отражены распределения электронов при температурах, отличных от нуля. Площадь криволинейного тре-

угольника, расположенного под кривой распределения правее значения W_F (площадка 2), пропорциональна числу электронов, перешедших на возбужденные уровни, а площадь такого же треугольника, расположенного слева от значения W_F над кривой распределения (площадка 1), пропорциональна числу электронов, ушедших с уровней, которые ранее были заполненными, то есть числу освободившихся под уровнем Ферми мест. Понятно, что площади этих двух треугольников одинаковы, так как с разных позиций они выражают одно и то же число электронов.

Следует отметить, что в диапазоне рабочих температур степень размытия кривой распределения электронов в металле очень невелика. Объясняется это тем, что тепловому возбуждению подвергаются только те электроны, которые расположены на энергетических уровнях, непосредственно примыкающих к уровню Ферми. Можно качественно оценить энергетическую глубину залегания уровней, подвергающихся возбуждению. Из молекулярной физики известно, что кинетическая энергия частиц, обусловленная тепловым движением, выражается так:

$$\frac{mv^2}{2} = \frac{3}{2} kT.$$

Следовательно, значение энергии, которую могут передать электронам испытывающие тепловые колебания атомы кристаллической решетки, по порядку величины равно kT . При комнатной температуре $kT \approx 0,025$ эВ, в то время как энергия Ферми для металлов при этой температуре лежит в диапазоне от 3 до 10 эВ. Поэтому оказывается, что в обычных условиях в переходах на более высокие энергетические уровни могут принимать участие не более 1 % всех свободных электронов. Причем это как раз те электроны, энергия которых близка к энергии Ферми. Что же касается электронов, заселяющих энергетические уровни, расположенные в глубине потенциальной ямы и удаленные от уровня Ферми больше чем на kT , то они не принимают участия в тепловом возбуждении, из-за чего распределение этих электронов остается таким же, как и при абсолютном нуле.

Физический смысл уровня Ферми. Обсуждая в § 6 способность твердых тел проводить электрический ток, мы пришли к выводу, что проводимость связана с возможностью перехода электронов на более высокие энер-

гетические уровни, то есть определяется возможностью получения электронами ускорения во внешнем электрическом поле. В металлах при $T > 0$ такая возможность имеется только у электронов, находящихся в области размытия функции распределения, так как реальные электрические поля не в состоянии вырвать электроны из глубины потенциальной ямы и перевести их на свободные уровни, энергия которых выше W_F (перейти же на соседние, более высоко расположенные уровни глубинные электроны не могут, потому что все эти уровни заняты). Следовательно, при $T > 0$ энергия Ферми имеет смысл наиболее вероятной или средней энергии электронов металла, могущих принять участие в проводимости при данной температуре. Эти электроны ответственны не только за создание электрической проводимости. Именно они определяют вклад электронной теплоемкости в общую теплоемкость кристалла и в значительной степени определяют теплопроводность кристалла.

Уровень Ферми в металлах практически не изменяет своего положения по мере повышения температуры. С ростом температуры степень возбуждения электронов растет, и они переходят на более высоко расположенные уровни. Одновременно с этим возбуждению подвергаются и все более глубоко расположенные уровни, имеющие меньшую энергию. Кривая распределения при $T_2 > T_1$ (см. рис. 50) «размывается» более сильно, чем при T_1 , но в равной степени вправо и влево. Поэтому средняя энергия электронов, принимающих участие в проводимости, остается практически неизменной. Это тем более справедливо, что между возбужденными уровнями идет постоянный обмен электронами.

§ 18. УРОВЕНЬ ФЕРМИ В ПОЛУПРОВОДНИКАХ

Вопрос об уровне Ферми в полупроводниках оказывается несколько более сложным, чем в случае металлов. Связано это в первую очередь с тем, что в полупроводниках на уровне Ферми, как правило, электронов нет. Но физический смысл уровня W_F (или μ) остается тем же, что и в случае металлов: это уровень, определяющий среднюю энергию электронов (или носителей вообще), могущих принять участие в проводимости. Именно

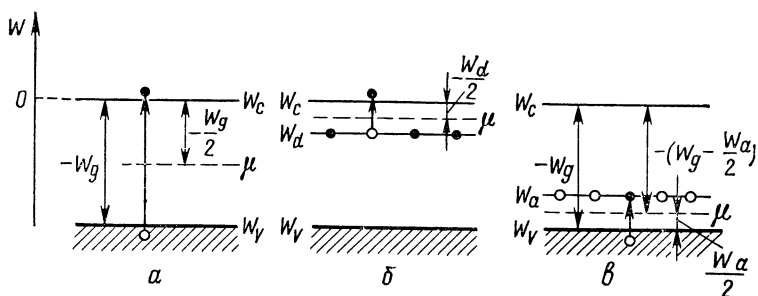


Рис. 51

поэтому уровень μ и в случае полупроводников называют уровнем Ферми.

Выясним расположение уровня Ферми в полупроводниках.

Собственный полупроводник. В теории полупроводников за начало отсчета энергии обычно принимают, как и в металлах, уровень, соответствующий дну зоны проводимости.

Тепловое возбуждение собственного полупроводника сопровождается переходами электронов из валентной зоны в зону проводимости. Перешедшие в зону проводимости электроны не находятся там постоянно, а спустя очень короткое время возвращаются обратно в валентную зону. При этом на смену им из валентной зоны приходят новые электроны. Благодаря такому обмену в проводимости в равной степени принимают участие как электроны, расположенные на нижних уровнях зоны проводимости, так и электроны, находящиеся на верхних уровнях валентной зоны. Энергия первых при указанном выше начале отсчета равна нулю, а энергия вторых равна $-W_g$ (знак «минус» означает, что положительные значения энергии от дна зоны проводимости откладываются вверх). Таким образом, средняя энергия электронов, принимающих участие в проводимости, равна $-\frac{W_g}{2}$, то есть

уровень Ферми в собственных полупроводниках располагается посередине запрещенной зоны (рис. 51, а).

Примесные полупроводники. В области достаточно низких температур, близких к абсолютному нулю, в электронном полупроводнике тепловое возбуждение мо-

жет перевести в зону проводимости только те электроны, которые находятся на примесном донорном уровне W_d , в то время как электроны, находящиеся на уровнях валентной зоны, не могут принять участия в проводимости, потому что энергия тепловых колебаний решетки в области низких температур оказывается недостаточной для перевода таких электронов в зону проводимости ($W_g \gg W_d$). Полагая, что электроны, перешедшие в зону проводимости, находятся вблизи ее дна и имеют энергию, близкую к нулю, можно считать (на тех же основаниях, что и в случае собственного полупроводника), что средняя энергия электронов, принимающих участие в проводимости, равна $-\frac{W_d}{2}$. Таким образом, уровень μ в донорном полупроводнике при низких температурах располагается в запрещенной зоне на расстоянии $\frac{W_d}{2}$ от дна зоны проводимости (см. рис. 51, б).

В полупроводниках *p*-типа для создания дырочной проводимости необходимо перевести электроны из валентной зоны на акцепторные уровни, находящиеся на расстоянии W_a от ее потолка. Рассуждения, аналогичные приведенным выше, позволяют сделать вывод, что уровень μ в дырочном полупроводнике в области низких температур располагается посередине между потолком валентной зоны и уровнем W_a акцепторной примеси. Поскольку отсчет энергии мы ведем от дна зоны проводимости, то

$$\mu = -W_g + \frac{W_a}{2}$$

(см. рис. 51, в).

Влияние температуры на положение уровня Ферми. В собственном полупроводнике положение уровня Ферми не зависит от температуры. Объясняется это тем, что процессы, определяющие проводимость собственного полупроводника, не меняют своего характера при изменении температуры. Конечно, с ростом температуры и увеличением энергии теплового возбуждения электроны будут переходить на все более высокие уровни зоны проводимости, достаточно удаленные от ее дна. Однако одновременно с этим тепловое возбуждение будет переводить в зону проводимости электроны, располагающиеся

на наиболее глубоких уровнях валентной зоны. Поэтому средняя энергия носителей заряда, принимающих участие в проводимости, остается неизменной, и уровень Ферми в собственном полупроводнике независимо от температуры сохраняет свое положение в середине запрещенной зоны.

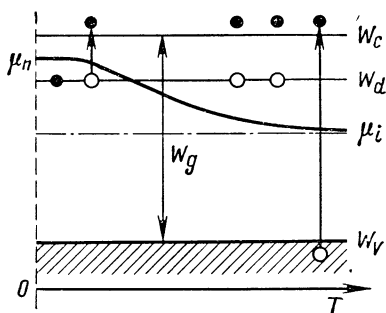


Рис. 52

Иная картина наблюдается в примесных полупроводниках. В них, как мы выяснили еще в § 10, с ростом температуры на смену примесной проводимости приходит собственная проводимость. Это и является причиной изменения положения уровня μ . Рассмотрим для определенности полупроводник n -типа.

В области низких температур наличие у электронного полупроводника проводящих свойств обусловлено переходами электронов в зону проводимости с донорных уровней. Поэтому и уровень Ферми расположен посередине между примесным уровнем и дном зоны проводимости. По мере повышения температуры происходит, как мы знаем, истощение примесей, а при $T > T_s$ донорный уровень оказывается практически пустым. Но в этой области температур на смену переходам в зону проводимости с донорного уровня все в большей степени приходят переходы из валентной зоны — начинает проявляться собственная проводимость. При высоких температурах, когда $T > T_i$, проводимость полупроводника практически полностью определяется переходами электронов из валентной зоны, из-за чего и средняя энергия электронов проводимости становится равной $-\frac{W_g}{2}$: уровень μ в этих условиях занимает положение в середине запрещенной зоны. Таким образом, мы видим, что по мере роста температуры уровень Ферми в электронном полупроводнике изменяет свое положение от $-\frac{W_d}{2}$ до $-\frac{W_g}{2}$ (рис. 52).

Аналогично изменяется с ростом температуры положение уровня Ферми и в дырочном полупроводнике:

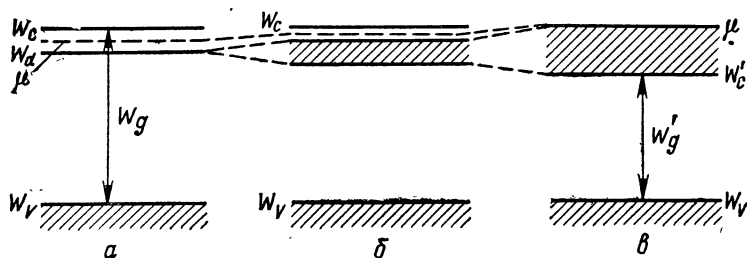


Рис. 53

располагаясь при $T=0$ посередине между акцепторным уровнем W_a и потолком валентной зоны, он при повышении температуры переходит в середину запрещенной зоны.

Уровень Ферми в вырожденных полупроводниках. В отличие от обычных, не очень сильно легированных полупроводников, у которых уровень Ферми лежит в запрещенной зоне, вырожденные полупроводники, в которых концентрация легирующей примеси велика, характеризуются расположением уровня μ в одной из разрешенных зон: в валентной зоне или в зоне проводимости. Выясним на примере вырожденного полупроводника n -типа, как по мере увеличения концентрации легирующей примеси происходит перемещение уровня Ферми из запрещенной зоны в зону проводимости.

Пока концентрация примеси не очень велика и примесный донорный уровень узок и расположен вблизи зоны проводимости, уровень Ферми находится в запрещенной зоне на расстоянии $-\frac{W_d}{2}$ от дна зоны проводимости (рис. 53, а). Когда же концентрация примеси возрастает настолько, что примесный уровень «размывается» в примесную зону, энергия активации электронов, находящихся на донорных уровнях, становится меньше, поскольку потолок размытой примесной зоны располагается ближе к зоне проводимости, чем примесный уровень в исходном состоянии. Благодаря этому и средняя энергия электронов проводимости становится меньше, и, следовательно, уровень Ферми оказывается ближе к дну зоны проводимости (рис. 53, б). При очень высокой концентрации примесных атомов донорная зо-

на настолько «размывается», что смыкается с зоной проводимости (рис. 53, в). При этом электроны с верхних уровней донорной зоны могут беспрепятственно перемещаться по пустым уровням зоны проводимости и полупроводник приобретает свойства металла (отпадает необходимость в тепловом возбуждении). Уровень Ферми, располагающийся при малой концентрации примесных атомов между потолком примесной зоны и дном зоны проводимости, после смыкания зон (в случае большой концентрации примесных атомов) сливается с верхним уровнем примесной зоны. Таким образом, в вырожденном полупроводнике уровень Ферми выполняет ту же роль, что и в металлах: он представляет собой самый верхний уровень, занятый электронами при $T=0$. Поскольку роль дна зоны проводимости после смыкания зон выполняет уже нижний уровень размывшейся примесной зоны (уровень W'_c на рисунке 53, в), то уровень Ферми в вырожденном полупроводнике n -типа оказывается в зоне проводимости.

Совершенно аналогичная картина наблюдается при увеличении концентрации акцепторной примеси в дырочном полупроводнике. Акцепторный уровень, размываясь (с ростом концентрации примеси) в примесную зону, смыкается с потолком валентной зоны. При этом уровень Ферми вырожденного дырочного полупроводника оказывается в глубине валентной зоны, и чем больше концентрация легирующей примеси, тем глубже в валентной зоне он располагается.

Уровень Ферми и работа выхода электронов в полупроводниках. Положение уровня Ферми определяет не только тип проводимости, но и специфические свойства вводимой примеси: чем меньше энергия активации примесных центров, тем ближе к соответствующей зоне располагается уровень Ферми. Кроме того, положение уровня Ферми характеризует концентрацию примесных центров и температуру полупроводника.

Таким образом, можно сказать, что положение уровня Ферми определяет электрофизические свойства полупроводника.

С расположением уровня Ферми связано также значение работы выхода электронов. На первый взгляд эта связь кажется довольно странной. Какую роль может играть для значения работы выхода уровень, на котором

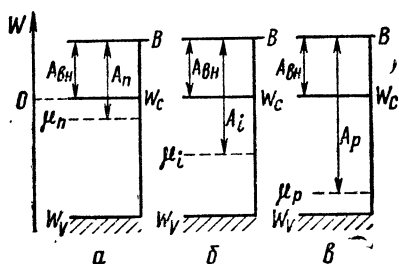


Рис. 54

в общем случае электронов вообще нет? Но тем не менее такая связь есть. И объяснить ее можно следующим образом.

Для выхода из полупроводника свободный электрон, находящийся у дна зоны проводимости, должен совершить работу, равную энергетическому расстоянию от дна

зоны проводимости до ее потолка, то есть до уровня, соответствующего энергии свободного электрона, бесконечно удаленного от полупроводника. Эту работу $A_{вн}$ (рис. 54) называют *внешней работой выхода*. Однако в невозбужденном полупроводнике электронов в зоне проводимости нет. Для того чтобы их создать, необходимо затратить определенную энергию на перевод электронов в зону проводимости с донорного уровня или из валентной зоны. Иначе говоря, необходимо затратить некоторую энергию на создание электронов проводимости. Так как средняя энергия электронов проводимости равна энергии Ферми, то есть энергетическому расстоянию от уровня Ферми до дна зоны проводимости, то работа, затраченная на создание электрона проводимости и его вывод с дна зоны проводимости за пределы полупроводника оказывается равной расстоянию от уровня Ферми до потолка зоны проводимости. Эта работа, как и в металлах, называется *термодинамической работой выхода* или просто *работой выхода*.

На рисунке 54 показаны значения работы выхода соответственно для донорного *а*, собственного *б* и акцепторного *в* полупроводников. Значение термодинамической работы выхода определяет поведение полупроводника во всех контактных явлениях и в процессах, связанных с выходом электронов за пределы кристалла.

Внешняя работа выхода $A_{вн}$ определяется природой самого кристалла и не зависит от типа вводимой примеси. Поэтому о различии значений термодинамической работы выхода в одинаковых полупроводниках, содержащих различные примеси (например, на границе раздела образцов из кремния с электронной и дырочной про-

водимостью), можно судить по расстояниям уровней Ферми от дна зоны проводимости, не учитывая положения потолка зоны проводимости и значения внешней работы выхода $A_{\text{вн}}$.

§ 19. КОНТАКТНАЯ РАЗНОСТЬ ПОТЕНЦИАЛОВ

Рассмотрим процессы, происходящие при сближении двух металлов, с различными работами выхода A_1 и A_2 до такого расстояния, чтобы между ними мог происходить обмен электронами. Такая ситуация возможна даже в отсутствие непосредственного контакта между металлами, так как в металлах всегда имеется определенное число электронов, способных выйти за пределы границ образца. Подавляющее большинство таких электронов при малых температурах, как мы уже видели, возвращается обратно в образец. Но если металлы сблизить настолько, что возвращающие силы, действующие у поверхности каждого из них, смогут захватить «чужие» электроны, далеко отошедшие от «своего» кристалла, то становится возможным довольно интенсивный обмен электронами между телами.

В первое время после сближения обмен этот оказывается далеко не равноправным. При одной и той же температуре тел поток электронов из металла, характеризующегося меньшей работой выхода A_1 , будет больше встречного потока, идущего из металла с большей работой выхода A_2 . Поэтому поверхность первого металла будет заряжаться отрицательно, а поверхность второго — положительно. Возникающая разность потенциалов препятствует дальнейшему преимущественному перетеканию электронов. Когда разность потенциалов U_k между телами достигнет значения, при котором $|e| U_k = A_2 - A_1$, число электронов, преодолевающих вакуумный промежуток между металлами во встречных направлениях, окажется одинаковым и рассматриваемая система металл — вакуум — металл придет в состояние динамического равновесия. Разность потенциалов U_k между металлами при установлении такого равновесия называется *контактной разностью потенциалов*.

Рассмотрим процесс образования контактной разности потенциалов на основе зонной теории.

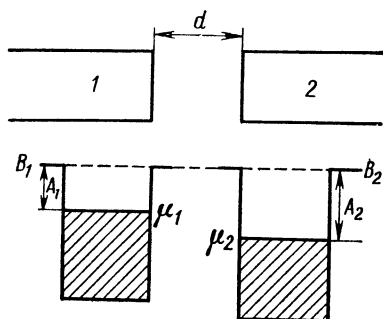


Рис. 55

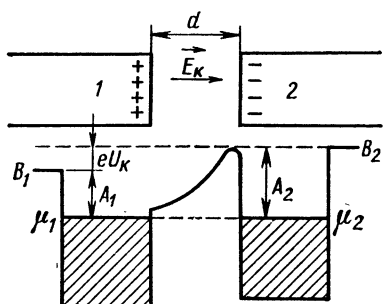


Рис. 56

На рисунке 55 показаны уровни Ферми и значения работы выхода для рассматриваемых двух металлов. Преимущественный переход электронов из металла 1 в металл 2 сопровождается изменением их потенциалов и соответствующим смещением энергетических уровней (по внешней шкале). Уровни металла 1, заряжающегося положительно, опускаются, а уровни металла 2, заряжающегося отрицательно, поднимаются. Изменение положения энергетических уровней будет происходить до тех пор, пока условия перехода электронов из одного металла в другой не окажутся одинаковыми. Так, для электронов, находящихся на

уровнях Ферми, условие равновесия выражается равенством

$$A_2 = A_1 + eU_k,$$

где eU_k есть та работа, которую должен совершить электрон, идущий из металла 1 в металл 2, преодолевая контактную разность потенциалов U_k . Равновесное состояние между металлами иллюстрирует рисунок 56.

Если до начала перехода электронов из металла 1 в металл 2 (рис. 55) между уровнями Ферми разных металлов имеется некоторое энергетическое расстояние, то после установления статистического равновесия эти уровни оказываются одинаковыми (по внешней энергетической шкале). В этом смысле процесс установления статистического равновесия между коллективами электронов двух металлов, приведенных в контакт, аналогичен установлению одинаковых уровней жидкости в сообщающихся сосудах.

Зная работы выхода A_1 и A_2 для обоих металлов, можно определить контактную разность потенциалов U_k :

$$U_k = \frac{A_2 - A_1}{e}.$$

Эта величина в зависимости от выбора пары металлов лежит в пределах от десятых долей вольта до нескольких вольт.

Выравнивание уровней Ферми при контакте двух тел является необходимым условием статистического равновесия свободных носителей не только в случае двух металлов, но и при контакте любых тел (полупроводников, диэлектриков).

Электрическое поле, создаваемое контактной разностью потенциалов при сближении металлов, локализуется непосредственно в месте их контакта. Для того чтобы убедиться в этом, оценим количество электронов, которое переходит из одного металла в другой при установлении равновесной контактной разности потенциалов. Предположим, что $U_k = 1$ В. Так как минимальное расстояние d , на которое можно сблизить два кристалла, не может быть меньше расстояния между атомами в кристаллической решетке, то примем это расстояние равным параметру кристаллической решетки $a_0 \approx 0,3$ нм. Контактирующие поверхности металлов можно рассматривать как обкладки плоского конденсатора. Напряженность электрического поля в зазоре после установления контактной разности потенциалов U_k определяется формулой

$$E_k = \frac{U_k}{d}.$$

С другой стороны, напряженность поля связана с поверхностной плотностью зарядов σ соотношением

$$E_k = \frac{\sigma}{\epsilon_0},$$

где ϵ_0 — электрическая постоянная вакуума. Поэтому число электронов n , переходящее из одного металла в другой через единицу поверхности контакта, может быть найдено следующим образом:

$$n = \frac{\sigma}{e} = \frac{\epsilon_0 E_k}{e} = \frac{\epsilon_0 U_k}{ed}.$$

Подставляя числовые значения, получаем:

$$n \approx 2 \cdot 10^{13} \text{ см}^{-2}.$$

В то же время число свободных электронов, приходящихся на 1 см² поверхности металла, равно по порядку величины 10¹⁵ см⁻². Следовательно, контактная разность потенциалов даже при минимально возможном зазоре создается за счет перехода с контактирующей поверхности одного металла на контактирующую поверхность другого всего 2 % свободных поверхностных электронов.

Именно поэтому контактное электрическое поле оказывается локализованным в контактном зазоре и почти не захватывает объемные области металлов.

§ 20. КОНТАКТ МЕТАЛЛА С ПОЛУПРОВОДНИКОМ

Состояние равновесия. Для конкретности рассмотрим сначала электронный полупроводник, у которого работа выхода электронов меньше, чем у контактирующего с ним металла. Пусть сначала между полупроводником и металлом существует незначительный вакуумный промежуток, достаточно тонкий для того, чтобы рассматриваемые кристаллы могли эффективно обмениваться электронами. После осуществления контакта между полупроводником и металлом на границе между ними происходят такие же процессы, как в случае контакта между двумя металлами. После выравнивания уровней Ферми возникает контактная разность потенциалов порядка нескольких вольт. Основное отличие от случая контакта между двумя металлами сводится к тому, что, поскольку в обычных, невырожденных полупроводниках концентрация свободных носителей оказывается много меньше, чем в металлах, может случиться, что обеднение электронами одного лишь наружного слоя полупроводника окажется недостаточным.

Оценим число ушедших из полупроводника электронов при расстоянии между контактирующими поверхностями в 1 мкм и в случае плотного контакта.

Пользуясь формулой, полученной в предыдущем параграфе, находим, что для установления контактной разности потенциалов, равной 1 В, при $d = 1$ мкм с каждого квадратного сантиметра контактной поверхности

полупроводника должно перейти в металл приблизительно $5 \cdot 10^7$ электронов. В полупроводнике средней степени легированности плотность свободных электронов при комнатной температуре обычно составляет величину порядка 10^{18} см^{-3} . Число же свободных электронов в наружном моноатомном слое окажется по порядку величины равным $(10^{15})^{2/3} \text{ см}^{-2} = 10^{10} \text{ см}^{-2}$. Таким образом, в этом случае для создания контактной разности потенциалов вполне достаточно даже менее 1 % электронов, имеющихся в моноатомном слое полупроводника. Ясно, что при этом контактное электрическое поле практически не заходит не только в металл, но и в полупроводник и полностью сосредоточено в вакуумном промежутке.

Запорный слой. Совершенно иная картина наблюдается при создании плотного контакта металла с полупроводником, при котором расстояние между ними по порядку величины равно постоянной кристаллической решетки a_0 (для германия $a_0 \approx 0,5 \text{ нм}$). В этом случае для создания контактной разности потенциалов 1В через каждый квадратный сантиметр поверхности контакта из полупроводника в металл должно перейти порядка 10^{13} электронов. Приход такого количества электронов в металл практически не скажется на его объемных свойствах, так как общее число перешедших электронов составляет приблизительно всего 1 % от числа свободных электронов, имеющихся в моноатомном слое металла; все приходящие электроны размещаются в ближайшем к контакту поверхностном слое металла. Совсем по-другому обстоит дело в полупроводнике. Поскольку в 1 см^2 моноатомного слоя полупроводника средней степени легированности имеется всего 10^{10} свободных электронов, то переход 10^{13} электронов может реализоваться, если перейдут все свободные электроны из 1000 атомных слоев приконтактной области полупроводника (рис. 57, а). Эта область, в которой после установления равновесия остаются лишь неподвижные некомпенсированные положительные ионы доноров, по своим свойствам представляет собой типичный диэлектрик. Толщина этой области в обычно используемых в приборах полупроводниках составляет $10^{-3} \div 10^{-5} \text{ см}$ и во много раз превосходит длину свободного пробега носителей. Из-за отсутствия свободных носителей и благодаря большой протяженности по сравнению с длиной свободного

пробега обедненная область обладает большим сопротивлением и поэтому называется *запорным слоем*.

Понятно, что в этом случае электрическое поле уже не локализуется в вакуумной промежутке на границе раздела тел, а существует также в запорном слое толщиной d_0 . Более того, поскольку ширина вакуумной промежутка во много раз меньше протяженности запорного слоя, то можно считать, что электрическое поле образуется именно в обедненном слое полупроводника.

Изгиб зон. Проникновение контактного поля в полупроводник и, следовательно, изменение потенциала полупроводника в разных сечениях на различные величины приводит к тому, что энергетические уровни и зоны смещаются в сторону больших или меньших значений энергии, причем эти смещения для разных сечений полупроводника различны. В результате этого происходит *изгиб энергетических уровней и зон* в приконтактной области полупроводника (рис. 57, б).

Поведение свободных носителей в электрическом поле, существующем в приконтактной зоне полупроводника, можно при качественном рассмотрении явлений определять на основе следующих наглядных представлений: электроны уподобляют тяжелым шарикам, а дырки — пузырькам газа в жидкости; находясь на искривленном энергетическом уровне, электрон как бы скатывается по этому уровню в сторону его понижения; дырка же как бы всплывает, перемещаясь вдоль уровня, на котором она находится, в сторону его повышения. Согласно такой модели в рассматриваемом нами случае контакта электроны, находящиеся в зоне проводимости вблизи границы раздела, должны скатиться по дну зоны, оставляя приконтактную область обедненной основными носителями. В то же время дырки в валентной зоне должны всплыть на поверхность, обогащая прилегающую к границе раздела область неосновными носителями. Так как концентрация неосновных носителей всегда много меньше концентрации основных, то приход дополнительных дырок в приконтактную область хотя и повышает их концентрацию, но тем не менее не может в заметной степени компенсировать уход основных носителей — электронов. Поэтому область запорного слоя и характеризуется высоким сопротивлением.

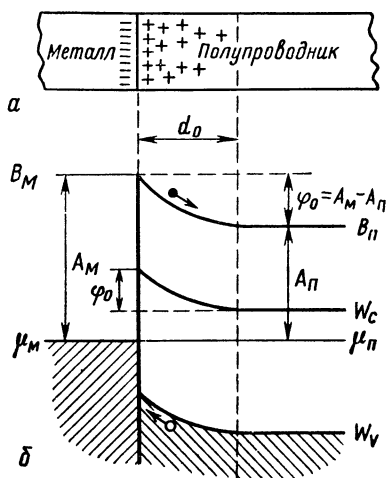


Рис. 57

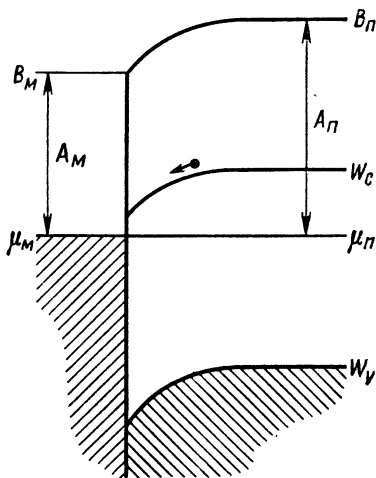


Рис. 58

Антизапорные слои. Образование запирающего слоя при контакте электронного полупроводника с металлом возможно только в том случае, если работа выхода электронов из металла A_M больше работы выхода из полупроводника A_P ($A_M > A_P$). Если же $A_M < A_P$, то наблюдается противоположная картина. В этом случае после создания контакта поток электронов из металла будет вплоть до установления динамического равновесия превалировать над встречным потоком электронов из полупроводника, из-за чего полупроводник будет заряжаться отрицательно, а металл — положительно. Направление электрического поля в приграничной зоне полупроводника и соответствующие изменения потенциала полупроводника в различных сечениях таковы, что энергетические уровни и зоны в полупроводнике, искривляясь, поднимаются по мере удаления от границы с металлом. Поэтому электроны, приходящие из объема полупроводника, «скатываются» к границе контакта и тем самым обогащают эту область основными носителями (рис. 58). Повышение концентрации основных носителей в приконтактном слое приводит к снижению его сопротивления, и поэтому такой слой называют *антизапорным*.

Аналогичные процессы протекают и при создании контакта металла с дырочным полупроводником. Отличие заключается лишь в том, что запирающий слой в этом случае возникает, когда работа выхода у металла меньше, чем у полупроводника ($A_M < A_P$). Когда же $A_M > A_P$, образуется антизапирающий слой.

§ 21. ВЫПРЯМЛЯЮЩИЕ СВОЙСТВА КОНТАКТА «МЕТАЛЛ — ПОЛУПРОВОДНИК»

Из рассмотренных видов контактов наибольший интерес представляют контакты, при создании которых образуются запирающие слои. Объясняется это тем, что именно такие контакты обладают ярко выраженной односторонней проводимостью. Если в замкнутой цепи имеется контакт с запирающим слоем, то при одной полярности включения источника сила тока велика (прямое направление), а в случае противоположной полярности включения источника сила тока оказывается во много раз меньшей (запирающее направление). Рассмотрим проявление этого эффекта на примере уже знакомого нам контакта полупроводника n -типа с металлом. Зонная схема такого контакта, находящегося в равновесном состоянии, приведена на рисунке 59, а.

Прежде всего обратим внимание на то, что в отличие от рисунка 57 на этой схеме не указаны энергетические уровни, соответствующие энергии электронов в вакууме (нет уровней B_M и B_P). Такое упрощение связано с тем, что при рассмотрении контактных явлений выход электронов в вакуум уже не представляет интереса, и, следовательно, нет необходимости указывать на схеме абсолютные значения термодинамической работы выхода A_M и A_P для металла и полупроводника. Разность же работ выхода контактирующих тел, определяющая условия установления равновесия и характеризующая свойства границы раздела, отражена на схеме высотой потенциального барьера $\phi_0 = A_M - A_P$.

Потенциальный барьер, возникающий на границе раздела, является регулятором потока электронов, идущих из полупроводника в металл. В начальный момент после образования контакта число электронов, проходящих за 1 с из полупроводника в металл, намного превосходит встречный поток, то есть число электронов, проходящих

за 1 с из металла в полупроводник. Это происходит потому, что работа выхода из рассматриваемого полупроводника меньше работы выхода из металла. Но затем по мере возрастания разности потенциалов и, следовательно, напряженности электрического поля, препятствующего переходу электронов из полупроводника в металл, происходит выравнивание обоих противоположно направленных потоков. При этом ток через контакт не течет.

Включение внешней разности потенциалов приводит к нарушению равновесия и к появлению тока через поверхность раздела. Сила тока зависит от полярности подключения источника и значения приложенной разности потенциалов.

Действие внешней разности потенциалов в прямом направлении. Рассмотрим сначала случай, когда приложенная к контакту «металл — полупроводник» внешняя разность потенциалов действует навстречу контактной разности потенциа-

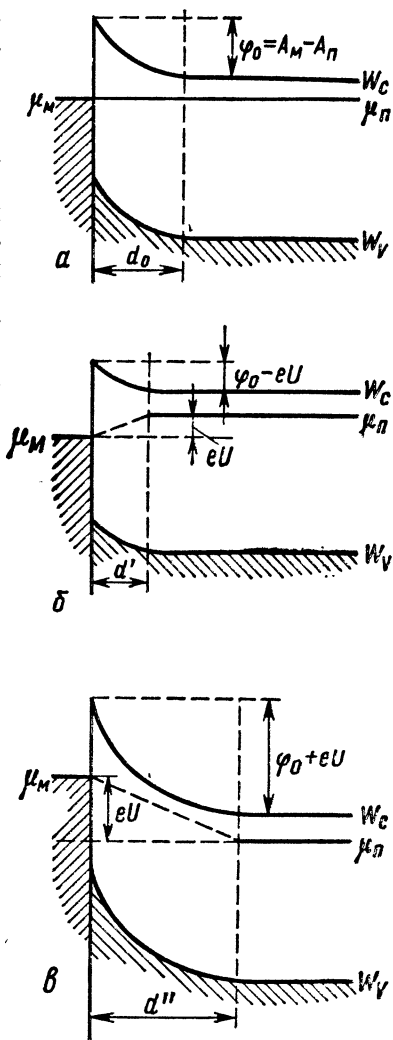


Рис. 59

(где U — приложенная разность потенциалов). Такое смещение уровней (рис. 59, б) приводит к уменьшению потенциального барьера, который должны преодолеть электроны, идущие из полупроводника в металл. Он теперь становится равным

$$\varphi = \varphi_0 - eU.$$

Уменьшение потенциального барьера φ приводит к резкому возрастанию числа электронов, переходящих из полупроводника в металл. В то же время потенциальный барьер на пути электронов, идущих из металла в полупроводник, сохраняет свою величину неизменной. Поэтому поток электронов из металла остается таким же, каким он был в отсутствие внешнего поля. Таким образом, баланс токов, текущих через контакт, оказывается нарушенным: поток электронов из полупроводника в металл во много раз превышает встречный поток электронов из металла. Нарушение баланса потоков приводит к появлению тока через контакт, направленного из металла в полупроводник, причем сила тока тем больше, чем больше разность потенциалов U , приложенная к контакту.

Благодаря тому что сопротивление объема полупроводника, и тем более металла, во много раз меньше сопротивления запорного слоя, приложенная разность потенциалов практически полностью падает на приконтактной области полупроводника. Это приводит к уменьшению в приконтактной области разности потенциалов и напряженности поля, выталкивающего электроны из этой области в объем полупроводника. Изгиб энергетических зон при этом, естественно, уменьшается, так же как уменьшается и протяженность области, обедненной основными носителями. Говоря иными словами, приложение внешней разности потенциалов в рассматриваемом случае приводит не только к снижению потенциального барьера φ , но и к уменьшению толщины запорного слоя d_0 и, следовательно, к уменьшению сопротивления контакта. Оба эти фактора способствуют увеличению силы тока, текущего через контакт.

В этом случае о внешней разности потенциалов говорят, что она приложена в *прямом*, или *пропускном*, направлении.

Действие внешней разности потенциалов в обратном направлении. Диаметрально противоположная картина наблюдается в исследуемом контакте, когда действия приложенной к нему внешней разности потенциалов и контактной разности потенциалов происходят в одном направлении. В этом случае говорят, что внешняя разность потенциалов приложена в *обратном*, или *запорном*, направлении.

Подведение такой разности потенциалов приводит к тому, что все энергетические уровни полупроводника, в том числе и уровень Ферми, теперь опускаются на величину eU (рис. 59, в) относительно положения, которое они имели в условиях равновесия. Высота потенциального барьера, препятствующего переходу электронов из полупроводника в металл, при этом увеличивается и становится равной

$$\varphi = \varphi_0 + eU,$$

а интенсивность потока электронов, идущих из полупроводника в металл, уменьшается. Поскольку потенциальный барьер на пути электронов, идущих из металла в полупроводник, при этом сохраняет свое значение неизменным, то баланс встречных потоков электронов через плоскость контакта нарушается: поток электронов, идущих из металла, будет больше встречного потока, благодаря чему через контакт начинает протекать ток, направленный от полупроводника к металлу.

Отметим, что нарушение равновесия встречных потоков электронов через контакт «металл — полупроводник» в случаях действия внешнего напряжения в прямом и запорном направлениях происходит по различным схемам. В первом случае равновесие нарушается из-за *в о з р а с т а н и я* потока электронов, идущих из полупроводника в металл. Ток, возникающий при этом в цепи, обусловлен преобладанием этого потока над встречным потоком электронов, остающимся таким же, как в отсутствие внешнего напряжения. Преобладание это тем больше, чем больше приложенная разность потенциалов. Поэтому с повышением внешней разности потенциалов в определенных пределах возрастание тока оказывается практически неограниченным.

В случае же, когда внешнее напряжение действует в запорном направлении, равновесие электронных потоков

нарушается из-за уменьшения потока электронов, идущих из полупроводника в металл. Поэтому возникновение тока связано с тем, что поток электронов, идущих из металла в полупроводник, не компенсируется встречным потоком. При этом чем больше запирающее напряжение, тем меньше компенсация. Когда же высота потенциального барьера на пути электронов, идущих из полупроводника в металл, станет настолько большой, что переход электронов в металл практически прекращается, сила тока через контакт достигнет максимального значения. Дальнейший рост запирающего напряжения уже не вызывает увеличения силы тока. Это объясняется тем, что высота потенциального барьера на пути электронов, идущих из металла в полупроводник, не зависит от приложенного запирающего напряжения.

Максимальный ток, проходящий через контакт в запирающем направлении, называется *током насыщения*. Плотность тока насыщения обозначается через I_s .

Наращению запирающего тока препятствует еще то, что внешняя разность потенциалов, распределяясь в основном в припограничной области полупроводника, складывается с контактной разностью потенциалов, благодаря чему усиливается электрическое поле, выталкивающее основные носители — электроны — в глубь полупроводника. Это приводит к обеднению электронами более широкой приконтактной области полупроводника и, следовательно, к увеличению толщины запирающего слоя и росту его сопротивления.

Вольт-амперная характеристика контакта «металл — полупроводник». Несимметричность поведения контакта «металл — полупроводник» в отношении зависимости силы тока от внешнего напряжения иллюстрирует вольт-амперная характеристика такого контакта, ход которой приведен на рисунке 60. Из рисунка видно, что рассматриваемый контакт обладает четко выраженной односторонней проводимостью: он неограниченно пропускает ток в прямом направлении и почти не пропускает в обратном.

Омический контакт. Односторонней проводимостью обладают далеко не всякие контакты металла с полупроводником, а только те, при создании которых возникает запирающий слой. Контакты, характеризующиеся образованием антизапирающего слоя, лишены этого свойства.

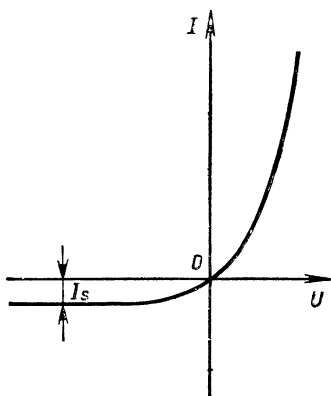


Рис. 60

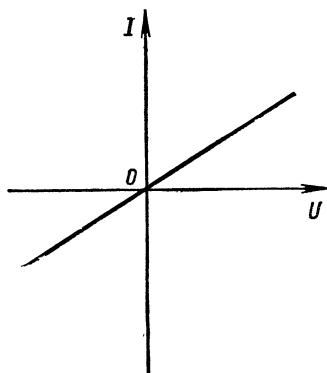


Рис. 61

Контакты, при образовании которых создаются антизапорные слои, нашли в радиотехнике и электронике не менее широкое использование для соединения различных устройств и приборов. Главное требование, предъявляемое к подсоединяющим контактам, заключается в том, чтобы они не искажали форму и характер сигнала. Для этого необходимо, чтобы их вольт-амперная характеристика была линейной. Как раз этому требованию и удовлетворяют контакты с антизапорным слоем. Для них выполняется закон Ома, и они поэтому получили название *омических*. Типичная вольт-амперная характеристика омического контакта приведена на рисунке 61.

§ 22. ЭЛЕКТРОННО-ДЫРОЧНЫЙ ПЕРЕХОД

В подавляющем большинстве полупроводниковых устройств, используемых в настоящее время в технике, главной частью является контакт двух примесных полупроводников, обладающих различным типом проводимости, так называемый *электронно-дырочный переход* или *p—n-переход*.

Способы получения p—n-переходов. Контакт между электронным и дырочным полупроводниками можно получить и путем прямого тесного соединения двух образцов, имеющих различные типы проводимости. Однако использовать такой контакт для создания прибора практически невозможно, так как наличие различных дефек-

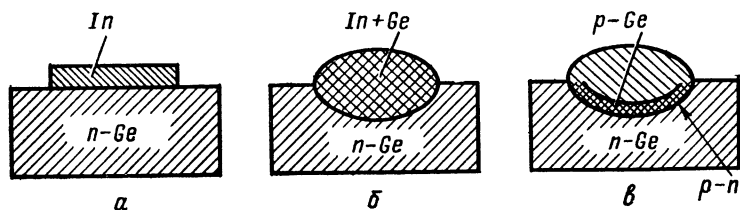


Рис. 62

тов и примесей, и в первую очередь окисной пленки, всегда присутствующей на поверхности полупроводника, вносит резкие изменения в свойства границы раздела полупроводников. Поэтому для получения $p-n$ -перехода с хорошо контролируемыми и постоянными свойствами необходимо создать его в виде внутренней границы раздела, на которой полупроводник одного типа переходил бы непрерывно в полупроводник другого типа. Методов изготовления $p-n$ -переходов в настоящее время довольно много, но мы ограничимся рассмотрением только двух из них: *метода сплавления* и *метода диффузии*.

При использовании метода сплавления на пластину из полупроводника n -типа кладут небольшой кусочек какого-либо трехвалентного металла, например индия (рис. 62, *а*), а затем помещают в печь, где в атмосфере инертного газа нагревают до температуры 550—600 °С. Индий при этой температуре плавится, и образующаяся капля растворяет в себе германий (рис. 62, *б*). Спустя некоторое время печь выключают, и при охлаждении из расплава начинает выпадать германий, содержащий примесные атомы индия. Если охлаждение происходит достаточно медленно, то выпадающий германий кристаллизуется в виде монокристалла, ориентация которого в точности совпадает с ориентацией монокристалла подложки. Однако в отличие от подложки, имеющей проводимость n -типа, вновь создаваемая область германия обладает проводимостью p -типа, и поэтому на границе раздела нерастворившейся части подложки и выкристаллизовавшегося из расплава участка образуется $p-n$ -переход (рис. 62, *в*). Вплавной $p-n$ -переход характеризуется почти скачкообразным изменением типа проводимости от электронной к дырочной и поэтому называется *резким $p-n$ -переходом*.

В противоположность к вплавному диффузионный переход отличается постепенным изменением типа проводимости от n -типа к p -типу, и поэтому его называют *плавным p — n -переходом*. Такой электронно-дырочный переход образуется в результате диффузии акцепторной примеси из газообраз-

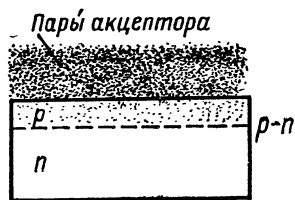


Рис. 63

ной или жидкой фазы в донорный полупроводник или донорной примеси в акцепторный полупроводник. Если в качестве подложки используется донорный полупроводник, то акцепторные атомы, проникая в процессе диффузии в глубь образца, сначала превращают его в компенсированный полупроводник (то есть нейтрализуют донорные примеси и придают образцу свойства собственного полупроводника), а затем по мере накопления примесных атомов — в дырочный. Степень легированности образующегося p -полупроводника, а также глубина проникновения акцепторных атомов в объем подложки определяются температурой и длительностью процесса диффузии. Граница, разделяющая преобразованную в p -область часть полупроводника от незатронутой диффузией подложки n -типа, как раз и представляет собой p — n -переход (рис. 63). Вполне естественно, что в данном случае граница раздела между акцепторным и донорным полупроводниками, представляющими фактически различные части одного и того же образца, не может быть резкой: изменение типа проводимости в диффузионных p — n -переходах происходит постепенно.

Различают симметричные и несимметричные p — n -переходы. Если концентрации легирующих примесей по разные стороны перехода одинаковы или достаточно близки друг к другу, то говорят о *симметричном p — n -переходе*. Если же концентрация одной из примесей, например донорной, много больше концентрации примеси по другую сторону p — n -перехода ($N_D \gg N_A$), то переход называется несимметричным. В технике в большинстве случаев используются *несимметричные p — n -переходы*. Но мы для простоты ограничимся рассмотрением симметричного электронно-дырочного перехода, считая $N_D = N_A$.

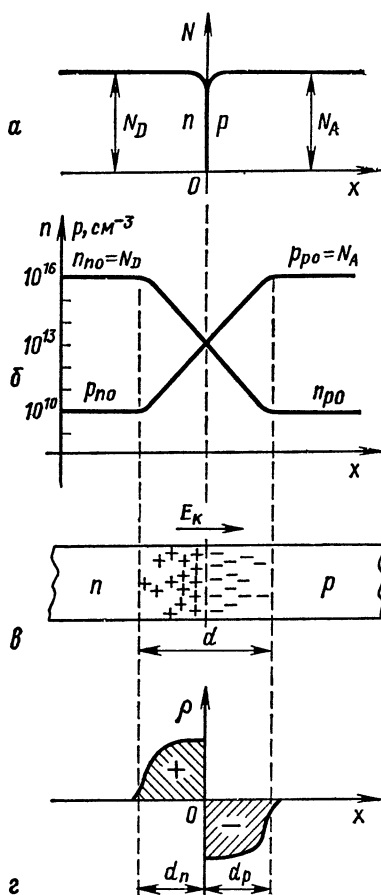


Рис. 64

Электронно-дырочный переход в состоянии равновесия. Рассмотрим резкий p — n -переход с симметричным распределением примесей по обе стороны от границы раздела. Пусть слева от границы (рис. 64, а) в электронном полупроводнике концентрация донорной примеси скачкообразно падает от значения N_D до 0, а справа в дырочном полупроводнике концентрация акцепторной примеси возрастает также скачкообразно от 0 до N_A . Основными носителями в полупроводнике n -типа являются электроны, а в полупроводнике p -типа — дырки. Появление большого количества основных носителей в соответствующих областях объясняется тепловым возбуждением и ионизацией донорных и акцепторных центров. При достаточно высокой (например, при комнатной) температуре практически все донорные и акцепторные примесные атомы оказываются ионизированными.

Поэтому вдали от границы раздела полупроводников концентрацию электронов в полупроводнике n -типа можно считать равной концентрации донорной примеси ($n_{n0} = N_D$), а концентрацию дырок в полупроводнике p -типа — равной концентрации акцепторных центров ($p_{p0} = N_A$). Индекс 0 указывает на то, что соответствующая концентрация относится к состоянию равновесия.

В каждой из контактирующих областей наряду с основными носителями имеется и определенное количество неосновных носителей: в n -области имеются дырки, а в p -области — электроны. Их концентрацию можно определить, используя закон действующих масс (см. § 10). Допустим, что концентрация примесей в каждой из областей равна 10^{16} см $^{-3}$. Такой же будет и концентрация основных носителей:

$$n_{no} = p_{po} = N_D = N_A = 10^{16} \text{ см}^{-3}.$$

Если считать, что при комнатной температуре концентрация носителей в собственном полупроводнике $n_i = 10^{13}$ см $^{-3}$, то для концентраций неосновных носителей, например электронов в дырочном полупроводнике, получим:

$$n_{po} = \frac{n_i^2}{p_{po}} = \frac{10^{26}}{10^{16}} = 10^{10} \text{ см}^{-3}.$$

Точно таким же будет и количество дырок в электронном полупроводнике. Таким образом, мы видим, что при переходе через границу раздела концентрации носителей одного типа (электронов или дырок) изменяются в 10^6 раз. Изменение концентрации носителей вблизи границы происходит однако не столь резко, как изменение концентрации атомов легирующих примесей, так как примесные атомы входят в состав кристаллической решетки и жестко связаны с ней, в то время как носители являются свободными и могут перемещаться по кристаллу. Изменение концентраций носителей в области контакта графически показано на рисунке 64, б.

Однако, каким бы плавным ни было изменение концентрации, полный перепад оказывается очень большим. Это приводит к довольно интенсивной диффузии основных носителей через p — n -переход: из n -области в p -область диффундируют электроны, а навстречу им — дырки. Результатом процесса диффузии является возникновение диффузионного тока через p — n -переход и контактной разности потенциалов.

Диффузионный ток. Электроны, приходящие в p -область, богатую дырками, довольно быстро рекомбинируют, поэтому концентрация их в этой области практически не повышается и остается на уровне равновесной. В то же время в n -области вместо ушедших электронов благодаря непрерывной тепловой генерации появляются

новые электроны. Таким образом, несмотря на диффузию электронов из n -области в p -область, различие их концентрации в этих областях остается неизменным. Поэтому неизменной остается и скорость диффузии электронов. То же относится и к процессу диффузии дырок из p -области в n -область. Такой установившийся направленный перенос носителей заряда, естественно, приводит к возникновению через p — n -переход тока, который называется *диффузионным током*. Подчеркнем, что диффузионный ток, создаваемый переходом дырок, складывается с диффузионным током, создаваемым встречным потоком электронов.

Контактная разность потенциалов. Ток проводимости. Приход в приконтактную область электронного полупроводника дырок из p -полупроводника и, что более существенно, уход из этой области большого числа электронов, оставляющих нескомпенсированными положительно заряженные неподвижные ионы донорной примеси, приводят к появлению у границы раздела в n -полупроводнике положительного заряда (рис. 64, *в*). Аналогично этому уход дырок из приконтактной области p -полупроводника, а также приход электронов из n -полупроводника создают у границы раздела в p -полупроводнике отрицательно заряженную область. Протяженность этих областей в n - и p -полупроводниках обозначим соответственно через d_n и d_p . Распределение плотности заряда в приграничных областях показано на рисунке 64, *г*. Между разноименно заряженными областями образуется контактная разность потенциалов U_k и электрическое поле, которое препятствует дальнейшему диффузионному перетеканию основных носителей. Вместе с тем это поле не только не препятствует, но и, наоборот, способствует переходу через границу раздела неосновных носителей. Действительно, электроны, подходящие к p — n -переходу из глубины p -полупроводника, и дырки, приходящие из глубины n -полупроводника, подхватываются контактным полем и выводятся соответственно в n -область и p -область. Таким образом, в противовес диффузионному току, возникающему за счет перехода через границу раздела основных носителей, возникает встречный ток, определяемый переходом через ту же границу неосновных носителей в противоположном направлении. Ток этот получил название *тока проводимости*. Сила тока проводи-

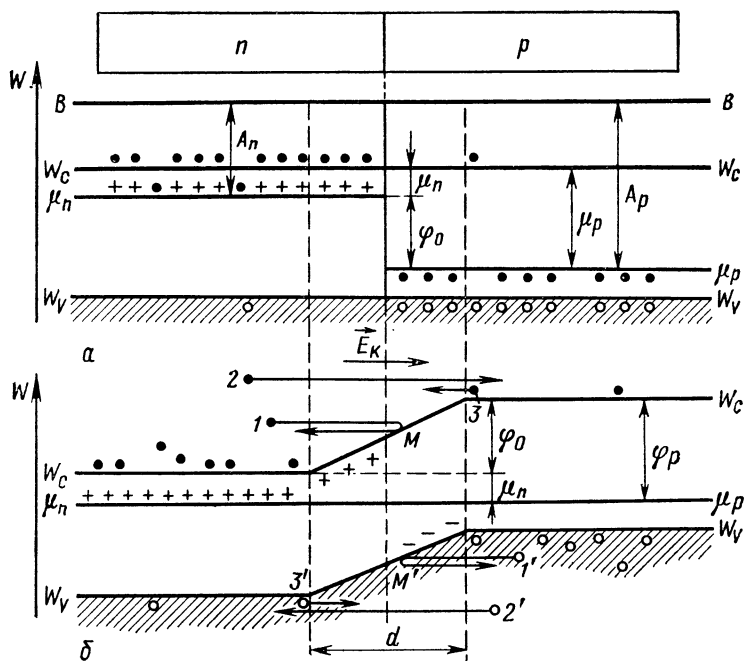


Рис. 65

мости практически не зависит от значения контактной разности потенциалов, а определяется лишь процессом тепловой генерации неосновных носителей и условиями перехода их из объема полупроводника к границе раздела.

Зонная структура p — n -перехода в состоянии равновесия. Зонные диаграммы n - и p -полупроводников до начала перехода носителей через границу приведены на рисунке 65, а. Уровни электронов в вакууме, оба дна зон проводимости и потолки валентных зон расположены на одной высоте, а уровень Ферми при комнатной температуре находится в n -области вблизи дна зоны проводимости, в p -области у вершины валентной зоны. Диффузионное перетекание основных носителей, приводящее к оголению ионизированных донорных и акцепторных атомов и к появлению нескомпенсированных зарядов в приконтактных областях, вызывает смещение энергетических

уровней. В n -области, заряжающейся положительно, все уровни смещаются вниз, а в p -области, заряжающейся отрицательно,—вверх. Перемещение уровней происходит до тех пор, пока, как уже было выяснено ранее, уровни Ферми не расположатся на одной высоте (рис. 65, б), что как раз и соответствует установлению состояния равновесия. Вследствие образования объемных разноименных электрических зарядов в приграничных областях по обе стороны от границы раздела изменятся соответственно потенциалы полупроводников и на участке p — n -перехода произойдет изгиб энергетических зон. Поскольку в глубинных областях полупроводников, достаточно удаленных от границы перехода, уровни Ферми сохраняют свое положение относительно остальных энергетических уровней, то изгиб энергетических зон приводит к возникновению потенциального барьера $\Phi_0 = eU_h$. Как видно из рисунка,

$$\Phi_0 = \mu_p - \mu_n,$$

или (что то же самое)

$$\Phi_0 = A_p - A_n.$$

Мы уже знаем, что положение уровня Ферми зависит от степени легированности полупроводника: чем сильнее легирован полупроводник, тем ближе к соответствующей разрешенной зоне располагается уровень Ферми. Например, чем больше донорной примеси введено в полупроводник, тем ближе к дну зоны проводимости разместится уровень Ферми. В предельном случае в невырожденных полупроводниках уровень Ферми в донорной области подходит вплотную к дну зоны проводимости, а в дырочной — к потолку валентной зоны. Таким образом, максимальное значение потенциального барьера $\Phi_{0, \max}$ на границе двух невырожденных разнотипных полупроводников равно ширине запрещенной зоны W_g . В реальных же случаях Φ_0 для p — n -перехода между невырожденными полупроводниками обычно меньше W_g . Например, для рассматриваемого нами германиевого p — n -перехода с $N_D = N_A = 10^{16} \text{ см}^{-3}$ потенциальный барьер Φ_0 имеет при комнатной температуре значение около 0,35 эВ, что приблизительно составляет половину ширины запрещенной зоны.

Контактное электрическое поле является своеобраз-

ным регулятором перехода носителей через p — n -переход. Далеко не все электроны, идущие из n -области в p -область, могут преодолеть потенциальный барьер ϕ_0 . Так, электрон, обозначенный на рисунке цифрой 1, может за счет своей избыточной кинетической энергии преодолеть только часть потенциального барьера. Дойдя до точки М, он растрчивает всю свою кинетическую энергию на совершение работы против сил контактного поля и останавливается, после чего он выводится этим полем обратно в объем n -полупроводника. Аналогичная ситуация имеет место в случае дырки, идущей из p -области и обозначенной цифрой 1'. Преодолеть потенциальный барьер могут лишь те основные носители, энергия которых больше ϕ_0 . Такая возможность представляется, например, электрону 2 и дырке 2'. Но высокоэнергетических электронов и дырок мало, так как основная часть электронов находится вблизи дна зоны проводимости, а основная часть дырок — вблизи потолка валентной зоны. Поэтому потоки основных носителей через область p — n -перехода невелики, хотя общее число этих носителей огромно.

По отношению к неосновным носителям действие контактного поля оказывается противоположным. Это видно и из рисунка 65, б. Электрон 3 и дырка 3', подходящие к p — n -переходу, подхватываются контактным полем и выводятся на противоположную сторону перехода. При этом их кинетическая энергия увеличивается за счет работы сил поля. Однако, несмотря на благоприятные условия перехода электронов из p -области в n -область и дырок из n -области в p -область, потоки этих неосновных носителей невелики, так как сравнительно мало общее число неосновных носителей в каждом из полупроводников.

Напряженность контактного электрического поля практически не сказывается на величине потоков неосновных носителей через границу. Независимо от значения напряженности контактного поля любой из неосновных носителей, попавших в область p — n -перехода, будет переправлен этим полем на противоположную сторону от перехода. Величина же каждого из потоков определяется числом неосновных носителей (электронов или дырок), подходящих к переходу, то есть тем, сколько их образуется в непосредственной близости от пере-

хода в процессе тепловой генерации. Именно поэтому сила тока проводимости не зависит от значения контактной разности потенциалов. Толщина слоя полупроводника, прилегающего непосредственно к p — n -переходу и участвующего в создании тока проводимости, определяется диффузионной длиной соответственно L_n для электронов в p -области и L_p для дырок в n -области, то есть средними расстояниями, которые проходят электроны и дырки за время их жизни. Электроны, возникающие благодаря тепловой генерации в объеме p -полупроводника на расстояниях от p — n -перехода, больших, чем L_n , и дырки, возникающие в n -полупроводнике на расстояниях от p — n -перехода, больших, чем L_p , участия в образовании тока проводимости не принимают, так как успевают прорекомбинировать, не дойдя до области p — n -перехода. Диффузионная длина обычно невелика и, например, для германия равна по порядку величины 10^{-2} см.

Так как сила тока, создаваемого диффузионными потоками основных носителей, зависит от значения потенциального барьера $\phi_0 = eU_k$, а сила тока проводимости, обусловленного потоками неосновных носителей, от значения потенциального барьера не зависит и оба эти тока имеют противоположные направления, то общая сила тока через p — n -переход равна нулю лишь при определенном значении потенциального барьера, при котором диффузионный ток равен току проводимости.

Электронно-дырочный переход — запорный слой. Контактное поле выталкивает подвижные носители из области p — n -перехода в глубинные области контактирующих полупроводников, где соответствующие носители являются основными: электроны — в n -полупроводник, а дырки — в p -полупроводник. Такое действие контактного поля не дает возможности подвижным носителям из глубинных областей полупроводников проникать в область p — n -перехода и восполнять дефицит основных носителей в этой области. Оставшиеся после ухода электронов нескомпенсированные и жестко связанные с кристаллической решеткой ионы доноров создают положительный объемный заряд в области n -полупроводника, прилегающей к границе контакта, а ушедшие через контакт дырки оставляют в p -области такой же по модулю, но отрицательный заряд нескомпенсированных ионизованных ак-

цепторов. Именно эти неподвижные заряды и создают контактное поле и контактную разность потенциалов.

Уход основной массы подвижных носителей из приконтактных областей и невозможность прихода в эти области новых подвижных носителей приводят к резкому возрастанию сопротивления p — n -перехода по отношению к остальному объему полупроводника. Поэтому область p — n -перехода всегда представляет собой запорный слой.

Толщины обедненных слоев (d_n — со стороны n -полупроводника и d_p — со стороны p -полупроводника) зависят от степени легированности каждой из контактирующих областей. В случае симметричного p — n -перехода (при $N_D = N_A$) толщины d_n и d_p равны между собой (см. рис. 64, $з$). Если же p — n -переход несимметричен, то $d_n \neq d_p$. Например, при $N_D > N_A$ толщина d_n обедненного слоя со стороны n -полупроводника меньше толщины d_p обедненного слоя со стороны дырочного полупроводника (рис. 66, $а$). Объясняется это тем, что в более легированном полупроводнике каждый монослой содержит большее число примесных центров и, следовательно, больше подвижных носителей, чем моноатомный слой менее легированного полупроводника. В этом случае говорят, что p — n -переход проникает больше в слаболегированную (более высокоомную) область полупроводника. Если же $N_D \gg N_A$, то p — n -переход практически полностью локализуется в высокоомной p -области, захватывая только очень тонкий, прилегающий непосредственно к границе контакта слой высоколегированного полупроводника n -типа (рис. 66, $б$).

Общая толщина p — n -перехода очень невелика. Например, при средней степени легированности p - и n -областей в кристалле германия толщина перехода составляет величину порядка 1 мкм.

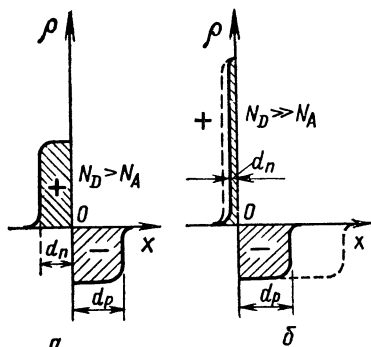


Рис. 66

§ 23. ЭФФЕКТ ВЫПРЯМЛЕНИЯ НА $p-n$ -ПЕРЕХОДЕ

Обратный ток. Выясним, как изменяются условия перехода носителей через $p-n$ -переход, когда к нему приложена некоторая внешняя разность потенциалов, или, как ее часто называют, напряжение смещения.

Пусть источник внешней разности потенциалов присоединен к $p-n$ -переходу таким образом, что плюс подан на n -область, а минус — на p -область. Такая полярность подключения внешнего источника обеспечивает создание в области $p-n$ -перехода дополнительного электрического поля, совпадающего по направлению с контактным электрическим полем. Как и в случае контакта металла с полупроводником, таким образом включенное напряжение смещения называется *обратным*. Так как область перехода обеднена подвижными носителями и обладает гораздо большим сопротивлением, чем остальная часть полупроводника, то приложенная внешняя разность потенциалов практически вся будет падать на запиорном слое и падением напряжения на остальных участках полупроводников можно пренебречь. Таким образом, напряжение U обратного смещения складывается с контактной разностью потенциалов U_k , благодаря чему потенциальный барьер в области $p-n$ -перехода увеличивается по сравнению с его равновесным значением на eU (рис. 67, б):

$$\phi' = eU_k + eU = \phi_0 + eU.$$

Сравнение рисунков 67, а и 67, б показывает, что внешняя разность потенциалов U смещает уровни в контактирующих областях на eU . Повышение потенциала n -области вызывает понижение относящихся к ней энергетических уровней, а понижение потенциала p -области обуславливает повышение соответствующих уровней. Сдвигаются, конечно, и уровни Ферми. Различие в положениях уровней Ферми в контактирующих областях после подачи напряжения смещения (рис. 67, б) указывает на нарушение состояния равновесия, имевшегося до приложенного внешнего смещения (рис. 67, а).

Возрастание потенциального барьера на eU приводит к уменьшению электронной и дырочной составляющих диффузионного тока, текущего через переход: чем больше приложенное обратное смещение, тем выше потен-

циальный барьер и тем меньше число основных носителей, способных преодолеть его.

При некотором достаточно большом значении напряжения смещения диффузионный ток через переход полностью исчезает.

Наряду с уменьшением диффузионного тока происходит увеличение толщины обедненного слоя на границе раздела и возрастание его сопротивления, так как увеличение напряженности результирующего электрического поля (по сравнению с напряженностью контактного электрического поля) приводит к усилению эффекта выталкивания основных носителей из области p — n -перехода.

Ток проводимости, создаваемый потоками неосновных носителей, на напряжение смещения практически не реагирует: возросшее электрическое поле в области p — n -перехода лишь увеличивает скорость переноса неосновных носителей через переход, не изменяя числа этих носителей. Не влияет на силу тока проводимости и возрастание толщины обедненного слоя, так как оба поля (поле смещения и контактное поле) способствуют переходу неосновных носителей через этот слой.

Таким образом, обратное напряжение смещения ограничивает поток основных носителей настолько, что диффузионный ток не может компенсировать ток проводимости. При больших обратных смещениях диффузионный ток стремится к нулю и весь ток, текущий через переход, определяется практически только током проводимости, создаваемым неосновными носителями. Так как предель-

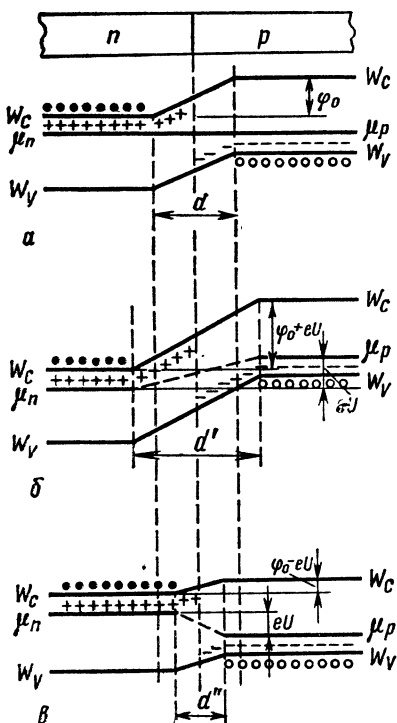


Рис. 67

ное значение силы тока проводимости не зависит от приложенного напряжения смещения, его называют *током насыщения* и обозначают через I_s (индекс s происходит от английского saturation — на с ы щ е н и е). Иногда этот ток называют *тепловым неуправляемым током*, что довольно точно отражает физический смысл.

Отметим, что симметричное утолщение области p — n -перехода по обе стороны от границы раздела наблюдается только в симметричном переходе. Когда же концентрация примесей в одной области больше концентрации примесей в другой (например, $N_D > N_A$), то расширение обедненной области произойдет в основном в сторону полупроводника, имеющего меньшую концентрацию легирующей примеси. Если степень легирования контактирующих областей отличается сильно, например, если $N_D \gg N_A$, то практически все возрастание толщины обедненного слоя приходится на слаболегированную область (на рис. 66, б эти изменения толщин слоев показаны пунктиром).

Прямой ток. Много большее нарушение равновесия наблюдается при подаче на p — n -переход прямого смещения (прямой разности потенциалов), когда положительный полюс источника внешнего напряжения присоединяется к p -области, а отрицательный полюс — к n -области. Смещение энергетических уровней контактирующих полупроводников приводит в этом случае к уменьшению потенциального барьера (рис. 67, в). Внешняя разность потенциалов U , вычитаясь из значения контактной разности потенциалов, приводит к снижению потенциального барьера до значения

$$\varphi'' = \varphi_0 - eU.$$

Это сопровождается увеличением числа электронов, способных преодолеть барьер (чем ближе к дну зоны проводимости, тем выше плотность заселенности энергетических уровней электронами). Увеличение числа основных носителей, преодолевающих барьер, обуславливает быстрый рост диффузионного тока. Кроме того, в этом случае поле смещения и контактное поле в области перехода имеют противоположные направления. Поэтому результирующее поле оказывается ослабленным, из-за чего сужается обедненная носителями область ($d'' < d$)

и уменьшается сопротивление p — n -перехода, что также способствует росту диффузионного тока.

Поскольку сила тока проводимости, сохраняющая неизменное значение и в случае прямого напряжения смещения, мала, то по мере возрастания этого напряжения и соответствующего увеличения диффузионного тока можно током проводимости пренебречь. Таким образом, ток, текущий через p — n -переход, в случае прямого напряжения смещения оказывается практически полностью обусловленным потоками основных носителей, то есть представляет собой диффузионный ток. Основные носители, преодолевшие p — n -переход и попавшие в соседнюю область, становятся в ней неосновными носителями, повышая тем самым концентрацию неосновных носителей в приконтактной области. Создающийся перепад концентрации вызывает диффузию избыточных неосновных носителей от p — n -перехода в глубь полупроводника, где они довольно быстро рекомбинируют. Чем больше прямое напряжение смещения, тем ниже потенциальный барьер и тем больше избыточная концентрация неосновных носителей в областях p — n -перехода. Поэтому с ростом прямого напряжения смещения увеличивается скорость процессов диффузии и рекомбинации и растет сила тока, текущего через переход. Когда напряжение смещения превысит контактную разность потенциалов, потенциальный барьер совсем исчезнет. Вместе с ним исчезнет и обедненная носителями область, после чего напряжение $U - U_k$ окажется распределенным по всему образцу. Дальнейшее увеличение прямой разности потенциалов обусловит рост тока в соответствии с законом Ома:

$$I = \frac{U - U_k}{R},$$

где R — сопротивление всего образца.

Инжекция носителей. Вызванное прямым напряжением смещения проникновение основных носителей через p — n -переход, сопровождающееся повышением концентрации неосновных носителей в прилегающих к переходу областях, называется *инжекцией неосновных носителей*. Отношение избыточной концентрации неосновных носителей в непосредственной близости от p — n -перехода к равновесной концентрации основных носителей называется *уровнем инжекции*.

Если p — n -переход симметричен, то число электронов Δn_p , инжектированных из n -области в p -область, равно числу дырок Δp_n , инжектированных из p -области в n -область. При несимметричном переходе число носителей, инжектированных из более высоколегированной области, будет превышать встречный поток, а отношение избыточных концентраций носителей, инжектированных в соответствующие области, определится отношением концентраций основных носителей:

$$\frac{\Delta n_p}{\Delta p_n} = \frac{n_{po}}{p_{po}}.$$

Если, например, в n -области концентрация атомов легирующей примеси в 1000 раз больше, чем в p -области, то и поток электронов из n -области в p -область будет в 1000 раз больше, чем встречный поток дырок. Так как снижение потенциального барьера, вызванное прямым напряжением смещения, в одинаковой степени облегчает как переход для дырок из p -области в n -область, так и для электронов из n -области в p -область, то дополнительный поток носителей будет больше из той области, в которой концентрация их больше.

При $n_{po} \gg p_{po}$ поток дырок, которые инжектируются в n -область из p -области, оказывается пренебрежимо малым по сравнению с встречным потоком электронов и можно считать, что весь диффузионный ток, текущий через p — n -переход, определяется только его электронной составляющей.

Область полупроводника, из которой происходит преимущественно инжекция, называется *эмиттерной областью* или просто *эмиттером*, а область, в которую происходит преимущественная инжекция, называется *базовой областью* или просто *базой*. В последнем примере эмиттером является n -область, а базой — p -область.

При несимметричном p — n -переходе как прямой, так и обратный ток создается практически носителями одного типа. Так, при $n_{po} \gg p_{po}$ как прямой ток, так и обратный ток создаются главным образом электронами, поскольку в высоколегированной области n -типа дырок (неосновных носителей) мало, а в слаболегированной p -области электронов гораздо больше.

Вольт-амперная характеристика p — n -перехода (рис. 68) имеет такой же вид, что и в случае контак-

та металл — полупроводник. Полный ток $I_{\text{пр}}$ через переход при прямом смещении, равный разности между диффузионным током $I_{\text{диф}}$ и током проводимости $I_{\text{пров}}$:

$$I_{\text{пр}} = I_{\text{диф}} - I_{\text{пров}},$$

быстро растет по мере увеличения напряжения смещения благодаря увеличению диффузионного тока при неизменном токе проводимости. При положительном напряжении смещения, превышающем 0,1 В, когда током проводимости можно пренебречь, нарастание силы

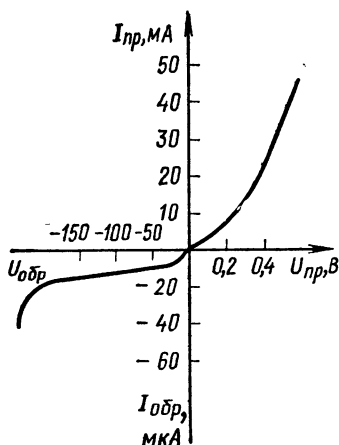


Рис. 68

тока через переход практически происходит по экспоненциальному закону. Существенная нелинейность характеристики на начальном участке прямого тока обусловлена уменьшением толщины p — n -перехода и снижением его сопротивления с ростом прямого напряжения смещения в области малых его значений. При значениях прямого напряжения смещения порядка нескольких десятых долей вольта и выше характеристика становится практически линейной, так как в этой области потенциальный барьер ϕ , а вместе с ним и обедненная носителями область исчезают и приложенная разность потенциалов оказывается распределенной по всей длине образца (выполняется закон Ома).

Обратный ток через переход всегда много меньше прямого. Поэтому при построении обратной ветви вольт-амперной характеристики перехода используют обычно другие масштабы: для силы тока — на 2—3 порядка больший, а для напряжения — на 2—3 порядка меньший (из-за этого в начале координат на вольт-амперной характеристике наблюдается излом, которого при построении в одинаковом масштабе нет). Обратный ток

$$I_{\text{обр}} = I_{\text{пров}} - I_{\text{диф}}$$

на начальном участке довольно быстро растет за счет резкого уменьшения диффузионного тока с увеличением

потенциального барьера. Однако уже при обратном напряжении смещения порядка 0,1—0,2 В рост этот прекращается. При больших напряжениях диффузионный ток падает практически до нуля и обратный ток становится равным току проводимости, который, как мы видели, почти не зависит от приложенного напряжения. Некоторый рост обратного тока на участке «насыщения» объясняется нагревом p — n -перехода самим током и другими побочными эффектами.

Выпрямительные свойства p — n -перехода характеризуются коэффициентом выпрямления, который определяется как отношение $\frac{I_{\text{пр}}}{I_{\text{обр}}}$ прямого тока к обратному току (при одинаковых по модулю напряжениях смещения). Обычно коэффициент выпрямления по порядку величины равен 10^5 — 10^7 . Это значит, что p — n -переход обладает практически односторонней проводимостью.

Влияние температуры на выпрямляющие свойства p — n -перехода. При комнатной температуре большинство полупроводников, в том числе германий и кремний, находятся в состоянии «истощения» примесей: все примесные центры ионизированы. В таком состоянии донорные полупроводники имеют электронную проводимость, а акцепторные — дырочную. При нагревании полупроводника по мере приближения его температуры к температуре собственной проводимости T_i все более интенсивным становится процесс возбуждения собственных атомов полупроводника, сопровождающийся одновременным рождением свободных электронов и дырок. При $T > T_i$ концентрация n_i собственных носителей становится больше концентрации «примесных» носителей, которые имеются в полупроводнике благодаря ионизации примесных центров. Полупроводник в этой области температур теряет свойства примесного полупроводника и превращается в собственный полупроводник. Его уровень Ферми при этом смещается в середину запрещенной зоны; исчезают потенциальный барьер и запирающий слой, из-за чего p — n -переход теряет свои выпрямляющие свойства.

Температура собственной проводимости T_i зависит от ширины запрещенной зоны W_g полупроводника: чем больше W_g , тем выше T_i . Поэтому и температурный предел работы p — n -переходов определяется характеристиками материалов, на базе которых они выполнены. В гер-

мании, ширина запрещенной зоны которого $W_g = 0,72$ эВ, p — n -переход работает до температуры около 75°C , а в кремнии, у которого $W_g = 1,12$ эВ, рабочая температура может доходить до 150°C .

§ 24. ПРОБОЙ p — n -ПЕРЕХОДА

Если напряжение обратного смещения непрерывно повышать, то в отсутствие ограничительного сопротивления в цепи перехода после достижения некоторого значения $U_{\text{проб}}$ (рис. 69) наблюдается резкое увеличение силы тока, происходит так называемый *пробой p — n -перехода*. Различают *электрический* и *тепловой* пробой. Электрический пробой не приводит к разрушению перехода, и если за ним не последует теплового пробоя, свойства p — n -перехода после снятия обратного напряжения восстанавливаются. Существуют два вида электрического пробоя: *лавинный* и *туннельный*.

Лавинный пробой. В основе этого вида пробоя лежит эффект лавинного размножения носителей в толще p — n -перехода. При некотором значении обратного напряжения смещения напряженность поля в p — n -переходе становится настолько большой, что неосновные носители, ускоряясь в нем, приобретают энергию, достаточную для ионизации нейтральных атомов полупроводника в переходе. В результате ионизации происходит лавинное нарастание числа носителей, создающих обратный ток. Лавинный пробой обычно наблюдается при обратных напряжениях смещения порядка десятков или сотен вольт. Он характерен для достаточно толстых p — n -переходов, в которых каждый из неосновных носителей, разгоняемый электрическим полем, вызывает акты ионизации атомов в обедненном слое много раз. В узких переходах обычно наблюдается туннельный пробой.

Туннельный пробой своим происхождением обязан так называемому *туннельному эффекту*. Возникает этот эффект благодаря непосредственному воздействию сильного электрического поля на атомы кристаллической решетки полупроводника в p — n -переходе. Под действием этого поля происходит разрыв валентной связи и электрон становится свободным носителем, переходя в межузельное пространство и оставляя на своем месте дырку. Зонная схема туннельного пробоя приведена на рисун-

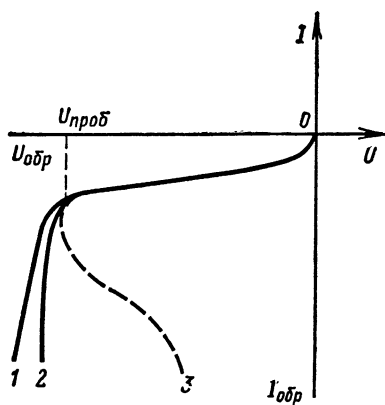


Рис. 69

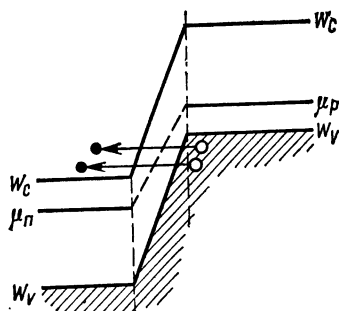


Рис. 70

ке 70. Электроны из валентной зоны p -полупроводника переходят, не изменяя свою энергию, в зону проводимости полупроводника n -типа, пересекая запрещенную зону p — n -перехода. Необходимым условием реализации туннельного перехода является незанятость в зоне проводимости n -полупроводника энергетического уровня, соответствующего энергии переходящего из p -области электрона.

Туннельный пробой наблюдается в тонких p — n -переходах, которые могут быть созданы только на границе раздела высоколегированных областей. Для того чтобы вызвать туннельный пробой, необходимо создать поле с напряженностью порядка 10^5 — 10^6 В/см. Поскольку туннельный пробой реализуется только в тонких переходах порядка 10^{-5} — 10^{-6} см, то для получения пробивных значений напряженности поля оказывается достаточной обратная разность потенциалов всего в несколько вольт.

Нарастание тока при туннельном пробое (кривая 2 на рисунке 69) происходит даже более круто, чем при лавинном пробое (кривая 1). Так как свойства p — n -перехода после электрического пробоя (лавинного или туннельного) восстанавливаются при выключении обратного напряжения смещения, то в технике во многих случаях p — n -переход используется именно в режиме пробоя (полупроводниковые стабилитроны, туннельные обращенные диоды и пр.).

Тепловой пробой. Если количество тепла, выделяющегося в $p-n$ -переходе, превышает количество тепла, отводимого от него, то разогрев перехода приводит к росту процесса генерации носителей и, следовательно, к увеличению силы тока, текущего через переход, что в свою очередь ведет к дальнейшему повышению температуры и т. д. В итоге такого лавинообразно развивающегося перегрева сила тока продолжает возрастать и при уменьшении напряжения (кривая 3 на рисунке 69); наступает разрушение материала полупроводника. Тепловой пробой может возникнуть самостоятельно, но может оказаться и следствием развивающегося электрического пробоя. Поэтому обычно в цепь $p-n$ -перехода последовательно включают ограничительный резистор, сопротивление которого подбирается так, чтобы сила тока не превосходила допустимого значения.

Поверхностный пробой. Лавинный или туннельный электрический пробой $p-n$ -перехода может происходить не только в объеме полупроводника, но и по его поверхности. На поверхностный пробой значительное влияние может оказать искажение электрического поля в $p-n$ -переходе поверхностными зарядами. Наличие поверхностного заряда связано с обрывом кристаллической решетки и с наличием в ней дефектов и примесей (особенно адсорбированных молекул воды). В определенных случаях поверхностный заряд приводит к сужению запорного слоя у поверхности и увеличению в приповерхностной области напряженности поля, благодаря этому пробой у поверхности начинается при меньших значениях обратного напряжения смещения, чем в объеме. Для уменьшения вероятности поверхностного пробоя применяют различные защитные покрытия, предотвращающие проникновение на поверхность $p-n$ -перехода влаги и различных активных примесей.

§ 25. ЭЛЕКТРИЧЕСКАЯ ЕМКОСТЬ $p-n$ -ПЕРЕХОДА

При подаче на $p-n$ -переход напряжения смещения он проявляет емкостные свойства.

Электронно-дырочный переход представляет собой обедненную подвижными носителями область полупроводника, близкую по своим свойствам к слою диэлектрика, в котором по разные стороны от некоторой плоскости

находятся объемно распределенные неподвижные электрические заряды противоположных знаков. В полупроводнике это заряды ионизированных примесных атомов: электронная область перехода заряжена положительно, а дырочная — отрицательно. Такая структура $p-n$ -перехода позволяет его в известной мере уподобить плоскому конденсатору. Включение обратного напряжения смещения приводит к дальнейшему вытеснению подвижных носителей из приграничных областей, увеличению толщины $p-n$ -перехода и числа нескомпенсированных неподвижных ионов примесей по обе стороны от границы раздела. Таким образом на изменение ΔU обратного напряжения смещения $p-n$ -переход откликается изменением заряда ΔQ . А это и есть проявление емкостных свойств. Величина

$$C_{\text{бар}} = \frac{\Delta Q}{\Delta U}$$

называется *барьерной емкостью*. Иногда ее называют *зарядовой емкостью* или просто *емкостью перехода*.

Как и в конденсаторах, барьерная емкость зависит от площади $p-n$ -перехода, диэлектрической проницаемости обедненного слоя и его толщины. В большинстве случаев площадь $p-n$ -перехода мала, и тем не менее емкость благодаря малой толщине обедненной области может быть очень большой. Варьируя толщину $p-n$ -перехода, можно изменять барьерную емкость от единиц до десятков тысяч пикофард на квадратный сантиметр. Важнейшей особенностью барьерной емкости является ее зависимость от напряжения смещения, поданного на пе-

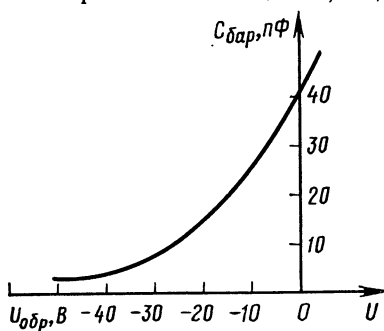


Рис. 71

реход. При увеличении обратного напряжения толщина запирающего слоя растет, и это приводит к уменьшению емкости. Ход изменения барьерной емкости от приложенного обратного напряжения можно проследить по графику, приведенному на рисунке 71. Зависимость барьерной емкости от напряжения используется в специальных полупроводниковых диодах, называемых *варикапами*. Варикапы широко используются в качестве конденсаторов переменной емкости для перестройки частот колебательных контуров. В выпрямительных диодах барьерная емкость играет вредную роль, потому что она шунтирует $p-n$ -переход, проводя как бы «в обход» часть переменного тока (особенно в области высоких частот).

Барьерная емкость проявляется и при прямом напряжении смещения, если $U < U_k$ (при таком напряжении запирающий слой еще существует).

Глава 5.

ПОЛУПРОВОДНИКОВЫЕ ПРИБОРЫ

§ 26. ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ

В настоящее время полупроводниковые приборы используются практически во всех областях электротехники и радиотехники. Однако, несмотря на чрезвычайное разнообразие этих приборов, в основе их, как правило, лежит работа обычного $p-n$ -перехода или системы из нескольких $p-n$ -переходов.

Полупроводниковый диод содержит лишь один $p-n$ -переход, к каждой из областей которого подведены с помощью омических контактов металлические вводы.

Выпрямительные диоды. Полупроводниковые диоды применяются в основном для выпрямления переменного тока. Простейшая схема использования полупроводникового диода в качестве выпрямляющего элемента показана на рисунке 72. Источник переменного напряжения $u \sim$, диод D и нагрузочный резистор R_n соединяются последовательно. Пропускное направление диода обозначено стрелкой (от анода к катоду).

Пусть напряжение на зажимах источника изменяется по синусоидальному закону (рис. 73, а). Во время положительного полупериода, когда на анод диода подан «+», а на катод «-», диод оказывается включенным в прямом направлении и через него проходит ток. При этом мгновенное значение силы тока i определяется мгновенным значением напряжения u на зажимах источника и сопротивлением нагрузки (сопротивление диода в пропускном направлении мало, и им можно пренебречь). Во время отрицательного полупериода ток через диод практически не течет. Таким образом, в цепи протекает пульсирующий ток, график которого приведен на рисунке 73, б. Таким же пульсирующим будет и напряжение

u_R на нагрузочном резисторе. Так как $u_R = iR_H$, то изменение напряжения u_R повторяет ход изменения тока i . Полярность напряжения, создаваемого на сопротивлении нагрузки, всегда одна и та же, и определяется она в соответствии с направлением пропускаемого тока: на конце сопротивления, обращенного к катоду, будет «+», а на противоположном конце «-».

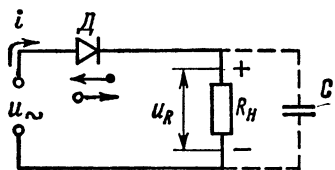


Рис. 72

Рассмотренная схема выпрямления является однополупериодной. Для уменьшения пульсаций выпрямленного напряжения используют сглаживающие фильтры. Наиболее простой метод сглаживания состоит в подключении параллельно нагрузочному резистору конденсатора C (на рисунке 72 он показан пунктиром).

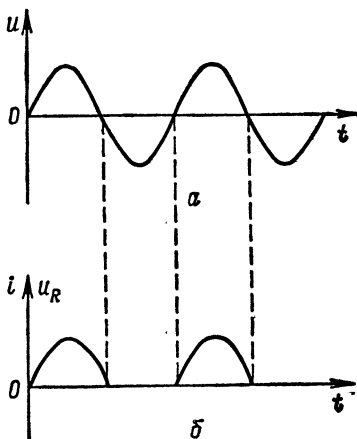


Рис. 73

Во время положительного полупериода часть тока, пропускаемого диодом, идет на зарядку конденсатора. Во время же отрицательного полупериода, когда диод заперт, конденсатор разряжается через R_H , создавая в нем ток в прежнем направлении. Благодаря этому пульсации напряжения на нагрузочном резисторе оказываются в значительной мере сглаженными.

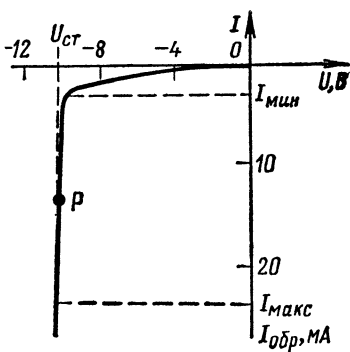


Рис. 74

Стабилитроны. Мы уже знаем, что при электрическом пробое $p-n$ -перехода сила тока через переход нарастает очень круто. Этот участок вольт-амперной характери-

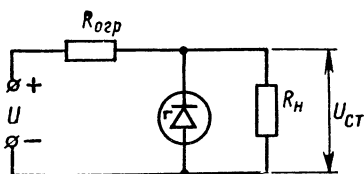


Рис. 75

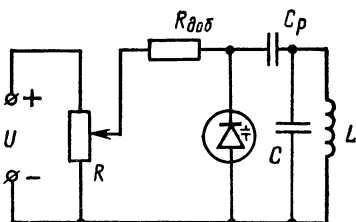


Рис. 76

стики может быть использован для стабилизации напряжения. Наилучшими для этой цели являются кремниевые диоды. На рисунке 74 приведена обратная ветвь характеристики такого диода, из которой видно, что при достижении напряжения пробоя очень незначительные изменения напряжения вызывают изменения силы тока через диод в очень широких пределах (в десятки раз).

На рисунке 75 приведена схема стабилизации напряжения. Диод-стабилитрон включается параллельно нагрузке R_H , на которой необходимо создать стабилизированное напряжение при наличии колебаний напряжения на зажимах источника. Как видно из схемы, в цепь последовательно включается ограничительный резистор. Его сопротивление $R_{огр}$ подбирается так, чтобы через стабилитрон шел ток, соответствующий средней части рабочего диапазона (точка P на рисунке 74). Если напряжение на зажимах источника, например, увеличивается, то будет увеличиваться и сила тока в цепи. Однако увеличение силы тока через стабилитрон (даже в значительных пределах) происходит при практически неизменном напряжении на нем, благодаря чему остается неизменным и напряжение на нагрузке R_H . Вместе с тем увеличение тока через ограничительный резистор приводит к возрастанию падения напряжения на нем. Таким образом, все изменение напряжения питания гасится ограничительным резистором, в то время как на сопротивлении нагрузки поддерживается постоянное напряжение $U_{ст}$.

На основе кремниевых $p-n$ -переходов в настоящее время изготавливаются стабилитроны с напряжением стабилизации от 1 до 300 В.

Варикапы. В режиме обратного смещения работают и диоды, используемые в качестве переменной емкости,—

варикапы. В отличие от обычных конденсаторов переменной емкости, управляемых механически, изменение емкости варикапов осуществляется путем изменения значения обратного напряжения смещения. Схема включения варикапа для перестройки частоты колебательного контура приведена на рисунке 76.

При изменении с помощью потенциометра R обратного напряжения, подаваемого на варикап, изменяется его барьерная емкость, благодаря чему изменяется и резонансная частота колебательного контура. Резистор $R_{доб}$, имея достаточно большое сопротивление, предотвращает шунтирование контура потенциометром R , а конденсатор C_p является разделительным. Если бы его не было, то варикап был бы накоротко замкнут катушкой L (по постоянной составляющей напряжения).

§ 27. ТУННЕЛЬНЫЕ ДИОДЫ

В большой семье полупроводниковых приборов имеет-ся группа приборов, у которых на определенном участке вольт-амперной характеристики увеличение напряжения ($\Delta U > 0$) сопровождается не ростом, а падением силы тока ($\Delta I < 0$). Такие участки вольт-амперной характеристики соответствуют *отрицательному сопротивлению*:

$$R = \frac{\Delta U}{\Delta I} < 0.$$

Наиболее распространенными и, пожалуй, наиболее интересными из всех приборов с отрицательным сопротивлением являются туннельные диоды. Идея использования туннельного эффекта для создания полупроводникового диода была высказана советскими учеными Я. И. Френкелем и А. Ф. Иоффе еще в 1932 году, но только в 1958 году японским инженером Л. Есаки был создан туннельный диод.

Изготовление туннельных диодов. Как и обычные выпрямительные диоды, туннельный диод может быть получен путем сплавления кусочка металла в пластинку полупроводника, например индия в германий n -типа. Иначе говоря, и для создания туннельного диода необходимо получить p — n -переход. Однако в отличие от изготовления обычных диодов для получения туннельного диода нужно использовать в качестве подложки полупроводник

с очень высокой степенью легированности, то есть с очень высокой концентрацией примеси. Если в обычных диодах концентрация примесей в полупроводнике, как правило, не превышает 10^{17} см^{-3} , то концентрация легирующей примеси в полупроводниках, используемых для создания туннельных диодов по порядку величины, равна $10^{19} \text{--} 10^{20} \text{ см}^{-3}$.

Особенности p — n -перехода между вырожденными полупроводниками. Полупроводники с такой концентрацией примеси, как мы уже видели ранее, являются вырожденными: у них уровни Ферми располагаются в области разрешенных зон (в вырожденном полупроводнике n -типа уровень Ферми лежит в области зоны проводимости, а в вырожденном полупроводнике p -типа — в области валентной зоны). Такое расположение уровней Ферми приводит к возникновению в контакте между вырожденными полупроводниками большой контактной разности потенциалов, превышающей почти вдвое значение контактной разности потенциалов в обычных диодах. Так как в туннельных диодах уровни Ферми лежат вне пределов запрещенной зоны, то у них потенциальный барьер на границе перехода всегда больше ширины запрещенной зоны. На рисунке 77, а приведена зонная схема двух высоколегированных вырожденных полупроводников (n -типа и p -типа) до контакта, а на рисунке 77, б — зонная схема p — n -перехода, образовавшегося после приведения полупроводников в контакт. Из рисунка 77, б видно, что при установлении равновесия между вырожден-

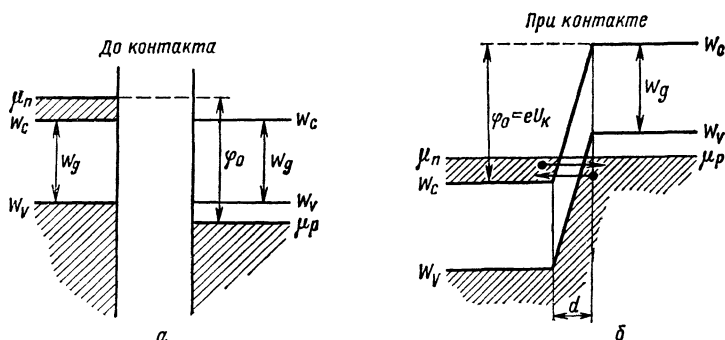


Рис. 77

ными n - и p -областями происходит перекрытие зон по внешней шкале энергии: дно зоны проводимости n -полупроводника располагается ниже потолка валентной зоны полупроводника p -типа. Таким образом, электроны, находящиеся, например, вблизи уровня Ферми в n - и p -областях, имеют одну и ту же энергию и переходу их из одной области в другую препятствует лишь зона запрещенных энергий, являющаяся для них некоторым потенциальным барьером.

Отличительной особенностью p — n -перехода между вырожденными полупроводниками является также его крайне малая толщина d — порядка 10^{-6} см. Дело в том, что благодаря высокой плотности свободных носителей их уход даже из небольшого пограничного слоя связан с образованием большого числа некомпенсированных заряженных донорных и акцепторных примесных центров, достаточного для возникновения равновесного потенциального барьера.

Туннельные переходы электронов в состоянии равновесия. Чрезвычайно малая толщина p — n -перехода в совокупности с перекрытием зон, благодаря которому по обе стороны перехода имеются области с одинаковыми разрешенными энергиями, создают благоприятные условия для туннельных переходов: электроны из зоны проводимости n -области переходят в валентную зону p -области, а электроны из валентной зоны p -области переходят в зону проводимости n -области (см. рис. 77, б). Конечно, для туннельного перехода электрона через барьер из одной области полупроводника в другую необходимо, чтобы по ту сторону барьера, куда переходит электрон, имелись свободные состояния. Но ведь уровень Ферми как раз тем и характеризуется, что вероятность его заполнения равна всего $1/2$. Поэтому для электронов, имеющих энергию, не слишком отличающуюся от энергии Ферми, всегда найдется место за потенциальным барьером p — n -перехода.

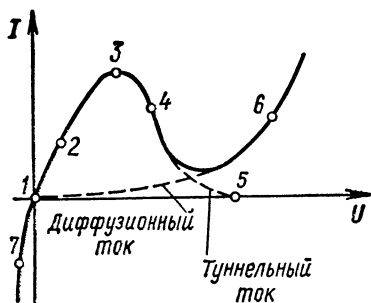


Рис. 78

При равновесии в отсутствие напряжения смещения число туннельных переходов электронов слева направо равно числу встречных переходов справа налево и суммарный туннельный ток равен нулю. Помимо туннельных переходов, в рассматриваемом диоде, конечно, существуют и надбарьерные переходы основных и неосновных носителей, создающих диффузионный ток и ток проводимости. Но, во-первых, в условиях равновесия и эти токи оказываются одинаковыми и направленными навстречу друг другу, так что в сумме они не дают тока. А во-вторых, в сравнении с числом туннельных переходов число надбарьерных переходов оказывается пренебрежимо малым. Итак, в отсутствие внешнего смещения ток через диод равен нулю, что соответствует началу координат на кривой вольт-амперной характеристики прибора (точка 1 на рисунке 78).

Поведение туннельного диода при подаче прямого напряжения смещения. Если на диод подать небольшое положительное смещение, то произойдет некоторое смещение энергетических зон, в результате чего потенциальный барьер на границе перехода немного понизится и незаполненная часть валентной зоны полупроводника *p*-типа разместится напротив заполненной области зоны проводимости *n*-полупроводника (рис. 79, а). При этом равновесие между туннельными переходами электронов слева направо и справа налево нарушится. Действительно, в области перекрытия заполненных частей зон указанные переходы компенсируют друг друга (пунктирные стрелки на рисунке), но переходы из верхней области заполненной части зоны проводимости *n*-полупроводника (жирная стрелка) уже не встречают встречного компенсирующего потока, так как расположенная напротив область валентной зоны *p*-полупроводника практически пуста. Возникающий нескомпенсированный поток электронов из полупроводника *n*-типа в полупроводник *p*-типа приводит к появлению прямого тока через диод (точка 2 на кривой рисунка 78).

Повышение положительного смещения приводит к все большему перекрытию заполненной области зоны проводимости *n*-полупроводника с пустой областью валентной зоны полупроводника *p*-типа, благодаря чему растет и туннельный ток через диод. Максимального значения (точка 3 на рисунке 78) он достигает тогда, когда уро-

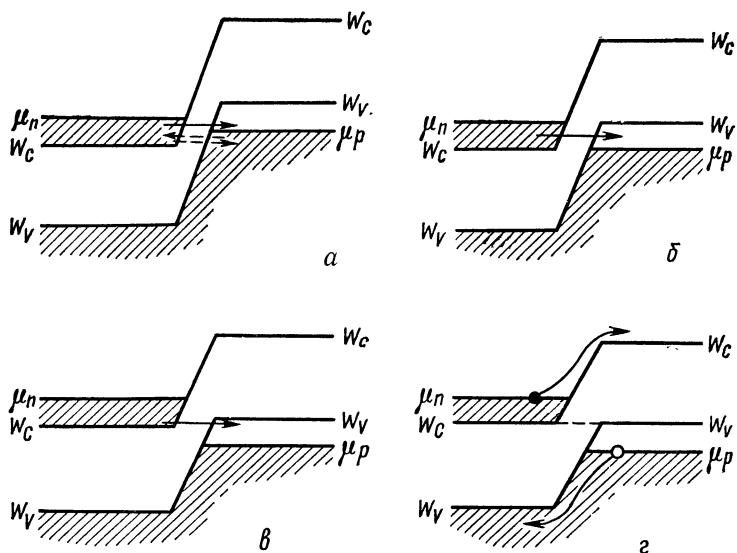


Рис. 79

вень Ферми n -полупроводника располагается напротив потолка валентной зоны p -области (рис. 79, б).

Дальнейшее повышение прямого напряжения сопровождается уже уменьшением перекрытия заполненной части зоны проводимости полупроводника n -типа и пустой части валентной зоны p -полупроводника, а значит, и ухудшением условий для переходов электронов из n -области в p -область (рис. 79, в). Электронам, расположенным в верхней части заполненной области зоны проводимости n -полупроводника, теперь противостоит зона запрещенных энергий p -полупроводника, из-за чего их переход в p -область становится невозможным. Таким образом, мы приходим к парадоксальному, на первый взгляд, явлению: увеличение разности потенциалов, приложенной к прибору в прямом направлении, сопровождается не увеличением, а уменьшением протекающего через него тока (точка 4 на рисунке 78). На вольт-амперной характеристике диода появляется падающий участок, которому соответствует отрицательное сопротивление.

Уменьшение туннельного тока по мере роста приложенного прямого напряжения будет продолжаться и

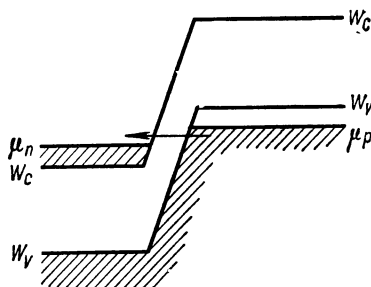


Рис. 80

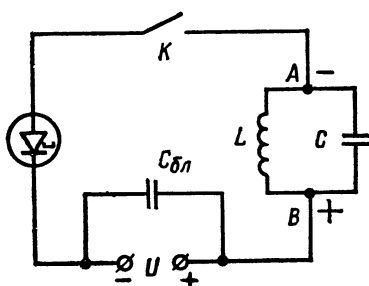


Рис. 81

далее вплоть до момента, когда дно зоны проводимости полупроводника n -типа окажется на одном уровне с потолком валентной зоны p -полупроводника (рис. 79, z). Туннельные переходы в такой ситуации становятся принципиально невозможными, и туннельный ток падает до нуля (точка $б$ на рисунке 78).

Однако, как видно из хода вольт-амперной характеристики (см. рис. 78), ток через диод не только не исчезает, но даже начинает расти по мере увеличения прямого напряжения. Объясняется это тем, что большое прямое напряжение смещения приводит к заметному снижению потенциального барьера на границе перехода. Благодаря этому увеличивается вероятность надбарьерного перехода носителей через границу раздела, то есть появляется возможность для надбарьерной инжекции электронов из n -полупроводника и дырок из p -области (см. рис. 79, z). Возникающий диффузионный ток, как и у обычных диодов, растет по мере увеличения прямого напряжения, все более снижающего потенциальный барьер на границе p — n -перехода (восходящий участок характеристики с точкой $б$).

Поведение туннельного диода при подаче обратного напряжения смещения. При включении обратного смещения преимущественными оказываются туннельные переходы электронов из валентной области p -полупроводника в зону проводимости полупроводника n -типа (на рисунке 80 — справа налево). Переходы эти оказываются ничем не ограниченными, и число их растет по мере увеличения обратного напряжения. Этим и объясняется быстрый рост обратного тока через диод (см. на

рисунке 78 участок вольт-амперной характеристики с точкой 7).

Генерирование незатухающих колебаний с помощью туннельного диода. Проиллюстрируем применение туннельных диодов на примере генерирования незатухающих колебаний. Используя отрицательное сопротивление туннельного диода, можно скомпенсировать положительное активное сопротивление какого-либо определенного участка электрической цепи и обеспечить усиление сигнала или генерацию колебаний. Так, если рабочая точка туннельного диода, включенного в цепь постоянного тока последовательно с колебательным контуром (рис. 81), находится на падающем участке вольт-амперной характеристики, то происходит восполнение потерь энергии в колебательном контуре и в нем возникают незатухающие колебания.

При замыкании ключа K в колебательном контуре возникают свободные колебания с малой амплитудой, которые в отсутствие туннельного диода вскоре бы затухли. Установим напряжение питания U схемы таким, чтобы рабочая точка диода находилась посередине участка отрицательного сопротивления вольт-амперной характеристики. В процессе электрических колебаний в контуре, возникающих после замыкания цепи, полярность точек A и B будет каждые полпериода меняться. Во время одного из полупериодов полярность этих точек окажется такой, как указано на рисунке. В этом случае напряжение, имеющееся на контуре, вычитается из напряжения питания, и общее напряжение прямого смещения на диоде уменьшается. Так как диод в выбранном нами режиме работает на участке отрицательного сопротивления, то уменьшение прямого напряжения смещения вызовет возрастание тока через диод, а следовательно, и во всей цепи. Когда же полярность зажимов контура (во время второго полупериода) станет противоположной, напряжение прямого смещения увеличится, а сила тока в цепи станет меньшей. Таким образом ток в цепи будет пульсирующим. Легко сообразить, что переменная составляющая этого тока совпадает по фазе с колебаниями напряжения на контуре. Это значит, что мощность электрического тока на участке цепи, образованном колебательным контуром, положительна ($\cos \varphi = 1$) и имеет место непрерывное пополнение энергии в

контур. За счет этого амплитуда колебаний в контуре увеличивается. Одновременно растут и потери энергии. Когда наступает равновесие между потерями энергии и ее пополнением, в контуре устанавливаются незатухающие колебания.

Как видно из схемы, генератор незатухающих электрических колебаний на туннельном диоде по своему устройству значительно проще лампового генератора.

В последнее время туннельные диоды нашли широкое применение в электронных вычислительных устройствах и других радиоэлектронных системах, требующих высокого быстродействия. Такое использование туннельных диодов объясняется их исключительно малой инерционностью (туннельный переход электронов через потенциальный барьер происходит всего за 10^{-12} — 10^{-14} с). Малая инерционность туннельных диодов позволяет применять их для генерирования и усиления колебаний сверхвысоких частот (вплоть до сотен гигагерц).

Туннельные диоды используются также в качестве быстродействующих переключателей (время переключения может быть доведена до 10^{-9} с). В электрической цепи туннельный диод работает как клапан, который при уменьшении напряжения прямого смещения открывается, а при увеличении этого смещения закрывается.

Обращенные диоды. Интересной разновидностью туннельных диодов являются так называемые обращенные диоды. Для их создания используются полупроводники с несколько меньшей степенью легированности, чем в случае обычных туннельных диодов (концентрация примеси, вводимой в полупроводник в этом случае, составляет примерно 10^{18} см⁻³). В таких полупроводниках уровни Ферми совпадают с границами разрешенных зон: в *n*-полупроводнике уровень Ферми совпадает с дном зоны проводимости, а в *p*-полупроводнике — с потолком валентной зоны. Если рассмотреть контакт таких полупроводников, находящийся в равновесном состоянии, то окажется, что перекрытия энергетических зон не происходит (рис. 82). Поэтому и туннельных переходов через границу раздела областей в отсутствие внешнего напряжения смещения нет. Не появляются они и при наличии прямого напряжения смещения, так как и в этом случае разрешенным энергиям электронов в одной области противостоит зона запрещенных энергий в другой области. По этой

причине прямой ток в диоде может быть обусловлен только надбарьерными переходами носителей. А поскольку потенциальный барьер на границе столь высоко легированных полупроводников достаточно велик (как видно из рисунка, он равен ширине запрещенной зоны полупроводника), то и сила прямого тока вплоть до весьма больших значений прямого напряжения смещения оказывается ничтожной (рис. 83). Практически она равна значению диффузионного тока, характерного для туннельных диодов вообще (пунктирная линия на рисунке 78).

Подача же на диод внешнего напряжения в обратном направлении приводит к появлению перекрытия разрешенных зон, которое с увеличением этого напряжения растет. При этом появляется возможность для туннельных переходов, число которых неограниченно увеличивается с ростом $U_{обр}$, как и в обычных туннельных диодах, из-за чего сила тока в запиорном направлении также быстро растет и становится несравненно большей силы тока в прямом направлении. Следовательно, в отношении зависимости проводимости от напряжения смещения свойства таких диодов противоположны свойствам обычных выпрямительных диодов, из-за чего эти диоды получили название *обращенных*. Они не имеют участка отрицательного сопротивления и поэтому не могут быть использованы для генерирования и усиления колебаний, но применяются в качестве детекторов в области очень высоких частот.

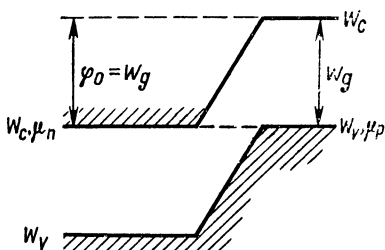


Рис. 82

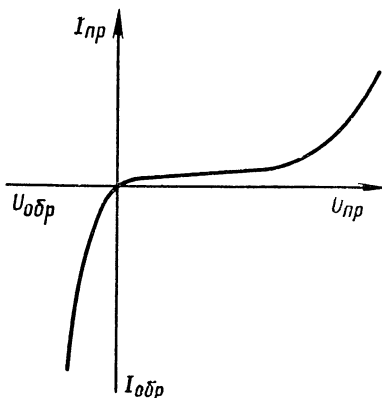


Рис. 83

§ 28. ТРАНЗИСТОРЫ

В отличие от полупроводниковых диодов *транзисторы* представляют собой полупроводниковые системы, состоящие уже из трех областей, разделенных между собой двумя p — n -переходами. Каждая из областей имеет свой вывод. Поэтому по аналогии с вакуумными триодами транзисторы часто называют *полупроводниковыми триодами*. И по назначению транзисторы аналогичны вакуумным триодам: основная область их использования — усиление электрических сигналов по напряжению и по мощности.

Для получения транзисторов в полупроводниковую монокристаллическую пластинку с определенным типом проводимости на двух ее противоположных гранях осуществляют вплавление или диффузионное проникновение примеси, сообщающей приповерхностным областям проводимость противоположного типа. Можно создать транзистор как p — n — p -типа, так и n — p — n -типа. Принципиальной разницы между ними нет. Просто главную роль в транзисторах p — n — p -типа играют дырки, а в транзисторах n — p — n -типа — электроны. Мы рассмотрим транзистор p — n — p -типа.

Плоскостной транзистор p — n -типа. Общий вид транзистора, полученного на основе германиевой пластинки n -типа, в которую с двух сторон на определенных участках вплавлен индий, создающий две p -области, приведен на рисунке 84. Средняя область транзистора называется *базой*, а крайние — *эмиттером* и *коллектором*. От базы крайние области отделены соответственно *эмиттерным* и *коллекторным p — n -переходами*. Площадь коллек-

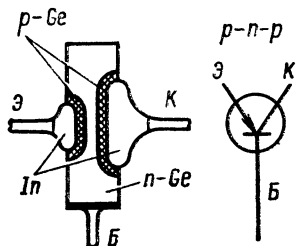


Рис. 84

торного перехода делается большей площади эмиттерного перехода. Этим обеспечивается более эффективное захватывание полем коллекторного перехода носителей, попадающих в базу из эмиттера. Схематически устройство плоскостного транзистора показано на рисунке 85, а. В состоянии равновесия, когда к транзистору не подключены источники пита-

ния, суммарные токи через оба перехода равны нулю, а уровни Ферми во всех трех областях находятся на одной высоте (рис. 85, б). На практике используются различные схемы включения транзисторов. Мы сначала остановимся на работе транзистора, включенного по так называемой схеме с общей базой.

В рабочем режиме к эмиттерному p — n -переходу приложено прямое напряжение (на эмиттер подается «+», а на базу «-»), обуславливающее снижение потенциального барьера и сужение перехода; на коллекторный же p — n -переход подается обратное смещение (на базу «+», а на коллектор «-»), вызывающее увеличение потенциального барьера и расширение перехода (рис. 86, а и б). Рассмотрим сначала процессы, протекающие в каждом из переходов, в отдельности.

Инжекция дырок в базу. Снижение потенциального барьера эмиттерного перехода приводит к ин-

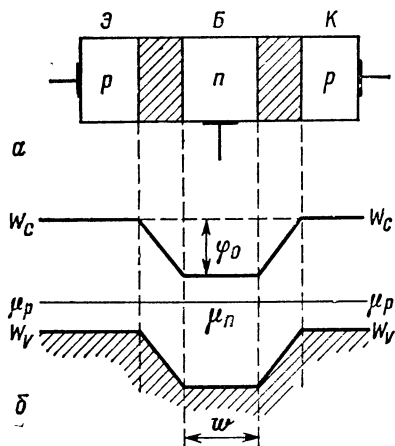


Рис. 85

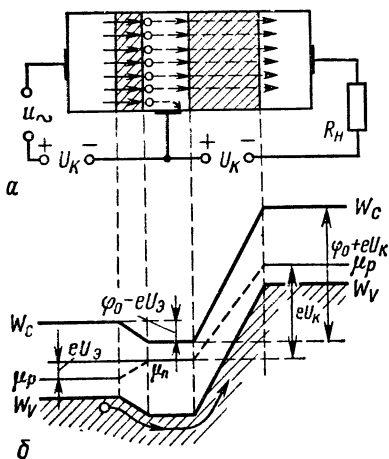


Рис. 86

тенсивной инжекции носителей и возникновению диффузионного тока. Электронная составляющая этого тока, определяемая переходом электронов из базы в эмиттер, оказывается ненужной (и даже вредной) в работе транзистора. Поэтому на практике стремятся свести ее к минимуму. Этого удастся достигнуть в том случае, когда степень легированности эмиттера намного превосходит

легированность базы. Благодаря тому что концентрация дырок в эмиттере несравненно больше концентрации электронов в базе, ($p_{po} \ll n_{no}$) число дырок, переходящих из эмиттера в базу, во много раз превосходит число электронов, идущих из базы в эмиттер, и можно считать, что весь диффузионный ток через эмиттерный переход создается только дырками, инжектируемыми из эмиттера в базу.

Интенсивная инжекция приводит к резкому возрастанию концентрации дырок (неосновных носителей) в базе у ее границы с эмиттером.

Обратимся к коллекторному переходу.

Коллекторный переход. На коллекторный p — n -переход подается обратное напряжение смещения. Поэтому (при выключенном питании эмиттерного перехода) через него течет только обратный ток, равный току насыщения i_s и создаваемый исключительно неосновными носителями: дырками, идущими из базы n -типа в коллектор, и электронами, идущими из коллектора p -типа в базу. В коллекторном переходе электронная составляющая тока также оказывается ненужной. Поэтому и здесь стараются ее уменьшить, создавая резкое различие в легированности контактирующих областей так, чтобы выполнялось условие $p_{po} \gg n_{no}$, при котором число электронов (неосновных носителей) в коллекторе оказывается относительно мало, и можно считать, что весь обратный ток через коллекторный переход создается дырками, идущими из базы в коллектор. Так как поле коллекторного перехода подхватывает дырки, идущие из базы, и выводит их в коллектор, то концентрация их в базе на границе с коллектором практически равна нулю.

Рассмотрим теперь ситуацию, когда включены оба p — n -перехода. Благодаря интенсивной инжекции дырок из эмиттера в базу их концентрация в базе на границе с эмиттером во много раз превышает равновесную, в то же время у границы базы с коллектором концентрация дырок близка к нулю. Такой резкий перепад в концентрациях приводит к интенсивной диффузии дырок через базу от эмиттера к коллектору. Этому способствует малая толщина базы: часть дырок, проходящих через базу, рекомбинирует в ней с электронами; поэтому, чем меньше толщина базы, тем меньшее число дырок рекомбини-

рует в ее объеме и тем большее число их доходит до коллекторного перехода. Если толщина базы w много меньше диффузионной длины дырок в n -области ($w \ll L_p$), то почти все дырки, инжектированные из эмиттера в базу, достигнут коллекторного перехода. Здесь, будучи подхваченными полем перехода, они переводятся в коллектор. Понятно, что дополнительный коллекторный ток, образованный переходом этих дырок, практически равен эмиттерному току. Протекавший через коллекторный переход обратный ток много меньше тока, создаваемого дырками, инжектированными из эмиттера в базу и проходящими далее через коллекторный переход. Поэтому коллекторный ток можно считать равным эмиттерному току ($I_K \approx I_3$).

Как транзистор усиливает? Равенство коллекторного тока эмиттерному лежит в основе усиливающего действия транзистора. Реализуется же это действие благодаря большому различию сопротивлений коллекторного и эмиттерного p — n -переходов, включенных в противоположных направлениях.

Эмиттерный переход, на который подано прямое напряжение смещения, имеет малое сопротивление, и падение напряжения на нем U_3 мало. На коллекторный же переход подается обратное напряжение смещения, и сопротивление этого перехода значительно больше. Поэтому в коллекторную цепь может быть включена высокоомная нагрузка, сопротивление которой R_K значительно больше сопротивления эмиттерного перехода. Поскольку эмиттерный и коллекторный токи одинаковы, то падение напряжения на высокоомной коллекторной нагрузке $U_K = I_K R_K \approx I_3 R_K$ окажется намного большим падения напряжения U_3 на эмиттерном переходе ($U_K \gg U_3$). В этом по существу и состоит эффект усиления по напряжению, обеспечиваемого транзистором.

Так как $U_K I_K \ll U_3 I_3$, то усиление происходит и по мощности: выходная мощность $P_{\text{вых}}$ на высокоомной нагрузке в коллекторной цепи значительно больше входной мощности $P_{\text{вх}}$, подводимой к эмиттерному переходу ($P_{\text{вых}} \gg P_{\text{вх}}$).

Способы включения транзисторов. Транзисторы можно включать тремя различными способами: по схеме с общей базой, по схеме с общим эмиттером и по схеме с общим коллектором.

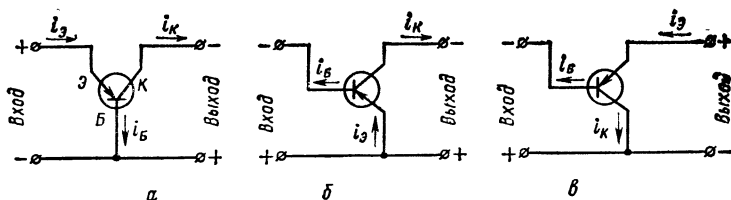


Рис. 87

При включении транзистора по схеме с общей базой (рис. 87, а) и эмиттерное, и коллекторное напряжения отсчитываются относительно базы. Цепь эмиттера при этом служит входной цепью, а цепь коллектора — выходной. Усилительное действие транзистора при его включении по такой схеме (см. также рис. 88, а) уже было рассмотрено выше.

На рисунке 87, б показано включение транзистора по схеме с общим эмиттером. В этом случае потенциалы базы и коллектора отсчитываются относительно эмиттера. Базовая цепь является входной, а коллекторная — выходной. Эта схема получила наиболее широкое распространение в различных транзисторных устройствах.

В случае схемы с общим коллектором (рис. 87, в) входной сигнал подается на переход «коллектор — база», а нагрузка включается между эмиттером и коллектором. Такая схема включения встречается значительно реже, чем предыдущие.

Работа транзистора в схеме с общим эмиттером. Рассмотрим работу усилительного каскада с использованием транзистора $p-n-p$ -типа, включенного по схеме с общим эмиттером (рис. 88, б). Источник усиливаемого сигнала включается в цепь перехода «эмиттер — база» последовательно с источником напряжения смещения. Поскольку эмиттерный

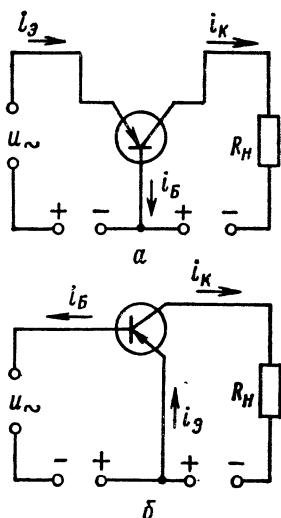


Рис. 88

переход включен в прямом направлении, то величина входного сопротивления оказывается сравнительно небольшой.

В цепь «эмиттер—коллектор» источник питания включается таким образом, что его напряжение оказывается для коллекторного перехода обратным. Сопротивление коллекторного перехода при таком подключении оказывается большим, и это позволяет в цепь коллектора включить высокоомную нагрузку R_n , на которой выделяется усиленный сигнал.

Под действием напряжения усиливаемого сигнала происходит изменение силы тока в цепи «эмиттер—база», что в свою очередь вызывает изменение сопротивления коллекторного перехода и, следовательно, перераспределение падения напряжения источника между транзистором и нагрузкой. Действительно, в тот полупериод, когда напряжение усиливаемого сигнала складывается с напряжением прямого смещения, подаваемого на эмиттерный переход, происходит увеличение инжекции носителей из эмиттера в базу, сопровождающееся ростом эмиттерного и, следовательно, коллекторного токов. С другой стороны, приток подвижных носителей (дырок) в обедненную область коллекторного перехода снижает его сопротивление, благодаря чему большая доля напряжения источника, включенного в коллекторную цепь, падает на сопротивлении нагрузки R_n . Небольшое увеличение прямого смещения на эмиттерном переходе сопровождается значительным увеличением разности потенциалов на сопротивлении нагрузки R_n . При обратной полярности входного сигнала наблюдается противоположная картина: сопротивление коллекторного перехода увеличивается и падение напряжения на нагрузке уменьшается.

При синусоидальном характере входного сигнала амплитуда переменной составляющей напряжения на сопротивлении нагрузки R_n оказывается в десятки и даже сотни раз большей амплитуды переменного напряжения на входе. В отличие от схемы с общей базой схема с общим эмиттером дает значительное усиление и по току, так как изменение тока коллектора, практически равное изменению тока эмиттера, во много раз больше изменения тока базы (входного тока усилительного каскада). Усиление по току, как и усиление по напряжению, про-

исходит в десятки и сотни раз. Одновременное усиление по току и напряжению приводит к реализации максимальных возможностей усиления по мощности. Коэффициент усиления по мощности, представляющий собой отношение выходной мощности к входной, может достигать в схеме с общим эмиттером нескольких тысяч и даже десятков тысяч.

§ 29. ПОЛУПРОВОДНИКОВЫЕ ИНЖЕКЦИОННЫЕ ЛАЗЕРЫ

Фотоны создают фотоны. Мы уже знаем, что полупроводники способны как поглощать, так и испускать свет.

Поглощение света с энергией квантов $h\nu \approx W_g$ связано в первую очередь с явлением внутреннего фотоэффекта. Электроны, находящиеся на уровнях валентной зоны, поглощая фотоны, переходят в зону проводимости (рис. 89, а), то есть из невозбужденного состояния в состояние возбуждения. В возбужденном состоянии они находятся недолго. Спустя некоторое время, продолжительность которого зависит от ширины запрещенной зоны полупроводника, они возвращаются в валентную зону, рекомбинируя при этом с дырками и испуская кванты света с энергией $h\nu$, приблизительно равной ширине запрещенной зоны W_g (рис. 89, б). Пока полупроводник не подвергается никакому внешнему воздействию, возврат электронов в валентную зону происходит с а м о п р о и з в о л ь н о, или, как говорят, *спонтанно*. Никто не может предсказать, когда произойдет акт спонтанной рекомбинации и какими свойствами будет обладать испущенный при этом фотон.

Однако, помимо спонтанной, возможна еще и *вынужденная рекомбинация*, происходящая благодаря воздействию внешнего излучения. Фотон, попавший в полупроводник, встречая на своем пути возбужденный электрон, может его как бы «подтолкнуть» и тем самым вынудить возвратиться в валентную зону. При этом сам первичный фотон, воздействуя на электрон, никак не меняет свои свойства, а фотон, рожденный при рекомбинации электрона с дыркой, по всем характеристикам в точности такой же, как первичный фотон: он имеет такую же частоту и энергию, такое же направление распространения и такую же поляризацию. На выходе из полу-

проводника уже нельзя отличить первичные фотоны от фотонов *вынужденного* (индуцированного) излучения (рис. 90, а, б).

Инверсия населенностей. Попадая в полупроводник, фотон может или поглотиться и исчезнуть, создав свободный электрон и дырку, или вызвать акт рекомбинации, сопровождающийся рождением нового фотона. Вероятности обоих процессов совершенно одинаковы.

Но это не означает, что эти процессы будут происходить с одинаковой интенсивностью. Соотношения между числом переходов электронов из валентной зоны в зону проводимости (поглощение света) и числом переходов электронов из зоны проводимости в валентную зону (испускание света) зависит не только от вероятностей отдельных переходов, но и от общего числа электронов, способных принять участие в каждом из этих видов переходов. Иными словами, число переходов, при которых свет поглощается, пропорционально количеству электронов, населяющих верхние энергетические уровни вблизи потолка валентной зоны, а число переходов, при которых свет испускается, пропорционально количеству электронов, населяющих нижние уровни у дна зоны проводимости. В общем случае, когда распределение электронов по энергетическим состояниям определяется только тепловым возбуждением, населенность электронами верхних уровней валентной зоны всегда много больше, чем населенность нижних уровней зоны проводимости. Поэтому в обычных условиях полупроводник поглощает свет.

Для того чтобы полупроводник мог не поглощать, а усиливать свет, необходимо, чтобы населенность электронами уровней, прилегающих к дну зоны проводимости, была больше населенности уровней, находящихся вблизи потолка валентной зоны. В этом случае при равенстве вероятностей встречных актов создания и рекомбинации

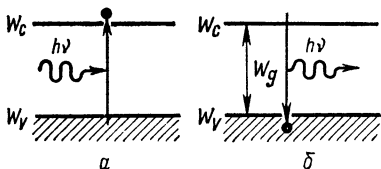


Рис. 89

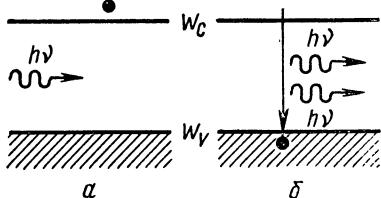


Рис. 90

электронно-дырочных пар общее количество рекомбинационных актов будет преобладать. Подобное распределение электронов по энергетическим уровням представляется обратным по отношению к распределению их при тепловом равновесии, и поэтому такое состояние полупроводника называется состоянием с *инверсией населенностей*.

Полупроводник, в котором большинство уровней в области дна зоны проводимости занято электронами, представляет собой вырожденный электронный полупроводник с характерным для таких полупроводников расположением уровня Ферми в самой зоне проводимости. В то же время полупроводник, характеризующийся малой населенностью электронами уровней, прилегающих к потолку валентной зоны (иными словами, характеризующийся большой населенностью этих уровней дырками), есть не что иное, как вырожденный дырочный полупроводник, у которого уровень Ферми расположен в валентной зоне. Таким образом, перевод полупроводника в активное состояние, характеризующееся инверсией населенностей, связан с созданием в нем одновременного вырождения как по электронам, так и по дыркам (рис. 91). В этом случае расстояние между уровнями Ферми для электронов μ_n и для дырок μ_p оказывается больше ширины запрещенной зоны:

$$\mu_n - \mu_p > W_g.$$

Это соотношение называется *условием инверсии населенностей* в полупроводнике.

Состояние полупроводника с одновременным вырождением и по электронам и по дыркам является идеальным для усиления света, энергия квантов которого лежит в интервале от $h\nu_{\min} = W_g$ до $h\nu_{\max} = \mu_n - \mu_p$. Действительно, кванты света, энергия которых лежит в этом диапазоне, не будут поглощаться таким полупроводником, так как, с одной стороны, их некому поглощать (электронов у потолка валентной зоны мало) и, с другой стороны, даже тому небольшому количеству электронов, которое может появиться в этой области, некуда переходить при поглощении кванта (уровни у дна зоны проводимости практически все заняты электронами). Вместе с тем в подобном состоянии имеются благоприятные условия для рекомбинации электронов, находя-

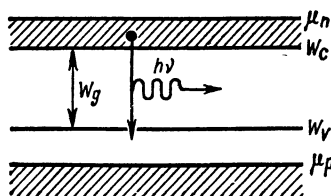


Рис. 91

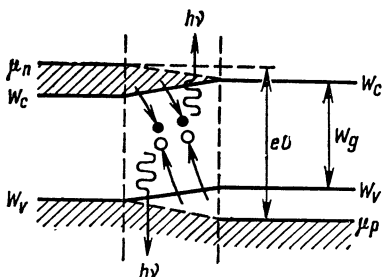


Рис. 92

щихся на заполненных уровнях зоны проводимости с дырками на пустых уровнях валентной зоны. В процессе такой рекомбинации испускаются кванты света, энергия которых лежит в указанном диапазоне.

Создание инверсии населенностей в контакте между вырожденными полупроводниками. Инверсию населенностей в полупроводнике можно получить различными путями: облучением полупроводника светом, бомбардировкой его быстрыми электронами, прямым электрическим возбуждением. Но мы остановимся на наиболее интересном способе — получении инверсии населенностей в p — n -переходе, образованном между полупроводниками, один из которых вырожден по электронам, а другой — по дыркам. Так как расстояние между уровнями Ферми у таких полупроводников больше W_g (см. рис. 77), то при создании p — n -перехода между ними возникающий потенциальный барьер $\phi_0 = \mu_n - \mu_p$ оказывается больше ширины запрещенной зоны W_g . Если приложить к переходу внешнюю прямую разность потенциалов, сравнимую с контактной разностью потенциалов, то это приведет к исчезновению потенциального барьера и к созданию вблизи границы раздела области, в которой реализуется условие инверсии населенностей (рис. 92):

$$eU = \mu_n - \mu_p > W_g.$$

Резкое снижение потенциального барьера приводит к интенсивной инжекции в p — n -переход электронов из n -области и дырок из p -области полупроводника. Поток этих носителей, встречаясь в p — n -переходе¹, рекомби-

¹ Этому способствует и малая толщина p — n -перехода, образованного двумя вырожденными полупроводниками.

нируют, излучая при этом свет. Чем выше прикладываемая разность потенциалов, тем больше ток, текущий через переход, и тем более интенсивной становится рекомбинация. Минимальная сила тока, при которой интенсивность рекомбинационного излучения становится сравнимой с интенсивностью поглощенного света в p — n -переходе, называется *пороговым током*. При токе, большем порогового, p — n -переход становится активной средой, которая усиливает свет, распространяющийся в плоскости p — n -перехода. В таком состоянии p — n -переход может быть использован в качестве *квантового усилителя* света, энергия квантов которого близка к ширине запрещенной зоны. Первичное усиливаемое излучение не обязательно должно приходить в p — n -переход извне, оно может создаваться и в самом переходе за счет спонтанных переходов. В этом случае p — n -переход превращается в *квантовый генератор*, или *лазер*. Поскольку в основе работы такого источника света лежит инжекция носителей в p — n -переход, то такой квантовый генератор получил название *полупроводникового инжекционного лазера*.

Для улучшения условий генерации света в любую лазерную систему вводят обратную связь, задача которой состоит в том, чтобы возвращать часть полученного излучения обратно в активную среду. Обычно для этих целей используют полупрозрачные зеркала. В полупроводниковых лазерах роль зеркал выполняют полированные грани самого кристалла, перпендикулярные плоскости p — n -перехода и параллельные между собой. Часть квантов света, отраженная от выходной грани и возвращенная обратно в кристалл, проходя через него, вызывает вынужденную рекомбинацию, при которой испускаются дополнительные кванты света, идентичные отраженным. Повторное отражение от другой грани приведет к еще большему усилению проходящего света и так далее. При многократном отражении может быть получено очень большое усиление того излучения, которое распространяется в плоскости p — n -перехода в направлении, перпендикулярном полированными отражающим граням кристалла (рис. 93).

Наиболее широкое распространение получили инжекционные лазеры на основе арсенида галлия (GaAs). Грани кристаллов этих лазеров имеют размеры порядка

0,2—1 мм, а толщина p — n -перехода составляет всего приблизительно 0,1 мкм. Толщина излучающего слоя несколько больше толщины перехода и достигает 1—2 мкм. Мощность излучения лазеров на арсениде галлия равна десяткам ватт. Инжекционные лазеры име-

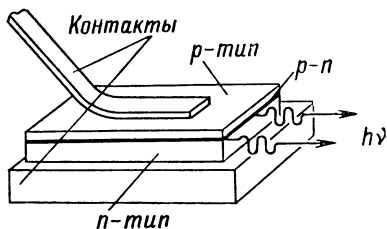


Рис. 93

ют большой коэффициент полезного действия (коэффициент преобразования электрической энергии в световую), который во многих случаях превышает 50—60 %. Большой КПД, миниатюрные размеры, простая конструкция и значительная мощность — все эти положительные стороны открывают широкую перспективу для использования инжекционных лазеров. Наиболее интересными областями применения инжекционных лазеров являются счетно-вычислительные устройства, локаторные установки с высокой разрешающей способностью, оптоэлектронные системы, устройства для беспроводной передачи энергии на расстояние, телевизионные системы и многое другое. Из логических элементов, в которых используются инжекционные лазеры, можно построить чисто оптическое вычислительное устройство с быстродействием 1 млрд. и более операций в секунду.

ПОЛУПРОВОДНИКИ СЕГОДНЯ И ЗАВТРА

Полупроводники стремительно ворвались в науку и технику нашего века. Колоссальная экономия в энергопотреблении, удивительная компактность аппаратуры за счет необычайно большой плотности упаковки элементов в схемах, высокая надежность позволили полупроводникам завоевать ведущее положение в электронике, радиотехнике и науке. Исследования в космосе, где так критичны требования к размерам, весу и энергозатратам, в настоящее время немыслимы без полупроводниковых устройств, которые, кстати, и энергию-то в автономном полете аппарата получают от солнечных батарей, работающих на полупроводниковых элементах.

Удивительные перспективы в развитии полупроводниковой техники открыла микроэлектроника. Применение микроэлектронной технологии позволило перейти от полупроводниковых приборов, собираемых из отдельных элементов, к созданию и применению так называемых интегральных схем. Современная технология дает возможность на 1 см^2 создать 10^5 элементов. Используя созданные в последние годы слоистые структуры (металл—нитрид—диэлектрик—полупроводник), можно на 1 мм^2 их поверхности получить 10^8 участков, несущих определенную информацию. Считывать эту информацию можно не только с помощью электрических сигналов, но и применяя лазерное излучение, получаемое от тех же полупроводниковых инжекционных лазеров.

Однако возможности полупроводников еще далеко не исчерпаны, и они ждут своих новых исследователей.

ОГЛАВЛЕНИЕ

Глава 1.

О ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ

§ 1. Как устроены атомы? Атомы водорода (4). § 2. Многоэлектронные атомы (12). § 3. Вырождение энергетических уровней в свободных атомах. Снятие вырождения под влиянием внешнего воздействия (15). § 4. Образование энергетических зон в кристаллах (17). § 5. Заполнение энергетических зон электронами (19). § 6. Деление твердых тел на проводники, полупроводники и диэлектрики (22)

Глава 2.

ЭЛЕКТРОПРОВОДНОСТЬ ТВЕРДЫХ ТЕЛ

§ 7. Силы связи в кристаллической решетке (25). § 8. Электрическая проводимость металлов (30). § 9. Электрическая проводимость полупроводников (34). § 10. Влияние температуры на концентрацию носителей заряда в полупроводниках (44). § 11. Зависимость электрической проводимости полупроводников от температуры (51)

Глава 3.

НЕРАВНОВЕСНЫЕ ПРОЦЕССЫ В ПОЛУПРОВОДНИКАХ

§ 12. Генерация и рекомбинация неравновесных носителей заряда (57). § 13. Диффузионные явления в полупроводниках (66). § 14. Фотопроводимость и поглощение света (70). § 15. Люминесценция (78)

Глава 4.

КОНТАКТНЫЕ ЯВЛЕНИЯ

§ 16. Работа выхода электронов из металла (84). § 17. Уровень Ферми в металлах и функция распределения Ферми-Дирака (88). § 18. Уровень Ферми в полупроводниках (92). § 19. Контактная разность потенциалов (99). § 20. Контакт металла с полупроводником (102). § 21. Выпрямляющие свойства контакта «металл-полупроводник» (106). § 22. Электронно-дырочный переход (111).

§ 23. Эффект выпрямления на p - n -переходе (122). § 24. Прибор p - n -перехода (129). § 25. Электрическая емкость p - n -перехода (131)

Глава 5.

ПОЛУПРОВОДНИКОВЫЕ ПРИБОРЫ

§ 26. Полупроводниковые диоды (134). § 27. Туннельные диоды (137). § 28. Транзисторы (146). § 29. Полупроводниковые инжекционные лазеры (152). Полупроводники сегодня и завтра (157)

Александр Михайлович Поляков

РАЗГАДАННЫЙ ПОЛУПРОВОДНИК

Редактор *Г. Р. Лисенкер*
Художественный редактор *В. М. Прокофьев*
Технический редактор *Н. Н. Махова*
Корректор *В. Г. Соловьева*

ИБ № 5767

Сдано в набор 03.09.80. Подписано к печати 27.03.81. 84×108¹/₃₂. Бумага тип. № 2. Гарнитура «Литературная». Печать высокая. Усл. печ. л. 8,4. Уч.-изд. л. 8,55. Тираж 100 000 экз. Заказ № 988. Цена 25 коп.

Ордена Трудового Красного Знамени издательство «Просвещение» Государственного комитета РСФСР по делам издательств, полиграфии и книжной торговли Москва, 3-й проезд Марьиной рощи, 41.

Полиграфкомбинат им. Я. Коласа. 220005, Минск, Красная, 23.

