

Министерство образования и науки Российской Федерации

САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Н.М. Гнучев

ЭЛЕКТРОННЫЕ ПРИБОРЫ
ФИЗИЧЕСКИЕ ОСНОВЫ ЭЛЕКТРОНИКИ

Учебное пособие

Санкт-Петербург
Издательство Политехнического университета
2013

УДК 53.043

Гнучев Н.М. Электроника и схемотехника. Электронные приборы. Физические основы электроники: Учеб. пособие / Н.М. Гнучев. – СПб.: Изд-во Политехн. ун-та. – 2013. – с.

Соответствует программам дисциплин Б3.Б.10 «Электроника и схемотехника. Электронные приборы» и Б3.Б.8 «Физические основы электроники», входящих в базовые части профессиональных циклов учебных планов подготовки бакалавров по направлениям 223200 «Техническая физика» и 210100 «Электроника и наноэлектроника».

Материал учебного пособия может быть использован также при прохождении практикума в научных лабораториях студентов в рамках магистерской подготовки по тем же направлениям (дисциплины М1.ДВ3 – направление 223200, М1.ДВ2 – направление 210100).

Пособие содержит описание физических основ твердотельной электроники и элементарную теорию электрических полупроводниковых переходов и предназначено для студентов, обучающихся по направлениям подготовки 223200 «Техническая физика» и 210100 «Электроника и наноэлектроника». Оно также будет полезно для студентов, изучающих дисциплину «Электроника» в рамках направлений подготовки 210400 «Радиотехника» и 210700 «Телекоммуникации» очной и очно-заочной формы обучения.

Печатается по решению редакционно-издательского совета Санкт-Петербургского государственного политехнического университета

ОГЛАВЛЕНИЕ

1. Электронная структура кристаллов.....	5
1.1. Энергетические уровни атомов.....	5
1.2. Энергетические зоны кристаллов.....	6
2. Электроны и дырки в полупроводниках.....	10
2.1. Собственная проводимость.....	10
2.2. Примесная проводимость. Полупроводник n-типа.....	12
2.3. Примесная проводимость. Полупроводник p-типа.....	13
2.4. Распределение электронов по энергиям в пределах энергетических зон.....	16
2.5. Равновесная концентрация свободных носителей в полупроводниках.....	20
2.6. Уровень Ферми на энергетической диаграмме собственного полупроводника.....	22
2.7. Взаимосвязь концентраций подвижных носителей и уровень Ферми в примесных полупроводниках.....	23
3. Движение носителей заряда в полупроводниках.....	25
3.1. Хаотическое движение.....	25
3.2. Дрейф подвижных носителей заряда. Дрейфовые токи.....	25
3.3. Диффузия подвижных носителей заряда. Диффузионные токи.....	26
3.4. Взаимосвязь диффузии и дрейфа заряженных частиц.....	27
3.5. Движение носителей в сильных электрических полях.....	28
4. Генерация и рекомбинация носителей заряда.....	29
4.1. Тепловая генерация.....	29
4.2. Ударная генерация.....	31
4.3. Полевая генерация.....	32
4.4. Световая генерация.....	33
4.5. Рекомбинация подвижных носителей заряда. Механизмы рекомбинации.....	34
4.6. Диффузионная длина неосновных носителей в примесных полупроводниках.....	37
4.7. Время жизни неосновных носителей в полупроводнике. Уравнение генерации- рекомбинации.....	38
5. Контактные явления в полупроводниках.....	39
5.1. Электронно-дырочный переход (p-n переход).....	39
5.2. P-n переход в равновесном состоянии.....	40

5.3. Прямое включение р-n перехода.....	47
5.4. Обратное включение р-n перехода.....	49
5.5. Вольтамперная характеристика идеализированного р-n перехода.....	52
5.6. Реальный электронно-дырочный переход..... Обратное смещение. Токи генерации и утечки.....	57
5.7. Прямое смещение реального р-n перехода. Ток рекомбинации и влияние сопротивления базы.....	59
5.8. Р-n переход на основе вырожденных полупроводников..	61
5.9. Пробой электронно-дырочного перехода.....	65
5.10. Ёмкость р-n перехода.....	69
5.11. Переходные процессы в р-n переходе. Установление прямого напряжения при подаче на переход скачка прямого тока.....	72
5.12. Установление постоянного обратного тока перехода при ступенчатом переключении напряжения с прямого на обратное.....	75
5.13. Переходы металл – полупроводник. Выпрямляющий контакт (барьер Шоттки).....	77
5.14. Омический контакт металл- полупроводник.....	82
5.15. Гетеропереходы.....	82
6. Фотоэлектрические явления и светоизлучение.....	85
6.1. Взаимодействие оптического излучения с полупроводниками. Фотопроводимость.....	85
6.2. Фоторезистивный эффект. Фоторезистор.....	86
6.3. Фотоэлектрический эффект в р-n переходе.....	88
6.4. Излучение света полупроводниками. Светодиоды.....	92

1. ЭЛЕКТРОННАЯ СТРУКТУРА КРИСТАЛЛОВ.

1.1. Энергетические уровни атомов

Атом любого вещества является весьма сложной физической системой. В рамках классического подхода его структура (в «нулевом» приближении – ядро с вращающимися вокруг него электронами) абсолютно неустойчива. Электроны, двигаясь по орбитам с центростремительным ускорением, должны терять энергию на электромагнитное излучение и неизбежно «упасть» на ядро.

Теория Бора позволила разрешить это противоречие. Согласно её постулатам, электроны движутся по стационарным орбитам вокруг ядра, не излучая электромагнитные волны и, следовательно, не теряют энергию. Кроме того, переход электрона на более «высокую» или более «низкую» орбиту возможен только при поглощении или испускании кванта с энергией $h\nu = E_m - E_n$, где E_m и E_n - значения энергий электрона в атоме, соответствующие двум различным состояниям.

Расположение электронов по энергетическим состояниям и электронным оболочкам в атомах должно соответствовать основному правилу атомной физики – принципу запрета Паули. Согласно этому принципу, в атоме в данном состоянии может находиться только один электрон, состояние другого должно отличаться хотя бы одним квантовым числом. Под уровнем энергии будем понимать объединение электронов в двух состояниях с противоположными ориентациями спинов. Тогда можно считать, что на одном уровне энергии может быть не более двух электронов, и на электронных оболочках атомов могут находиться электроны различных энергетических уровней. На основе всех этих закономерностей построена периодическая система элементов.

Учитывая всё, сказанное выше, проследим, как изменяется количество электронов на первых двух атомных «орбитах». На K-оболочке, ближайшей к ядру атома, может находиться не более двух электронов (один энергетический уровень – 1s-состояние). На L-оболочке может находиться не более восьми электронов, из которых два – в 2s-состоянии и шесть – в состоянии 2p; в этом состоянии объединены 3 уровня энергии. Так, например, в атоме про-

стейшего металла – лития – 3 электрона (рис. 1.1), из которых два – на К-оболочке (1s), и третий – на L-оболочке в состоянии 2s. В атоме следующего металла – бериллия – 4 электрона, по два в состояниях 1s и 2s. Затем, начиная с атома бора вплоть до атома инертного газа – неона, происходит заполнение 2p-состояний (от одного до шести электронов). Предшествующий неону в таблице Менделеева атом простейшего галогена – фтора имеет на L-оболочке семь электронов (два – в 2s-состоянии и пять – в состоянии 2p).

1.2. Энергетические зоны кристаллов

При объединении отдельных атомов вещества в кристалл все электроны образуют систему, в которой действуют законы квантовой механики. Согласно принципу Паули, электроны, занимающие один и тот же энергетический уровень в отдельных атомах, не могут находиться в кристалле в абсолютно одинаковых энергетических состояниях. Например, в кристалле лития, составленном из N атомов, 1s-электроны не могут занимать один и тот же уровень энергии. Поэтому при образовании кристалла происходит расщепление энергетических уровней, и образуется энергетическая зона (1s-зона лития), состоящая из N очень близко (с интервалом 10^{-20} – 10^{-21} эВ) расположенных уровней. Таким же образом в кристалле образуются и остальные энергетические зоны. Ясно, что при такой «тесноте» уровней распределение электронов по энергиям в пределах зоны можно считать не дискретным, а квазинепрерывным.

Принцип формирования энергетических зон в кристалле лития показан на рис. 1.1. В зоне 1s все $2N$ вакантных мест заполнены электронами, а зона 2s оказывается заполненной лишь наполовину: на $2N$ мест приходится только N электронов. Наличие частично заполненной энергетической зоны – характерный признак всех металлов, обеспечивающий их электропроводность.

Электроны частично заполненной зоны – это «свободные» электроны в металле. В отсутствие внешнего электрического поля их импульсы распределены в пространстве кристалла изотропно, что соответствует их хаотическому тепловому движению. При появлении электрического поля возникает асимметрия в пространственном распределении импульсов электронов, и как следствие,

дрейф всей совокупности электронов в направлении, противоположном вектору электрического поля.

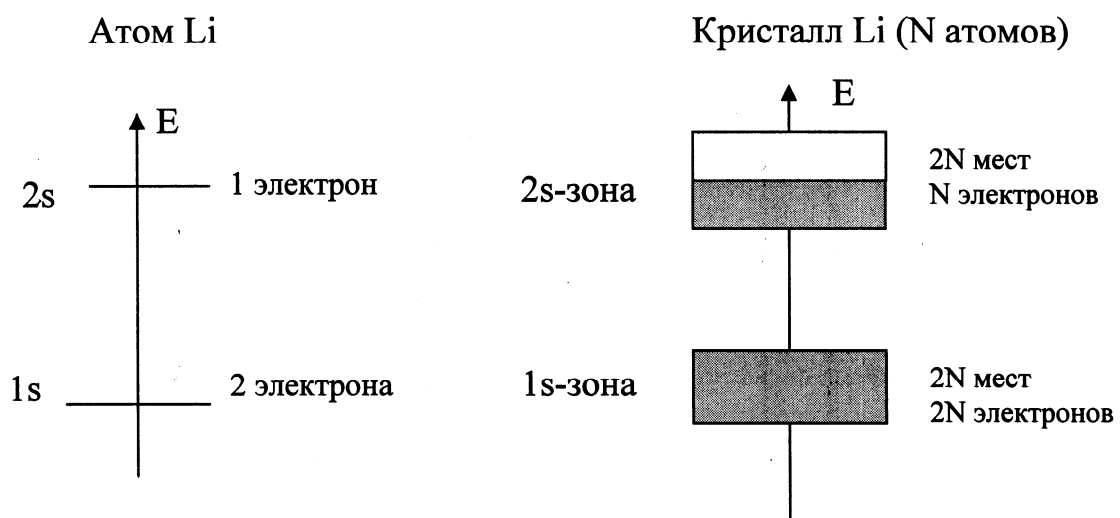


Рис. 1.1. Упрощенный принцип образования энергетических зон в кристалле лития, составленном из N атомов.

□ – незанятые электронами части энергетических зон; ■ – состояния, занятые электронами в зоне.

Рассмотрим теперь, каким образом образуются энергетические зоны в кристалле простейшего диэлектрика – фторида лития LiF . Будем считать, что кристалл составлен из N атомов лития и N атомов фтора (рис.1.2). В соответствии с рассмотренными выше принципами 1s-уровень лития, 1s- и 2s- уровни фтора расщепляются в энергетические зоны. В каждой из этих зон будет по $2N$ мест, полностью заполненных электронами.

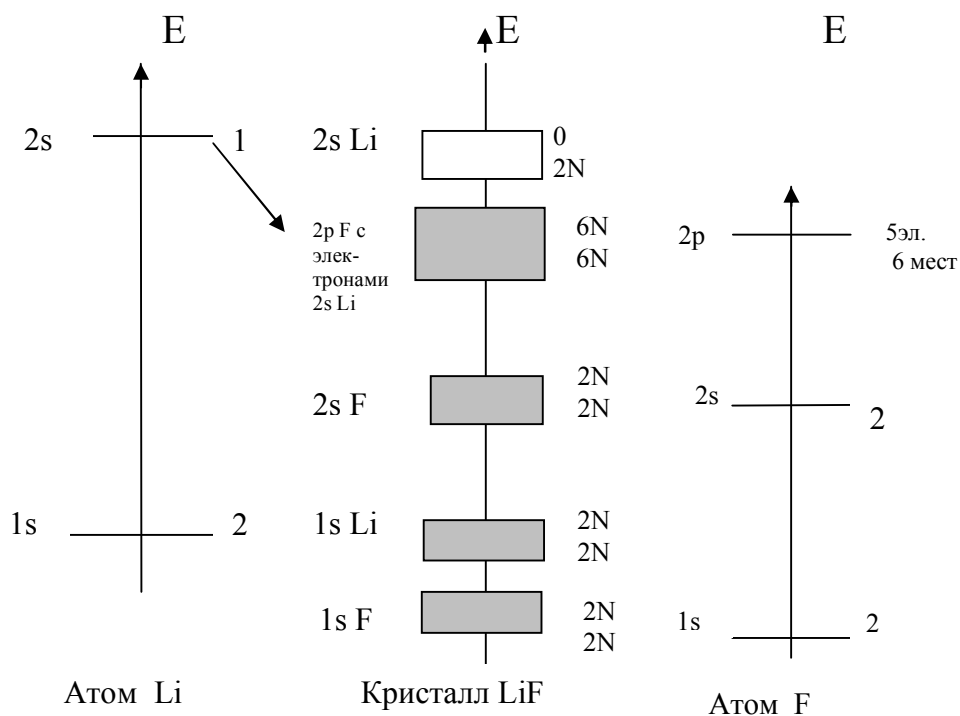


Рис. 1.2. Упрощенный принцип образования энергетических зон в кристалле диэлектрика - LiF, составленном из N атомов Li и N атомов F. Справа от зон (снизу вверх) указаны: число мест для электронов и число электронов в зоне.

□ – энергетическая зона $2s$ Li, не занятая электронами;
 ■ - полностью заполненные зоны.

При образовании зон $2p$ фтора и $2s$ лития в кристалле диэлектрика LiF происходит перераспределение электронов. $2s$ -электроны лития переходят на нижележащие свободные уровни $2p$ -зоны фтора, в которой изначально имеется N свободных мест. В результате $2p$ -зона заполняется полностью, в $2s$ -зоне Li все $2N$ состояний остаются незанятыми, и между свободными и занятыми электронными энергетическими состояниями возникает интервал запрещенных значений энергии – «запрещенная зона».

В качестве одномерной энергетической диаграммы используют самые верхние зоны энергий электронов, содержащие занятые и свободные энергетические уровни. В нашем примере для лития – это частично заполненная $2s$ -зона, для LiF – полностью заполненная зона $2p$ фтора и зона свободных состояний $2s$ Li. Таким образом, простейшая одномерная энергетическая диаграмма металла

(рис.1.3) – это квазинепрерывный ряд занятых и свободных состояний.

Простейшая энергетическая диаграмма диэлектрика (рис.1.4) состоит из двух зон: валентной (полностью заполненной) и зоны проводимости, в которой при $T = 0$ нет электронов. Эти зоны отделены друг от друга интервалом запрещенных для электронов значений энергии – «запрещенной зоной» ΔE_z . Физический смысл этой величины очевиден: это – наименьшая энергия, которую необходимо сообщить связанным электронам для перевода их в «свободное» состояние в кристалле, когда они могут обеспечивать электропроводность.

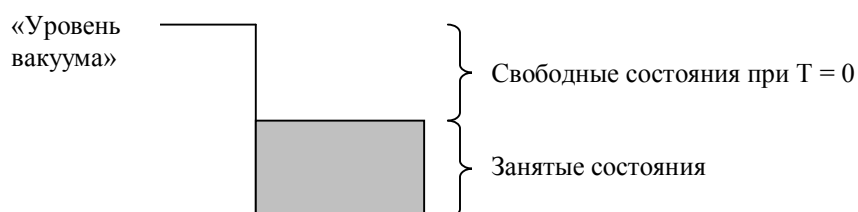


Рис. 1.3. Простейшая одномерная энергетическая диаграмма металла. «Уровень вакуума» – энергия, которой обладает электрон, вышедший из металла и находящийся в состоянии покоя с нулевой кинетической энергией вблизи его поверхности (формально – в бесконечно удалённой точке пространства).

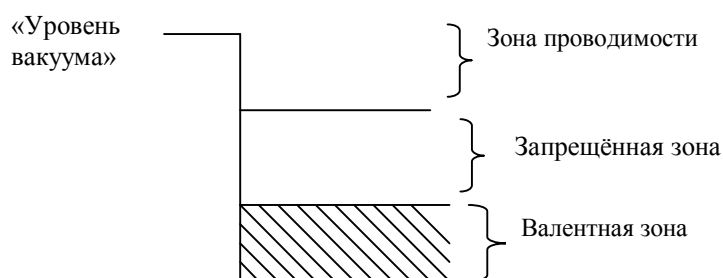


Рис. 1.4. Простейшая одномерная энергетическая диаграмма идеального диэлектрика или беспримесного (собственного) полупроводника.

2. ЭЛЕКТРОНЫ И ДЫРКИ В ПОЛУПРОВОДНИКАХ

2.1. Собственная проводимость

Рассмотрим кристалл идеально чистого четырехвалентного полупроводника, например, кремния. Кристаллическая решетка таких веществ построена на основе ковалентной (гомеополярной) связи; при этом, как известно, происходит обобществление валентных электронов соседних атомов.

При сообщении кристаллической решетке некоторой энергии, например за счет нагрева кристалла, отдельные электроны могут покинуть валентные связи и превратиться в свободные носители заряда. Величина этой энергии должна быть равной или большей, чем энергия связи валентного электрона. Так как изначально электрон был общим для двух атомов (Si), то нельзя сказать, что ионизирован один из этих атомов. Уход электрона приведет к частичной ионизации двух соседних атомов. Поэтому появляющийся при этом положительный единичный заряд, равный по абсолютной величине заряду электрона, относят не к тому или иному атому, а к дефектной связи, оставленной электроном. Такой положительный заряд принято

называть дыркой (рис.2.1).

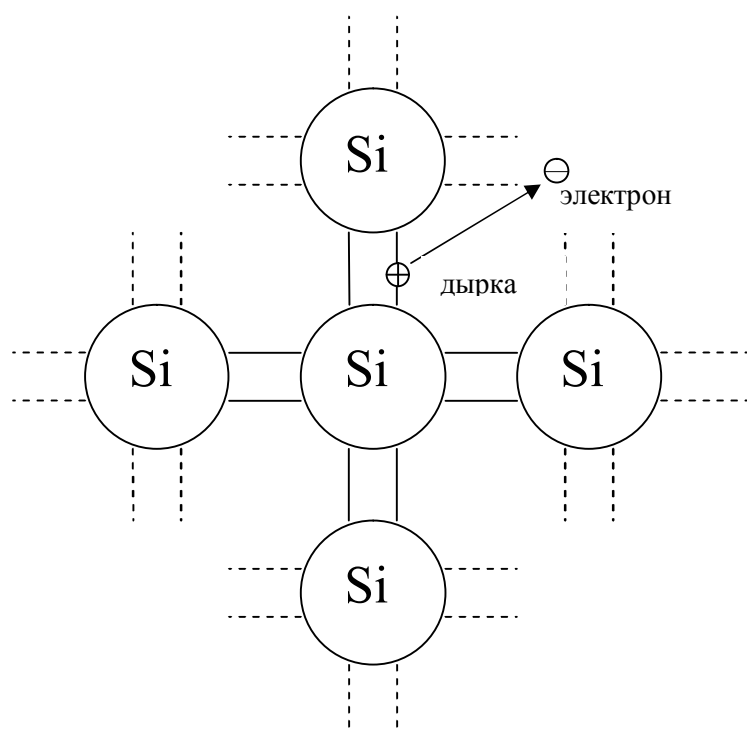


Рис. 2.1. Образование пары «электрон-дырка».

Итак, с уходом электрона в одной из валентных связей появляется вакансия, которую может замещать один из валентных электронов соседних связей. При таком переходе электрона

из заполненной связи в дефектную происходит перемещение дырки в обратном направлении. Этот процесс взаимного перемещения электронов и дырок носит случайный, беспорядочный характер. Однако если поместить кристалл полупроводника в электрическое поле, то более вероятными будут такие переходы электронов из связи в связь, при которых дырка (положительный заряд) передвигалась бы в направлении вектора напряженности электрического поля.

Таким образом, разрыв валентной связи за счет передачи энергии кристаллу извне приводит к появлению пары свободных носителей заряда – электрона и дырки. Электропроводность, возникающая в кристалле полупроводника за счет нарушения валентных связей, называется собственной электропроводностью. При этом концентрация подвижных носителей – электронов и дырок – одинакова.

Энергетическая диаграмма идеального собственного полупроводника по своей структуре точно такая же, как на рис. 1.4. Имеется лишь одно отличие – в ширине запрещенной зоны. Полупроводником считается (весьма условно!) вещество, имеющее ширину запрещенной зоны $\Delta E_z < 3-3,5$ эВ. Характерные значения величины ΔE_z для наиболее распространённых полупроводников приведены в таблице 2.1.

Таблица 2.1.

Полупроводник	Ширина запрещенной зоны, эВ
Германий (Ge)	0,66
Кремний (Si)	1,12
Арсенид галлия (GaAs)	1,42
Фосфид галлия (GaP)	2,27
Нитрид галлия (GaN)	3,39

2.2. Примесная проводимость. Полупроводник n-типа

Если вместо одного из атомов четырехвалентного полупроводника (Si, Ge) ввести атом вещества 5-й группы периодической системы элементов (Sb, P), то этот примесный атом образует только 4 ковалентные связи с четырьмя соседними атомами. Пятый электрон оказывается связанным со своим атомом значительно слабее. Для того чтобы оторвать его от атома и превратить в свободный носитель заряда, требуется значительно меньше энергии, чем для высвобождения электрона из валентной связи атома основного вещества. Такие примесные атомы называются донорами.

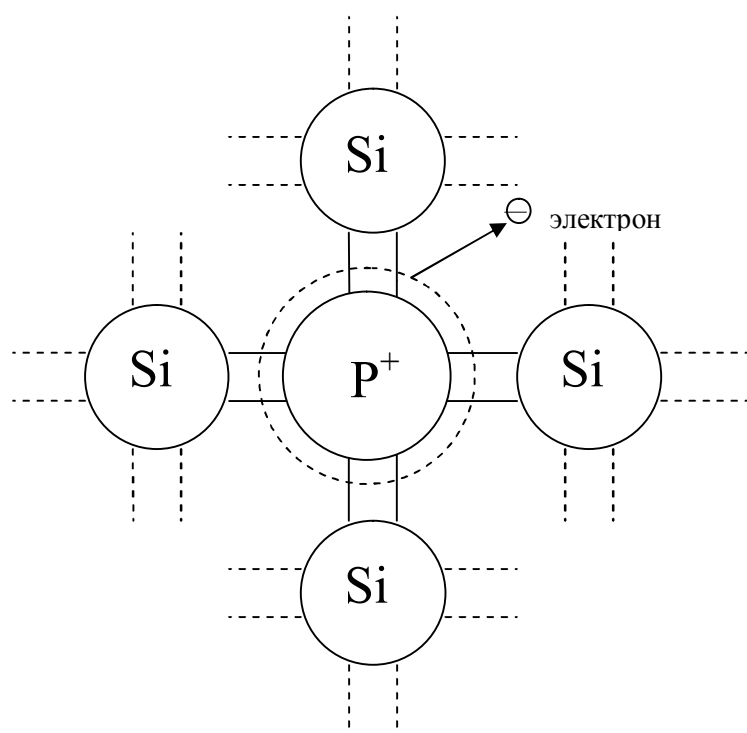


Рис. 2.2. Возникновение «свободного» электрона в полупроводнике n-типа за счет ионизации атома донорной примеси при повышении температуры. Штриховой круг около атома фосфора обозначает

локализацию 5-го электрона при 0 К.

В рамках зонной модели это значит, что для того, чтобы перевести такой электрон в зону проводимости, необходимо значительно меньшее количество энергии, чем ширина запрещенной зоны ΔE_z . Иными словами донорный уровень, на котором будут находиться такие электроны в невозбужденном состоянии, должен располагаться в запрещенной зоне вблизи дна зоны проводимости (рис.2.3). При температурах, близких к абсолютному нулю, все избыточные электроны находятся на этом уровне (т.е. «привязаны» к своим примесным атомам), но уже при незначительном повышении температуры они получают достаточное количество энергии для пере-

хода в зону проводимости. При этом атомы примеси фосфора превращаются в положительные ионы P^+ .

Увеличивая концентрацию примесных атомов, мы тем самым увеличиваем количество свободных электронов в полупроводнике. Количество дырок при этом не возрастает; оно определяется концентрацией собственных электронно-дырочных пар при данной температуре. В таком кристалле полупроводника ток в основном переносится электронами. Электроны в этом случае будут основными носителями, дырки – неосновными.

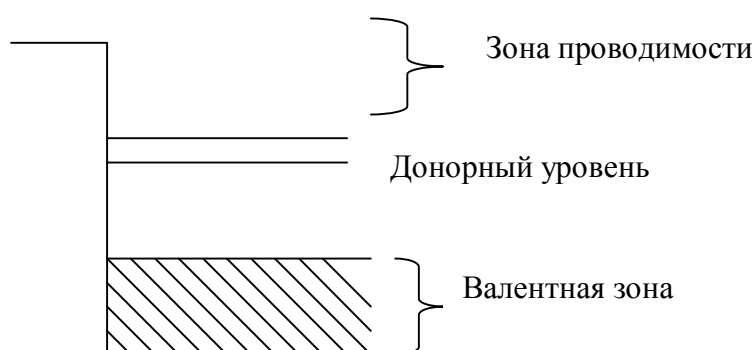


Рис.2.3. Энергетическая диаграмма донорного полупроводника при невысоких концентрациях примеси ($< 10^{18}$ атомов/см³).

Донорные примеси обеспечивают электронную проводимость полупроводника. Такие полупроводники называются донорными, электронными или полупроводниками n-типа.

2.3. Примесная проводимость. Полупроводник p-типа

Внешняя оболочка таких атомов содержит три валентных электрона. Например, в кристаллической решетке кремния трехвалентный атом бора (B) образует только три заполненные валентные связи. Четвертая связь остается незаполненной, но заряда она никакого не несет. Атом B и соседние атомы Si являются электрически нейтральными.

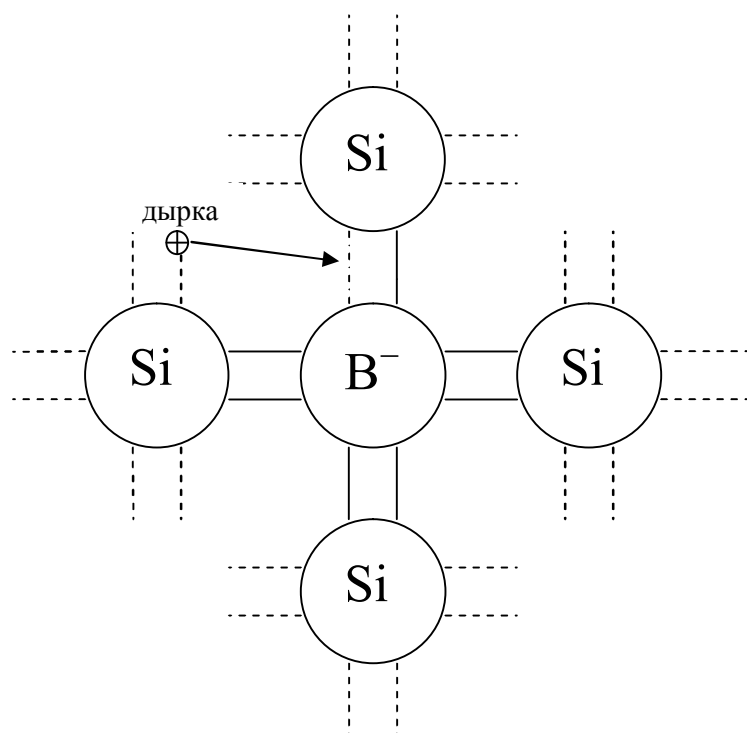


Рис. 2.4. Возникновение дырки в полупроводнике р- типа за счет ионизации трехвалентного атома примеси при повышении температуры. Стрелкой показан переход валентного электрона атома кремния к атому примеси, приводящий к его ионизации.

При небольшом тепловом возбуждении электрон с одной из соседних заполненных валентных связей (рис. 2.4) может перейти в эту связь. При этом во внешней оболочке атома В появляется лишний электрон; атом бора превращается в отрицательный ион, а в связи, откуда ушел электрон, появляется дырка, несущая (условно) эффективный положительный заряд $+e$. Такие примесные атомы называются акцепторами.

Вводя в Ge и Si примеси акцепторов – атомов III группы – мы получаем возможность повышать концентрацию только дырок. Такой полупроводник будет обладать, в основном, дырочной проводимостью. Дырки будут основными носителями, электроны – неосновными.

Так как переход электронов из валентных связей к атому примеси не требует больших энергий, сравнимых с ΔE_3 , то введение атомов трехвалентной примеси в решетку кремния или германия

приводит к появлению незанятого при $T=0$ К акцепторного уровня E_A в запрещенной зоне вблизи потолка валентной зоны (рис.2.5).

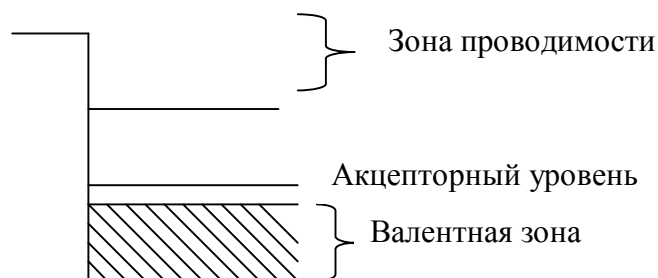


Рис. 2.5. Энергетическая диаграмма акцепторного полупроводника при невысоких концентрациях примеси ($< 10^{18}$ атомов/см³).

При температурах, близких к абсолютному нулю, акцепторный уровень остается свободным. При небольшом повышении температуры один из валентных электронов покидает валентную зону и занимает этот уровень, оставляя после себя в валентной зоне свободный уровень – «дырку».

Таким образом, акцепторные примеси, принимая валентные электроны на акцепторные уровни, приводят к появлению дырочной проводимости. Полупроводники, в которых основными носителями заряда являются дырки, называются акцепторными, дырочными, или полупроводниками p-типа.

Концентрация примесей в полупроводниках обычно невысока: 10^{-5} – 10^{-3} % (10^{15} – 10^{17} атомов/см³). Иначе говоря, при концентрации 10^{16} см⁻³ на 10^6 атомов вещества приходится один атом примеси. Это означает, что при равномерном распределении по объему полупроводника отдельные примесные атомы отделены друг от друга расстояниями в 10^6 – 10^8 постоянных решетки, т. е. их можно считать изолированными и не взаимодействующими друг с другом. Поэтому при таких концентрациях энергетические состояния всех атомов примеси характеризуются единым энергетическим уровнем.

В ряде случаев концентрацию примесей в полупроводниках доводят до 10^{-1} – 1% (10^{19} – 10^{20} атомов/см³). При этом примесные атомы образуют систему взаимодействующих частиц, и дискретный

энергетический уровень примеси расщепляется в примесную зону. При мелкозалегающих уровнях примеси эта зона перекрывается с соседней зоной (валентной зоной р-полупроводника или зоной проводимости полупроводника n-типа). При этом для подвижных носителей образуется непрерывная энергетическая диаграмма, характерная для металлов. Такие примесные полупроводники называются вырожденными; их электрические свойства приближаются к свойствам металлов (подробнее см. раздел 5.8).

2.4. Распределение электронов по энергиям в пределах энергетических зон

Для того, чтобы выяснить, каков характер распределения электронов по энергиям в пределах данной энергетической зоны, необходимо знать, во-первых, как распределены электроны по энергетическим уровням («по состояниям») в зоне, т. е. функцию плотности заполнения состояний электронами $\frac{dN}{dZ}$, и во-вторых, по какому закону эти уровни размещены внутри зоны, или, иначе говоря, какова функция плотности состояний $\frac{dZ}{dE}$. Перемножив эти два выражения, мы и получим функцию распределения электронов по энергиям:

$$\frac{dN}{dZ} \cdot \frac{dZ}{dE} = \frac{dN}{dE} = f(E) \quad (2.1)$$

Для начала обратимся к энергетической диаграмме металла. Она представляет собой так называемую «открытую зону», в которой отсутствует верхняя граница энергетических состояний электронов. Электроны в твердых телах представляют собой квантово-механическую систему, поэтому распределение таких электронов по состояниям подчиняется квантовой статистике Ферми-Дирака. Вероятность того, что уровень с данной энергией E при данной температуре T будет занят электроном, описывается функцией Ферми:

$$\frac{dN}{dZ} = F_{\text{кв}}(E) = \frac{1}{e^{\frac{E-E_F}{kT}} + 1} \quad (2.2)$$

Рассмотрим конкретный вид функции Ферми при $T = 0$ К. Для случая $E > E_F$ $F_{\text{кв}}(E) = 0$; при $E < E_F$ $F_{\text{кв}}(E) = 1$, наконец, при $E = E_F$ функция Ферми не определена. При $T > 0$ и $E = E_F$ функция определена и равна $F_{\text{кв}}(E) = \frac{1}{2}$.

График функции Ферми приведён на рис. 2.6.

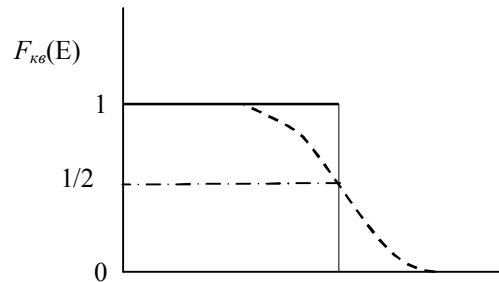


Рис. 2.6. Функция Ферми при $T = 0$ (сплошная кривая) и при $T > 0$ (штриховая линия).

Таким образом, при температуре абсолютного нуля электроны металла занимают все энергетические уровни в интервале $0 \leq E \leq E_F$. Уровни, расположенные выше E_F , при $T = 0$ свободны. Уровень, соответствующий энергии электрона, равной E_F , называется уровнем Ферми. Физический смысл уровня Ферми в металле ясен: это – наибольшее значение энергии, которую может иметь свободный электрон в металле при температуре абсолютного нуля.

При повышении температуры резкая граница распределения Ферми при $E = E_F$ «размывается»: вероятность нахождения электрона на уровнях, примыкающих к E_F со стороны бóльших энергий, становится отличной от нуля, и, соответственно, вероятность нахождения электрона на уровнях, примыкающих к E_F со стороны меньших энергий, становится меньше единицы. Поэтому, часть электронов с энергиями, меньшими E_F , получив энергию порядка kT , переходят на уровни энергии, лежащие выше уровня Ферми.

Теперь обратимся к функции плотности состояний $\frac{dZ}{dE}$. Требуется определить закон изменения числа энергетических состояний Z внутри малого интервала энергий ΔE в зависимости от расположения этого интервала на оси энергии. Эта задача решена в физике твердого тела:

$$S_{\text{кв}}(E) = \frac{dZ}{dE} = A\sqrt{|E - E_{zp}|}, \quad (2.3)$$

где $A = \frac{4\pi}{h^3}(2m^*)^{3/2}$, E_{zp} – величина, соответствующая началу отсчета энергии, h – постоянная Планка, m^* – эффективная масса электрона¹⁾.

Для металла $E_{zp} = 0$, и распределение электронов проводимости по энергиям описывается выражением:

$$\frac{dN}{dE} = \Phi_{\text{кв}}(E) = \frac{4\pi}{h^3}(2m^*)^{3/2} \frac{\sqrt{E}}{e^{\frac{E-E_F}{kT}} + 1} \quad (2.4)$$

График функции распределения электронов по энергиям в металле имеет вид, представленный на рис.2.7.

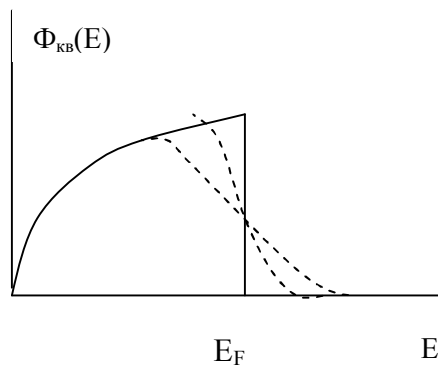


Рис. 2.7. Распределение электронов проводимости металла по энергиям при $T = 0$ (сплошная кривая) и при $T > 0$ (штриховые линии). С ростом температуры происходит размытие границы при $E = E_F$: «хвост» распределения увеличивается.

¹⁾Введение эффективной массы m^* вместо массы свободного электрона связано с тем, что электроны в кристалле находятся в периодическом поле кристаллической решётки. Электроны перемещаются в кристалле под действием двух сил: внешней силы \mathbf{F} , например, при наличии внешнего электрического поля, и сил, действующих со стороны решётки. Такое движение можно описать как движение свободного электрона под действием только внешней силы \mathbf{F} , и тогда взаимосвязь между приложенной внешней силой и ускорением записывается в форме второго закона Ньютона: $\mathbf{F} = m^* \cdot \mathbf{a}$, где m^* – эффективная масса электрона. Этот подход применяется в физике твёрдого тела также и для описания движения дырок в кристалле полупроводника.

В собственном полупроводнике рассматриваются состояния электронов в валентной зоне и в зоне проводимости. Валентная зона – «закрытая», т.е. имеет верхнюю и нижнюю границы; при $T = 0$ все уровни в ней заняты электронами. Зона проводимости является открытой, и при $T = 0$ полностью свободна. В закрытой и полностью заполненной зоне вид функции плотности состояний $\frac{dZ}{dE} = A\sqrt{|E - E_{cp}|}$ и, соответственно, распределение электронов по энергиям известны только на краях зоны вблизи от граничных значений энергии.

Положение уровня Ферми на энергетической диаграмме собственного полупроводника определяется из условия равенства концентраций электронов и дырок (см. ниже). Расчет показывает, что уровень Ферми и, соответственно, граница функции Ферми E_F в этом случае расположены посередине запрещённой зоны. Расположение уровня Ферми на энергетической диаграмме и изменения функций распределения $F_{кв}(E)$, $S_{кв}(E)$ и $\Phi_{кв}(E)$ с ростом температуры показаны на рис.2.8.

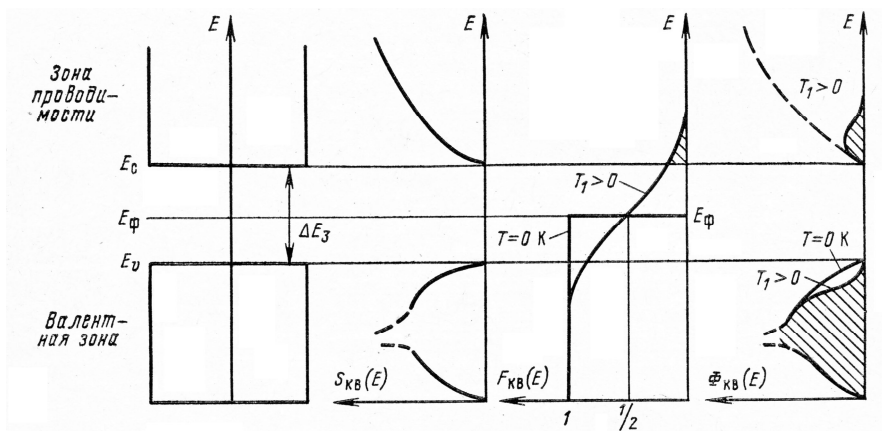


Рис. 2.8. Положение уровня Ферми и изменение функций распределения с температурой в собственном (беспримесном) полупроводнике.

Заштрихованные площади под графиками соответствуют занятым электронным состояниям. При $T > 0$ часть электронов, занимающих верхние энергетические уровни валентной зоны, переходят на уровни зоны проводимости (см. правую часть рисунка). Освобождённые в результате таких переходов состояния соответствуют появлению уровней дырок у потолка валентной зоны.

2.5. Равновесная концентрация свободных носителей в полупроводниках

Очевидно, что равновесную концентрацию электронов в зоне проводимости полупроводника n_0 можно определить как интеграл от функции $\Phi_{\text{кв}}(E)$. В этом случае $E_{zp}=E_c$, т.е. значение E_{zp} в формуле (2.3) соответствует энергии электрона у дна зоны проводимости.

$$N = n_0 = \int_{E_c}^{\infty} \frac{A\sqrt{E-E_c}}{e^{\frac{E-E_F}{kT}} + 1} dE \quad (2.5)$$

Этот интеграл в общем случае не имеет аналитического решения. Однако для собственных полупроводников, а также для примесных с относительно невысокой концентрацией доноров и акцепторов ($10^{16}-10^{18} \text{ см}^{-3}$), т. е. для невырожденных полупроводников, можно получить аналитическое выражение для n_0 .

Для таких полупроводников величина $E - E_F$ составляет как минимум десятые доли эВ, в то время как при комнатной температуре $kT = 0,026$ эВ (1 электронвольт (эВ) равен $1,6 \cdot 10^{-19}$ Дж). Поэтому единицей в знаменателе можно пренебречь, и тогда интеграл преобразуется к виду:

$$n_0 = \frac{4\pi}{h^3} (2m^*)^{\frac{3}{2}} \int_{E_c}^{\infty} \sqrt{E-E_c} \cdot e^{-\frac{E-E_F}{kT}} dE \quad (2.6).$$

Это упрощение означает, что электроны в зоне проводимости собственных и примесных невырожденных полупроводников с очень хорошим приближением подчиняются классической статистике Максвелла-Больцмана.

Введя новую переменную $x = \frac{E-E_c}{kT}$, получаем выражение, в котором имеется табличный интеграл, равный $\frac{\sqrt{\pi}}{2}$:

$$n_0 = \frac{4\pi}{h^3} (2m^*kT)^{\frac{3}{2}} \cdot e^{-\frac{E_c-E_F}{kT}} \int_0^{\infty} \sqrt{x} \cdot e^{-x} dx.$$

Окончательно, имеем:

$$n_0 = \frac{2}{h^3} (2\pi m^*kT)^{\frac{3}{2}} \cdot e^{-\frac{E_c-E_F}{kT}} = N_c e^{-\frac{E_c-E_F}{kT}} \quad (2.7)$$

Коэффициент при экспоненте $N_c = \frac{2}{h^3} (2\pi m_n^* kT)^{\frac{3}{2}}$ – эффективная плотность состояний в зоне проводимости. При условии $E_F \rightarrow E_C$ величина N_c – это максимально возможная концентрация электронов в зоне проводимости невырожденного полупроводника. Следовательно, экспонента $e^{-\frac{E_C - E_F}{kT}}$ имеет смысл вероятности нахождения свободных электронов в зоне проводимости.

Для дырок в валентной зоне справедливы те же упрощения, и их концентрация в невырожденных полупроводниках описывается выражением, аналогичным (2.7):

$$p_0 = N_V e^{-\frac{E_F - E_V}{kT}}; N_V = \frac{2}{h^3} (2\pi m_p^* kT)^{\frac{3}{2}} \quad (2.8)$$

Здесь E_V – уровень, соответствующий началу отсчета энергии для дырок в валентной зоне. N_V – эффективная плотность состояний в валентной зоне, т.е. максимально возможная концентрация дырок в валентной зоне невырожденного полупроводника, и, соответственно, экспоненциальный множитель имеет смысл вероятности нахождения дырок в валентной зоне.

В собственных и компенсированных полупроводниках число электронов проводимости равно числу дырок.

$$n_0 = p_0 = n_i, \text{ или } n_i^2 = n_0 \cdot p_0 \quad (2.9)$$

Подставив сюда выражения (2.7) и (2.8) и, учитывая, что $E_C - E_V = \Delta E_z$, получим:

$$n_i = \sqrt{n_0 p_0} = \sqrt{N_C N_V} \cdot e^{-\frac{\Delta E_z}{2kT}} \quad (2.10)$$

Отсюда видно, что, как и следовало ожидать, концентрация носителей в собственном полупроводнике сильно зависит от двух параметров: ширины запрещённой зоны ΔE_z и температуры, причём температурная зависимость определяется тем, что величина T стоит в знаменателе экспоненты. Собственная концентрация подвижных носителей в полупроводниках при температуре 300 К является табличной величиной; её значения n_i для некоторых полупроводников приведены в таблице 2.2.

Таблица 2.2.

Полупроводник	ΔE_z , эВ	n_i , см ⁻³
Ge	0,66	$2,4 \cdot 10^{13}$
Si	1,12	$1,4 \cdot 10^{10}$
GaAs	1,42	$1,8 \cdot 10^6$

Из данных таблицы ясно видна резкая зависимость величины собственной концентрации подвижных носителей от ширины запрещенной зоны полупроводника (т. е. от энергии связи валентных электронов).

2.6. Уровень Ферми на энергетической диаграмме собственного полупроводника

Концентрации дырок и электронов в собственном или компенсированном полупроводнике одинаковы: $n_o = p_o$:

$$N_C e^{\frac{E_{F_i} - E_C}{kT}} = N_V e^{\frac{E_V - E_{F_i}}{kT}} \quad (2.11);$$

откуда следует, что

$$E_{F_i} = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln \frac{N_V}{N_C} \quad (2.12),$$

или,

$$E_{F_i} = \frac{E_C + E_V}{2} + \frac{3}{4} kT \ln \frac{m_p^*}{m_e^*} \quad (2.13).$$

Это справедливо практически для любых температур, так как эффективные массы электронов проводимости и дырок примерно одинаковы: $m_p^* \sim m_e^*$. Таким образом,

$$E_{F_i} \cong \frac{E_V + E_C}{2} \quad (2.14)$$

Можно считать с хорошим приближением, что уровень Ферми в собственных полупроводниках при любой температуре находится посередине запрещенной зоны.

2.7. Взаимосвязь концентраций подвижных носителей и уровень Ферми в примесных полупроводниках

Для примесного невырожденного полупроводника, как следует из соотношений (2.7), (2.8) и (2.10), произведение равновесных концентраций основных и неосновных носителей оказывается всегда равным квадрату собственной концентрации носителей в полупроводнике без примесей. Введём следующие обозначения: n_n, p_p – концентрации основных носителей, n_p, p_n – концентрации неосновных носителей в n- и p-полупроводнике соответственно. Тогда:

$$n_n p_n = p_p n_p = n_i^2 \quad (2.15)$$

Это соотношение не зависит от концентрации примесей в полупроводнике.

Рассмотрим причину этого в рамках зонной модели. Возрастание концентрации основных носителей, например, дырок при увеличении количества акцепторов, приводит к тому, что увеличивается количество свободных энергетических уровней электронов у потолка валентной зоны. Следовательно, для появления в зоне проводимости неосновных носителей – электронов – необходимо сообщить связанным валентным электронам энергию, превышающую ширину запрещенной зоны в тем большей степени, чем выше уровень легирования p-полупроводника. Поэтому концентрация неосновных носителей при увеличении концентрации примесей (и, соответственно, основных носителей) будет пропорционально уменьшаться.

Для нахождения положения уровня Ферми в примесных полупроводниках необходимо решить уравнения, соответствующие условию равенства концентраций основных носителей (n_n, p_p) и равновесных концентраций (n_0, p_0). Это справедливо с очень хорошей точностью, так как вклад собственных носителей пренебрежимо мал. Формулы получаются очень громоздкими, но для наиболее важных случаев возможны упрощения. Так, для области низких температур, когда ещё не все примесные атомы превратились в ионы (см. разделы 2.2. и 2.3), получаются следующие формулы:

Для полупроводника n-типа:

$$E_{F_n} = \frac{E_C + E_D}{2} - \frac{kT}{2} \ln 2 \frac{N_C}{N_D} \quad (2.16)$$

Для полупроводника р-типа:

$$E_{F_p} = \frac{E_V + E_A}{2} + \frac{kT}{2} \ln 2 \frac{N_V}{N_A} \quad (2.17)$$

Из этих соотношений следует, что при температуре абсолютно нуля уровень Ферми в полупроводнике n-типа располагается посередине между дном зоны проводимости и донорным уровнем, а в р-полупроводнике – посередине между потолком валентной зоны и акцепторным уровнем:

$$E_{F_n} = \frac{E_C + E_D}{2}, \quad E_{F_p} = \frac{E_V + E_A}{2} \quad (T = 0 \text{ К}) \quad (2.18)$$

С ростом температуры уровень Ферми в n-полупроводнике смещается вниз. При комнатной температуре, когда все атомы донорной примеси ионизованы, он располагается в верхней части запрещенной зоны ниже донорного уровня. Полная ионизация доноров означает, что при комнатной температуре донорный уровень не занят электронами.

В р-полупроводнике при тех же условиях уровень Ферми поднимается вверх, и при комнатной температуре находится в нижней части запрещенной зоны выше акцепторного уровня. При этом все электроны, перешедшие к атомам акцепторов и превратившие их в отрицательные ионы, находятся на акцепторном уровне. Общий ха-

раактер изменения положения уровня Ферми показан на рис.2.9.

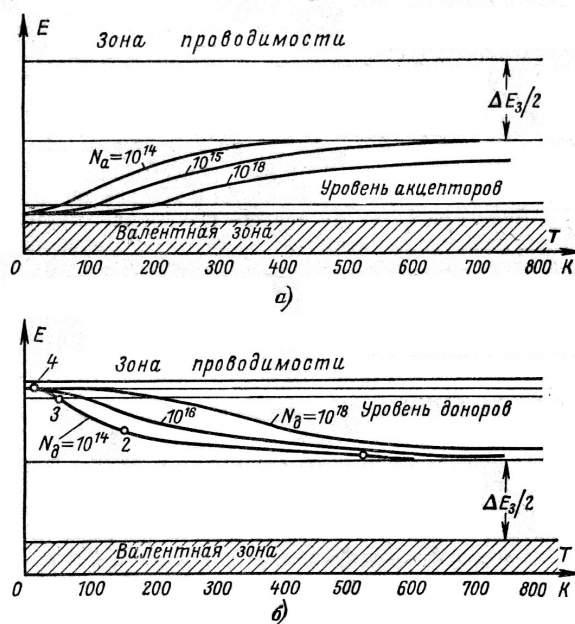


Рис. 2.9. Изменение положения уровня Ферми в акцепторном (а) и донорном (б) полупроводниках с ростом температуры для различных концентраций примесей.

При относительно высоких температурах (500-600 К) уровень Ферми в полупроводнике n-типа E_{F_n} снижается к середине запрещенной зоны, а уровень Ферми в полупроводнике р-типа E_{F_p} поднимается к середине запрещенной

зоны (см. рис. 2.9). Это связано с тем, что при таких условиях из-за резкого увеличения концентрации собственных электронно-дырочных пар в примесных полупроводниках (т. е. из-за возрастания скорости тепловой генерации) начинает преобладать собственная проводимость.

3. ДВИЖЕНИЕ НОСИТЕЛЕЙ ЗАРЯДА В ПОЛУПРОВОДНИКАХ

3.1. Хаотическое движение

В отсутствие электрического поля подвижные носители в кристалле находятся в хаотическом движении. Оно характеризуется тем, что направления перемещений частиц равновероятны, то есть направленное движение отсутствует. Распределение их по энергиям определяется законами квантовой статистики Ферми-Дирака (см. соотношение 2.4). В частном случае полупроводника с невысокой концентрацией примесей оно преобразуется в классическое энергетическое распределение Максвелла-Больцмана (2.6).

3.2. Дрейф подвижных носителей заряда. Дрейфовые токи

Направленное движение носителей заряда под действием электрического поля называется дрейфом. Электроны, получая ускорение в электрическом поле, приобретают на длине свободного пробега дополнительную кинетическую энергию. В результате, совокупность хаотически движущихся электронов будет перемещаться в направлении, противоположном вектору напряженности электрического поля $\vec{\varepsilon}$. Такое движение электронов в полупроводнике приводит к появлению дрейфового тока:

$$J_{n\varepsilon} = env_{cp} = en \frac{v_{cp}}{\varepsilon} \varepsilon = en\mu_n \varepsilon = \sigma_n \varepsilon \quad (3.1)$$

где σ_n - удельная проводимость, μ_n - подвижность электронов, т.е. величина, равная средней скорости их движения в электрическом поле с напряженностью, равной единице: ($\mu = \frac{v_{cp}}{\varepsilon}$).

Аналогичным образом появится дырочная составляющая дрейфового тока:

$$J_{p\varepsilon} = \sigma_p \varepsilon = ep\mu_p \varepsilon \quad (3.2)$$

Знаки в правых частях (3.1) и (3.2) одинаковы, т.к. противоположные направления движения разноименно заряженных частиц соответствуют току одного направления.

Полная плотность дрейфового тока при наличии свободных электронов и дырок равна:

$$J_{\varepsilon} = J_{n\varepsilon} + J_{p\varepsilon} = e(n\mu_n + p\mu_p)\varepsilon \quad (3.3)$$

3.3. Диффузия подвижных носителей заряда. Диффузионные токи

Диффузия – это направленное движение частиц (нейтральных или заряженных) из-за их неравномерного распределения по объему, т. е. из-за наличия градиента концентрации. Движущей силой диффузии является хаотическое тепловое движение частиц.

Поток частиц через плоскость раздела двух элементарных объемов с концентрациями N_1 и N_2

$$\Phi = -D \frac{dN}{dx}; \quad \frac{dN}{dx} = \frac{N_1 - N_2}{2\ell_{cp}} \quad (3.4)$$

Частицы взаимодействуют в среднем в пределах длины свободного пробега.

$$D = \frac{1}{3} v_{cp} \ell_{cp} \quad \text{- коэффициент диффузии}$$

Тогда плотности диффузионного тока электронов и дырок:

$$J_{nДиф} = eD_n \frac{dn}{dx}; \quad J_{pДиф} = -eD_p \frac{dp}{dx}. \quad (3.5)$$

Для полупроводника, содержащего свободные электроны и дырки, плотность суммарного диффузионного тока:

$$J_{Диф} = J_{nДиф} + J_{pДиф} = e \left(D_n \frac{dn}{dx} - D_p \frac{dp}{dx} \right) \quad (3.6)$$

Физической причиной этих токов является неравномерность распределения концентраций электронов и дырок по объёму полупроводника. Эти токи существуют в отсутствие внешнего электрического поля. Диффузионные токи электронов и дырок могут быть как противоположно направленными (при одинаковом направлении градиентов их концентраций), так и сонаправленными (если направления градиентов противоположны).

В большинстве полупроводниковых приборов величины токов обусловлены как дрейфовым, так и диффузионным движением носителей. Поэтому, плотность полного тока:

$$\vec{J} = \vec{J}_\varepsilon + \vec{J}_{\text{Диф}} \quad (3.7)$$

3.4. Взаимосвязь диффузии и дрейфа заряженных частиц

Диффузионное и дрейфовое движение заряженных частиц, и в частности электронов и дырок в полупроводнике, взаимосвязано. Диффузионные процессы выравнивания концентрации сопровождаются перераспределением зарядов, следовательно, при этом возникает электрическое поле $\vec{\varepsilon}$. Это вызовет дрейфовое движение частиц в противоположном направлении. В равновесном состоянии плотности этих потоков равны.

Рассматривая эту задачу, можно вывести формулу, связывающую коэффициент диффузии с подвижностью носителей (соотношение Эйнштейна):

$$D = \frac{\mu kT}{e} \quad (3.8)$$

Примером проявления такой взаимосвязи является полупроводник с неравномерным распределением примеси по объёму

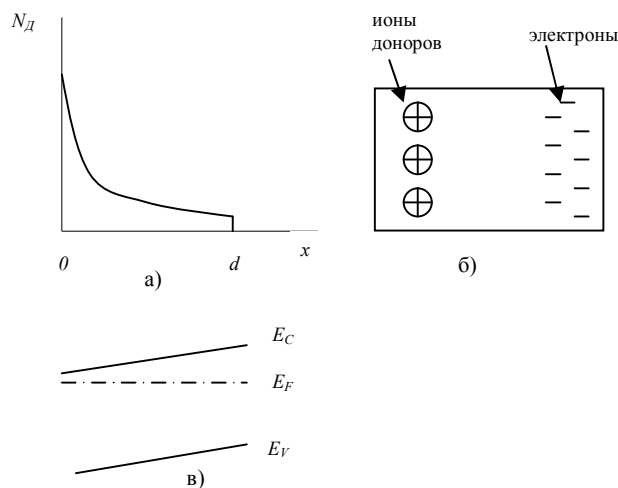


Рис.3.1. Полупроводник n-типа с неравномерным по объёму распределением примеси. а) изменение концентрации доноров по глубине; б) «встроенное» электрическое поле в полупроводнике, обусловленное разделением зарядов после установления равновесия;

в) энергетическая диаграмма полупроводника n-типа для такого случая.

Пусть в полупроводнике n-типа создано неравномерное по глубине распределение доноров, а значит и электронов проводимости (см. рис. 3.1, а). Тогда электроны начнут диффундировать в область с меньшей концентрацией донорной примеси, и в полупроводнике

произойдёт разделение зарядов. Область, куда пришли дополнительные электроны, приобретает избыточный отрицательный заряд, а на месте ушедших электронов остаётся некомпенсированный положительный заряд ионов доноров (рис. 3.1, б). Возникает диффузионное электрическое поле и встречный дрейфовый ток электронов. Когда суммарный ток окажется равным нулю, установится равновесное состояние; при этом в полупроводнике будет существовать так называемое «встроенное» электрическое поле.

Энергетическая диаграмма такого полупроводника показана на рис. 3.1, в. Она содержит наклонные зоны, отражающие наличие электрического поля в кристалле и горизонтальный уровень Ферми, свойственный равновесным системам. Подвижные носители, попадающие извне в такой полупроводник (например, за счет диффузии), будут ускоряться или замедляться в зависимости от знака их заряда.

3.5. Движение носителей в сильных электрических полях

В электрических полях с величиной напряженности $\varepsilon \leq 10^3 \text{ В/см}$ носители на длине свободного пробега приобретают относительно малую энергию. Поэтому их распределение по энергетическим уровням соответствует распределению при данной температуре кристаллической решетки. Иными словами дрейфовая скорость частиц в несильных электрических полях много меньше средней тепловой скорости: $(v_{др} \ll v_{cp} = \sqrt{\frac{3kT}{m}})$.

Следовательно, в слабых электрических полях подвижность не зависит от напряженности поля, т. е. полупроводник подчиняется закону Ома (см. начальный участок кривой на рис. 3.2).

При $\varepsilon > 10^3 \text{ В/см}$ средняя энергия носителей уже соответствует температуре, превышающей температуру кристаллической решетки. Число столкновений с атомами решетки увеличивается; это приводит к замедлению нарастания дрейфовой скорости и к уменьшению подвижности при увеличении напряженности электрического поля. Поэтому при $10^3 \text{ В/см} < \varepsilon < 10^4 \text{ В/см}$ полупроводник обладает свойством нелинейного сопротивления.

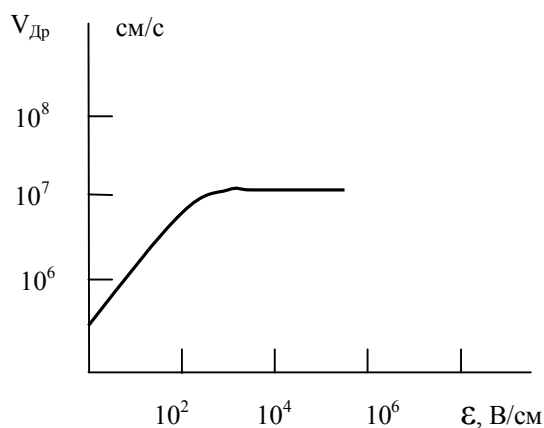


Рис.3.2. Зависимость дрейфовой скорости подвижных носителей от напряженности электрического поля.

При $\varepsilon > 10^4 \text{ В/см}$ дрейфовая скорость $v_{\text{др}}$ оказывается не зависящей от напряженности электрического поля ε . В полупроводнике

наблюдается насыщение тока. Напряженность поля, при котором начинается этот эффект, называется критической. Для большинства полупроводников $\varepsilon_{\text{кр}} \sim 10^4 \text{ В/см}$.

4. ГЕНЕРАЦИЯ И РЕКОМБИНАЦИЯ НОСИТЕЛЕЙ ЗАРЯДА

Генерацией называют процесс образования свободных носителей в полупроводниках при различных внешних воздействиях. Возникновение свободных носителей требует затраты энергии, необходимой для освобождения электронов из связанного состояния. Виды генерации определяются характером передачи энергии:

1. Тепловая генерация – при повышении температуры до некоторого критического значения.
2. Ударная генерация или ударная ионизация: энергия передается в виде кинетической энергии движущихся частиц.
3. Полевая генерация (или ионизация) – за счет энергии электрического поля.
4. Световая генерация – за счет передачи связанным электронам энергии фотонов при облучении светом.

4.1. Тепловая генерация

Процесс тепловой генерации электронно-дырочных пар в собственном полупроводнике и тепловой генерации основных носителей в примесных полупроводниках обсуждался в разделе 2. Рас-

смотрим более подробно, как происходит тепловая генерация подвижных носителей в полупроводнике n-типа.

На рис. 4.1 приведена температурная зависимость концентрации электронов проводимости. На кривой имеется три характерных участка, отражающих различные состояния донорного полупроводника.

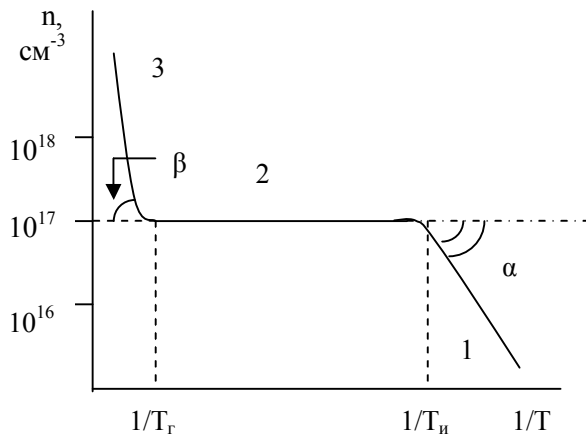


Рис. 4.1. Зависимость концентрации электронов полупроводника n-типа (в логарифмическом масштабе) от обратной температуры. Концентрация примеси равна 10^{17} см^{-3} .

Часть 1 графика соответствует области низких температур, при которых происходит ионизация атомов донорной примеси (см. раздел 2.2). Концентрация электронов растёт с повышением температуры до значения, равного концентрации доноров 10^{17} см^{-3} и затем на участке 2 остается неизменной. Этот переход происходит при температуре T_i – температуре истощения примесей. При достижении этой температуры все атомы доноров оказываются ионизованными.

В физике полупроводников выводится формула для концентрации электронов в полупроводнике n-типа в данном интервале температур:

$$n_n = \sqrt{\frac{N_C N_D}{2}} \cdot e^{-\frac{E_C - E_D}{kT}} \quad (4.1)$$

Таким образом, линейность участка 1 отражает экспоненциальную зависимость концентрации электронов проводимости, образовавшихся в результате ионизации доноров, от температуры. Тангенс угла наклона α этого прямолинейного участка к оси $1/T$ пропорционален энергии ионизации доноров $\Delta E_D = E_C - E_D$.

Горизонтальная часть кривой (2) соответствует интервалу рабочих температур полупроводника, в котором генерация собственных электронно-дырочных пар практически отсутствует, и концентра-

ция электронов остается постоянной. Температура T_g , выше которой начинается резкий рост концентрации электронов (участок 3), соответствует началу генерации собственных электронно-дырочных пар в полупроводнике. Тангенс β – угла наклона прямой на этом участке – пропорционален ширине запрещенной зоны (см. формулу 2.10).

4.2. Ударная генерация

Это явление наблюдается в полупроводнике при наличии сильных электрических полей ($\sim 10^5 - 10^6$ В/см). В полупроводниковых приборах такие поля наблюдаются в областях электронно-дырочных переходов или других неоднородностей. Электроны на длине свободного пробега разгоняются в электрическом поле и приобретают кинетическую энергию, достаточную для разрыва валентных связей атомов в решетке полупроводника. Если размеры кристалла значительно больше длины свободного пробега электрона, то произойдет лавинное размножение подвижных носителей: число вновь рожденных электронно-дырочных пар будет возрастать в геометрической прогрессии.

Для рассмотрения процесса ударной генерации в рамках зонной модели обратимся к рис. 4.2, а. При приложении к кристаллу полупроводника разности потенциалов края энергетических зон становятся наклонными – так же, как и в случае неравномерного распределения примеси по объёму (см. раздел 3.4). Зонная диаграмма строится в координатах $E - eU(x) = f(x)$, где x – расстояние от поверхности вглубь кристалла.

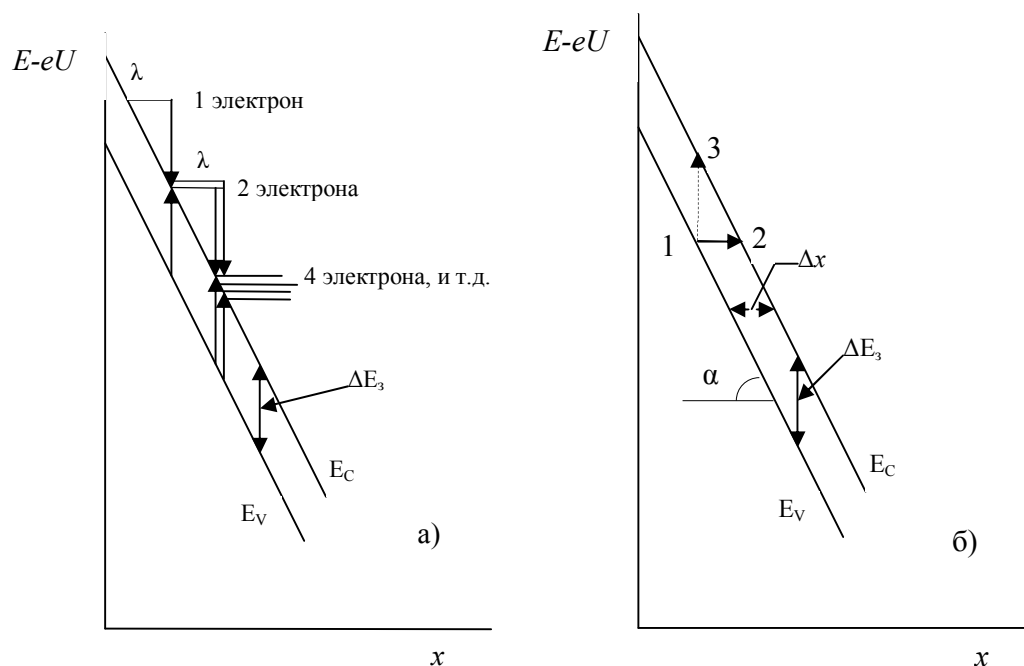


Рис. 4.2. Процесс ударной генерации (а) и полевой генерации (б) в полупроводнике в рамках зонной модели кристалла. λ – средняя длина свободного пробега электрона в кристалле полупроводника, Δx – ширина энергетического барьера для электронов потолка валентной зоны.

Пусть внешняя разность потенциалов достаточна для того, чтобы было выполнено условие для начала процесса ударной генерации. Тогда энергия электрона относительно дна зоны проводимости на длине свободного пробега λ будет возрастать (см. верхнюю часть рис 4.2, а), и в конце пути станет равной ширине запрещенной зоны. Обладая такой энергией, электрон может разорвать валентную связь атома полупроводника, при этом образуется новая электронно-дырочная пара. В результате, на дне зоны проводимости окажутся уже два электрона: один – потерявший энергию на разрыв валентной связи, и второй – появившийся в результате разрыва связи. Оба электрона, ускорившись на пути λ , так же потеряют энергию на разрыв двух связей, и на уровнях дна зоны проводимости окажутся теперь четыре электрона, и т. д.

4.3. Полевая генерация

Полевая генерация наблюдается в еще более сильных электрических полях ($> 10^6$ В/см), и принципиально отличается от ударной

генерации. В основе этого явления лежит квантовомеханический туннельный эффект.

Известно, что вероятность прохождения частицы через узкий потенциальный барьер без изменения своей энергии

$$P = e^{-\frac{4\pi}{h}\sqrt{2m^*\Delta E}\Delta x}$$

Здесь ΔE и Δx – высота и ширина потенциального барьера, m^* – эффективная масса электрона (см. раздел 2.3).

В нашем случае $\Delta E = \Delta E_3$, и, учитывая, что $\operatorname{tg} \alpha = \frac{e\Delta U}{\Delta x} = e\varepsilon = \frac{\Delta E_3}{\Delta x}$

(см. треугольник 1-2-3 рис.4.2, б), получим: $\Delta x = \frac{\Delta E_3}{e\varepsilon}$.

Таким образом, вероятность полевой генерации определяется следующим выражением:

$$P = e^{-\frac{4\pi}{he\varepsilon}\sqrt{2m^*(\Delta E_3)^3}} \quad (4.2).$$

В результате смещения энергетических зон (см. рис.4.2.,б) энергия электрона в точке 1 на потолке валентной зоны равна энергии электрона в точке 2 на дне зоны проводимости. Эти два состояния разделяет треугольный энергетический барьер с высотой ΔE_3 и шириной Δx . В очень сильных полях Δx мало и сравнимо с длиной волны электрона в точке 1. Если уровень 2 свободен, то в таких условиях возможен туннельный переход электрона из точки 1 в точку 2.

Вероятность туннельного перехода электронов из валентной зоны сквозь потенциальный барьер 1-3-2 в зону проводимости и обратно одинакова. Однако переход электронов из валентной зоны преобладает, поскольку их там значительно больше, чем в зоне проводимости. Поэтому и наблюдается увеличение концентрации носителей.

4.4. Световая генерация

Различают биполярную и однополярную световую генерацию. В первом случае энергия $h\nu$ поглощенного фотона расходуется на разрыв валентной связи собственных атомов полупроводника. Разрыв валентной связи может произойти лишь при условии, что энергия фотона превышает ширину запрещенной зоны: $h\nu > \Delta E_3$. При

этом образуются и электроны, и дырки, т. е. возникают пары разноименно заряженных подвижных частиц.

Однополярная генерация происходит при фотоионизации примесных атомов при низких температурах. Энергия поглощаемого фотона при этом должна быть больше энергии ионизации доноров ΔE_D для полупроводников n-типа или больше энергии ионизации акцепторов ΔE_A для полупроводников p-типа. Образуются соответственно либо только электроны, либо дырки.

Примеси с мелкозалегающими уровнями при комнатной температуре почти полностью ионизованы. Поэтому в этих условиях однополярная генерация может проявляться при наличии в полупроводнике чужеродных атомов, в которых энергия связи электронов относительно велика, и соответствующие энергетические уровни лежат в запрещенной зоне далеко от ее границ. Иными словами, в запрещенной зоне должны быть глубокозалегающие примесные уровни.

4.5. Рекомбинация подвижных носителей заряда. Механизмы рекомбинации

Рекомбинация – это процесс возвращения свободных носителей заряда в связанное состояние. Рекомбинация может происходить различными способами; при этом необходимо, чтобы одновременно выполнялись законы сохранения энергии и импульса. Тот или иной механизм рекомбинации реализуется в зависимости от того, каким образом расходуется энергия, выделившаяся при исчезновении электронно-дырочной пары.

Процесс рекомбинации может быть одноступенчатым и двухступенчатым. В первом случае происходит межзонная, или непосредственная рекомбинация: свободный электрон с энергетического уровня у дна зоны проводимости переходит на один из свободных верхних энергетических уровней валентной зоны; при этом выделяется энергия, равная ширине запрещенной зоны полупроводника. Это соответствует заполнению электроном проводимости ранее незанятой валентной связи атома полупроводника.

Если энергия выделяется в виде кванта электромагнитного (светового) излучения: $h\nu = \Delta E_z$, то происходит излучательная рекомбинация. Ясно, что закон сохранения энергии выполняется: ве-

личина выделившейся энергии $E_c - E_v$, равна энергии кванта светового излучения. Для выполнения закона сохранения импульса при излучательной рекомбинации необходимо, чтобы векторная сумма импульсов электрона и дырки была равна импульсу фотона. Так как его величина $\frac{h\nu}{c}$ очень мала (близка к нулю), то для выполнения закона сохранения импульса необходимо, чтобы электрон и дырка оказались в одной точке кристалла с равными по модулю и противоположно направленными импульсами. Вероятность такого события, и, следовательно, вероятность межзонной излучательной рекомбинации невелика; в лучшем случае она достигает десятых долей процента.

Если выделившаяся при заполнении электроном валентной связи энергия передается решетке кристалла и идёт на увеличение амплитуды тепловых колебаний атомов полупроводника, возникает безызлучательная межзонная рекомбинация. Как известно из физики твёрдого тела, кристаллическая решетка может воспринимать энергию лишь малыми порциями порядка нескольких kT , т. е. нескольких сотых долей эВ. Величина выделившейся при рекомбинации энергии ΔE_z для различных полупроводников, по крайней мере, на порядок больше, и поэтому вероятность межзонной безызлучательной рекомбинации также мала.

Основной механизм рекомбинации подвижных носителей в полупроводниках с относительно широкой запрещенной зоной (например, в арсениде галлия) – рекомбинация через ловушки. Этот процесс протекает в два этапа. В примесных полупроводниках последовательность переходов зависит от типа проводимости. Например, в полупроводнике n-типа уровни ловушек, как правило, находятся ниже уровня Ферми в запрещенной зоне и заняты электронами. Поэтому первым шагом рекомбинации является переход электрона с уровня ловушки на энергетический уровень валентной зоны вблизи её потолка (рис. 4.3, а, переход 1). Иначе говоря, происходит захват дырки – неосновного носителя – уровнем ловушки. На втором этапе электрон проводимости переходит на освободившийся уровень ловушки.

В полупроводнике p-типа уровни ловушек, расположенные выше уровня Ферми, в основном, свободны, и первым этапом ре-

комбинации является захват электрона (неосновного носителя) ловушкой (рис. 4.3, б). Вторая стадия рекомбинации – переход электрона с энергетического уровня ловушки на свободный уровень в валентной зоне.

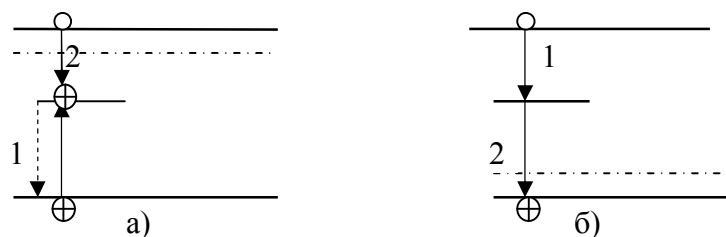


Рис. 4.3. Рекомбинация через ловушки в примесных полупроводниках.

а) для полупроводника n-типа, б) для полупроводника p-типа.

Как видно, первый этап рекомбинации в примесных полупроводниках – захват неосновного носителя атомом-ловушкой. Этот этап определяет преобладающую часть времени жизни неосновных носителей: их концентрация на много порядков меньше концентрации основных носителей, и столкновения с атомами ловушек происходят относительно редко.

На втором этапе рекомбинации происходит захват основного носителя. В n-полупроводнике на уровень ловушки захватывается электрон (рис. 4.3,а, стрелка 2), в полупроводнике p-типа на уровень ловушки захватывается дырка (стрелка 2 на рис. 4.3, б). Это – относительно быстрый процесс, так как концентрация основных носителей велика.

Роль рекомбинационных ловушек могут играть специальные примесные атомы или ионы (Cu, Au, Cd, S, N₂, O₂), а также различные несовершенства кристалла полупроводника: незаполненные узлы решетки, трещины и т.п.

Поверхностная рекомбинация – разновидность рекомбинации через ловушки. Из-за большого количества дефектов на поверхности полупроводника этот процесс идет значительно интенсивнее. Скорость поверхностной рекомбинации сильно зависит от качества обработки поверхности, т. е. от числа дефектов на единице площади. Например, для шлифованной поверхности кремния скорость рекомбинации в несколько тысяч раз больше, чем для поверхности после химического травления.

4.6. Диффузионная длина неосновных носителей в примесных полупроводниках

Этот важный параметр можно определить как среднюю длину пути, который проходит неосновной носитель за время своей жизни, т. е. от момента возникновения за счет генерации до момента исчезновения при рекомбинации. Для нахождения математической формулы, определяющей диффузионную длину, необходимо решить уравнение диффузии для стационарного случая, когда концентрация неосновных носителей в любой точке кристалла не изменяется со временем.

Пусть на поверхности полупроводника р-типа или на границе раздела с другим веществом (координата $x = 0$) создается постоянная во времени избыточная концентрация неосновных носителей $\Delta n_p(0)$. Тогда в результате диффузии их концентрация в толще полупроводника при $x > 0$ будет больше равновесной. При удалении от поверхности вглубь кристалла избыточные неосновные носители будут рекомбинировать, их концентрация будет уменьшаться, и в предельном случае при $x \rightarrow \infty$ $\Delta n_p(x) \rightarrow 0$ и $n_p \rightarrow n_{p0}$.

Уравнение, описывающее диффузионное движение электронов в полупроводнике р-типа с учетом рекомбинации имеет вид:

$$D_n \frac{d^2 \Delta n_p}{dx^2} - \frac{\Delta n_p}{\tau_n} = 0 \quad (4.3)$$

Здесь D_n – коэффициент диффузии электронов, x – координата вглубь кристалла, τ_n – время жизни неосновных носителей в полупроводнике р-типа. Решая данное уравнение при указанных выше граничных условиях, получаем:

$$\Delta n_p(x) = \Delta n_p(0) \cdot e^{-\frac{x}{\sqrt{D_n \cdot \tau_n}}}; \quad n_p(x) = n_{p0} + \Delta n_p(0) \cdot e^{-\frac{x}{\sqrt{D_n \cdot \tau_n}}} \quad (4.4)$$

Знаменатель показателя экспоненты имеет размерность длины, и величина

$$L_n = \sqrt{D_n \cdot \tau_n} \quad (4.5)$$

называется диффузионной длиной неосновных носителей – электронов в полупроводнике р-типа. Аналогично рассматривается и величина $L_p = \sqrt{D_p \cdot \tau_p}$ – диффузионная длина дырок в n-полупроводнике. Из (4.4) следует, что диффузионная длина – это расстояние

от границы вглубь кристалла, на котором концентрация диффундирующих неосновных носителей уменьшается в e раз.

4.7. Время жизни неосновных носителей в полупроводнике. Уравнение генерации-рекомбинации

Время жизни неосновных носителей – это средний промежуток времени от момента их появления в кристалле за счет какого-либо механизма генерации до рекомбинации. Иначе говоря, это – средне-статистическое время существования их в свободном состоянии. Генерация и рекомбинация подвижных носителей (основных и неосновных) происходят в полупроводнике непрерывно, поэтому скорость изменения их концентрации определяется разностью скоростей этих процессов. Наибольший интерес с точки зрения степени влияния на работу полупроводниковых приборов представляют неосновные носители.

Рассмотрим полупроводник p -типа. Изменение концентрации неосновных носителей при внешнем воздействии (например, при резком нагреве, при облучении светом) можно в грубом приближении описать следующим уравнением:

$$\frac{d\Delta n_p}{dt} = g - r \quad (4.6)$$

Здесь g – скорость генерации, r – скорость рекомбинации, т. е. число неравновесных неосновных носителей, возникающих при генерации и исчезающих при рекомбинации за 1 секунду в единице объёма ($\text{см}^{-3} \cdot \text{с}^{-1}$).

Скорость рекомбинации можно представить в виде:

$$r = \frac{\Delta n_p}{\tau_n} \quad (4.7),$$

где Δn_p – избыточная концентрация, а τ_n – время жизни неосновных носителей в полупроводнике p -типа. После прекращения внешнего воздействия начинается процесс установления равновесия: концентрация неравновесных носителей уменьшается за счет рекомбинации.

Если в некоторый момент времени внешнее воздействие на полупроводник прекратится, то $g = 0$ и уравнение (4.6) примет вид:

$$\frac{d\Delta n_p}{dt} = -\frac{\Delta n_p}{\tau_n} \quad (4.8)$$

Решив это уравнение с учетом начального условия (при $t = 0$ $\Delta n_p = \Delta n_p(0)$), получим:

$$\Delta n_p(t) = \Delta n_p(0) \cdot e^{-\frac{t}{\tau_n}} \quad (4.9)$$

Видно, что по прошествии промежутка времени τ_n концентрация избыточных неосновных носителей уменьшается за счет рекомбинации в e раз. Таким образом, время жизни неосновных носителей является одним из важнейших параметров, определяющих скорость убывания их концентрации, или – как принято говорить – скорость их рассасывания.

5. КОНТАКТНЫЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

При идеальном контакте двух проводящих тел на границе возникают переходные слои различной природы. Контактные явления в полупроводниках определяют принцип действия всех активных элементов полупроводниковой электроники: диодов, полевых и биполярных транзисторов, тиристоров. Переходные слои на границах раздела, которые существуют в полупроводниковых активных элементах, возникают на контактах различного типа:

- 1) контакты между полупроводниками одной природы с разными или одинаковыми типами примесных атомов (р-п переходы, переходы $n-n^+$, $p-p^+$);
- 2) контакты между полупроводниками различной природы (гетеропереходы),
- 3) контакты металл – полупроводник n - или p -типа;
- 4) контакты металл-диэлектрик-полупроводник.

В полупроводниковой электронике такие контакты часто называют электрическими переходами, так как при установлении равновесного состояния на границе раздела веществ возникает электрическое поле.

5.1. Электронно-дырочный переход (р-п переход)

Р-п переход – это переходный слой между двумя областями полупроводника с разным типом электропроводности, в котором существует электрическое поле. Такой переход нельзя создать пу-

тем механического контакта двух кристаллов полупроводника, так как поверхности всегда загрязнены чужеродными атомами, окислами и т.п. Переходы создаются с помощью специальных технологических приемов, например, методом диффузии примесей или методом сплавления.

Переходы различаются по двум основным параметрам:

- по характеру распределения концентрации примесей у границы раздела,
- по соотношению концентраций примесных атомов в смежных полупроводниковых областях.

В первом случае характерными величинами являются: ширина области изменения концентрации примеси Δx у границы раздела и ширина перехода l . Если $\Delta x \ll l$, переход называется резким, а при условии $\Delta x \sim l$ образуется плавный р-п переход.

Если концентрации примесных атомов в р- и n-областях одинаковы, то переход называется симметричным. Несимметричный р-п переход образуется полупроводниками с различными концентрациями примесей и, следовательно, с различными концентрациями основных и неосновных носителей. Если концентрации основных носителей отличаются более чем на порядок, то переход называется односторонним. Резкому и плавному симметричному переходу могут соответствовать распределения концентрации примесей, показанные на рис. 5.1.

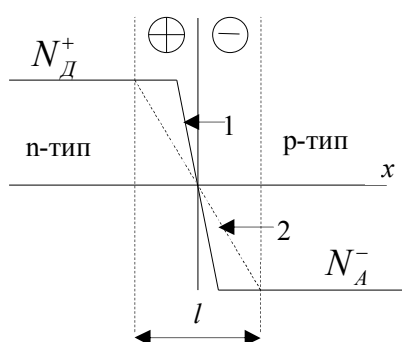


Рис. 5.1. Одно из возможных распределений концентраций примесей в резком (1) и плавном (2) р-п переходе. l – ширина области пространственного заряда ионов примесей (ширина перехода).

5.2. Р-п переход в равновесном состоянии

В этом разделе рассматривается идеализированный р-п переход, для которого действуют следующие ограничения:

- Концентрация примесей в смежных областях не превышает 10^{18} см^{-3} , т. е. переход изготовлен на основе невырожденных полупроводников.
- Токи малы и не вызывают заметного падения напряжения в толще полупроводника. Вследствие этого электрическое поле сосредоточено только в области р-п перехода.
- Омические контакты к р- и п-областям идеальны, т.е. около них в полупроводнике существует всегда равновесная концентрация носителей заряда.
- В области перехода не происходят процессы генерации и рекомбинации носителей.
- Поверхностные явления несущественны.

Установление равновесного состояния на р-п переходе происходит следующим образом. При идеальном контакте двух кристаллов одного и того же полупроводника с различным типом электропроводности из-за градиента концентрации основных носителей возникает их диффузия через плоскость раздела. Дырки из р-области, примыкающей к границе, начинают диффундировать в п-область; диффузия электронов протекает в обратном направлении. В результате нарушается условие электронейтральности в областях, примыкающих к границе. В р-области после ухода из нее дырок у границы раздела остаются некомпенсированные отрицательные ионы акцепторов, а в п-области – положительные ионы доноров. Образуется область объемного заряда, состоящая из двух разноименно заряженных слоев, в которой возникает электрическое поле ε_k , и между п- и р-областями устанавливается контактная разность потенциалов φ_k . Потенциал п-области становится положительным по отношению к потенциалу р-области.

Возникшее электрическое поле ε_k создаёт потенциальный барьер на пути диффундирующих основных носителей, и скорость их перемещения в смежную область уменьшается. Вместе с тем под действием ε_k возникает дрейфовое движение неосновных носителей заряда. Дырки движутся из п-области в р-область; электроны – в противоположном направлении.

Таким образом, в процессе установления равновесия на переходе через границу проходят встречные потоки одноименно заряженных частиц, т.е. текущие навстречу друг другу токи:

$$J_{Диф} = J_{nДиф} + J_{pДиф} \quad (\text{основные носители})$$

$$J_{Др} = J_{pДр} + J_{nДр} \quad (\text{неосновные носители})$$

Равновесие на переходе установится при условии, когда напряженность контактного электрического поля ε_k достигнет такой величины, при которой диффузионный ток основных носителей полностью компенсируется встречным дрейфовым током неосновных носителей. Полный ток через переход при этом становится равным нулю:

$$J = J_{Диф} + J_{Др} = 0$$

Переход обеднен подвижными носителями зарядов, поэтому сопротивление его значительно выше сопротивления объемов полупроводника, лежащих за пределами слоя толщиной l .

Энергетическая диаграмма симметричного р-п перехода, находящегося в равновесии, приведена на рис. 5.2. За пределами области объемного заряда двойного слоя ионов примесей электрическое поле отсутствует. Поэтому энергетические зоны тут горизонтальны. В пределах ширины перехода энергетические зоны имеют наклон, отражающий величину напряженности контактного электрического поля. Так как потенциал n-области положителен по отношению к потенциалу р-области, то на диаграмме энергетические зоны для n-области расположены ниже, чем для р-области.

Поскольку в равновесном состоянии уровень Ферми должен быть единым для всей системы, высота установившегося потенциального барьера $e\varphi_k$, препятствующего диффузии основных носителей, равна разности энергий Ферми р- и n- областей. Таким образом:

$$e\varphi_k = E_{F_n} - E_{F_p} \quad (5.1)$$

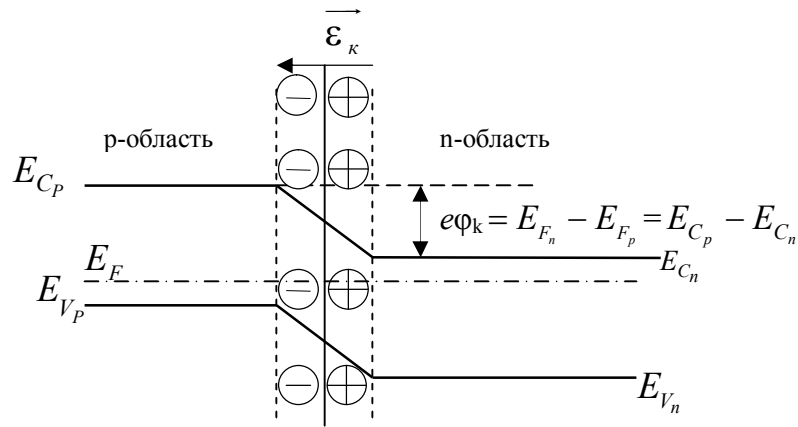


Рис. 5.2. Энергетическая диаграмма идеального симметричного электронно-дырочного перехода в состоянии равновесия

Будем считать, что объемные заряды в переходе обусловлены только ионизованными атомами доноров и акцепторов, и на основе этой идеальной модели рассмотрим основные физические параметры резкого p-n перехода в равновесном состоянии.

Для нахождения зависимости высоты потенциального барьера от концентрации основных и неосновных носителей воспользуемся соотношением (2.7) для концентрации электронов. В полупроводнике n-типа для основных носителей имеем:

$$n_n = N_C e^{-\frac{E_{C_n} - E_F}{kT}} \quad (5.2)$$

Концентрация электронов – неосновных носителей в p-полупроводнике – равна:

$$n_p = N_C e^{-\frac{E_{C_p} - E_F}{kT}} \quad (5.3)$$

Из рис. 5.2 видно, что потенциальный барьер для диффузии электронов в p-область равен разности уровней дна зоны проводимости в p- и n-областях:

$$e\phi_k = E_{C_p} - E_{C_n}; \text{ тогда } E_{C_p} = e\phi_k + E_{C_n} \quad (5.4)$$

Подставляя (5.4) в выражение (5.3), получим:

$$n_p = N_C e^{-\frac{E_{C_n} + e\phi_k - E_F}{kT}} = N_C e^{-\frac{E_{C_n} - E_F}{kT}} \cdot e^{-\frac{e\phi_k}{kT}} = n_n e^{-\frac{e\phi_k}{kT}} \quad (5.5)$$

Окончательно имеем:

$$e\phi_k = kT \ln \frac{n_n}{n_p} \quad (5.6)$$

Аналогичным образом можно выразить $e\varphi_k$ через концентрации дырок:

$$e\varphi_k = kT \ln \frac{p_p}{p_n} \quad (5.7)$$

Оценим величину потенциального барьера $e\varphi_k$ для кремниевого несимметричного р-п перехода при комнатной температуре ($T = 300$ К). Пусть концентрация дырок в р-области равна 10^{15} см^{-3} , тогда концентрация неосновных носителей-электронов n_p в соответствии с формулой (2.15) и таблицей (2.1) будет равна 10^5 см^{-3} . Концентрацию основных носителей n_n в полупроводнике n-типа выберем равной 10^{17} см^{-3} . Тогда в соответствии с формулой (5.6) получим:

$$e\varphi_k[\text{эВ}] = 2,6 \cdot 10^{-2} \cdot 2,3 \cdot \lg \frac{10^{17}}{10^5} \approx 0,72 \text{ эВ}$$

Такие же расчеты для германиевого р-п перехода с $N_A = N_D = 10^{16} \text{ см}^{-3}$ дают величину $e\varphi_k \approx 0,32 \text{ эВ}$.

Для нахождения контактного электрического поля ε_k необходимо решить уравнение Пуассона для конкретного характера распределения заряда и данных граничных условий. Будем считать, что в пределах перехода имеются постоянные концентрации ионов примесей: $+eN_D$ при изменении координаты от $-l_n$ до 0 и $-eN_A$ на отрезке $[0; l_p]$ (см. рис 5.3, а).

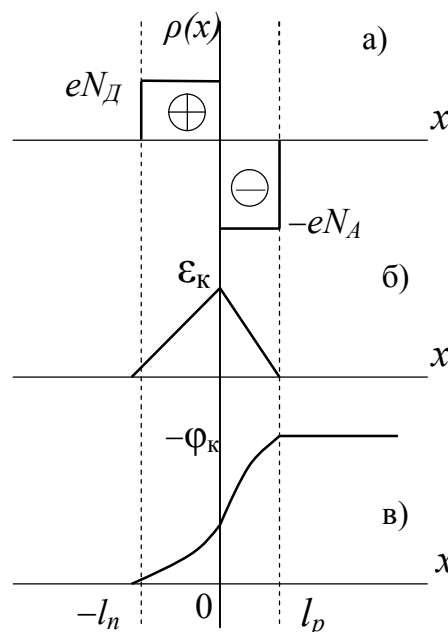


Рис. 5.3. Изменение параметров идеального резкого р-п перехода в пределах его ширины: а) заряда $\rho(x)$, б) величины контактного электрического поля ε_k ; в) потенциала φ_k .

В такой модели необходимо решить одномерное уравнение Пуассона для двух интервалов: $-l_n \leq x \leq 0$ и $0 \leq x \leq l_p$:

$$\frac{d^2\varphi}{dx^2} = -\frac{\rho(x)}{\varepsilon \cdot \varepsilon_0} \quad (5.8)$$

$$1. -l_n \leq x \leq 0: \frac{d^2\varphi}{dx^2} = -\frac{eN_D}{\varepsilon \cdot \varepsilon_0} \quad 2. 0 \leq x \leq l_p: \frac{d^2\varphi}{dx^2} = \frac{eN_A}{\varepsilon \cdot \varepsilon_0}$$

Граничное условие:

$$\left. \frac{d\varphi}{dx} \right|_{x=-l_n} = 0 \quad (5.9)$$

$$\frac{d\varphi}{dx} = -\frac{eN_D}{\varepsilon \cdot \varepsilon_0} x + C$$

Граничное условие:

$$\left. \frac{d\varphi}{dx} \right|_{x=l_p} = 0 \quad (5.10)$$

$$\frac{d\varphi}{dx} = \frac{eN_A}{\varepsilon \cdot \varepsilon_0} x + C$$

Константы интегрирования определяются из условий (5.9) и (5.10):

$$\frac{d\varphi}{dx} = -\frac{eN_D}{\varepsilon \cdot \varepsilon_0} (x + l_n)$$

$$\frac{d\varphi}{dx} = \frac{eN_A}{\varepsilon \cdot \varepsilon_0} (x - l_p)$$

Как известно, электрическое поле равно взятому с обратным знаком градиенту потенциала. В одномерном случае $\varepsilon_k = -\frac{d\varphi}{dx}$, и зависимости величины контактного электрического поля от координаты в рассматриваемых интервалах будут выражаться следующим образом:

$$-l_n \leq x \leq 0: \varepsilon_{k1} = \frac{eN_D}{\varepsilon \cdot \varepsilon_0} (x + l_n) \quad (5.11)$$

$$0 \leq x \leq l_p: \varepsilon_{k2} = -\frac{eN_A}{\varepsilon \cdot \varepsilon_0} (x - l_p) \quad (5.12)$$

Изменение величины контактного электрического поля ε_k в пределах перехода показано на рис. 5.3, б. Максимальная величина ε_k наблюдается на границе между р- и п-областями, т. е. при $x = 0$:

$$\varepsilon_{k \text{ макс}} = \varepsilon_{k1}(0) = \varepsilon_{k2}(0) = \frac{eN_D}{\varepsilon \cdot \varepsilon_0} l_n = \frac{eN_A}{\varepsilon \cdot \varepsilon_0} l_p \quad (5.13)$$

Отсюда вытекает очень важное соотношение:

$$\frac{l_n}{l_p} = \frac{N_A}{N_D} \quad (5.14)$$

Видно, что глубина проникновения контактного поля в полупроводник определяется степенью его легирования: область перехода проникает в полупроводник на тем большую глубину, чем меньше в нём концентрация примеси.

Для нахождения зависимости ширины перехода l от разности потенциалов на переходе и концентрации примесей в смежных областях необходимо вычислить второй интеграл уравнения (5.6):

$$\varphi_k = \int_{-l_n}^0 \left(\frac{d\varphi}{dx} \right) dx + \int_0^{l_p} \left(\frac{d\varphi}{dx} \right) dx$$

Эта величина равна площади под графиком $\varepsilon_k(x)$ на рис. 5.3,б:

$$\begin{aligned} \varphi_k &= \frac{1}{2} \varepsilon_{k \max} \cdot l_n + \frac{1}{2} \varepsilon_{k \max} \cdot l_p = \frac{1}{2} \frac{eN_D}{\varepsilon\varepsilon_0} l_n^2 + \frac{1}{2} \frac{eN_A}{\varepsilon\varepsilon_0} l_p^2 = \\ &= \frac{e}{2\varepsilon\varepsilon_0} (N_D l_n^2 + N_A l_p^2) \end{aligned} \quad (5.15)$$

Учитывая соотношение (5.14), выразим l_n и l_p через полную ширину перехода l :

$$l = l_n + l_p = l_n + l_n \frac{N_D}{N_A} = \frac{N_A + N_D}{N_A} l_n;$$

Отсюда:
$$l_n = \frac{N_A}{N_A + N_D} l \quad (5.16)$$

Аналогично:
$$l_p = \frac{N_D}{N_A + N_D} l \quad (5.17)$$

Подставляя (5.16) и (5.17) в выражение (5.15), получим:

$$\varphi_k = \frac{el^2}{2\varepsilon\varepsilon_0} \cdot \frac{N_D N_A}{N_D + N_A} \quad (5.18)$$

Таким образом, ширина равновесного р-п перехода выражается формулой:

$$l = \sqrt{\frac{2\varepsilon\varepsilon_0 \varphi_k}{e} \cdot \frac{N_D + N_A}{N_D \cdot N_A}} \quad (5.19)$$

Для односторонних переходов получаются более простые выражения:

$$N_A \gg N_D: l \approx l_n \approx \sqrt{\frac{2\varepsilon\varepsilon_0 \varphi_k}{eN_D}}; \quad (5.20)$$

$$N_D \gg N_A: l \approx l_p \approx \sqrt{\frac{2\varepsilon\varepsilon_0 \varphi_k}{eN_A}} \quad (5.21)$$

Из формул (5.19) и (5.20, 5.21) следует, что ширина р-п перехода определяется главным образом величинами концентраций при-

месей в смежных областях: при повышении уровня легирования ширина перехода уменьшается. Величина контактной разности потенциалов на переходе также оказывает влияние на его ширину. Увеличение φ_k происходит при использовании полупроводников с более широкой запрещенной зоной. Это можно легко увидеть, построив соответствующие энергетические диаграммы р-п переходов. Ширина р-п перехода при этом также возрастает. Для плавного р-п перехода зависимость его ширины от разности потенциалов на переходе выражена слабее: $l \sim \sqrt[3]{\varphi_k}$.

Проведем оценку ширины резкого одностороннего кремниевого перехода. Пусть уровень легирования п-области по крайней мере на порядок меньше, чем в р-области, и $N_D = 10^{16} \text{ см}^{-3}$, $\varphi_k \approx 0,7 \text{ В}$. Относительная диэлектрическая проницаемость кремния имеет величину $\varepsilon \approx 10$. Подставляя эти величины в формулу (5.20), получим:

$$l \approx l_n \approx \sqrt{\frac{2 \cdot 10 \cdot 8,85 \cdot 10^{-12} \cdot 7,2 \cdot 10^{-1}}{1,6 \cdot 10^{-19} \cdot 10^{22}}} \approx \frac{2\sqrt{2}}{10^7} \approx 2,8 \cdot 10^{-7} \text{ (м)} = 0,28 \text{ мкм}$$

Отсюда следует, что контактная разность потенциалов 0,72 В приходится на расстояние 0,28 мкм, т. е. величина напряженности контактного электрического поля на таком переходе составляет $\sim 2,5 \cdot 10^6 \text{ В/м}$. (!).

5.3. Прямое включение р-п перехода

Пусть к р-п переходу подключен источник напряжения U_{np} плюсом к р-области, а минусом к п-области, т. е. полярностью, противоположной контактной разности потенциалов φ_k . Сопротивление перехода гораздо больше сопротивления объемов р- и п-областей, поэтому практически всё внешнее напряжение оказывается приложенным к переходу. В результате потенциальный барьер для диффузии основных носителей уменьшается до величины $e(\varphi_k - U_{np})$, равновесное состояние нарушается, и через переход течет диффузионный ток основных носителей: дырок из р-области в п-область и электронов – из п-области в р-область.

Диффузионное движение основных носителей через р-п переход в смежную область, где они становятся неосновными, называется инжекцией неосновных носителей заряда.

С уменьшением барьера уменьшится ширина перехода:

$$l_{np} = \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k - U_{np})}{e} \cdot \frac{N_D + N_A}{N_D \cdot N_A}} \quad (5.22)$$

Соответственно, для односторонних переходов:

$$l_{np} \approx l_n \approx \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k - U_{np})}{eN_D}}; l_{np} \approx l_p \approx \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k - U_{np})}{eN_A}} \quad (5.23)$$

Введем понятие уровня инжекции:

$$\delta = \frac{\Delta p_n}{n_{n_o}} = \frac{\Delta n_p}{p_{p_o}} \quad (5.24)$$

Уровень инжекции равен отношению приращения концентрации неосновных носителей к равновесной концентрации основных носителей заряда.

Низким уровнем инжекции ($\delta \ll 1$) будем считать такой, при котором концентрация инжектированных неосновных носителей намного меньше равновесной концентрации основных носителей. Средний уровень: $\delta \sim 1$ (разница между Δn , Δp_n и p_{p_o} , n_{n_o} мала). Высокий уровень инжекции: $\delta \gg 1$; при этом концентрация инжектированных неосновных носителей намного больше равновесной концентрации основных носителей.

На энергетической диаграмме прямо смещённого перехода (рис. 5.4) уровень Ферми р-области располагается ниже уровня Ферми n-области, разность между ними равна eU_{np} . Стрелками показаны направления диффузионного движения электронов и дырок.

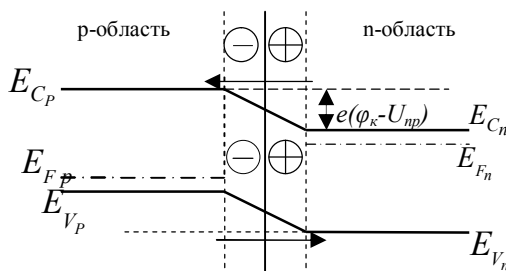


Рис. 5.4. Энергетическая диаграмма р-п перехода, смещённого в прямом направлении

В результате инжекции у границ перехода образуются избыточные концентрации неосновных носителей. Дырки диффундируют от границы вглубь n-области, постепенно рекомбинируя с электронами; такие же процессы будут происходить с участием электронов в р-области. На границе возникают следующие условия: за одинаковые промежутки времени число приходящих на границу основных носителей равно числу уходящих от границы вглубь р- и n- областей. Их избыточные

концентрации на границе поддерживаются постоянными. Поэтому (см. раздел 4.6), если размеры областей значительно больше диффузионных длин для неосновных носителей, то концентрации инжектированных неосновных носителей будут уменьшаться по экспоненциальному закону при удалении от границы вглубь p- и n- областей (рис. 5.5.)

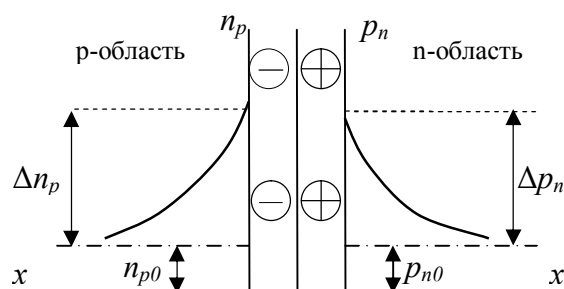


Рис.5.5. Распределение концентраций инжектированных неосновных носителей по глубине p- и n-областей при прямом смещении симметричного перехода

5.4. Обратное включение p-n перехода

Пусть к переходу подключён внешний источник напряжения $U_{обр}$ минусом к p-области, а плюсом к n-области, т.е. полярностью, совпадающей с контактной разностью потенциалов φ_k . При этом равновесие на переходе также нарушается: потенциальный барьер для диффузии основных носителей возрастает до величины $e(\varphi_k + U_{обр})$ (см. рис.5.6.). Поскольку величина $U_{обр}$ составляет обычно десятки-сотни вольт, диффузия основных носителей через переход прекращается, и их ток уменьшается до нуля. Вместе с тем условия для протекания через переход дрейфового тока неосновных носителей становятся ещё более благоприятными, так как ускоряющая их разность потенциалов увеличилась на $U_{обр}$.

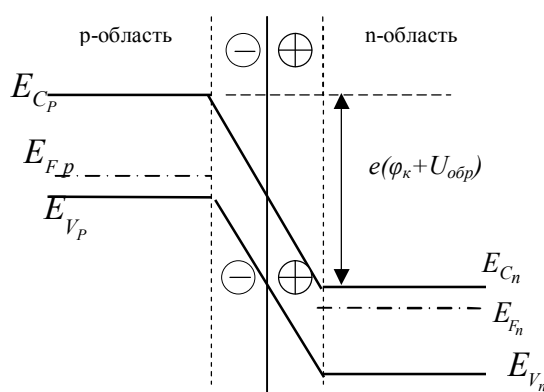


Рис.5.6. Энергетическая диаграмма p-n перехода, смещённого в обратном направлении

Таким образом, при обратном смещении через р-п переход протекает дрейфовый ток неосновных носителей заряда: дырок из п-области в р-область и электронов из р-области в п-область. Этот ток намного меньше прямого диффузионного тока из-за того, что концентрация неосновных носителей очень мала по сравнению с концентрацией основных носителей. Например, для симметричного кремниевого перехода с концентрацией основных носителей в смежных областях $\sim 10^{16} \text{ см}^{-3}$ концентрация неосновных носителей будет $\sim 10^4 \text{ см}^{-3}$ (см. формулу 2.15).

При повышении потенциального барьера ширина перехода увеличится:

$$l_{обп} = \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k + U_{обп})}{e} \cdot \frac{N_A + N_D}{N_A \cdot N_D}} \quad (5.25)$$

Соответственно, для односторонних переходов:

$$l_{обп} \approx l_n \approx \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k + U_{обп})}{eN_D}}; l_{обп} \approx l_p \approx \sqrt{\frac{2\varepsilon\varepsilon_o(\varphi_k + U_{обп})}{eN_A}} \quad (5.26)$$

Втягивание неосновных носителей полем р-п перехода в смежные области, где они становятся основными, называется экстракцией. Из-за экстракции неосновных носителей из областей, прилегающих к границам перехода, их концентрации здесь значительно уменьшатся. Оценим величину граничных концентраций неосновных носителей при обратном смещении перехода.

В случае обратного смещения в формуле (5.6) для контактной разности потенциалов равновесного перехода необходимо заменить φ_k на $\varphi_k + U_{обп}$. Концентрация основных носителей на границе будет такой же, как в объёме (n_n), а концентрация неосновных носителей снизится до значения $n_p(0)$. Следовательно, для обратного смещения:

$$e(\varphi_k + U_{обп}) = kT \ln \frac{n_n}{n_p(0)} \quad (5.27)$$

Отсюда после преобразований получим: $n_p(0) = n_n \cdot e^{-\frac{e\varphi_k}{kT}} \cdot e^{-\frac{eU_{обп}}{kT}}$.

Учитывая, что первые два множителя дают выражение для концентрации неосновных носителей в объёме р-области (см. формулу (5.5)) имеем:

$$n_p(0) = n_p \cdot e^{-\frac{eU_{обр}}{KT}} \quad (5.28)$$

Величины обратных напряжений на переходе могут составлять десятки и сотни вольт. Таким образом, граничные концентрации неосновных носителей при обратном смещении перехода снижаются до значений, близких к нулю; при удалении от границ перехода вглубь р- и n-областей значения n_p и p_n приближаются к равновесным значениям n_{p0} и p_{n0} (рис. 5.7).

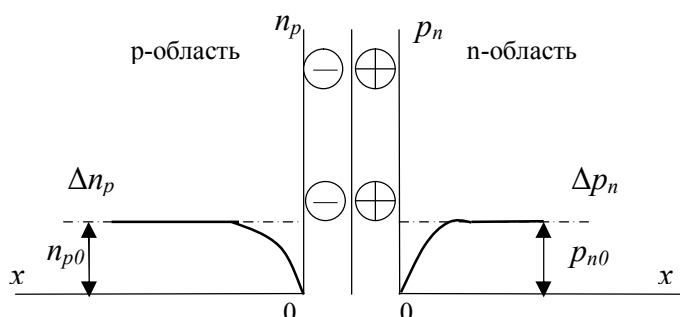


Рис. 5.7. Распределение концентраций неосновных носителей по глубине р- и n-областей при обратном смещении симметричного перехода

Таким образом, вблизи границ перехода появляются градиенты концентраций неосновных носителей, и число уходящих в смежные области неосновных носителей за счёт экстракции оказывается равным числу приходящих из объема к границе перехода за счет диффузии.

С увеличением обратного напряжения р-n переход расширяется, разность потенциалов на нем возрастает, но значение тока через переход остается постоянным. Это объясняется тем, что градиенты концентрации неосновных носителей у границ перехода практически не изменяются, т.к. параметры диффузии (n_p, p_n, L_p, L_n) не зависят от приложенного внешнего напряжения. Поэтому ток, текущий через идеальный переход при обратном включении за счет экстракции, называется обратным током насыщения или тепловым током I_0 . Величина I_0 может измениться лишь при изменении температуры перехода за счет увеличения или уменьшения концентраций неосновных носителей.

5.5. Вольтамперная характеристика идеализированного р-п перехода

Для облегчения расчетов выберем идеальную одномерную модель перехода, (см. раздел 5.1), что позволит пренебречь падением напряжения в толще р- и п-областей и не учитывать процессы генерации и рекомбинации в пределах самого перехода.

В случае прямого смещения через переход протекают диффузионные токи основных носителей. Рассмотрим электронную составляющую прямого тока – диффузию электронов из п-области в р-область. Ось x в пределах р-области направим от границы перехода внутрь. В обозначениях формул $e = |e| = 1,6 \cdot 10^{-19}$ Кл. Тогда:

$$I_n = -e \cdot S \cdot D_n \left. \frac{dn_p}{dx} \right|_{x=0} \quad (5.29)$$

При инжекции в р-области возникает избыточная концентрация неосновных носителей (см. формулу (4.4)). Поэтому производную от концентрации по координате в (5.29) можно выразить следующим образом:

$$\left. \frac{dn_p}{dx} \right|_{x=0} = -\frac{\Delta n_p(0)}{L_n}$$

Следовательно, диффузионный ток электронов при прямом смещении равен:

$$I_n = e \cdot S \cdot D_n \frac{\Delta n_p(0)}{L_n} \quad (5.30)$$

Необходимо выяснить, каким образом избыточная концентрация неосновных носителей $\Delta n_p(0)$ зависит от приложенного внешнего прямого напряжения. Воспользуемся выведенной выше формулой (5.6) для контактной разности потенциалов на переходе. При прямом смещении перехода концентрация основных носителей на границе в п-области будет такой же, как в объеме (n_{n_0}), а концентрация неосновных носителей увеличится до значения $n_{p0} + \Delta n_p(0)$. Следовательно, при прямом смещении пониженный потенциальный барьер для диффузии электронов равен:

$$e\varphi_k - eU = kT \ln \frac{n_{n_0}}{n_{p0} + \Delta n_p(0)} \quad (5.31)$$

Преобразуем это выражение: $-\frac{e\varphi_{\kappa}}{kT} + \frac{eU}{kT} = \ln \frac{n_{p_0} + \Delta n_p(0)}{n_{n_0}}$.

Переходя к соотношению для экспонент, получим:

$$n_{n_0} e^{-\frac{e\varphi_{\kappa}}{kT}} \cdot e^{\frac{eU}{kT}} = n_{p_0} + \Delta n_p(0).$$

Отсюда следует зависимость избыточной концентрации неосновных носителей – электронов в р-области – от прямого напряжения:

$$\Delta n_p(0) = n_{p_0} \left(e^{\frac{eU}{kT}} - 1 \right) \quad (5.32)$$

Аналогично, зависимость избыточной концентрации дырок в п-области от прямого напряжения:

$$\Delta p_n(0) = p_{n_0} \left(e^{\frac{eU}{kT}} - 1 \right) \quad (5.32, a)$$

Подставляя (5.32) в (5.29) получим зависимость электронной составляющей прямого тока р-п перехода с площадью S от приложенного напряжения:

$$I_n = e \cdot S \cdot \frac{D_n}{L_n} n_{p_0} \left(e^{\frac{eU}{kT}} - 1 \right) \quad (5.33)$$

Проведя те же действия для дырок – неосновных носителей в р-области – можно получить зависимость дырочной составляющей прямого тока от приложенного напряжения:

$$I_p = e \cdot S \cdot \frac{D_p}{L_p} p_{n_0} \left(e^{\frac{eU}{kT}} - 1 \right) \quad (5.34)$$

Так как прямой ток через переход – это сумма диффузионных токов основных носителей, то формула вольтамперной характеристики для прямого включения р-п перехода имеет вид:

$$I = e \cdot S \left(\frac{D_n}{L_n} n_{p_0} + \frac{D_p}{L_p} p_{n_0} \right) \left(e^{\frac{eU}{kT}} - 1 \right) = I_0 \left(e^{\frac{eU}{kT}} - 1 \right) \quad (5.35)$$

При обратном смещении через переход протекает малый дрейфовый ток неосновных носителей в противоположном направлении; его величина поддерживается практически постоянной за счет диффузии неосновных носителей из объёма областей к границам перехода (см. раздел 5.4). Поэтому очевидно, что движущей силой, поддерживающей обратный ток, является диффузия, т. е. тот же меха-

низм, на основе которого выведена формула (5.35). Не выполняя соответствующих весьма объёмных математических преобразований, отметим только их конечный результат: формула (5.35) является также и вольтамперной характеристикой для обратного включения р-п перехода. Общий вид вольтамперной характеристики показан на рис. 5.8.

Рассмотрим формулу вольтамперной характеристики р-п перехода (5.35) и оценим величину показателя экспоненты при прямом смещении. Величины прямых напряжений составляют от десятых долей до единиц Вольт; для комнатных температур величина $kT \approx 2,6 \cdot 10^{-2}$ эВ. Следовательно, единицей в скобке можно пренебречь, и вольтамперная характеристика для прямого включения перехода хорошо описывается следующим выражением:

$$I_{np} = I_0 \cdot e^{\frac{eU}{kT}} \quad (5.36)$$

При обратном смещении напряжение отрицательно и обычно составляет десятки-сотни Вольт. Поэтому в (5.35) можно пренебречь величиной экспоненты по сравнению с единицей, и обратный ток идеализированного р-п перехода равен по величине тепловому току I_0 :

$$I_{обп} = -I_0 \quad (5.37)$$

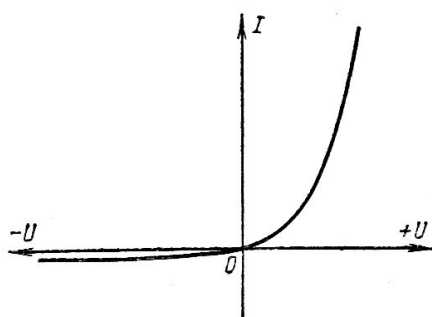


Рис. 5.8. Вольтамперная характеристика идеализированного р-п перехода. Ток при обратном смещении очень мал и не зависит от напряжения

Таким образом, обратный ток идеального перехода описывается формулой:

$$I_0 = e \cdot S \left(\frac{D_n}{L_n} n_{p0} + \frac{D_p}{L_p} p_{n0} \right) \quad (5.38)$$

Вольтамперная характеристика р-п перехода зависит от температуры р-п перехода и материала, из которого он изготовлен (см.

рис. 5.9). Видно, что обратный ток переходов резко возрастает с увеличением температуры, и его величина при данном обратном напряжении для кремниевого перехода значительно меньше, чем для германиевого. Прямой ток при этом также растёт, однако в процентном отношении его прирост гораздо меньше, чем прирост обратного тока. Видно также, что прямая ветвь вольтамперной характеристики кремниевого диода сдвинута в область более высоких прямых напряжений.

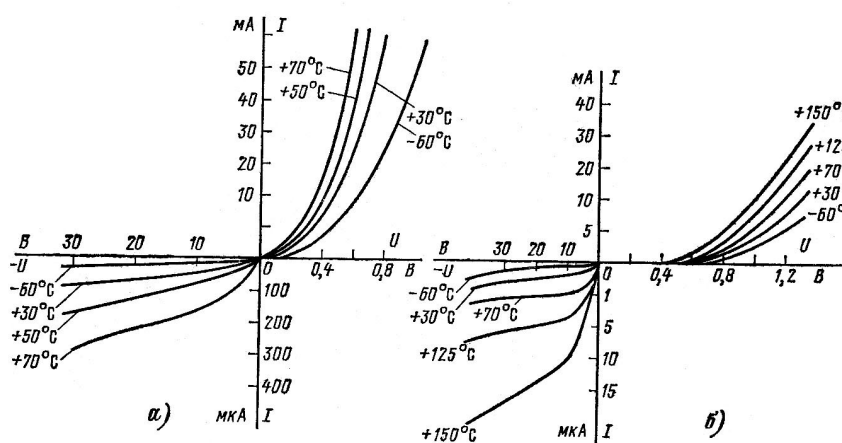


Рис.5.9. Зависимость вольтамперных характеристик германиевого (а) и кремниевого (б) диодов от температуры

Физическую причину этих зависимостей можно выяснить, преобразовав выражение (5.38) с использованием соотношения (2.15):

$$n_{p_0} = \frac{n_i^2}{p_{p_0}}; \quad p_{n_0} = \frac{n_i^2}{n_{n_0}}$$

Подставив эти величины в (5.38), и учитывая, что собственная концентрация подвижных носителей n_i зависит от температуры и ширины запрещённой зоны полупроводника (2.10), получим:

$$I_0 = e \cdot S \cdot N_c N_v \left(\frac{D_n}{L_n p_{p_0}} + \frac{D_p}{L_p n_{n_0}} \right) e^{-\frac{\Delta E_g}{kT}} \quad (5.39)$$

В полученной формуле (5.39) явным образом прослеживается зависимость обратного тока идеального перехода и от температуры, и от разновидности полупроводникового материала. Наблюдается экспоненциальный рост обратного тока с увеличением температуры и такое же резкое уменьшение его величины при увеличении шири-

ны запрещённой зоны. Это – прямое следствие того, что величина обратного тока определяется тепловой генерацией неосновных носителей заряда в смежных областях р-п перехода.

При рассмотрении этих же зависимостей для тока при прямом включении перехода можно впасть в ошибку при использовании формулы (5.36): из неё на первый взгляд следует, что прямой ток должен уменьшаться с ростом температуры, что не соответствует реальному поведению вольтамперных характеристик. Это противоречие возникает из-за того, что не учитывается зависимость I_0 от температуры (5.39).

С учётом этого факта зависимость прямого тока перехода от температуры имеет вид:

$$I_{np} = e \cdot S \cdot N_C N_V \left(\frac{D_n}{L_n p_{p_0}} + \frac{D_p}{L_p n_{n_0}} \right) e^{-\frac{\Delta E_3}{kT}} \cdot e^{\frac{eU}{kT}} = C \cdot e^{-\frac{\Delta E_3 - eU}{kT}} \quad (5.40)$$

Так как величины прямых напряжений таковы, что eU всегда меньше ширины запрещённой зоны ΔE_3 , то показатель экспоненты в формуле (5.40) отрицателен, и прямой ток увеличивается с ростом температуры. Физической причиной этого является прирост концентрации основных носителей за счёт тепловой генерации.

Увеличение прямого тока происходит также из-за уменьшения высоты потенциального барьера р-п перехода для их диффузии при повышении температуры. Преобразуем формулу (5.6) для потенциального барьера с учётом соотношений (2.10) и (2.15):

$$e\phi_k = kT \ln \frac{n_{n_0}}{n_{p_0}} = kT \ln \frac{n_{n_0} p_{p_0}}{n_i^2} = kT \ln(n_{n_0} p_{p_0}) - kT \ln N_C N_V - kT \ln \left(e^{-\frac{\Delta E_3}{kT}} \right)$$

Отсюда получаем зависимость величины потенциального барьера от температуры:

$$e\phi_k = \Delta E_3 - kT \ln \frac{N_C N_V}{n_{n_0} p_{p_0}} \quad (5.41)$$

Видно, что потенциальный барьер для диффузии основных носителей уменьшается с ростом температуры, что приводит к росту прямого тока через переход.

При использовании полупроводников с более широкой запрещённой зоной потенциальный барьер перехода $e\phi_k$ возрастает. Это приводит к уменьшению прямого тока при данном прямом напряжении. В результате вольтамперная характеристика перехода на ос-

нове полупроводника с более широкой запрещённой зоной сдвигается вправо, в область более высоких прямых напряжений. Так, например, резкое увеличение прямого тока для германиевого р-п перехода ($\Delta E_g = 0,66$ эВ) начинается при $U \approx 0,28$ В, а для кремниевого перехода ($\Delta E_g = 1,12$ эВ) – при $U \approx 0,7$ В.

5.6. Реальный электронно-дырочный переход. Обратное смещение. Токи генерации и утечки

Реальные р-п переходы, как правило, несимметричные, и в подавляющем большинстве случаев их можно считать односторонними. Полупроводник с большей концентрацией примеси, и соответственно, с более высокой концентрацией основных носителей называется эмиттером, второй полупроводник – с меньшей концентрацией основных носителей – базой.

При рассмотрении физических процессов в реальном р-п переходе необходимо учитывать все факторы, которые остались «за бортом» при анализе идеализированного перехода, а именно:

- процессы генерации электронно-дырочных пар и их рекомбинации в пределах ширины перехода;
- падение напряжения на р- и п-областях при протекании прямого тока;
- поверхностные явления, в частности, токи утечки при больших обратных напряжениях.

Будем считать, что вблизи омических контактов в р- и п-полупроводниках всегда существует равновесная концентрация носителей заряда, и, следовательно, влияние этих контактов на свойства р-п перехода несущественно. Выясним, как изменится вольт-амперная характеристика перехода при учёте указанных выше факторов.

Тепловая генерация электронно-дырочных пар происходит не только в р- и п-областях, но и в самом р-п переходе. Образовавшиеся электроны и дырки разделяются электрическим полем перехода и перебрасываются в области, где они становятся основными носителями. Возникающий при этом дрейфовый ток называется током генерации; он совпадает по направлению с обратным током насыщения I_0 идеализированного перехода.

Следовательно, ток генерации определяется полным числом носителей, образовавшихся в р-п переходе в единицу времени. Число подвижных носителей, возникающих в единице объёма в единицу времени, т.е. скорость генерации – это отношение n_i/τ , где τ – время жизни носителей в области перехода, n_i – собственная концентрация подвижных носителей. Генерация происходит в объёме р-п перехода $S \cdot l$ (S – площадь перехода, l – его ширина). Тогда ток генерации можно представить в следующем виде:

$$I_G = eSl \frac{n_i}{\tau} \quad (5.42)$$

Отсюда следует, что ток генерации увеличивается с ростом обратного напряжения из-за расширения перехода (см. формулы 5.26).

Оценим долю тока генерации в полном обратном токе реального перехода. Для этого рассмотрим отношение тока генерации к тепловому току I_0 (5.39). Это отношение оказывается пропорциональным собственной концентрации носителей:

$$\frac{I_G}{I_0} \sim \frac{1}{n_i} \sim e^{\frac{\Delta E_g}{2kT}} \quad (5.43)$$

Видно, что доля тока генерации в полном обратном токе тем больше, чем шире запрещённая зона полупроводника и чем ниже температура. Так, для кремниевого диода средней мощности при комнатной температуре и при $U_{обр} = 10$ В ток генерации превышает тепловой ток на 5 порядков: $I_G \approx 10^{-9}$ А, $I_0 \approx 10^{-14}$ А.

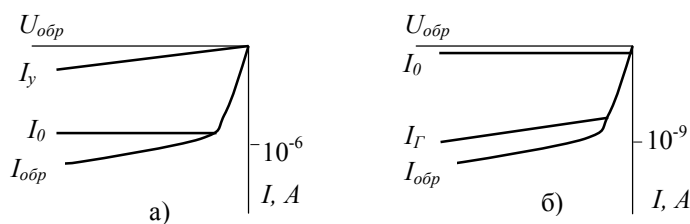


Рис. 5.10. Составляющие обратных токов реальных р-п переходов, изготовленных из германия

(а) и кремния (б). Указаны порядки величин обратных токов для диодов средней мощности

Второй составляющей обратного тока является ток утечки. Этот ток возникает из-за того, что реальные переходы всегда имеют участки, выходящие на поверхность кристалла. Поэтому между смежными областями перехода могут образовываться проводящие по-

верхностные «дорожки». Ток утечки подчиняется закону Ома и почти не зависит от температуры. В кремниевых р-п переходах токи утечки очень малы, так как поверхность кристалла покрыта тонким защитным оксидным слоем.

Из рис. 5.10 видно, что величину обратного тока германиевого перехода определяют, в основном, тепловой ток I_0 и ток утечки I_y , в то время как в кремниевом переходе обратный ток практически полностью обусловлен током генерации I_G . Величина полного обратного тока кремниевых переходов на несколько порядков меньше, чем германиевого перехода с такими же параметрами.

5.7. Прямое смещение реального р-п перехода. Ток рекомбинации и влияние сопротивления базы.

В реальном р-п переходе при прямом смещении происходит также и рекомбинация подвижных носителей. Рис. 5.11 поясняет механизм её возникновения. Электроны п-области и дырки р-области с энергиями, меньшими, чем высота барьера для диффузии основных носителей, могут проникать в область обеднения и рекомбинировать друг с другом.

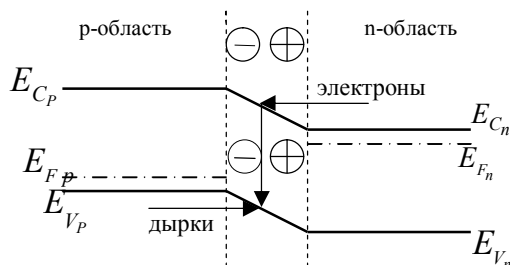


Рис. 5.11. Механизм возникновения тока рекомбинации в реальном р-п переходе

При этом электроны уходят из п-области, а дырки – из р-области, и направление их движения совпадает с направлением прямого тока через переход. Вследствие этого в реальном р-п переходе возникает дополнительный прямой ток – ток рекомбинации.

Ток рекомбинации определяется количеством основных носителей, проникших в область перехода и исчезнувших за счёт рекомбинации во всём его объёме в единицу времени. Следовательно, он должен сильно зависеть от высоты потенциального барьера и от температуры. Ток рекомбинации можно определить по следующей упрощенной формуле:

$$I_{рек} = eSl \frac{n_i}{\tau} e^{\frac{eU}{2kT}} \quad (5.44)$$

Отсюда видно, что ток рекомбинации, как и диффузионный ток (5.36), экспоненциально возрастает с увеличением прямого напряжения. Вместе с тем, при сравнении формул (5.36) и (5.44) видно, что зависимость $I_{рек}$ от прямого напряжения более слабая: показатель экспоненты в (5.44) вдвое меньше. Таким образом, ток рекомбинации является основной составляющей прямого тока при малых прямых напряжениях. Так, например, для кремниевого р-п перехода при $U < 0,4$ В в прямом токе преобладает ток рекомбинации, а при $U > 0,4$ В – диффузионный ток основных носителей.

При протекании через реальный переход больших прямых токов необходимо учитывать падение напряжения не только на р-п переходе, но и на объёмном сопротивлении базы: $U_{\delta} = I \cdot r_{\delta}$. Величина r_{δ} обычно находится в пределах от десяти до ста Ом. Таким образом, учитывая формулу для вольтамперной характеристики перехода в прямом включении (5.36), можно записать:

$$U = U_{p-n} + U_{\delta} = \frac{kT}{e} \ln \frac{I}{I_0} + I \cdot r_{\delta} \quad (5.45)$$

Это означает, что при достаточно больших прямых токах падение напряжения на базе U_{δ} может превысить падение напряжения на р-п переходе, и на вольтамперной характеристике появится почти линейный участок (см. рис. 5.12, кривая 2).

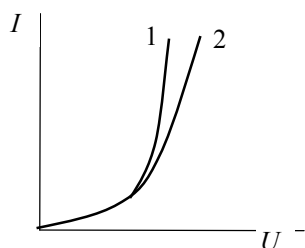


Рис. 5.12. Прямая ветвь вольтамперной характеристики р-п перехода. 1 – для идеального перехода ($r_{\delta} = 0$), 2 – изменённая за счет падения напряжения на объёмном сопротивлении базы

Проведём оценку степени влияния падения напряжения U_{δ} на форму вольтамперной характеристики. Пусть $r_{\delta} = 20$ Ом и прямой ток составляет 100 мА. Тогда $U_{\delta} = I \cdot r_{\delta} = 10^{-1} \cdot 20 = 2$ В (!). На кривых $I(U)$ для кремниевых диодов Д210 прямой ток 100 мА наблюдается при напряжении $U = U_{p-n} + I \cdot r_{\delta} \sim 0,85$ В. Такое расхождение объясняется модуляцией сопротивления базы: при больших прямых токах доля инжектированных в базу диода неосновных носителей оказывается очень высокой, и сопротивление базы уменьшается в не-

сколько раз. Этот эффект проявляется как в статическом, так и в импульсном режиме при высоких уровнях инжекции.

5.8. P-n переход на основе вырожденных полупроводников

Если концентрация примесей в полупроводнике очень высока ($\sim 10^{20}$ см⁻³), то примесные атомы располагаются в кристаллической решетке на относительно малых расстояниях друг от друга, и в отличие от обычных полупроводников образуют систему взаимодействующих частиц. Следствием этого является расщепление единого примесного уровня энергии в энергетическую зону, которая при таких высоких концентрациях примеси перекрывается с зоной разрешенных состояний. В полупроводнике n-типа примесная зона сливается с зоной проводимости, в p-полупроводнике – с валентной зоной. Таким образом, при очень высоких концентрациях примесей дно зоны проводимости n-полупроводника – это низший уровень примесной донорной зоны, а потолок валентной зоны полупроводника p-типа – это самый верхний уровень примесной зоны акцепторов. При этом уровень Ферми располагается в зоне разрешенных энергетических состояний, и для основных носителей образуется непрерывная энергетическая диаграмма, характерная для металлов. Полупроводники, обладающие такими свойствами, называются вырожденными.

P-n переход на основе вырожденных полупроводников также обладает особыми свойствами. Энергетическая диаграмма такого перехода приведена на рис. 5.13.

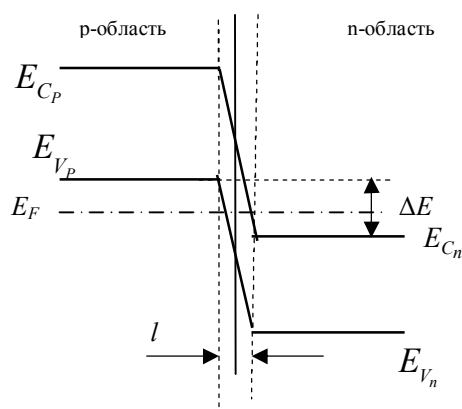


Рис. 5.13. Энергетическая диаграмма p-n перехода на основе вырожденных полупроводников в равновесном состоянии

Видно, что на энергетической диаграмме образуется полоса энергий электронов ΔE , в пределах которой уровни валентной зоны p-области находятся на-

против уровней зоны проводимости n-области. Ширина разделяющего их потенциального барьера l очень мала (\sim десятки нм) из-за высокой концентрации примесей (5.19), и вследствие этого возникают переходы электронов в обоих направлениях без изменения энергии за счет туннельного эффекта (см. раздел 4.3). В состоянии равновесия встречные туннельные токи электронов равны, и суммарный туннельный ток равен нулю.

Подача на переход внешнего напряжения нарушает равенство встречных туннельных токов. При увеличении прямого смещения туннельный ток возрастает, достигает максимума, а затем спадает до нуля. Рассмотрим причину такой зависимости с использованием энергетических диаграмм.

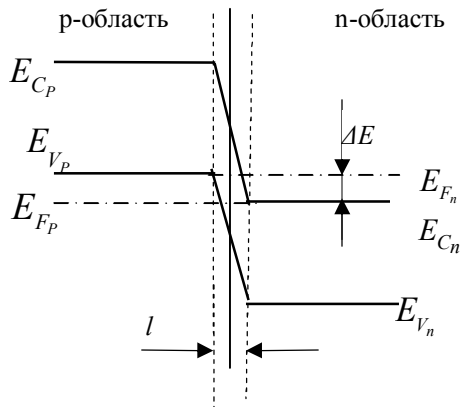


Рис.5.14. Энергетическая диаграмма прямо смещенного p-n перехода на основе вырожденных полупроводников для случая максимального туннельного тока

Для объяснения такого характера изменения прямого туннельного тока следует учесть, что максимум плотности состояний электронов в зоне проводимости находится примерно посередине между её дном E_{Cn} и уровнем Ферми. Максимум плотности состояний дырок в валентной зоне располагается примерно посередине между уровнем Ферми и её верхней границей E_{Vp} .

При увеличении прямого смещения число туннельных переходов электронов из зоны проводимости n-полупроводника на уровни дырок в валентной зоне p-полупроводника растёт, так как всё большее число свободных состояний в валентной зоне оказывается при той же энергии, что и занятые состояния в зоне проводимости. Число туннельных переходов электронов в обратном направлении соответственно уменьшается. Наибольшее значение прямого туннельного тока соответствует максимальному перекрытию по энергии свободных уровней у потолка валентной зоны p-полупроводника и

занятыми электронами уровней зоны проводимости полупроводника n-типа.

Энергетическая диаграмма на рис. 5.14 соответствует именно такому случаю; при этом уровень Ферми полупроводника p-типа совпадает с нижней границей зоны проводимости, а уровень Ферми полупроводника n-типа находится при той же энергии, что и верхняя граница валентной зоны.

Дальнейшее увеличение прямого смещения приводит к тому, что перекрытие зон по энергиям ΔE уменьшается, при совпадении границ E_{C_n} и E_{V_p} оно исчезает и прямой туннельный ток становится равным нулю. При последующем возрастании прямого напряжения начинают происходить те же процессы, что и в обычном p-n переходе (см. раздел 5.3), и прямой ток опять растёт, но только за счет инжекции неосновных носителей.

При увеличении обратного смещения (рис. 5.15) всё большее число занятых электронами уровней валентной зоны p-полупроводника оказывается при тех же значениях энергии, что и свободные уровни зоны проводимости полупроводника n-типа. Ширина разделяющего их потенциального барьера по-прежнему очень мала, и поэтому наблюдается резкое возрастание туннельного тока, протекающего в обратном направлении – из полупроводника p-типа в полупроводник n-типа.

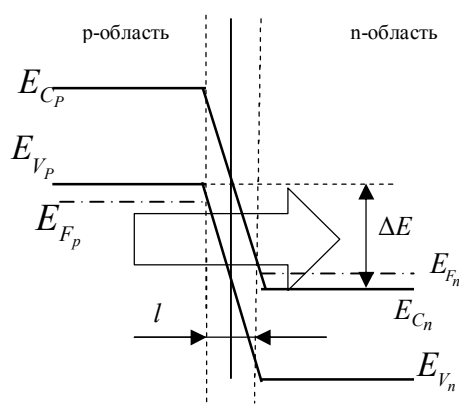


Рис.5.15. Энергетическая диаграмма p-n перехода на основе вырожденных полупроводников при обратном смещении. Стрелка указывает направление перемещения электронов за счет туннельного эффекта

Очевидно, что вольтамперная характеристика (ВАХ) p-n перехода на основе вырожденных полупроводников (рис 5.16, а, кривая 3) должна резко отличаться от таковой для обычного перехода (рис.5.16). Она получается как результат наложения двух характе-

ристик: для туннельного тока и для обычного р-п перехода (кривые 2 и 1 на рис. 5.16, а, соответственно).

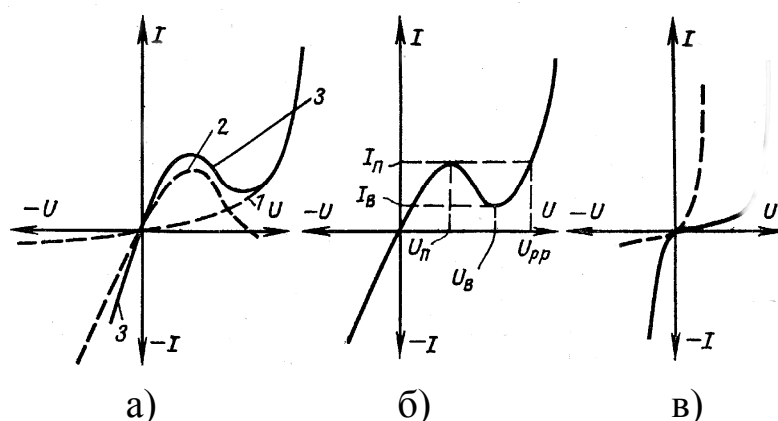


Рис. 5.16. Вольтамперные характеристики р-п переходов на основе вырожденных полупроводников: а, б – туннельного диода, в – обращенного диода (сплошная линия)

Прямая ветвь вольтамперной характеристики содержит область отрицательного дифференциального сопротивления – падающий участок кривой, который отражает уменьшение перекрытия зон по энергиям ΔE . Рост прямого тока после прохождения минимального значения имеет ту же причину, что и для обычного перехода на основе невырожденных полупроводников (так называемая инжекционная ветвь вольтамперной характеристики). При обратном смещении, как указано выше, туннельный ток меняет направление и резко возрастает с увеличением напряжения на переходе.

На основе таких переходов изготавливаются туннельные диоды. Туннельный диод – это диод на основе вырожденного полупроводника, в котором туннельный эффект приводит к появлению на вольтамперной характеристике при прямом напряжении участка отрицательного дифференциального сопротивления. Для их изготовления используют германий или арсенид галлия с высокой концентрацией примесей $\sim 10^{20} \text{ см}^{-3}$. Рабочей частью вольтамперной характеристики туннельного диода является её спадающий участок с отрицательной дифференциальной проводимостью (рис. 5.16, б) в интервале напряжений $U_n < U < U_B$. Характерное время туннелирования $10^{-13} - 10^{-14} \text{ с}$; кроме того, туннельные диоды работают при очень малых прямых напряжениях (до появления инжекционного тока), и накопление избыточных неосновных носителей в базе не происходит. Поэтому туннельные диоды практически безинерцион-

ны и могут работать на частотах до сотен ГГц (мм-диапазон). В зависимости от варианта схемы устройства с туннельным диодом он может служить либо усилителем, либо переключателем с рекордно высокой граничной частотой.

На рис 5.16, б показаны параметры туннельного диода: I_n , U_n – ток и напряжение пика, I_0 , U_0 – ток и напряжение впадины, U_{pp} – напряжение раствора вольтамперной характеристики, т. е. напряжение, при котором ток на инжекционной ветви равен току пика.

Для вырожденных полупроводников существует понятие «критическая концентрация примесей», при которой уровень Ферми n-полупроводника совпадает с дном зоны проводимости, а в р-полупроводнике – с потолком валентной зоны. Если для изготовления р-n перехода используются полупроводники с критической концентрацией примесей ($\sim 10^{19} \text{см}^{-3}$), то на его равновесной энергетической диаграмме не будет наблюдаться начального перекрытия зон ΔE . Поэтому в идеальном случае при прямом смещении туннельный ток будет отсутствовать, и будет наблюдаться, как и в обычном переходе, только инжекционный прямой ток. Перекрытие зон возникнет только при обратном смещении, и обратная ветвь его вольтамперной характеристики будет такой же, как и у туннельного диода (рис. 5.16, в). Такие переходы используются в обращённых диодах. Их название объясняется тем, что, в противоположность обычным диодам, обратный ток намного больше прямого – вплоть до прямого напряжения, при котором начинается инжекция неосновных носителей. Благодаря очень высокому быстродействию, эти диоды используются как детекторы и смесители в области сверхвысоких частот.

В реальных обращённых диодах из-за сложности изготовления полупроводников, имеющих критическую концентрацию примесей, на прямом участке характеристики, как правило, наблюдается небольшой «туннельный максимум», однако величины прямых туннельных токов очень малы.

5.9. Пробой электронно-дырочного перехода

Пробой р-n перехода – это явление резкого увеличения обратного тока при достижении некоторой критической величины обратного напряжения. В зависимости от физических явлений, приводя-

щих к пробоям, различают лавинный, туннельный и тепловой пробои.

Физическая причина лавинного пробоя – ударная генерация электронно-дырочных пар (см. раздел 4.2, рис. 4.2, а) в области р-п перехода, где приложено электрическое поле, обусловленное контактной разностью потенциалов и внешним напряжением. Образовавшиеся электронно-дырочные пары ускоряются электрическим полем перехода и перебрасываются в те области, где они становятся основными носителями. Легко видеть, что направление их тока совпадает с направлением обратного тока через переход. Таким образом, при достижении напряжения лавинного пробоя $U_{\text{проб.лав.}}$ – порога ударной генерации – наблюдается резкое возрастание обратного тока перехода.

Для возникновения лавинного размножения носителей заряда необходимо, чтобы на ширине перехода укладывалось большое число длин свободного пробега. Поэтому лавинный пробой развивается в относительно широких переходах, изготовленных на основе слаболегированных полупроводников с концентрацией примесей не более 10^{17} см^{-3} .

Основные факторы, влияющие на величину $U_{\text{проб.лав.}}$: ширина запрещенной зоны полупроводника ΔE_z , концентрация примесей в базе и температура перехода.

Ширина запрещенной зоны, как известно, определяется величиной энергии связи валентных электронов. Поэтому в переходах, изготовленных из полупроводников с большей ΔE_z , при прочих равных условиях напряжение лавинного пробоя выше.

Концентрация примесей в базе определяет ширину перехода l , и, следовательно, величину напряженности электрического поля в нём ($\mathcal{E} = \mathcal{E}_k + U_{\text{обр}}/l$). При снижении концентрации примесей ширина перехода увеличивается, что приводит к уменьшению \mathcal{E} и, соответственно, к увеличению напряжения лавинного пробоя. Влияние вышеуказанных факторов показано на рис.5.17.

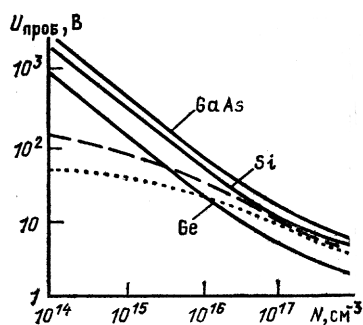


Рис.5.17. Зависимость напряжения лавинного пробоя р-п перехода от концентрации примесей в базе и от материала, из которого изготовлен переход. Наибольшая величина ΔE_3 (1,4 эВ) – у арсенида галлия, наименьшая (0,66 эВ) – у германия

При увеличении температуры р-п перехода напряжение лавинного пробоя возрастает. Физической причиной этого является уменьшение средней длины свободного пробега подвижных носителей в полупроводнике, при этом уменьшается и энергия, которую носитель может приобрести на этой длине в электрическом поле. Поэтому для того, чтобы электрон проводимости, ускоряясь в электрическом поле на меньшей длине пробега, набрал энергию, достаточную для ионизации атома полупроводника, необходимо увеличить обратное напряжение на переходе.

Для оценки влияния температуры на напряжение пробоя перехода используется температурный коэффициент напряжения:

$$\alpha = \frac{\Delta U}{U_{\text{проб}} \cdot \Delta T} \quad (1/\text{K}) \quad (5.46)$$

Для лавинного пробоя этот коэффициент положителен, т.к. ΔT и ΔU имеют одинаковые знаки.

На рис. 5.18 приведены вольтамперные характеристики двух разновидностей электрического пробоя кремниевого р-п перехода. Видно, что для лавинного пробоя (кривая 1) характерен резкий рост обратного тока при практически неизменном обратном напряжении.

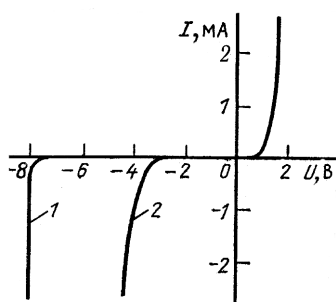


Рис. 5.18. Вольт амперные характеристики кремниевых р-п переходов на участках электрического пробоя. 1 – лавинный пробой, 2 – туннельный пробой

Туннельный пробой р-п перехода – электрический пробой, обусловленный явлением полевой генерации электронно-дырочных пар

(см. раздел 4.3, рис.4.2, б). При подаче обратного напряжения $U_{обр}$ на переход величина напряженности электрического поля в нём может достигать величин $\sim 10^6$ В/см и более, при которых становится возможным туннельный переход электронов с энергетических уровней валентной зоны на уровни зоны проводимости. Очевидно, что для достижения достаточно большой вероятности туннельного переноса электронов сквозь потенциальный барьер перехода необходимо, чтобы его ширина l была очень малой. Это достигается тем, что для изготовления таких переходов используются высоколегированные невырожденные полупроводники с концентрациями примесей более 10^{18} см $^{-3}$.

Из рис. 5.19 видно, что в этом случае при обратном смещении перехода на его энергетической диаграмме образуется полоса энергий электронов $\Delta E_{тун}$, в пределах которой заняты электронами уровни валентной зоны р-области напротив свободных уровней зоны проводимости n-области. Ширина потенциального барьера для туннелирования Δx меньше ширины перехода l и уменьшается с ростом обратного напряжения. Туннельный пробой начинается при таком значении $U_{обр}$, при котором уровень дна зоны проводимости n-области опускается ниже потолка валентной зоны р-полупроводника, т.е. появляется энергетическая полоса $\Delta E_{тун}$.

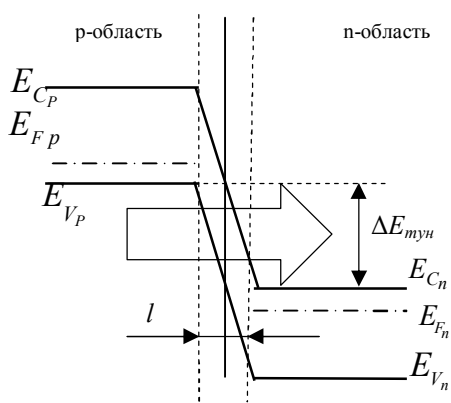


Рис. 5.19. Энергетическая диаграмма, поясняющая причину возникновения туннельного пробоя при обратном смещении р-п перехода. Стрелкой показано преимущественное направление перемещения электронов за счёт туннельного эффекта

В отличие от лавинного пробоя, туннельный пробой р-п перехода происходит при малых обратных напряжениях (3-5 В), кроме того, соответствующая вольтамперная характеристика более пологая (кривая 2 на рис. 5.18).

При увеличении температуры р-п перехода напряжение туннельного пробоя уменьшается, т. е. обратный ток при данном напряжении возрастает. Это связано с тем, что при увеличении температуры происходит небольшое уменьшение ширины запрещенной зоны полупроводника. Следствием этого является уменьшение ширины барьера Δx и увеличение туннельного тока. Следовательно, температурный коэффициент напряжения пробоя α (5.46) в этом случае отрицателен. Эффект выражен слабо, поэтому величины изменения напряжения значительно меньше, чем для лавинного пробоя.

Тепловой пробой р-п перехода – такой, развитие которого обусловлено выделением тепла вследствие прохождения тока через переход. Если количество тепловой энергии, выделяемой в переходе больше тепла, отводимого от него, то его температура повышается, концентрация неосновных носителей и обратный ток возрастают. Это приводит к дальнейшему увеличению температуры за счет выделения дополнительного джоулевого тепла; процесс нарастания температуры перехода и обратного тока продолжается, и в конечном счете это может привести к разрушению перехода. Напряжение теплового пробоя зависит от многих факторов: от величины теплового тока I_0 , удельного сопротивления полупроводника, условий теплоотвода, и т. д.

5.10. Емкость р-п перехода

Наличие равных по величине и противоположных по знаку зарядов по обе стороны границы между р- и п-областями можно рассматривать как электрическую емкость р-п перехода. Эти заряды могут иметь разную природу. Во-первых, это ионы доноров и акцепторов, во-вторых – при прямом смещении перехода – инжектированные неосновные носители у границ р- и п-областей.

Однако полную аналогию с электростатической ёмкостью, определяемой как отношение заряда, накопленного на одном из проводников к возникшей между ними разности потенциалов, провести нельзя. На обычном конденсаторе при наличии разности потенциалов между обкладками накапливается электрическая энергия, тогда как в равновесном р-п переходе этого не происходит. Влияние ёмкости р-п перехода на свойства электрических схем, в которых ис-

пользуются активные полупроводниковые элементы, проявляется только при относительно быстрых изменениях напряжения на р-п переходе. В зависимости от того, каково смещение перехода (прямое или обратное), ёмкостные свойства перехода определяются различными физическими процессами.

При обратном смещении р-п перехода двойной электрический слой на границе между р- и п-областями образован положительными ионами доноров и отрицательными ионами акцепторов.

Ёмкость, связанная с перераспределением зарядов неподвижных ионов примесей, называется барьерной ёмкостью р-п перехода. Ёмкостная реакция перехода проявляется в том, что при изменении обратного напряжения изменяется ширина перехода и величина заряда некомпенсированных ионов примесей. При этом к постоянному обратному току добавляется переменная ёмкостная составляющая: $i_{\sim} = C_{\sigma} \frac{dU}{dt}$. Физическая причина возникновения этого тока – изменение объёмных зарядов ионов примесей со временем, поэтому $i_{\sim} = \frac{dQ}{dt} = \frac{dQ}{dU} \cdot \frac{dU}{dt} = C_{\sigma} \frac{dU}{dt}$. Отсюда следует, что барьерная ёмкость не определяется формулой электростатики, а является дифференциальной величиной:

$$C_{\sigma} = \frac{dQ}{dU} \quad (5.47)$$

Для резкого одностороннего р-п перехода можно получить простую формулу, отражающую зависимость барьерной ёмкости от обратного напряжения и концентрации примесей. Пусть для данного перехода $N_A \gg N_D$, тогда используя формулы (5.47) и (5.26) с учётом того, что $U_{обр} \gg \phi_k$, получим:

$$Q = eN_D S \cdot l_{об} = eN_D S \sqrt{\frac{2\varepsilon\varepsilon_0 U_{обр}}{eN_D}};$$

$$C_{\sigma} = eN_D S \sqrt{\frac{2\varepsilon\varepsilon_0}{eN_D}} \cdot \frac{1}{2\sqrt{U_{обр}}} = \varepsilon\varepsilon_0 S \sqrt{\frac{eN_D}{2\varepsilon\varepsilon_0 U_{обр}}} = \frac{\varepsilon\varepsilon_0 S}{l_{об}} \quad (5.48)$$

Видно, что барьерная ёмкость может быть представлена формулой для плоского конденсатора, где роль расстояния между обкладками играет ширина перехода. Соответственно, величина C_{σ} растёт с увеличением концентрации примесей из-за сужения перехода и

уменьшается с ростом обратного напряжения, вызывающим противоположный эффект.

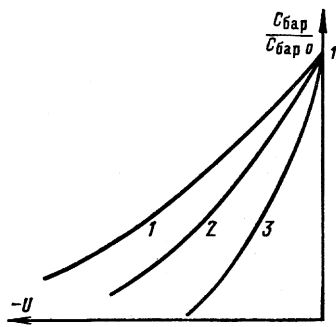


Рис 5.20 Зависимость барьерной ёмкости р-п перехода от обратного напряжения. 1 — для плавного перехода, 2 — для резкого перехода, 3 — для перехода с особым распределением примесей.

Скорость изменения барьерной ёмкости при увеличении обратного напряжения зависит от характера распределения примесей в пределах ширины перехода. Как следует из (5.48) и рис.5.20, для резкого перехода $C_{\sigma} \sim \frac{1}{\sqrt{U_{обр}}}$; для плавного перехода, как показы-

вают расчеты, эта зависимость более слабая: $C_{\sigma} \sim \frac{1}{\sqrt[3]{U_{обр}}}$.

На основе барьерной ёмкости перехода изготавливаются специальные диоды — варикапы. Это диоды с электрически изменяемой ёмкостью. Как правило, для изготовления варикапов используются переходы с более сложным характером распределения примесей, что позволяет получить большую крутизну вольт-фарадной характеристики (кривая 3 рис. 5.20).

Прямое смещение перехода вызывает инжекцию неосновных носителей, и в прилегающих к переходу частях р- и п-областей возникает двойной электрический слой, обусловленный избыточными электронами р-области и избыточными дырками п-области. Ёмкость, связанная с перераспределением зарядов инжектированных неосновных носителей при изменении прямого напряжения называется диффузионной ёмкостью р-п перехода. Изменение этих зарядов определяется зависимостью избыточной концентрации неосновных носителей от приложенного напряжения (см. формулы 5.32 и 5.32,а), поэтому диффузионная ёмкость при невысоких частотах изменения напряжения пропорциональна прямому току: $C_{диф} \sim I_{пр}$. Вольт-фарадная характеристика для диффузионной ёмкости $C_{диф} = f(U_{пр})$ имеет ту же форму, что и вольтамперная характеристика прямо смещенного перехода.

На высоких частотах, когда период переменного прямого напряжения становится меньше времени жизни неосновных носителей, заряд избыточных неосновных носителей в течение малого периода не успевает измениться. Следовательно, диффузионная ёмкость с ростом частоты будет уменьшаться практически до нуля. Поэтому, на очень высоких частотах, а также при резких «ступенчатых» изменениях напряжения на переходе её влияние будет не заметно.

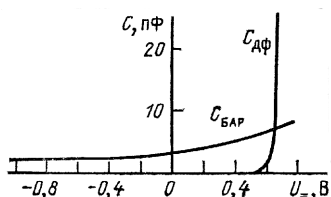


Рис. 5.21 Вольт-фарадные характеристики р-п перехода при прямом и обратном смещении.

Из рис. 5.21 видно, что барьерная ёмкость проявляется также и при прямом смещении; она играет определяющую роль до тех значений прямого напряжения, при которых начинается инжекция и, соответственно, резкий рост прямого тока. Например, для перехода на основе арсенида галлия преобладающее влияние барьерной ёмкости при прямом смещении наблюдается вплоть до напряжений 0,8-0,9 В.

5.11. Переходные процессы в р-п переходе. Установление прямого напряжения при подаче на переход скачка прямого тока

Если на переход подаётся не постоянное, а быстро меняющееся переменное напряжение, то при увеличении его частоты выпрямляющие свойства перехода ухудшаются: величина обратного тока в течение части отрицательного полупериода становится соизмеримой с током во время положительного полупериода. Наиболее ярко инерционность р-п перехода проявляется при подаче на него ступенчатых (импульсных) сигналов, переключающих переход скачком с прямого смещения на обратное и наоборот.

Переходными называются процессы установления постоянных токов и напряжений на р-п переходе при подаче на него ступенчато

меняющихся сигналов. Характер этих процессов зависит от амплитуды импульсного сигнала, т.е. от уровня инъекции, а также от свойств источника сигнала. Если внутреннее сопротивление источника значительно больше сопротивления перехода, то источник может поддерживать постоянный ток при любых изменениях параметров внешней цепи. Такой тип источника называется генератором тока. При обратном соотношении указанных сопротивлений источник при тех же условиях может поддерживать постоянным напряжением на выходе и поэтому называется источником напряжения.

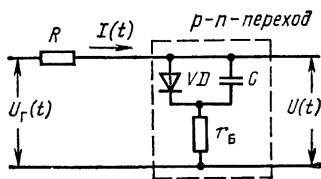


Рис. 5.22. Эквивалентная схема р-п перехода

Для рассмотрения переходных процессов используем схему на рис. 5.22. Напряжение от генератора импульсов $U_G(t)$ через резистор R подаётся на цепь, состоящую из «идеального диода», ёмкости перехода C , включённой параллельно диоду, и последовательно включённого сопротивления базы r_B . Из множества вариантов выберем два случая: установление прямого напряжения на переходе и установление постоянного обратного тока перехода.

Пусть в некоторый момент $t^{(1)}$ ток перехода скачком меняется от нуля до I_{np} (рис. 5.23, а). Влияние диффузионной ёмкости на продолжительность переходного процесса будет пренебрежимо малым (см. раздел 5.3).

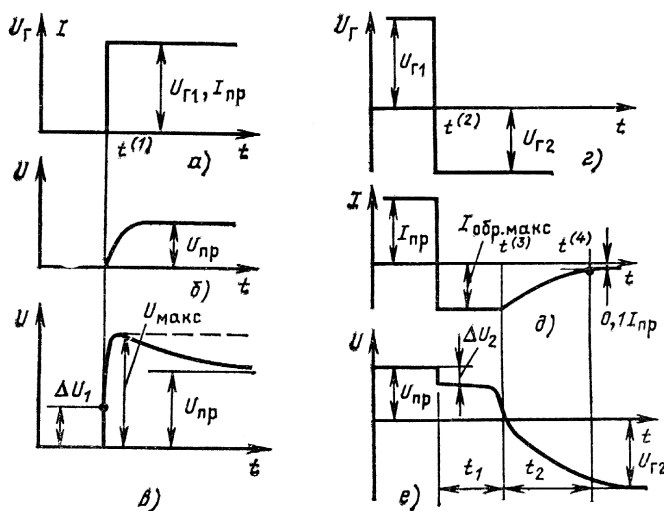


Рис. 5.23. Процесс установления прямого напряжения (а-в) и восстановления стационарного обратного тока и напряжения (г-е) на р-п переходе

При низком уровне инжекции ($\delta \ll 1$) влияние падения напряжения на объёмном сопротивлении базы, т. е. низколегированной области, можно не учитывать, поэтому наблюдается плавное нарастание прямого напряжения с последующим выходом на постоянное значение $U_{пр}$ (рис. 5.23, б). Запаздывание прямого напряжения на переходе связано с конечностью времени установления новой, меньшей ширины перехода, т.е. со временем перезарядки барьерной ёмкости.

При большом прямом токе $\delta \gg 1$, и характер переходного процесса меняется (рис. 5.23, в). Во-первых, из-за падения напряжения на объёмном сопротивлении базы наблюдается мгновенный скачок прямого напряжения ΔU_I . Далее процесс идёт так же, как и при низком уровне инжекции: вследствие конечного времени перезарядки барьерной ёмкости прямое напряжение нарастает и достигает наибольшего значения $U_{макс}$. Скорость нарастания напряжения на переходе гораздо выше, чем в случае низкого уровня инжекции, так как при большом прямом токе время заряда барьерной ёмкости становится очень малым. Например, при токе 10 мА и ёмкости 5 пФ время заряда $C_b U_{пер} / I_{пр}$ окажется равным 0,4 нс. Здесь $U_{пер}$ – прямое напряжение, соответствующее участку резкого роста прямого тока. Для кремниевого перехода эта величина составляет $\sim 0,8$ В.

На последней стадии переходного процесса прямое напряжение на переходе снижается до стационарного значения $U_{пр}$. Это связано с тем, что из-за высокого уровня инжекции концентрация подвижных носителей в базе резко возрастает, что приводит к уменьшению её объёмного сопротивления. Этот эффект называется модуляцией сопротивления базы. Следствием этого является уменьшение падения напряжения на сопротивлении базы и установление на переходе прямого напряжения $U_{пр} < U_{макс}$.

Поскольку диффузионная ёмкость не проявляется при установлении прямого напряжения на переходе, инерционность переключения определяется только временем перезарядки барьерной ёмкости. Вследствие этого время этого переходного процесса при низком и высоком уровнях инжекции оказывается очень малым, порядка единиц–десятков наносекунд.

5.12. Установление постоянного обратного тока перехода при ступенчатом переключении напряжения с прямого на обратное

Пусть в момент времени $t^{(2)}$ выходное напряжение генератора импульсов скачком меняется с прямого $U_{Г1}$ на обратное $U_{Г2}$ (рис. 5.23, г). Поскольку в данном случае источник является генератором напряжения, то при постоянном уровне $U_{Г2}$ протекает переходный процесс: обратный ток и связанное с ним падение напряжение на переходе с течением времени приходят к своему стационарному значению.

Скачок напряжения вызывает резкую смену полярности тока через переход, величина броска тока $\Delta I = (U_{Г1} + U_{Г2})/R$. Через переход начинает протекать обратный ток $I_{обр.макс}$, намного превышающий величину обратного тока при постоянном обратном смещении (рис. 5.23, д). Этот большой ток обусловлен тем, что в момент переключения в базе очень высока концентрация избыточных неосновных носителей, накопленных при прямом смещении, и при смене полярности напряжения происходит их экстракция через p-n переход. В момент переключения падение напряжения на переходе скачком уменьшается на величину $\Delta U_2 = \Delta I \cdot r_b$, но он остаётся смещенным в прямом направлении. Причина такого необычного явления в том, что на переходе по обе стороны от границы имеется двойной электрический слой избыточных неосновных носителей, и распределение их концентраций соответствует прямому смещению перехода (см. кривую для $t \leq t^{(2)}$ на рис 5.24).

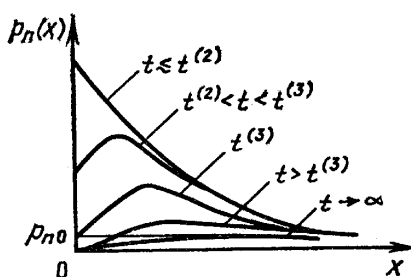


Рис. 5.24. Изменение распределения концентрации неосновных носителей в базе со временем после переключения

Время переходного процесса принято разделять на два этапа: t_1 — участок высокой обратной проводимости и t_2 — стадию восстановления постоянного значения обратного тока (рис. 5.23, e). В течение всего промежутка времени t_1 концентрация неосновных носителей в базе превышает равновесную (рис. 5.24). Такое состояние является

причиной того, что величина обратного тока соизмерима с величиной прямого тока, при этом обратный ток почти постоянен. Это явление обусловлено диффузионным механизмом движения накопленных неосновных носителей к переходу. Видно, что градиент их концентрации (наклон кривых вблизи перехода на рис. 5.24) приблизительно одинаков на всем протяжении отрезка времени t_1 . Это обеспечивает постоянную скорость поступления неосновных носителей базы на границу перехода и, соответственно, постоянство обратного тока за счет экстракции на стадии высокой обратной проводимости.

Данную стадию можно рассматривать как разряд диффузионной емкости перехода, происходящий из-за рассасывания накопленных неосновных носителей. На рис. 5.23, *e*, видно, что при этом прямое напряжение на переходе, которое можно считать напряжением на диффузионной емкости, постепенно уменьшается.

Как видно из рис. 5.24, в момент времени $t^{(3)}$ избыточные неосновные носители на границе базовой области с p - n -переходом исчезают ($\Delta p_n(0) = 0$), при этом падение напряжения на переходе становится равным нулю (рис. 5.23, *e*). Далее, на всем протяжении участка t_2 продолжается экстракция оставшихся у границ перехода неосновных носителей, т. е. заканчивается разряд диффузионной емкости. Одновременно происходит перестройка обедненного слоя к состоянию обратного смещения, что соответствует перезарядке барьерной емкости перехода. При этом напряжение на переходе и обратный ток стремятся к стационарным значениям $U_{Г2}$ и $I_{обр}$.

Кроме перечисленных выше явлений, в течение всего времени восстановления в базе диода происходит рекомбинация избыточных неосновных носителей, дающая дополнительный вклад в сокращение времени восстановления.

Основной параметр, характеризующий инерционность перехода при ступенчатом изменении напряжения с прямого на обратное — время восстановления обратного сопротивления. Наименьшие значения времени восстановления составляют единицы наносекунд. Оно отсчитывается от момента переключения до момента,

когда обратный ток, уменьшаясь, достигнет значения $0,1 I_{пр}$. Этот промежуток времени равен $t_1 + t_2$ (рис. 5.23, *e*).

Время восстановления не является универсальным параметром. Его значение зависит от характеристик цепи, в которую включается диод. На стадии высокой обратной проводимости сопротивление p – n -перехода мало, и ток определяется практически полностью номиналом резистора R , включенного последовательно с переходом: $I_{обр. макс} \approx U_{Г2}/R$ (см. рис. 5.22). При уменьшении величины R обратный ток возрастает, и время, необходимое для рассасывания накопленного заряда избыточных неосновных носителей, также уменьшается.

Для увеличения диапазона частот, в котором могут работать диоды на основе p – n -переходов, необходимо обеспечивать как можно меньшее время восстановления. Для этого используют переходы малой площади, низкий уровень инжекции; при изготовлении диодов вводят в базу примесь атомов золота – эффективных электронных ловушек для увеличения скорости рекомбинации. Активно используются диоды Шоттки (диоды на основе выпрямляющего контакта металл-полупроводник, см. раздел 5.13), которые обладают очень малой инерционностью.

5.13. Переходы металл – полупроводник. Выпрямляющий контакт (барьер Шоттки)

Контакты металл-полупроводник широко применяются в полупроводниковой электронике. Во-первых, они обеспечивают соединение полупроводникового прибора с элементами внешней цепи. Такие контакты должны обладать линейной вольтамперной характеристикой. Во-вторых, на их основе созданы быстродействующие диоды, обладающие очень малой инерционностью и работающие на высоких частотах. При соединении металла и полупроводника могут образоваться две разновидности переходов, или, как принято говорить – контактов:

- выпрямляющий контакт с «односторонней» проводимостью;
- омический контакт; ток и падение напряжения на нём приблизительно соответствуют закону Ома.

Вид контакта определяется следующими факторами: типом проводимости полупроводника; соотношением работ выхода металла и полупроводника; концентрацией примеси в полупроводнике; знаком и величиной поверхностного заряда.

Осуществим идеальный контакт между металлом и полупроводником *n*-типа, работа выхода которого меньше работы выхода электронов из металла: $e\varphi_M > e\varphi_n$ (рис. 5.25, а, б). При этом предполагается, что заряд на поверхности полупроводника отсутствует: $Q_{нов} = 0$.

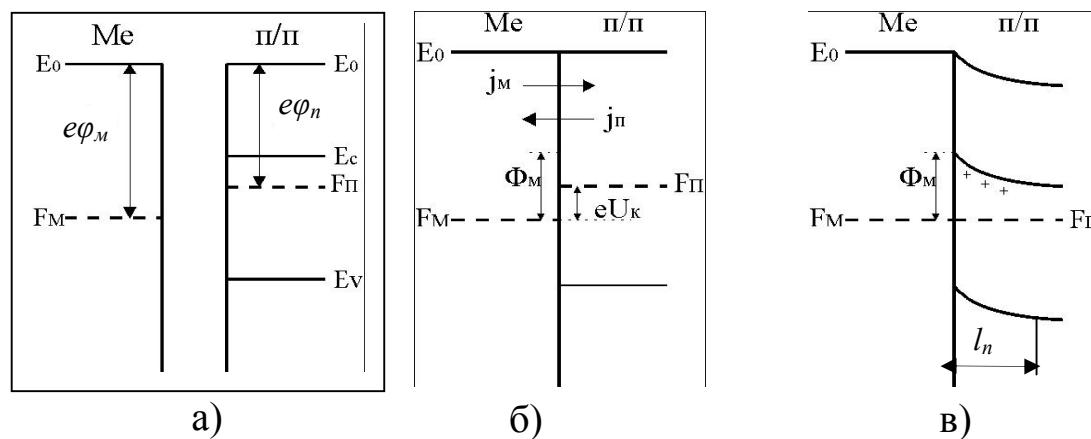


Рис. 5.25. Энергетические диаграммы металла и *n*-полупроводника:

а) до контакта, б) непосредственно после соприкосновения, в) после установления равновесия на контакте. $e\varphi_M$ – работа выхода металла, $e\varphi_n$ – работа выхода полупроводника, eU_k – контактная разность потенциалов, т. е. потенциальный барьер для электронов полупроводника, Φ_M – потенциальный барьер для электронов металла

При таком контакте энергетически выгодным оказывается переход электронов с уровней зоны проводимости полупроводника в металл (ток j_n). Они занимают свободные энергетические уровни в металле выше уровня Ферми и сообщают ему отрицательный заряд. Из-за ухода электронов в металл в приконтактной области полупроводника образуется слой, обедненный основными носителями шириной l_n , в котором остается некомпенсированный положительный заряд ионов доноров. Образующееся на контакте электрическое поле препятствует дальнейшему переходу электронов из *n*-

полупроводника в металл и одновременно ускоряет электроны, движущиеся в противоположном направлении (ток j_m). В результате на переходе устанавливается равновесие, при котором встречные токи электронов j_m и j_p уравниваются и суммарный ток равен нулю. При равновесии в системе образуется единый уровень Ферми (рис. 5.25, в). Обеднённый слой l_n лежит целиком в приграничной области полупроводника, его удельное сопротивление намного больше, чем в объёме.

При подключении прямого внешнего напряжения U (плюс – на металл, минус – на полупроводник) высота потенциального барьера для электронов зоны проводимости полупроводника снижается, ширина области обеднения уменьшается ($L_n < l_n$), и через переход течет ток основных носителей – электронов проводимости полупроводника – в металл (рис. 5.26, а).

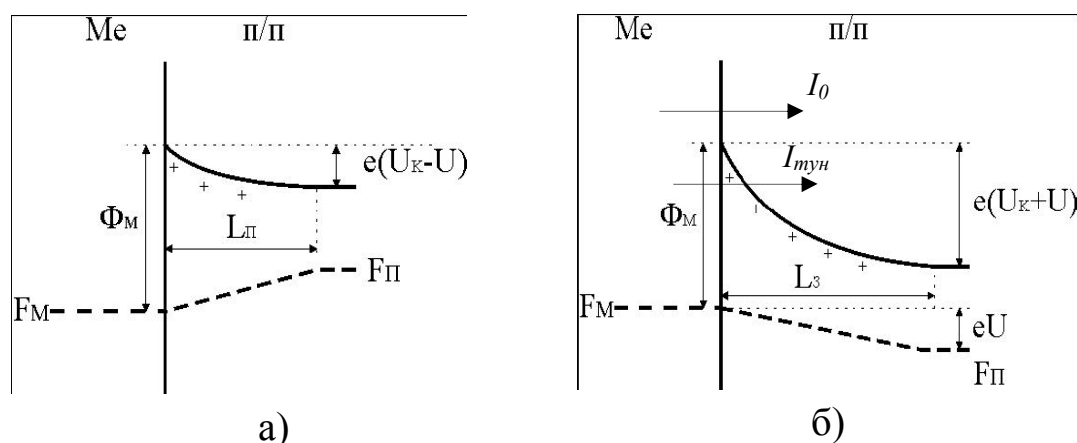


Рис. 5.26. Энергетические диаграммы барьера Шоттки при прямом (а) и обратном (б) включении

Важнейшим преимуществом барьера Шоттки перед p-n переходом является то, что при прямом включении в нём не происходит накопление избыточных неосновных носителей – электронов за счёт инжекции, так как «базой» в этом случае является металл, в котором концентрация электронов на несколько порядков выше, чем в полупроводнике.

При подключении обратного напряжения $U_{обр}$ («плюс» на полупроводник, «минус» на металл) область обеднения расширяется ($L_z > l_n$), потенциальный барьер для основных носителей повышается пропорционально величине $U_{обр}$, и прямой ток электронов уменьшается практически до нуля. Малый обратный ток I_0 противополо-

ложного направления образуется электронами, переходящими из металла в полупроводник. Их энергия должна быть достаточной для преодоления потенциального барьера для электронов, движущихся в этом направлении (рис. 5.26, б). Этот потенциальный барьер Φ_m равен разности энергий, соответствующих дну зоны проводимости полупроводника на границе раздела и уровню Ферми в металле (рис. 5.26, а, б); его высота в первом приближении не зависит от величины и полярности приложенного к переходу напряжения. Величина этого «надбарьерного» тока I_0 выражается (подобно току термоэлектронной эмиссии) следующей формулой:

$$I_0 = S \cdot A \cdot T^2 e^{-\frac{e\Phi_m}{kT}} \quad (5.49),$$

где S – площадь перехода, A – размерная константа, (~ 120 А/(см²·К)). Этот ток на несколько порядков меньше тока при прямом включении.

Сравним величины обратного тока для барьера Шоттки (металл–полупроводник n-типа) и p-n перехода, изготовленных на основе кремния. Пусть площадь каждого из переходов равна 10^{-4} см², концентрация донорной примеси $N_D = 10^{15}$ см⁻², $T = 300$ К. Тогда обратный ток для барьера Шоттки будет составлять $\sim 10^{-9}$ А, а та же величина для p-n перехода будет равна 10^{-13} А. Видно, что обратный ток выпрямляющего контакта металл-полупроводник на много порядков превышает тепловой ток p-n перехода.

Формула, описывающая вольтамперную характеристику барьера Шоттки, а также вид его вольтамперной характеристики совпадают с данными для p-n перехода (см. формулу 5.35 и рис. 5.8). Однако их сравнение при указанных выше условиях показывает также, что помимо большего обратного тока наблюдается различие и при прямом смещении (рис. 5.27).

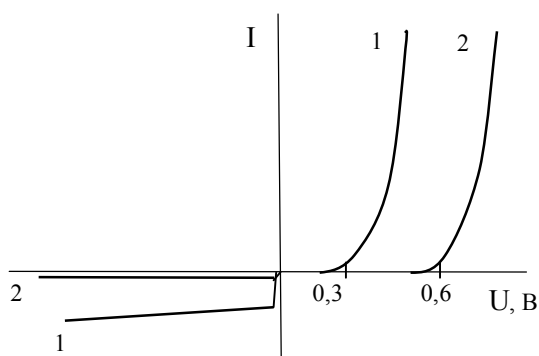


Рис.5.27. Вольтамперные характеристики барьера Шоттки алюминий – n-кремний (1) и кремниевого p-n перехода (2) при одинаковых площадях контактов и концентрациях примеси

Прямая ветвь вольтамперной характеристики барьера Шоттки располагается в области меньших прямых напряжений. Это обусловлено большей – по сравнению с р-п переходом – величиной его обратного тока I_0 , а также меньшей величиной потенциального барьера для электронов полупроводника n-типа на границе с металлом.

Кроме того, поскольку прямой ток выпрямляющего контакта – это ток электронов из полупроводника n-типа в металл, то при больших прямых токах можно пренебречь падением напряжения на малом сопротивлении металла. Вследствие этого вольтамперная характеристика барьера Шоттки в отличие от р-п перехода описывается экспонентой в более широком интервале допустимых прямых напряжений. Аналогичная картина «односторонней проводимости» получается и при контакте металла с полупроводником р-типа, у которого работа выхода больше, чем у металла.

Приведённое выше описание физических процессов на контакте при подаче обратного напряжения справедливо для концентраций примесей в полупроводнике не выше 10^{17} см^{-3} . При этом потенциальный барьер для электронов, движущихся из металла в полупроводник, достаточно широк, и можно пренебречь туннельным переносом электронов в том же направлении (на рис.5.26, б ток $I_{\text{тун}}$ пренебрежимо мал).

Предположение об отсутствии поверхностного заряда в действительности не выполняется. В полупроводнике n-типа всегда существует отрицательный поверхностный заряд; его плотность зависит от вещества, а также от состояния поверхности. Так, для GaAs n-типа среднее значение плотности поверхностного заряда, отнесенное к заряду электрона, составляет 10^{13} см^{-2} , для кремния n-типа эта величина равна 10^{10} – 10^{11} см^{-2} . Отрицательный поверхностный заряд вытесняет электроны из приграничной области вглубь полупроводника. В результате, у его поверхности даже в отсутствие контакта с металлом образуется область обеднения, и энергетические зоны у поверхности изгибаются вверх. В этом случае – особенно при высокой концентрации отрицательного поверхностного заряда – может оказаться, что тип контакта не будет зависеть от свойств металла, так как его роль как приёмника электронов из по-

лупроводника при установлении равновесия не будет определяющим фактором.

Таким образом, при наличии в полупроводнике n-типа высокой плотности отрицательного поверхностного заряда независимо от соотношения работ выхода металла и полупроводника при их контакте может образоваться барьер Шоттки.

5.14. Омический контакт металл – полупроводник

Существуют две разновидности контакта, вольтамперная характеристика которых, близка к закону Ома.

Во-первых, в случае контакта металла с n-полупроводником, когда работа выхода металла меньше, чем работа выхода полупроводника, энергетически выгодным оказывается переход электронов металла на свободные энергетические уровни зоны проводимости полупроводника. При этом в полупроводнике вблизи границы с металлом образуется слой с повышенной концентрацией основных носителей. Такой контакт металла с полупроводником называется омическим, так как он обладает двухсторонней, примерно одинаковой проводимостью. При подключении напряжения той или иной полярности изменяется лишь степень обогащения основными носителями приконтактной области полупроводника.

Во-вторых, барьер Шоттки при концентрации примеси в полупроводнике, превышающей 10^{17} см^{-3} превращается также в омический контакт. При таких концентрациях примеси ширина области обеднения на границе с металлом значительно уменьшается, и при подключении обратного напряжения образуется острый и узкий потенциальный барьер для электронов металла. Это приводит к тому, что величина туннельного тока электронов $I_{\text{тун}}$ из металла в полупроводник при обратном смещении (см. рис. 5.26, б) становится такой же, как и величина тока при прямом смещении.

5.15. Гетеропереходы

Гетеропереходом называется переход, образованный при контакте двух полупроводников с различной шириной запрещенной зоны. Поскольку в гетеропереходах используются разные полупроводники, необходимо, чтобы у этих материалов был один и тот же

тип кристаллической решетки. При этом значения постоянной решетки и температурного коэффициента расширения должны быть практически одинаковыми. При несоблюдении этих условий на контакте возникнут большие механические напряжения, которые приведут к его разрушению. Очевидно, что количество материалов для гетеропереходов ограничено. Наиболее распространенными из них являются германий и арсенид галлия (переход Ge-GaAs), арсенид галлия и трехкомпонентный твердый раствор GaAlAs (переход GaAs-GaAlAs), фосфид индия и трехкомпонентный раствор InGaAs (переход InP-In_{0,53}Ga_{0,47}As).

На основе гетеропереходов изготавливаются высокоэффективные полупроводниковые излучатели и фотоприемники (фотодиоды, светодиоды и лазеры). Гетероструктуры с последовательным чередованием слоёв разных полупроводников очень малой толщины (не более 10-15 нм) – сверхрешётки – находят ещё более широкое применение в полупроводниковой наноэлектронике.

Гетеропереходы, образованные полупроводниками с одинаковым типом проводимости (p⁺-p или n⁺-n), называются изотипными. При использовании полупроводников с разным типом проводимости формируются анизотипные гетеропереходы.

Рассмотрим анизотипный гетеропереход, у которого ширина запрещенной зоны n-полупроводника больше, чем у p-полупроводника. Такой переход образуют, например, n-GaAs и p-Ge, его энергетическая диаграмма показана на рис. 5.28, а.

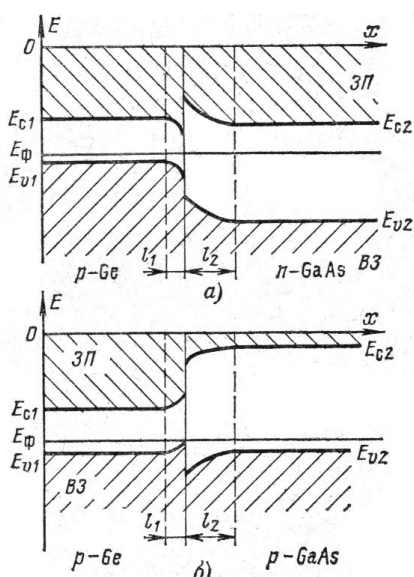


Рис.5.28. Энергетическая диаграмма гетеропереходов: а) анизотипного p-Ge – n-GaAs б) изотипного p-Ge – p-GaAs

При идеальном контакте полупроводников с различной шириной запрещенной зоны ΔE_z происходит установление равновесного состоя-

ния с единым уровнем Ферми. Физические процессы, протекающие при этом – те же, что и для случая обычного p-n перехода. Однако из-за разницы в величинах ΔE_z на линии раздела p- и n-областей возникают разрывы границ энергетических зон. В случае контакта p-Ge – n-GaAs (рис. 5.28, а) при разрыве границы дна зоны проводимости образуется пик, высота которого равна разности энергий $E_{C GaAs} - E_{C Ge}$; при разрыве границы валентной зоны образуется дополнительная ступенька высотой $E_{V Ge} - E_{V GaAs}$. Если в контакт приводятся те же полупроводники с противоположным соотношением типа проводимости (n-Ge и p-GaAs), то характер разрыва зон на границе изменяется на противоположный.

Появление разрывов границ зон приводит к тому, что потенциальные барьеры для диффузии основных носителей в смежные области оказываются не одинаковыми. Для гетероперехода, у которого ширина запрещенной зоны n-полупроводника больше, чем у p-полупроводника (рис. 5.28, а) потенциальный барьер для диффузии электронов n-GaAs в p-Ge гораздо меньше, чем для диффузии дырок в обратном направлении. Поэтому при приложении прямого напряжения будет преобладать инжекция электронов, даже если p-область имеет одинаковую или бóльшую концентрацию примесей. В случае гетероперехода n-Ge – p-GaAs будет преобладать инжекция дырок в n-Ge. Таким образом, в отличие от обычного p-n перехода, гетеропереход всегда обладает важным свойством: односторонней инжекцией из широкозонного полупроводника в узкозонный, практически не зависящей от уровня легирования p- и n-областей. Именно это свойство гетеропереходов используется для создания эффективных источников когерентного (лазеры) и некогерентного (светодиоды) излучения. Расчеты показывают, что для обеспечения односторонней инжекции достаточна разница в ширине запрещенной зоны полупроводников $\sim 0,1$ эВ.

Вольтамперная характеристика идеального гетероперехода описывается той же формулой (5.35), что и для обычного p-n перехода. В реальных гетеропереходах вблизи границы раздела сосредоточены дефекты кристаллической решетки различной природы, которые являются центрами генерации или рекомбинации. Вследствие этого обратный ток реального гетероперехода, в основном, является

током генерации, а доля тока рекомбинации в прямом токе оказывается значительно большей, чем в обычном переходе.

Изотипные гетеропереходы, в отличие от переходов между полупроводниками одной природы, например, $n^+-Si - n-Si$, являются выпрямляющими. В них, как и в анизотипных переходах, существует заметная разница в величинах потенциальных барьеров для диффузии основных носителей. Как видно из рис. 5.28, б, в гетеропереходе p^+-p потенциальный барьер для электронов существенно выше потенциального барьера для дырок.

6. ФОТОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ И СВЕТОИЗЛУЧЕНИЕ

6.1. Взаимодействие оптического излучения с полупроводниками. Фотопроводимость

Характер взаимодействия света с полупроводниками может быть различным. Он определяется свойствами материала полупроводника и длиной волны оптического излучения, т. е. энергией фотонов. В разделе 4.4 была рассмотрена генерация электронно-дырочных пар при облучении полупроводника светом (биполярная световая генерация). Это явление находит очень широкое применение в полупроводниковой электронике. На его основе изготавливаются различные фотоэлектрические приёмники: фоторезисторы, фотодиоды, фототранзисторы и фототиристоры.

При облучении полупроводников светом с энергией квантов, меньшей, чем ширина запрещённой зоны (как правило, в инфракрасной части спектра), генерация электронно-дырочных пар не происходит. Концентрация подвижных носителей в этом случае увеличивается только за счёт однополярной генерации подвижных электронов или дырок при ионизации примесных атомов. Такой механизм фотопроводимости может использоваться в фотоприёмниках инфракрасного излучения, но при низких рабочих температурах (~ 80 К), при которых мала скорость тепловой генерации.

При таких же температурах работают датчики теплового излучения – болометры. Однако, физический механизм фотопроводимости в этом случае принципиально иной: кванты света поглощаются подвижными носителями заряда, например, электронами, и они занимают более высоко лежащие свободные энергетические уровни в

зоне проводимости. В результате увеличивается подвижность электронов μ (см. раздел 3.2), и ток через полупроводник возрастает.

Кроме того, существуют также разновидности поглощения квантов света полупроводниками, не приводящие к изменению их проводимости (решеточное и экситонное поглощение).

Рассмотрим более подробно физические процессы, при которых фотопроводимость возникает за счёт биполярной световой генерации.

6.2. Фоторезистивный эффект. Фоторезистор

Фоторезистивным эффектом называется изменение сопротивления полупроводника под действием света. В отсутствие освещения полупроводник обладает темновой проводимостью, обусловленной равновесными концентрациями подвижных носителей:

$$\sigma_0 = e(n_0\mu_n + p_0\mu_p) \quad (6.1)$$

При облучении полупроводника светом с энергией квантов, превышающей ширину запрещённой зоны, появляются неравновесные электронно-дырочные пары. Концентрация подвижных носителей изменяется до величин $n_0 + \Delta n$ и $p_0 + \Delta p$, при этом проводимость полупроводника оказывается равной

$$\sigma = e[(n_0 + \Delta n)\mu_n + (p_0 + \Delta p)\mu_p] \quad (6.2)$$

Таким образом, фотопроводимость равна:

$$\sigma_{\phi n} = \sigma - \sigma_0 = e(\Delta n\mu_n + \Delta p\mu_p) \quad (6.3)$$

Увеличение концентрации подвижных носителей при облучении светом вызывает уменьшение электрического сопротивления и, соответственно, рост тока через полупроводник. На основе этого явления изготавливаются простейшие фотоприёмники – фоторезисторы.

Чувствительным элементом фоторезистора является кристаллическая пластинка полупроводника или полупроводниковая плёнка на изолирующей подложке. Во внешнюю цепь элемент подключается через омические контакты. В качестве светочувствительных материалов, как правило, применяют полупроводниковые соединения типа $A^{II}B^{VI}$ (CdS, CdSe – видимый диапазон) или $A^{IV}B^{VI}$ (PbS, PbSe – инфракрасный диапазон).

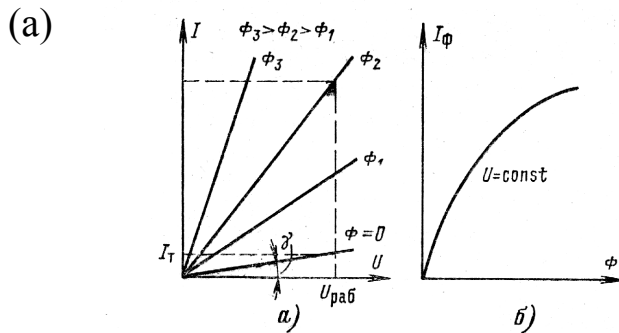


Рис. 6.1. Вольтамперная и энергетическая (б) характеристики фоторезистора.

Вольтамперная характеристика (рис.6.1, а) отображает зависимость тока через фоторезистор от приложенного напряжения; параметром семейства является величина светового потока. Строго говоря, эта зависимость не линейна. В области малых напряжений сказывается влияние неидеальных омических контактов. При больших напряжениях из-за разогрева фоторезистора заметно увеличивается темновая проводимость σ_0 (см. формулу 6.1) и, соответственно, темновой ток, и скорость нарастания фототока уменьшается. Однако в области рабочих напряжений характеристики линейны, их наклон определяет сопротивление фоторезистора при данном световом потоке. В частности, $\operatorname{ctg} \gamma$ равен темновому сопротивлению фоторезистора, его величина обычно находится в пределах $10^6 - 10^8$ Ом. При облучении светом сопротивление фоторезистора снижается на 3–4 порядка.

Энергетическая характеристика фоторезистора (рис. 6.1, б) линейна для невысоких значений светового потока Φ . При увеличении Φ рост фототока замедляется: значительный рост концентрации подвижных носителей приводит к увеличению скорости рекомбинации.

Важнейшим параметром фоточувствительного прибора является токовая чувствительность – отношение фототока к некоторой величине, характеризующей падающий световой поток. Для фоторезисторов в качестве такого параметра используют удельную интегральную чувствительность $S_{\Phi_{инт.уд.}}$, (мА/В·лм) к световому потоку:

$$S_{\Phi_{инт.уд.}} = \frac{I_{\Phi}}{\Phi \cdot U} \quad (6.4)$$

Как следует из семейства ВАХ (рис.6.1, а) чувствительность фоторезистора зависит от того, при каком напряжении она измеряется. Поэтому, для исключения влияния напряжения используют

удельную чувствительность $S_{\Phi_{инт.уд.}}$, приведённую к 1 вольту напряжения. Её величина для различных типов фоторезисторов колеблется в пределах от десятых долей до сотен единиц мА/В·лм.

При включении фоторезистора в измерительную цепь как составной части делителя (рис. 6.2, а) выходным сигналом является изменение падения напряжения. Если световой поток сформирован в виде импульсов, то на сопротивлении нагрузки R_n появится импульс напряжения с амплитудой $\Delta u_{вых} = \Delta I_{\Phi} \cdot R_n$, повторяющий форму импульса тока фоторезистора. В этом случае наиболее удобным параметром является вольтовая чувствительность S_U (В/лм) при данном рабочем напряжении:

$$S_U = \frac{\Delta u_{вых}}{\Phi} \quad (6.5)$$

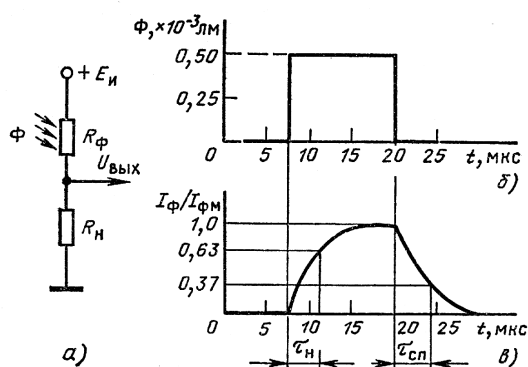


Рис.6.2. Измерительная цепь фоторезистора (а) и искажения, вносимые при преобразовании импульсного светового сигнала (б).

При облучении фоторезистора прямоугольными световыми импульсами форма выходного импульса искажается (рис. 6.2, в): ток нарастает и уменьшается с запаздыванием относительно светового импульса. Инерционность фоторезистора связана со временем жизни неравновесных носителей и определяется двумя параметрами: временем нарастания τ_n и временем спада $\tau_{сп}$. Эти параметры тесно связаны с граничной частотой $f_{гр}$. Она обратно пропорциональна времени жизни носителей, её средние значения составляют 10^4 Гц. Таким образом, чем больше время жизни неравновесных носителей, тем более инерционен фоторезистор.

6.3. Фотоэлектрический эффект в р-п переходе

При облучении р-п перехода светом с энергией квантов, превышающей ширину запрещенной зоны, во всей области поглощения происходит биполярная световая генерация подвижных носителей. Дальнейшая «судьба» этих электронов и дырок зависит от того, в каком месте они появились.

1. Если электронно-дырочные пары возникли непосредственно в переходе с шириной l , то они разделяются контактным электрическим полем и переходят в те области, где становятся основными носителями.

2. Если световая генерация происходит в частях р- и n-областей, прилегающих к границам перехода шириной порядка диффузионной длины неосновных носителей (L_p и L_n , см. раздел 4.6), то электронно-дырочные пары также разделяются. Основные носители остаются в месте возникновения электронно-дырочной пары, а неосновные диффундируют в область р-n перехода и затем за счет экстракции переходят в смежную область, где также становятся основными носителями.

3. Электронно-дырочные пары, образовавшиеся в глубине р- и n-областей, в течение времени жизни не попадают в контактное электрическое поле перехода и рекомбинируют.

Таким образом, при постоянном уровне облучения в области шириной $l + L_p + L_n$ происходят следующие процессы:

Ток, обусловленный электронами и дырками из разделённых полем перехода электронно-дырочных пар (фототок I_ϕ) совпадает по направлению с дрейфовым током неосновных носителей I_0 . Следовательно, через переход протекает суммарный дрейфовый ток $I_\phi + I_0$. Концентрации электронов в n-области и дырок в р-области повышаются, происходит частичная компенсация объёмного заряда неподвижных ионов примесей. Вследствие этого потенциальный барьер перехода понижается, и возникает диффузионный ток основных носителей $I_{\partial\phi}$, направленный навстречу дрейфовому току. Увеличиваясь по мере снижения барьера, диффузионный ток становится равным встречному дрейфовому току. В результате на переходе при постоянном уровне облучения устанавливается равновесие, при котором суммарный ток $I = I_{\partial\phi} - I_\phi - I_0$ равен нулю, и на переходе существует фото-ЭДС или напряжение холостого хода U_{xx} , обусловленное накопленными основными носителями в р- и n-областях.

Как было установлено ранее (см. раздел 5.5), диффузионный ток основных носителей при данном прямом напряжении (в нашем случае U_{xx}) равен $I_{\partial\phi} = I_0 \exp\left(\frac{eU_{xx}}{kT}\right)$. Для равновесного состояния I_ϕ

$= I_{\partial\phi} - I_0$, следовательно $I_{\phi} = I_0 \exp\left(\frac{eU_{xx}}{kT} - 1\right)$, и напряжение холостого хода представляется в виде:

$$U_{xx} = \frac{kT}{e} \ln\left(1 + \frac{I_{\phi}}{I_0}\right) \quad (6.6)$$

На энергетической диаграмме освещённого перехода в режиме холостого хода (рис. 6.3, б) видно, что световой поток смещает уровни Ферми р- и n-областей на величину eU_{xx} из-за повышения концентраций основных носителей, что приводит к снижению потенциального барьера на ту же величину.

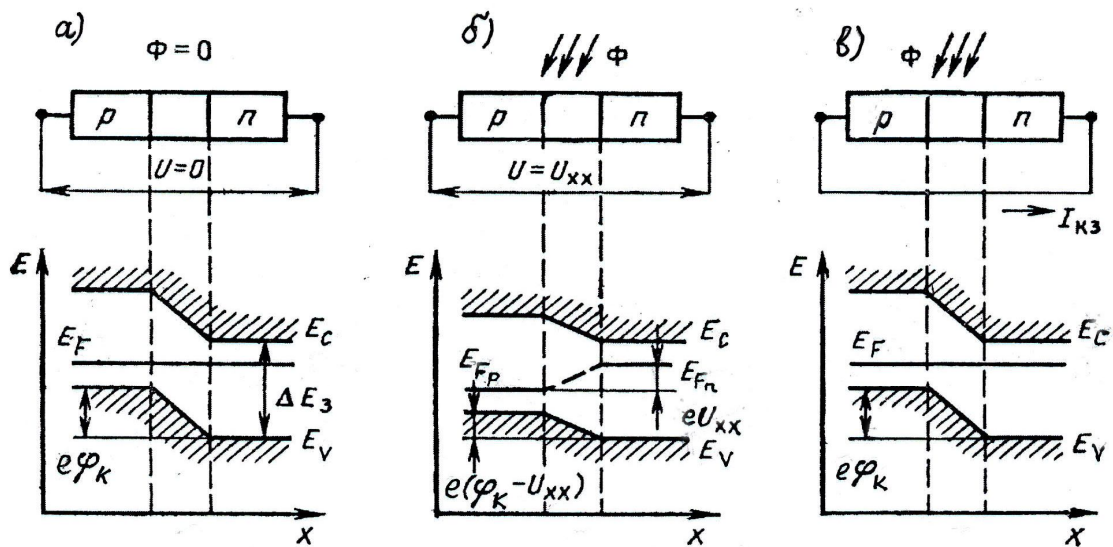


Рис. 6.3. Энергетические диаграммы р-п перехода: а) в отсутствие освещения; б) при освещении в режиме холостого хода, в) при освещении в режиме короткого замыкания.

При замыкании р- и n-областей (рис. 6.3, в) через переход течёт ток короткого замыкания:

$$I_{\phi} = I_{кз} = eSG(l + L_n + L_p) \quad (6.7),$$

где G – скорость биполярной световой генерации, S – площадь перехода. При коротком замыкании перехода генерируемые светом носители не накапливаются в смежных областях и не компенсируют объёмный заряд ионов примесей. Они образуют дрейфовый ток короткого замыкания во внешней цепи. В силу этих причин энергетическая диаграмма для данного случая полностью подобна равновес-

ной диаграмме p-n перехода при отсутствии освещения (см. рис. 6.3,а, в).

Вольтамперная характеристика p-n перехода, облучённого светом, (рис. 6.4) пересекает ось прямых напряжений в точке $U = U_{xx}$ и ось токов при $I = I_{кз}$, т. е. она смещена вниз относительно ВАХ в отсутствие освещения тем больше, чем выше значение светового потока. На характеристике при данном значении светового потока можно выделить две области, в которых возможно практическое применение фотоэффекта в p-n переходе. Область обратных напряжений (III) – режим фотодиода, область IV – режим фотоэлемента.

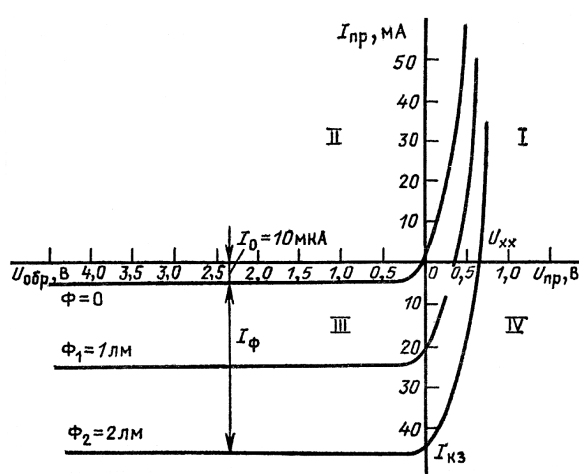


Рис. 6.4. Пример вольтамперной характеристики p-n перехода при разных уровнях освещенности.

В режиме фотоэлемента (IV) p-n переход является источником электрической энергии. При включении на-

грузки параллельно переходу протекающий в цепи фототок обеспечивает полезную мощность, выделяющуюся за счёт падения напряжения на нагрузочном сопротивлении. Такие элементы на основе кремния собираются в последовательно-параллельно соединённые блоки и служат основой для изготовления солнечных батарей.

При обратном смещении перехода и освещении его светом работают широко распространённые фотоприёмники – фотодиоды. Их вольтамперная характеристика представляет собой часть ВАХ перехода, облучаемого светом, лежащую в области III (рис. 6.4). Ток фотодиода образован неосновными носителями, концентрация которых увеличивается пропорционально световому потоку при освещении. Так же, как и для фоторезистора, фототок I_{ϕ} равен общему току I за вычетом темнового тока I_0 . Основным параметром фотодиода – интегральная токовая чувствительность S_I , определяемая отношением фототока либо к световому потоку (мА/лм), либо к мощности светового облучения (мА/Вт):

$$S_{I\Phi} = I_{\Phi} / \Phi; S_{IP} = I_{\Phi} / P \quad (6.8)$$

В отличие от фоторезистора, $S_{I\Phi}$ и S_{IP} не зависят от напряжения, так как фототок при изменении напряжения на фотодиоде остаётся постоянным.

6.4. Излучение света полупроводниками. Светодиоды

Излучающие полупроводниковые приборы преобразуют электрическую энергию в энергию оптического излучения в узком диапазоне длин волн. Их принцип действия основан на явлении излучательной рекомбинации. Для того чтобы поддерживать излучение света из полупроводника, необходимо создавать условия для поддержания высокой избыточной концентрации подвижных носителей. Наиболее простой и очевидный способ – использование явления инжекции неосновных носителей в прямо смещенном p-n переходе.

Полупроводниковые переходы могут быть источниками когерентного и некогерентного излучения. В первом случае это – инжекционные лазеры, во втором – светодиоды. В данном разделе будут рассмотрены физические механизмы некогерентного светового излучения, обусловленного самопроизвольной излучательной рекомбинацией носителей заряда, инжектированных через p-n переход.

Наибольший интерес представляет излучение света в видимой области, поэтому необходимо использовать полупроводники с широкой запрещённой зоной (1,8–3,4 эВ). Основные полупроводниковые материалы для светодиодов: нитрид галлия GaN ($\Delta E_z = 3,39$ эВ), фосфид галлия GaP ($\Delta E_z = 2,27$ эВ), арсенид галлия ($\Delta E_z = 1,4$ эВ), карбид кремния ($\Delta E_z = 2,36$ – $3,05$ эВ, в зависимости от типа кристаллической решетки). Используются также тройные соединения, например $\text{GaAs}_{1-x}\text{P}_x$, в которых путём изменения соотношения атомов мышьяка и фосфора в интервале значений $0 \leq x \leq 0,45$ можно изменять ширину запрещённой зоны для прямых излучательных переходов от 1,4 эВ до 2,0 эВ.

Поскольку вероятность такой рекомбинации невелика (см. раздел 4.5), важнейшей задачей является увеличение доли излучательных переходов между уровнями или энергетическими зонами. Эта

задача решается, прежде всего, путём использования полупроводников, в которых происходит межзонная излучательная рекомбинация. Такие полупроводники называются прямозонными; в них рекомбинация электрона и дырки с испусканием кванта света происходит без участия кристаллической решетки, т. е. без изменения импульса системы электрон-дырка (см. раздел 4.5).

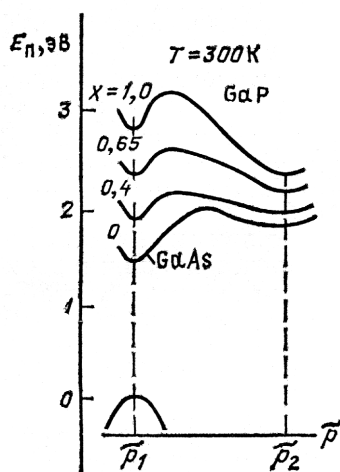


Рис. 6.5. Зависимость энергии электронов от квазиимпульса для арсенида галлия, фосфида галлия и тройного соединения $\text{GaAs}_{1-x}\text{P}_x$

Модель плоских энергетических зон не пригодна для отображения физической картины процессов излучательной и безызлучательной рекомбинации. Для этого необходимо учесть зависимость энергии, соответствующей краям зон, от квазиимпульса частиц. Из рис. 6.5 следует, что арсенид галлия GaAs – прямозонный полупроводник: абсолютный минимум энергии дна зоны проводимости и максимум энергии потолка валентной зоны находятся при одном и том же значении квазиимпульса \tilde{p}_1 . Напротив, фосфид галлия GaP – не прямозонный полупроводник, так как абсолютные экстремумы краёв зон находятся при разных значениях квазиимпульса \tilde{p}_1 и \tilde{p}_2 . Здесь же хорошо видно, что трёхкомпонентное вещество $\text{GaAs}_{1-x}\text{P}_x$ остаётся прямозонным полупроводником только при увеличении доли фосфора от 0 до 0,45. Несмотря на то, что в GaP невозможны прямые переходы, он широко используется как материал для изготовления светодиодов видимой области спектра. Для этого в полупроводник вводят чужеродные атомы для образования ловушек – центров излучательной рекомбинации. Эти атомы образуют ловушки донорного (S , O_2) и акцепторного типов (Cd , Zn , Cu). Такие же примеси вводятся в трёхкомпонентный полупроводник $\text{GaAs}_{1-x}\text{P}_x$ при $x > 0,45$. При этом открывается возможность излучательной рекомбинации с участием ловушек за счёт прямых переходов на конечной стадии ступенчатой

рекомбинации (см. рис. 4.3) и за счёт таких же переходов между уровнями ловушек донорного и акцепторного типов. В результате на основе фосфида галлия удалось создать светодиоды, дающие излучение в зелёной и красной части спектра.

В настоящее время для изготовления светодиодов часто используют гетеропереходы. Их основное преимущество – односторонняя инжекция из широкозонного полупроводника в узкозонный, что позволяет существенно увеличить интенсивность излучательной рекомбинации.