

3.7. Свободные электроны в кристаллической решетке твердого тела

Любое твердое тело состоит из атомов. Если атомы в твердом теле располагаются периодически, то такое тело называется кристаллическим. Если же этого не имеет места, то оно называется аморфным. Таким образом, у кристаллических тел наблюдается дальний порядок расположения атомов, у аморфных он отсутствует. В настоящее время микроскопическая теория кристаллического состояния в общих чертах разработана достаточно хорошо, микроскопическая теория аморфного состояния находится в стадии развития, и ее результаты все еще недостаточны, чтобы их можно было бы излагать в курсе общей физики.

Любое кристаллическое твердое тело состоит из огромного числа периодически расположенных в пространстве атомов. Для описания их совместного движения необходимо составить число уравнений, равное полному числу частиц, входящих в состав твердого тела. Но в такой постановке задача оказывается невыполнимой. Поэтому для описания движения системы частиц используются многочисленные соответствующие опыту упрощающие предположения.

Рассмотрим эту задачу, применяя для ее решения методы квантовой механики.

В общем виде уравнение Шредингера для системы частиц, входящих в состав кристаллической решетки твердого тела, имеет вид

$$\hat{H}\psi(\vec{r}, \vec{R}) = \varepsilon \psi(\vec{r}, \vec{R}), \quad (3.7.1)$$

где

$$\varepsilon = \varepsilon_e + \varepsilon_{я}. \quad (3.7.2)$$

В (3.7.1) \hat{H} – оператор Гамильтона всех частиц твердого тела, $\psi(\vec{r}, \vec{R})$ – волновая функция всей системы частиц, состоящей из электронов и ядер, \vec{r}, \vec{R} – соответственно координаты электронов и ядер, $\varepsilon_e, \varepsilon_{я}$ – соответственно полная энергия всех электронов и ядер.

Уравнение (3.7.1) в общем виде в принципе решить невозможно. Поэтому для дальнейшего анализа используют различные приближения.

7.1. Адиабатическое приближение

Из опыта известно, что масса электрона всегда значительно меньше массы ядра любого атома. Поэтому при одной и той же кинетической энергии скорость электрона многократно больше скорости ядра. Следовательно, при рассмотрении этой системы частиц можно кон-

статировать, что по сравнению с электронами ядра атомов можно принять за неподвижные или перемещающимися с очень малыми скоростями. Это позволяет считать движение электронов и ядер независимыми. В теории вероятностей имеется теорема: вероятность двух независимых событий равна произведению вероятностей каждого из них. Поэтому плотность вероятности такого сложного события равна произведению плотностей вероятности каждого из них:

$$\psi^2(\vec{r}, \vec{R}) = \psi^2(\vec{r}) \psi^2(\vec{R}). \quad (3.7.3)$$

Тогда результирующая волновая функция будет равна произведению исходных волновых функций:

$$\psi(\vec{r}, \vec{R}) = \psi(\vec{r}) \psi(\vec{R}) = \psi_e \psi_j. \quad (3.7.4)$$

Так как энергия – аддитивная величина, то оператор Гамильтона также аддитивен:

$$\hat{H} = \hat{H}_e + \hat{H}_j. \quad (3.7.5)$$

Подставим значения из (3.7.2)–(3.7.5) в уравнение (3.7.1) и разделим переменные:

$$\begin{aligned} (\hat{H}_e + \hat{H}_j) \psi_e \psi_j &= (\varepsilon_e + \varepsilon_j) \psi_e \psi_j; & \frac{\hat{H}_e \psi_e}{\psi_e} + \frac{\hat{H}_j \psi_j}{\psi_j} &= \varepsilon_e + \varepsilon_j; \\ \frac{\hat{H}_e \psi_e}{\psi_e} - \varepsilon_e &= \varepsilon_j - \frac{\hat{H}_j \psi_j}{\psi_j} = 0; & \hat{H}_e \psi_e &= \varepsilon_e \psi_e; \hat{H}_j \psi_j = \varepsilon_j \psi_j. \end{aligned}$$

Теперь уравнение Шредингера (3.7.1) распалось на два независимых уравнения: одно – для электронов, другое – для ядер. Однако и эти уравнения все еще оказываются слишком сложными для их решения. Поэтому используется следующее приближение.

7.2. Приближение самосогласованного поля

В этом приближении используются следующие допущения.

1. В процессе образования кристаллической решетки валентный электрон отрывается от своего атома.

2. Все электроны, кроме валентных, вместе с положительно заряженными ядрами создают в кристалле результирующее электрическое поле, в котором происходит движение валентных электронов. А так как ядра атомов твердого тела расположены в пространстве периодически, то движение валентных электронов происходит в периодическом поле кристаллической решетки.

3. Валентные электроны не взаимодействуют между собой, поэтому они образуют ЭФГ. С учетом принятых допущений потенциальная

энергия валентного электрона в электрическом поле кристаллической решетки (псевдопотенциал валентного электрона)

$$V(\vec{r}, \vec{k}) = U(\vec{k}) + U_c(\vec{r}), \quad (3.7.7)$$

где $U(\vec{k})$ – потенциальная энергия валентного электрона в электрическом поле ядер; $U_c(\vec{r})$ – потенциальная энергия валентного электрона в электрическом поле всех остальных электронов, кроме валентного, т.е. потенциальная энергия самосогласованного поля.

Применив такой подход, можно многочастичную задачу свести к одночастичной, для которой уравнение Шредингера имеет вид:

$$\Delta\psi + \frac{2m_e}{\hbar^2} (\epsilon - V(\vec{r}, \vec{k}))\psi = 0. \quad (3.7.8)$$

Введение самосогласованного поля позволяет все валентные электроны твердого тела рассматривать как систему невзаимодействующих частиц, т.е. как ЭФГ. Но и в этом случае задача все еще оказывается достаточно сложной. Поэтому вводятся два приближения, относящиеся к двум предельным случаям: приближение почти свободных электронов и приближение сильной связи.

7.3. Приближение почти свободных электронов

В этом приближении псевдопотенциал $V(\vec{r}, \vec{k})$ записывают в следующем виде:

$$V(\vec{r}, \vec{k}) = V_0 + V_1(\vec{k}), \quad (3.7.9)$$

где $V_0 = \text{const}$, $V_1(\vec{k})$ – периодическая функция, $V_1(\vec{k}) \ll V_0$; $V_1(\vec{k}) = V_1(\vec{k} + \vec{a})$, \vec{a} – вектор трансляции, вектор переноса координаты от одного узла кристаллической решетки к ближайшему узлу решетки в прямом направлении.

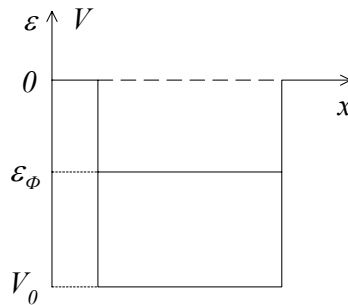


Рис. 3.7.1. Зона разрешенных энергий

Тогда $\varepsilon - V(\vec{r}, \vec{k}) \approx \varepsilon - V_0$.

На рис. 3.7.1 в качестве примера изображена зона разрешенных значений полных энергий свободных электронов в виде потенциальной ямы с плоским дном. При $T = 0$ К в интервале энергий $V_0 \leq \varepsilon \leq \varepsilon_{\text{Ф}}$ все уровни полностью заполнены электронами, а в интервале энергий $\varepsilon_{\text{Ф}} < \varepsilon \leq 0$ все уровни свободны. При $T > 0$ К эти уровни заполняются в соответствии с распределением Ферми-Дирака. Указанное приближение хорошо выполняется для металлов.

7.4. Приближение сильной связи

В этом приближении предполагается, что $U_c(\vec{r}) \ll U(\vec{k})$, и поэтому в соответствии с (3.7.7) имеем: $V(\vec{r}, \vec{k}) \approx U(\vec{k})$. Кроме того

$$\varepsilon = \varepsilon_a + \varepsilon(\vec{k}), \quad (3.7.10)$$

ε_a – собственное значение полной энергии электрона в изолированном атоме, $\varepsilon(\vec{k})$ – дополнительная энергия электрона, возникающая в электрическом поле кристаллической решетки, $\vec{k} = \vec{k}(\vec{r})$ – волновой вектор, $k = 2\pi / \lambda$, λ – длина волны де Бройля.

Анализ показывает, что решение уравнения Шредингера возможно лишь для определенных значений полных энергий электрона, значит существуют разрешенные и запрещенные значения полной собственной энергии. В отсутствии внешних полей составляющая полной энергии электрона при движении в направлении координатной оси Ox

$$\varepsilon_0 = \frac{p_x^2}{2m_e} = \frac{\hbar^2}{2m_e} k_x^2. \quad (3.7.11)$$

Наличие внешних воздействий делают зависимость (3.7.11) более сложной, и ее в общем виде можно записать лишь функционально.

На рис. 3.7.2 изображены графики полной энергии свободного электрона в отсутствии внешних воздействий (пунктирная линия) и с учетом действия электрического поля кристаллической решетки (сплошные линии). Если графики спроектировать на плоскость $\varepsilon - y$, то получим зоны разрешенных значений энергии, разделенные зонами запрещенных значений энергии.

Интервал значений проекций волновых векторов на ось Ox

$$-\frac{2\pi}{d} \leq k_x \leq \frac{2\pi}{d}$$

называется первой зоной Бриллюэна, интервалы значений

$$-2\frac{2\pi}{d} \leq k_x \leq -\frac{2\pi}{d} \quad \text{и} \quad \frac{2\pi}{d} \leq k_x \leq 2\frac{2\pi}{d}$$

называется 2-ой зоной Бриллюэна и т.д., d – постоянная кристаллической решетки.

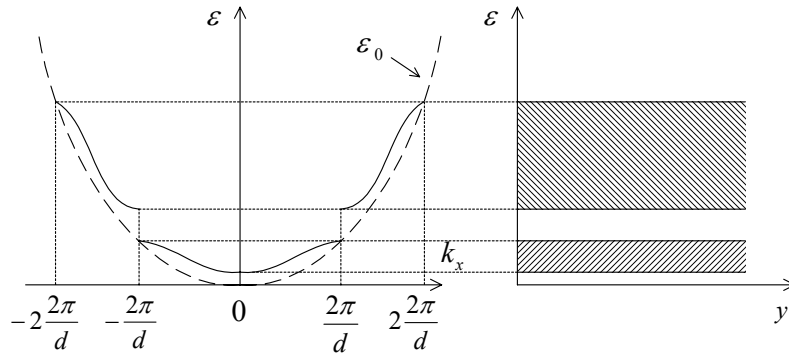


Рис. 3.7.2. Зависимость энергии электрона от волнового числа

Из рис. 3.7.2 видно, что ширина зон разрешенных значений энергий возрастает с ростом порядкового номера зоны Бриллюэна. Поэтому ширина каждой верхней зоны всегда больше соседней нижней зоны.

3.8. Зонная теория твердых тел

Из квантово-механической теории твердого состояния следует, что как в приближении почти свободных электронов (слабая связь валентного электрона с кристаллической решеткой), так и в приближении сильной связи у твердого тела возникают зоны разрешенных значений энергии, разделенные зонами запрещенных энергий. Рассмотрим более подробно физическую причину возникновения энергетических зон.

Среди металлов наиболее простое объяснение высокой электропроводности на основе зонной теории можно дать металлическому Li . Атом Li содержит 3 электрона в состояниях $1s^2 2s^1$. Если расстояние между соседними атомами превышает 0,35 нм, то на таких расстояниях атомы взаимодействуют очень слабо, в пространстве их электроны разделены потенциальными барьерами. В состоянии $1s$ находятся два электрона, в состоянии $2s$ – один электрон. В процессе кристаллизации расстояния между ближайшими атомами постепенно уменьшают-

ся до тех пор, пока они не окажутся равными постоянной кристаллической решетки (0,35 нм). На этих расстояниях электронные оболочки $2s$ соседних атомов перекрываются. Возникает зона разрешенных значений энергии $2s$. В результате взаимодействий между электронами в зоне $2s$ возникает $2N$ уровней полной энергии: N – число атомов в кристаллической решетке, 2 – два направления собственного момента импульса электронов. Но так как число электронов в состоянии $2s$ равно N , то из всех уровней полной энергии заполненными окажутся лишь $N/2$. Что же касается электронов $1s$, то их число равно $2N$, и все уровни этого состояния окажутся заполненными полностью (рис. 3.8.1).

Электроны $1s$ у соседних атомов разделены потенциальными барьерами весьма малой толщины. Причина возникновения потенциального барьера следующая. Электрон $1s$ находится в электрическом поле ядра и второго электрона $1s$, т.е. в электрическом поле, создаваемом эффективным зарядом $q^* = 3e - 1e = 2e$, где e – заряд электрона, $3e$ – заряд ядра. Электрическое поле заряда q^* обладает сферической симметрией. Поэтому валентный электрон $1s$ находится в электрическом поле кулоновских сил и обладает потенциальной энергией в электрическом поле остова атома:

$$U_{\text{Li}} = -\frac{(Z-2)e^2}{4\pi\epsilon_0 r}. \quad (3.8.1)$$

Электроны $1s$ находятся в непрерывном движении и туннельным эффектом могут преодолевать потенциальный барьер между атомами. А так как согласно Паули в одном состоянии не могут находиться два и большее число электронов, то в зоне $1s$ возникает число уровней полной энергии, равное числу атомов в кристаллической решетке твердого тела. Электроны $2s$ согласно принципу Паули заполняют лишь половину зоны $2s$. Тогда в зоне $2s$ имеется большое число свободных уровней полной энергии. Если теперь в таком кристалле существует электрическое поле, то в нем возникает направленное движение свободных электронов $2s$ – электрический ток. Такой кристалл является хорошим проводником, а в твердом состоянии – металлом.

Наличие частично незаполненной электронами зоны разрешенных значений энергии является необходимым условием возникновения тока проводимости. Только в этом случае под действием существующего в твердом теле электрического поля свободные электроны могут участвовать в направленном движении, только тогда возникают физические условия, при которых под действием поля на расстоянии, рав-

ном средней длине свободного пробега, электрон приобретает дополнительную скорость направленного движения, дополнительную кинетическую энергию.

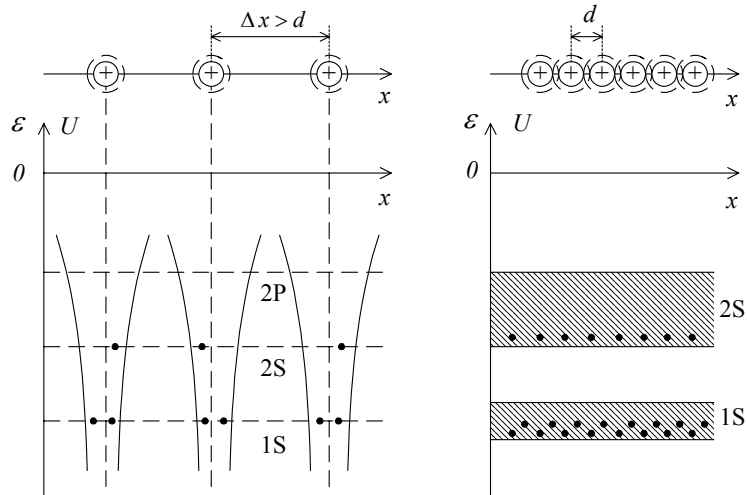


Рис. 3.8.1. Электрон в периодическом поле

Теперь рассмотрим твердое тело, образованное атомами Mg, у которого распределение электронов по состояниям $1s^2 2s^2 2p^6 3s^2$. У кристаллического Mg зона $3s$ заполнена полностью, зона $3p$ свободна. Однако зоны $3s$ и $3p$ частично перекрываются. Поэтому в зоне $3p$ всегда будет достаточное число уровней полной энергии, на которые под действием различных возбуждений могут переходить электроны с нижних уровней на верхние (рис. 3.8.2).

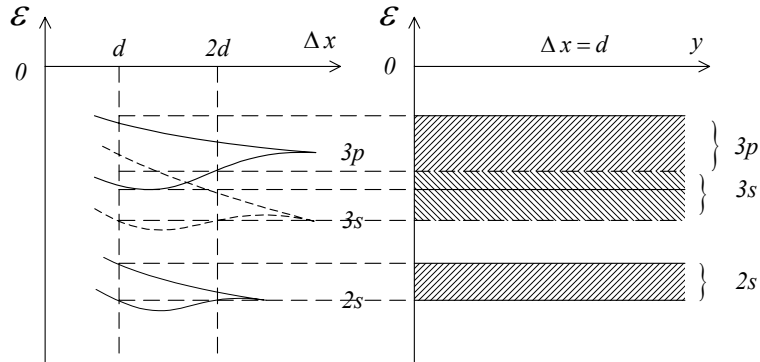


Рис. 3.8.2. Перекрывание энергетических зон

При наличии в твердом теле созданного внешним источником электрического поля эти электроны будут участвовать в образовании тока проводимости. Значит, кристаллический Mg должен обладать хорошей электропроводностью, должен быть хорошим проводником электрического тока.

Теперь предположим, что валентная зона заполнена полностью, но не перекрывает расположенную выше зону проводимости. Тогда величина электропроводности твердого тела зависит от ширины запрещенной зоны $\Delta\varepsilon$. Если $\Delta\varepsilon < 1\text{эВ}$, то в результате столкновений с фонами кристаллической решетки значительное число электронов валентной зоны приобретают дополнительную энергию и переходят в зону проводимости (рис. 3.8.3). К таким веществам, в частности, относятся Si и Ge. Так, у Si электроны атома по состояниям распределены следующим образом: $1s^2, 2s^2, 2p^6, 3s^2, 3p^2$. Валентными электронами являются электроны $3s$ и $3p$.

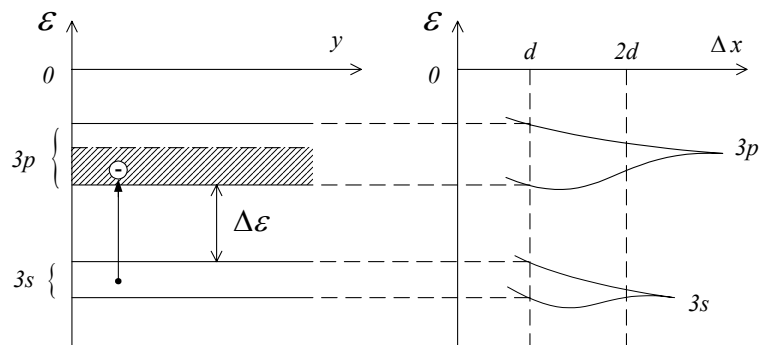


Рис. 3.8.3. Образование зоны проводимости

При комнатной температуре в результате столкновений с фонами кристаллической решетки некоторые из валентных электронов получают тепловое возбуждение, достаточное для перехода из валентной зоны в зону проводимости. В результате в зоне проводимости будет находиться значительное число электронов, которые под действием внешнего электрического поля будут перемещаться, создавая ток проводимости. Такой кристалл называется собственным полупроводником. В отличие от металлов, у собственных полупроводников с ростом температуры растет число электронов в зоне проводимости и уменьшается электрическое сопротивление. Но при $T = 0\text{ К}$ в зоне

проводимости свободных электронов нет, и собственный полупроводник превращается в диэлектрик. На рис. 3.8.3 изображена эволюция графиков полной энергии собственного полупроводника с изменением расстояния между соседними атомами Si : видно, как по мере уменьшения расстояния между атомами уровни $3s$ и $3p$ постепенно расщепляются, чтобы на расстоянии $\Delta x = d$ превратиться в зону.

На рис. 3.8.3 видно также, что в результате переходов некоторого числа электронов из валентной зоны $3s$ в зону проводимости $3p$ часть зоны проводимости заполняется электронами. Поскольку в зоне проводимости имеется множество свободных уровней, то под действием даже слабых возбуждений свободные электроны способны участвовать в токе проводимости. Но так как число свободных электронов в зоне проводимости на много порядков меньше числа свободных электронов у металлов, то и электрическое сопротивление собственного полупроводника будет значительно больше электрического сопротивления металла.

Если же $\Delta \epsilon > 1 \text{ эВ}$, то энергии фононов теперь уже будет недостаточно, чтобы при столкновении с ними свободные электроны получили энергию достаточную для перехода из валентной зоны в зону проводимости. В результате свободные электроны в зоне проводимости практически отсутствуют, свободных носителей заряда в ней нет. Такие вещества электрический ток не проводят. Это диэлектрики.

Наконец, может быть предельный случай, когда $\Delta \epsilon \approx 0 \text{ эВ}$. Это означает, что валентная зона и зона проводимости соприкасаются между собой, причем зона проводимости даже при $T = 0 \text{ К}$ должна быть частично заполнена электронами (рис. 3.8.4).

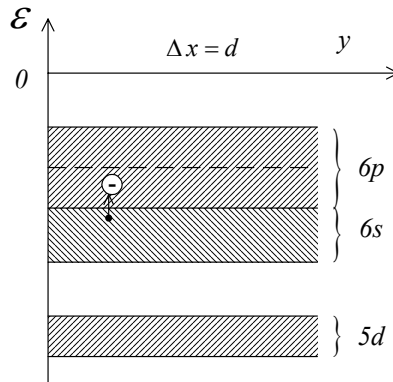


Рис. 3.8.4. Энергетические зоны висмута

Примером может служить Bi , у которого внешние электроны находятся в состояниях: $5d^{10} 6s^2 6p^3$. При $T = 0 \text{ K}$ зона $6s$ заполнена полностью, зона $6p$ заполнена частично. При $T > 0 \text{ K}$ вследствие переходов электронов из зоны $6s$ зона $6p$ заполняется дополнительно. Так как зоны $6p$ и $6s$ соприкасаются ($\Delta\varepsilon = 0 \text{ эВ}$), то подобные переходы имеют место при любой температуре. Такое вещество обладает высокой электропроводностью, но его электрическое сопротивление с ростом температуры уменьшается в области высоких температур и увеличивается в области низких. Это связано с тем, что при высоких температурах в зону проводимости в значительном количестве поступают электроны из валентной зоны. Но так как при высокой температуре энергия фононов достаточно велика, следовательно, и возбуждения электронов валентной зоны при столкновении с фононами оказываются значительными, что позволяет электронам переходить в зоны проводимости с более глубоких слоев валентной зоны. Такое вещество при высоких температурах ведет себя как полупроводник. Наоборот, при низких температурах поступление электронов из валентной зоны в зону проводимости невелико, влияние их на изменение концентрации свободных электронов зоны проводимости незначительно. Поэтому при низких температурах такое вещество ведет себя как металл.

Твердые тела, обладающие металлическими и полупроводниковыми свойствами, называются полуметаллами.

3.9. Эффективная масса квазиэлектрона в твердом теле

Под действием приложенной разности потенциалов в проводнике возникает электрический ток, т.е. направленный поток электронов. С микроскопической точки зрения это означает, что под действием электрического поля в проводнике возникает направленное ускоренное движение электронов, в результате на расстоянии, равном средней длине свободного пробега, электрон приобретает добавочную энергию – энергию возбуждения. Наличие энергии возбуждения наглядно проявляется на примере автоэлектронной эмиссии, когда вблизи поверхности металла под действием внешнего поля возникает потенциальный барьер конечной толщины. Электроны проводимости с определенной, хотя и весьма малой, вероятностью могут преодолеть его туннельным эффектом и вылететь в вакуум. Но вылететь может не любой электрон, а только тот, который получил элементарное возбуждение $\Delta\varepsilon$.

У полупроводников элементарные возбуждения электронов называются квазиэлектронами. Они наделяются скоростью, квазиимпульсом, энергией и эффективной массой.

По определению, квазиимпульс равен

$$\vec{p} = \eta \vec{k}. \quad (3.9.1)$$

Тогда сила, действующая на квазиэлектрон,

$$\vec{F} = \frac{d\vec{p}}{dt} = \eta \frac{d\vec{k}}{dt}. \quad (3.9.2)$$

Энергия квазиэлектрона

$$\varepsilon = \frac{p^2}{2m_e^*} = \frac{\eta}{2m_e^*} k^2, \quad (3.9.3)$$

где m_e^* – эффективная масса квазиэлектрона. Скорость квазиэлектрона

$$\vec{v} = \frac{\eta}{m_e^*} \vec{k}. \quad (3.9.4)$$

Тогда производная от полной энергии по волновому числу с учетом неравенства

$$2 \gg \frac{k}{m_e^*} \frac{dm_e^*}{dk}, \quad (3.9.5)$$

$$\frac{d\varepsilon}{dk} = \frac{\eta^2}{m_e^*} k.$$

Если квазиэлектрон находится в электрическом поле напряженностью \vec{E} , то на квазиэлектрон действует сила

$$F = eE = m_e^* \frac{dv}{dt}. \quad (3.9.6)$$

Из совместного решения уравнений (3.8.4) и (3.8.5) получаем:

$$v = \frac{1}{\eta} \frac{d\varepsilon}{dk}, \quad (3.9.7)$$

$$\frac{1}{m_e^*} = \frac{1}{\eta} \frac{dv}{dk}. \quad (3.9.8)$$

На рис. 3.9.1 слева изображены графики полной энергии ε , скорости v , эффективной массы m_e^* квазиэлектрона, а справа изображена зона разрешенных значений энергии для первой зоны Бриллюэна. Из рис. 3.9.1 следует, что вблизи дна зоны эффективная масса m_e^* положительна, вблизи верхней части зоны эффективная масса отрицатель-

на, и где-то в средней части зоны эффективная масса испытывает разрыв непрерывности.

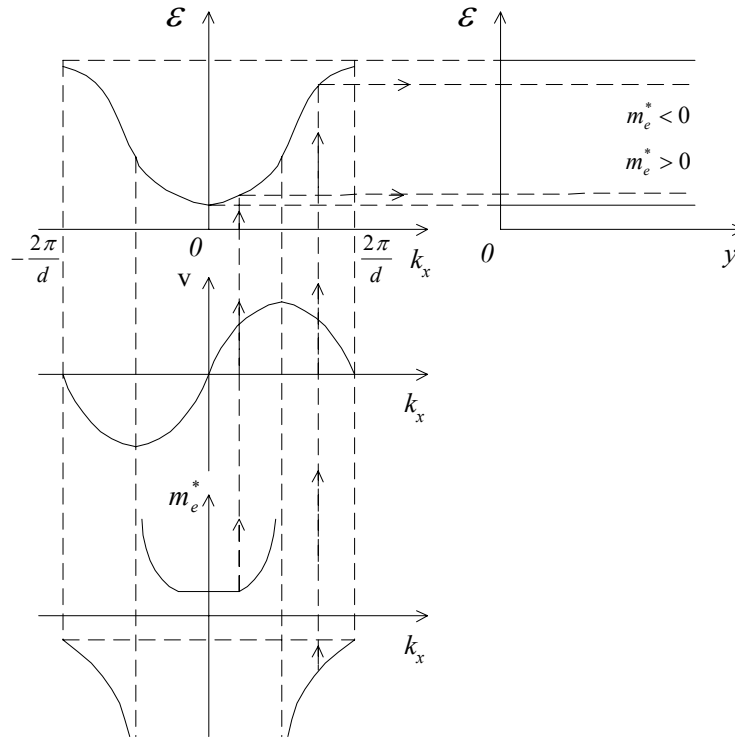


Рис. 3.9.1. К определению эффективной массы электрона

Подставив в (3.9.2) значение k из (3.9.4), получим

$$F = m_e^* \frac{dv}{dt}, \quad (3.9.9)$$

где

$$\frac{1}{m_e^*} = \frac{1}{\eta^2} \frac{d^2 \varepsilon}{dk^2}.$$

Следовательно, квазиэлектрон можно рассматривать как свободную частицу, движение которой в кристаллической решетке подчиняется 2-му закону Ньютона. Появление у квазиэлектрона эффективной массы вызвано влиянием периодического поля кристаллической решетки на движущийся в зоне проводимости валентный электрон.

Эффективная масса квазиэлектрона – это не обычная масса электрона, а всего лишь коэффициент пропорциональности между внешней силой F и ускорением $\frac{dv}{dt}$, определяющий составляющую направленного движения электрона на фоне его хаотического движения. Эффективная масса квазиэлектрона не определяет ни запаса энергии, ни гравитационные, ни инерционные свойства электрона. Больше того, в общем случае направление действия силы и ускорения в пространстве могут и не совпадать. Единственным оправданием введения такого понятия является привычность и простота законов классической механики, которые с введением указанного коэффициента можно достаточно просто использовать для анализа особенностей движения свободного электрона в периодическом поле кристаллической решетки при наличии в ней внешнего электрического поля, а также возможность ее экспериментального измерения. Кроме того, в целом ряде случаев величина измеряемой эффективной массы квазиэлектрона имеет тот же порядок величины, что и масса свободного электрона. Это показывает, что эффективная масса квазиэлектрона в какой-то мере отображает реальные инерционные свойства электрона. Проведенное рассмотрение справедливо лишь в приближении сильной связи. Опыт показывает, что в ряде случаев графики дисперсии $\epsilon(k)$ для полупроводников действительно имеют вид параболических зависимостей.

3.10. Дырочная проводимость

Для простоты рассуждений предположим, что в твердом теле имеется всего две зоны разрешенных значений энергии. Одна полностью заполнена электронами, другая полностью свободна, их разделяет зона запрещенных энергий. Теперь удалим электрон с одного из верхних уровней нижней зоны, для этого сообщим ему энергию, достаточную для перехода в верхнюю зону. Тогда в верхней зоне будет находиться электрон, а в нижней зоне освободится уровень полной энергии. Возникший свободный уровень, возникшее вакантное место в пределах заполненной зоны называется дыркой.

Электрон, попавший в верхнюю зону из нижней, под действием электрического поля будет перемещаться, при этом на расстоянии, равном средней длине свободного пробега, он будет постепенно накапливать дополнительную энергию, будет постепенно переходить на более высо-

кие уровни. В заполненной зоне существует свободный уровень полной энергии, там также возникает направленное движение электронов. Происходит процесс, аналогичный процессу перемещения электрона в верхней зоне, когда под действием поля электрон переходит с нижнего уровня на верхний свободный. В результате такого перехода в заполненной зоне освобождается нижележащий уровень полной энергии. Возникший вакантный уровень полной энергии наделяется свойством частицы – это квазичастица и называется дыркой.

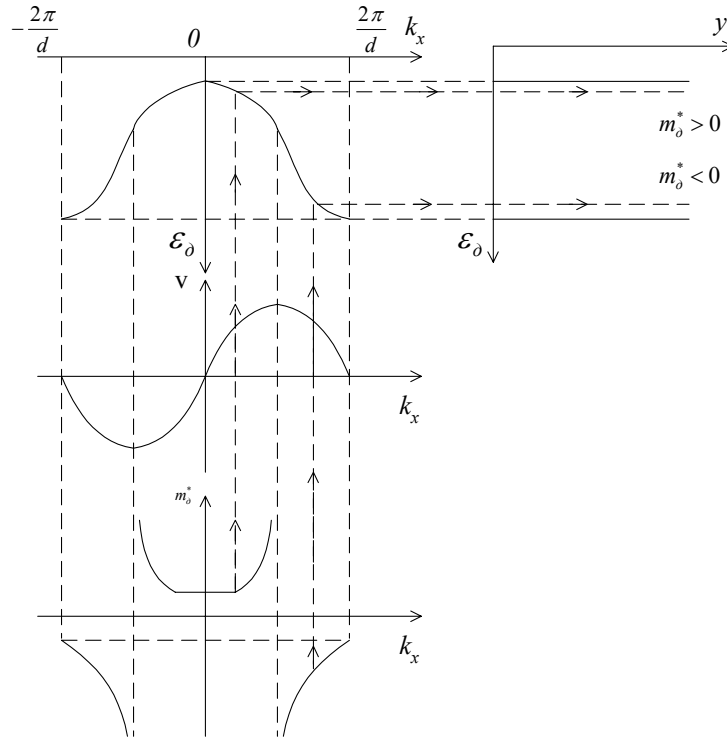


Рис. 3.10.1. К определению эффективной массы дырок

По мере возбуждения электронов, сопровождающегося переходами на выше расположенные соседние уровни, дырка постепенно будет опускаться в нижнюю часть заполненной зоны. Минимальным значением энергии дырка будет обладать на верхней границе заполненной зоны. В геометрическом пространстве дырка перемещается в направлении, обратном направлению движения электрона. Следовательно, под действием внешнего электрического поля дырка будет перемещаться точно так же,

как перемещалась бы положительно заряженная частица под действием того же поля. Стало быть, дырка как бы обладает положительным зарядом, равным заряду электрона, и, так как она – квазичастица, то также обладает эффективной массой $m_{\text{д}}^*$. Минимальным значением полной энергии дырка обладает на верхней границе заполненной зоны, поэтому отсчет энергии для дырок идет сверху вниз от верхней границы заполненной зоны. На рис. 3.10.1 по аналогии с квазиэлектронами приведены графики полной энергии ϵ , скорости v и эффективной массы $m_{\text{г}}^*$ дырок в первой зоне Бриллюэна, а также соответствующая заполненная зона разрешенных энергий. Из приведенных графиков видно, что в верхней части заполненной зоны эффективная масса дырки $m_{\text{д}}^*$ положительна, в нижней части зоны она отрицательна.

Если эти понятия применить к собственному полупроводнику, то окажется, что в зоне проводимости электрон обладает некоторой энергией возбуждения, т. е. там возникает квазиэлектрон, у которого минимум энергии находится на дне зоны проводимости и отсчет энергии ведется снизу вверх, а в валентной зоне – дырка, для нее энергия растет сверху вниз (рис. 3.10.2). Под действием электрического поля они перемещаются в противоположных направлениях: дырка в валентной зоне, электрон - в зоне проводимости.

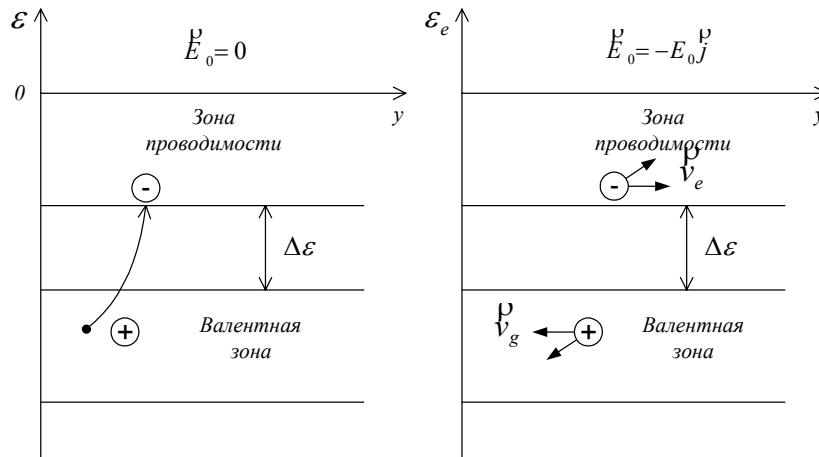


Рис. 3.10.2. Электронно-дырочная пара

Итак, в собственном полупроводнике сразу возникают две квазичастицы – квазиэлектрон и дырка. В нем по-прежнему структурной еди-

ницей материи является электрон, а структурными единицами движения являются квазиэлектроны и дырки, причем в геометрическом пространстве квазиэлектрон совпадает с электроном, что же касается дырок, то они с электроном пространственно разделены. Дырки прямого физического смысла не имеют. Тем не менее, на основе использования понятия дырок можно достаточно просто объяснить некоторые физические особенности полупроводников. Можно предложить такой наглядный образ дырки. При тепловом возбуждении одна из связей между соседними атомами разрывается, участвовавший в образовании связи электрон становится электроном проводимости. Затем в результате флуктуации тепловых колебаний ионов решетки связь разрывается в другой части кристалла, а освободившийся валентный электрон занимает место в предыдущей разорванной связи, и т.д. Каждая разорванная связь обладает положительным зарядом, энергией, эффективной массой. Если теперь в кристалле существует электрическое поле \vec{E}_0 , то под действием поля электроны и дырки будут двигаться в противоположные стороны, возникает ток проводимости.

3.11. Электрохимический потенциал в собственных полупроводниках

Пусть система, состоящая из N частиц, характеризуется постоянной температурой и объемом $(T, V) = \text{const}$. Такую систему можно описать, используя различные макроскопические параметры, в данном случае используя изохорно-изотермный потенциал, или, что тоже самое, свободную энергию:

$$F = U - TS, \quad (3.11.1)$$

U – внутренняя энергия, S – энтропия системы частиц, T – температура. По физическому смыслу свободная энергия является мерой той работы, которую может совершить система, находясь в изохорно-изотермном состоянии при переходе системы из одного фазового состояния в другое. Тогда химическим потенциалом называется величина, равная изменению свободной энергии при изменении числа незаряженных частиц в системе на единицу в изохорно-изотермном процессе:

$$\mu = \left(\frac{\partial F}{\partial N} \right)_{T,V}. \quad (3.11.2)$$

Если же в объеме находится система заряженных частиц, то свободная энергия системы заряженных частиц

$$F_q = F + N \varepsilon_n, \quad (3.11.3)$$

$\varepsilon_n = q\varphi$, ε_n – потенциальная энергия частицы с зарядом q , φ – потенциал электростатического поля, N – полное число частиц в системе.

По определению электрхимический потенциал заряженной частицы

$$\mu_{\circlearrowleft} = \left(\frac{\partial F_q}{\partial N} \right)_{V,T}, \quad (3.11.4)$$

$$\mu_{\circlearrowleft} = \mu + q\varphi.$$

Разделим полный объем V на два вспомогательных объема V_1 и V_2 . Тогда в объеме V_1 будет находиться N_1 частиц, а в объеме V_2 будет находиться N_2 частиц. Частицы движутся хаотически, поэтому между объемами V_1 и V_2 происходит обмен частицами, т.е. имеет место диффузионный контакт (рис. 3.11.1). Полное число частиц в объеме $N = N_1 + N_2 = \text{const}$. Следовательно, увеличение числа частиц в объеме V_1 означает убыль их числа в объеме V_2 : $dN_1 = -dN_2$. В общем случае изменение свободной энергии

$$dF = \mu_1 dN_1 + \mu_2 dN_2.$$

Но в состоянии термодинамического равновесия $dF = 0$, поэтому

$$\mu_1 dN_1 = -\mu_2 dN_2. \quad (3.11.5)$$

После подстановки в (3.11.5) $dN_1 = -dN_2$ получаем

$$\mu_1 = \mu_2 = \mu. \quad (3.11.6)$$

Итак, при термодинамическом равновесии химические потенциалы в различных частях системы частиц равны между собой.

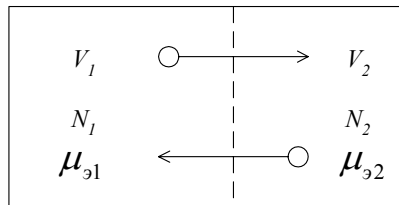


Рис. 3.11.1. Термодинамическое равновесие

Пусть теперь в объеме находятся N заряженных частиц. Тогда в общем случае

$$dF_q = \mu dN + \varepsilon_n dN.$$

Но при наличии термодинамического равновесия, когда $dF_{\bar{q}}=0$, $dN_1 = -dN_2$, имеем

$$\mu_{\bar{q}1} = \mu_{\bar{q}2} = \mu_{\bar{q}}. \quad (3.11.7)$$

Таким образом, при термодинамическом равновесии в различных частях пространства электрохимические потенциалы равны между собой. Эта особенность электрохимического потенциала позволяет использовать его для определения условия, при котором в системе имеет место термодинамическое равновесие. Неравенство между собой электрохимических потенциалов в различных микрообъемах указывает на протекание различных процессов переноса в пространстве.

Определим теперь положение уровня электрохимического потенциала в собственном полупроводнике, в котором существует система квазиэлектронов и дырок. Направления отсчета энергии у них противоположные: у квазиэлектронов начало отсчета энергии – на дне зоны проводимости, у дырок – на верхней границе валентной зоны.

Электрохимический потенциал – это энергетическая величина, определяющая состояние термодинамического равновесия в системе частиц или квазичастиц. Так как в собственном полупроводнике одновременно существуют квазиэлектроны и дырки, то электрохимический потенциал такой системы квазичастиц будет равен среднему значению его составляющих:

$$\mu_{\bar{q}} = \frac{\mu_{\bar{q}e} + \mu_{\bar{q}d}}{2}. \quad (3.11.8)$$

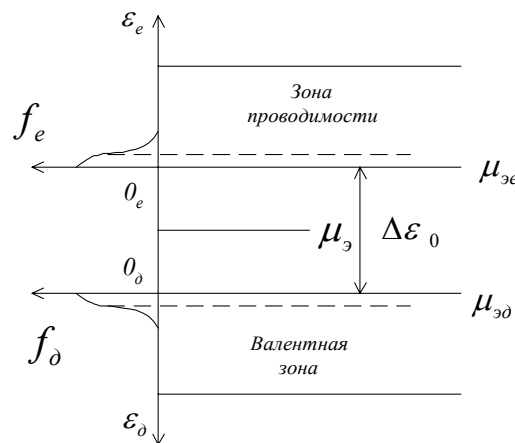


Рис. 3.11.2. Зоны собственного полупроводника

На рис. 3.11.2 изображены расположение зон в собственном полупроводнике, где f_e, f_d – функции Ферми-Дирака для квазиэлектронов и дырок:

$$f_e = \left(\exp\left(\frac{\varepsilon_e - \mu_{\varepsilon e}}{kT}\right) + 1 \right)^{-1}, \quad f_d = \left(\exp\left(\frac{\varepsilon_d - \mu_{\varepsilon d}}{kT}\right) + 1 \right)^{-1}. \quad (3.11.9)$$

Для вычислений функций распределения f_e и f_d необходимо знать электрохимический потенциал квазичастиц.

Ранее в разделе 3.5 было получено выражение для химического потенциала свободного электрона:

$$\mu = \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{kT}{\mu_0} \right)^2 \right), \quad (3.11.10)$$

$$\mu_0 = \frac{h^2}{8m^*} \left(\frac{3n}{\pi} \right)^{\frac{2}{3}}. \quad (3.11.11)$$

Формулы (3.11.10) и (3.11.11) полностью применимы для квазиэлектронов и дырок. Из опыта известно, что концентрация квазиэлектронов в металле примерно в 10^6 раз больше концентрации квазичастиц в собственных полупроводниках. Как видно из формулы (3.11.10), для вычисления химического потенциала необходимо знать эффективную массу квазичастиц. В табл. 3.11.1 приводятся относительные значения эффективных масс квазиэлектронов и дырок в Si и Ge, а также ширина запрещенной зоны $\Delta\varepsilon_0$ у каждого из этих полупроводников.

	$\frac{m_e^*}{m_e}$	$\frac{m_d^*}{m_e}$	$\Delta\varepsilon_0,$ ЭВ
<i>Si</i>	1,08	0,37	1,12
<i>Ge</i>	0,56	0,59	0,66

Табл. 3.11.1. Относительные значения эффективных масс квазиэлектронов, m_e^* , и дырок, m_d^* , и ширина запрещенной зоны, $\Delta\varepsilon_0$ в *Si* и *Ge*

Подставив найденные значения в (3.11.10), (3.11.11), получим:

$$\mu_{\varepsilon e} \approx 0, \quad \mu_{\varepsilon d} \approx 0.$$

Это означает, что функции Ферми-Дирака f_c и f_g для квазичастиц соответственно локализованы вблизи границ зоны проводимости и валентной зоны.

Из рис. 3.11.2 видно, что

$$\mu_{эд} = \mu_{эе} - \Delta\varepsilon_0. \quad (3.11.12)$$

Подставим $\mu_{эд}$ из (3.11.12) в (3.11.8):

$$\mu_{э} = -\frac{\Delta\varepsilon_0}{2}. \quad (3.11.13)$$

Итак, уровень электрохимического потенциала в собственном полупроводнике для системы, состоящей из квазиэлектронов и дырок, проходит через середину запрещенной зоны (рис. 3.11.2). Более строгий анализ показывает, что положение уровня электрохимического потенциала незначительно меняется с изменением температуры. Это связано с тем, что зависимость для квазиэлектронов более сильная, чем аналогичная зависимость для дырок.

3.12. Примесные полупроводники

Электропроводность собственных полупроводников очень сильно зависит от концентрации примесей, введенных в кристаллическую решетку. Так, введение в Si всего лишь 0,001% бора увеличивает электропроводность в 1000 раз. Повышенная электропроводность, обусловленная примесями, называется примесной проводимостью, а сами полупроводники – примесными полупроводниками.

Для выяснения механизма действия примесей на проводимость полупроводников рассмотрим влияние фосфора P и бора B на свойства Si . Чистый Si является собственным полупроводником. Атом Si имеет порядковый номер в периодической системе элементов $Z = 14$, поэтому в состав его электронной оболочки должно входить 14 электронов, распределенные по состояниям следующим образом: $1s^2 2s^2 2p^6 3s^2 3p^2$. Такая система электронов позволяет Si иметь кристаллическую решетку, в которой каждый атом окружен четырьмя ближайшими соседями. Связь между ними ковалентная и создается электронами $3s3p$. На рис. 3.12.1 изображена плоская модель кристаллической решетки.

Предположим, что некоторое число атомов Si заменена атомами фосфора P , у которого $Z=15$, а электроны распределены по состояниям $1s^2 2s^2 2p^6 3s^2 3p^3$. Значит, атом P имеет валентность, равную 5, она определяется электронами $3s^2 3p^3$. Внедренный атом P для образова-

ния связи с четырьмя ближайшими атомами использует 4 валентных электрона, пятый валентный электрон в образовании связи не участвует, он первоначально движется вокруг атома P (рис. 3.12.1), постепенно удаляясь от него. Тепловые столкновения с электронами и фононами кристаллической решетки приводят к тому, что электрон переходит в возбужденное состояние, а расстояние между электроном и остовом атома P увеличивается настолько, что между ними находится один или несколько атомов Si . Эти атомы экранируют электрическое поле остова атома P , в результате связь между электроном и остовом ослабляется настолько, что достаточно тепловых столкновений ионов Si с электроном, чтобы он оторвался от своего атома и превратился в электрон проводимости.

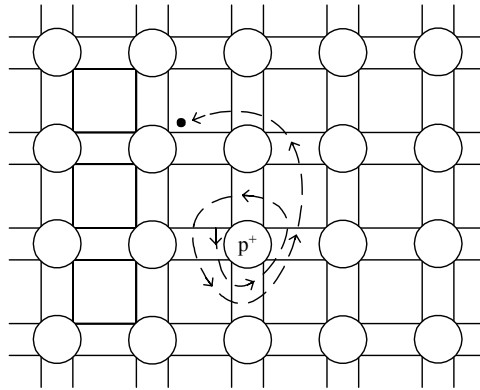


Рис. 3.12.1. Кристаллическая решетка кремния с примесью

Указанный наглядный процесс означает, что внедренный атом примеси создает в запрещенной зоне уровень полной энергии, на котором первоначально находится валентный электрон атома P . Этот уровень располагается вблизи дна зоны проводимости на расстоянии $\Delta\epsilon_{\text{дон}} = 0,044$ эВ и называется донорным уровнем, так как он является основным поставщиком электронов в зону проводимости (заполненный электроном уровень-донор). Действительно, в результате возбуждения атома P тепловыми колебаниями решетки валентный электрон $3r$ отрывается от атома и переходит с донорного уровня в зону проводимости (рис. 3.12.2), при этом атом P превращается в положительно заряженный ион, прочно связанный с ближайшими ионами кристаллической решетки. Положительный заряд иона под действием электрического поля в

образовании тока проводимости участвовать не может, в образовании тока проводимости будут участвовать электроны, перешедшие в зону проводимости как с примесных уровней, так и из валентной зоны. Ширина запрещенной зоны у Si $\Delta\varepsilon_0 = 1,12$ эВ.

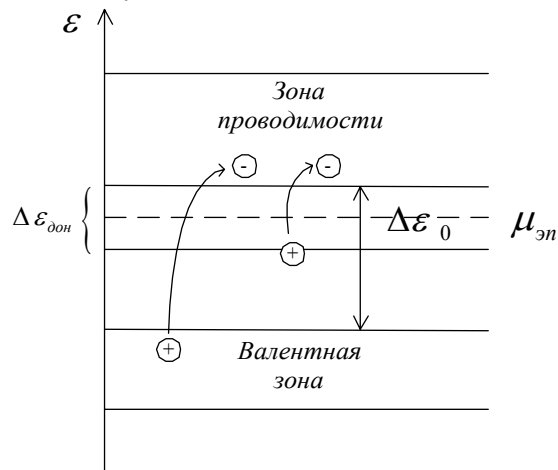


Рис. 3.12.2. Примесная электронная проводимость

Итак, концентрация электронов в зоне проводимости в основном определяется концентрацией атомов примеси. Возникшие же в результате переходов в валентной зоне дырки имеют концентрацию на много порядков меньше, чем концентрация электронов в зоне проводимости, и практически не влияют на величину электропроводности при комнатной температуре. Такой полупроводник обладает электронной проводимостью и называется полупроводником n-типа, в нем основными носителями заряда являются электроны, а неосновными носителями заряда являются дырки.

Предположим теперь, что в кристаллической решетке Si некоторое число атомов замещено атомами бора B , у которого электроны по состояниям распределены следующим образом: $1s^2 2s^2 2p^1$. Валентным электроном является электрон в состоянии $2p$. Для образования связей с четырьмя ближайшими соседними атомами Si атому B не хватает одного электрона. Его можно заимствовать у атома Si . Опыт показывает, что для отрыва валентного электрона $2p$ от атома Si требуется энергия активации $\Delta\varepsilon_{акт} = 0,046$ эВ, которая может быть получена от фононов кристаллической решетки. Разорванная в кристаллической решетке связь обладает свойством положительного заряда, равного заряду электрона, она не локализована возле какого-либо конкретного атома Si , а вследствие флук-

туаций колебаний ионов кристаллической решетки случайным образом перемещается от одного атома к другому (рис. 3.12.3).

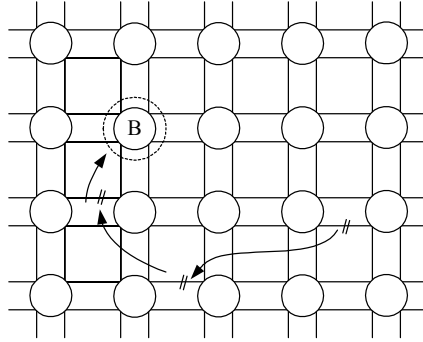


Рис. 3.12.3. Дырочная примесь

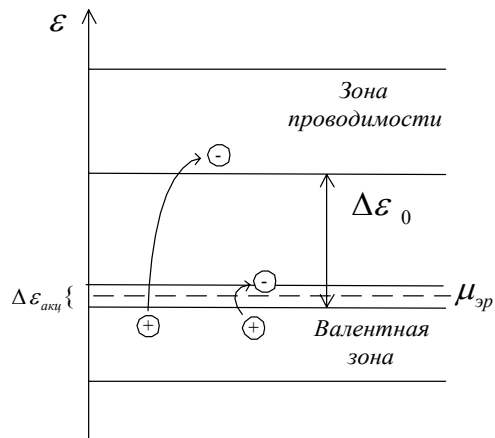


Рис. 3.12.4. Примесная дырочная проводимость

Внедрение в кристаллическую решетку атомов B означает, что в непосредственной близости к верхней границе валентной зоны возник примесный свободный уровень (рис. 3.12.4). Близость этого уровня к заполненной валентной зоне приводит к тому, что уже при сравнительно низких температурах электроны из валентной зоны переходят на примесный уровень. Кроме того, наряду с переходами на примесные уровни, в результате возбуждений некоторое число электронов переходит из валентной зоны в зону проводимости, но их concentra-

ция там чрезвычайно мала. Возникшая в валентной зоне разорванная между соседними атомами Si связь обладает свойством квазичастицы и называется дыркой. Дырка имеет положительный заряд, эффективную массу, энергию. Но так как концентрация дырок в валентной зоне во много раз больше концентрации квазиэлектронов в зоне проводимости, то основными носителями заряда в таких полупроводниках являются дырки, не основными – квазиэлектроны.

Такие полупроводники называются полупроводниками p -типа.

Таким образом, в кристаллическом Si все трехвалентные атомы примеси создают дырочный, а все пятивалентные атомы примеси – электронный тип проводимости. В отличие от собственной проводимости, которая имеет место в собственном полупроводнике и осуществляется одновременно квазиэлектронами и дырками, примесная проводимость полупроводника обусловлена в основном носителями заряда одного знака: квазиэлектронами в случае донорной примеси, дырками – в случае акцепторной примеси; эти носители заряда называются основными. Кроме них, примесный полупроводник содержит и неосновные носители заряда: полупроводник n -типа – дырки, полупроводник p -типа – квазиэлектроны. Концентрация их, как правило, значительно меньше концентрации основных носителей заряда. Поэтому доля, вносимая ими в электропроводность, во много раз меньше доли, вносимой основными носителями заряда.

3.13. Электропроводность полупроводников

Электропроводность вещества – это одна из тех немногих характеристик твердого тела, которую сравнительно легко можно определить из опыта, например, непосредственным измерением электрического сопротивления. Для оценки величины электропроводности любого тела используются такие физические величины, как электрическое сопротивление, удельное электрическое сопротивление, удельная электропроводность.

Если полупроводник однородный и имеет по длине постоянное поперечное сечение, то электрическое сопротивление образца вычисляется по формуле:

$$R = \rho \frac{\lambda}{S}, \quad (3.13.1)$$

где $\rho = \sigma^{-1}$ – удельное сопротивление, σ – удельная электропроводность.

3.13.1. Собственный полупроводник

Пусть в собственном полупроводнике существует электрическое поле напряженностью E_0 , под действием которого в нем возникает направленное движение свободных зарядов – электрический ток. Опыт показывает, что при малых напряженностях электрического поля средняя скорость направленного движения носителей заряда пропорциональна напряженности электрического поля:

$$\langle v \rangle = b E_0, \quad (3.13.2)$$

где b – подвижность, $\langle v \rangle$ – средняя скорость носителя заряда.

Уравнение (3.13.2) позволяет сформулировать следующее определение подвижности носителей заряда: подвижностью носителей заряда называется величина, численно равная средней скорости направленного движения, которую носители заряда того или иного знака приобретают в электрическом поле напряженностью 1В/м.

Из раздела "Электричество" известно, что плотность электрического тока

$$j = en \langle v \rangle, \quad (3.13.3)$$

где n – концентрация носителей заряда, e – заряд электрона, $e < 0$. Отсюда после подстановки значения $\langle v \rangle$ из (3.13.2) в (3.13.3) получаем:

$$j = enb E_0. \quad (3.13.4)$$

Но, поскольку в собственном полупроводнике находятся носители заряда двух знаков, то результирующая плотность тока будет зависеть от концентрации квазиэлектронов n_e и дырок n_d . В собственном полупроводнике всегда $n_e = n_d$. Что же касается подвижностей квазиэлектронов и дырок, то они в общем случае не равны между собой, поскольку эффективные массы квазиэлектронов и дырок различны по величине. Тогда результирующая плотность тока

$$j = j_e + j_g = en_e(b_e + b_d)E_0 \quad (3.13.5)$$

Запишем закон Ома в дифференциальной форме:

$$j = \sigma E_0. \quad (3.13.6)$$

Из уравнений (3.13.5) и (3.13.6) находим удельную электропроводность:

$$\sigma = en_e(b_e + b_d). \quad (3.13.7)$$

Найдем теперь температурную зависимость удельной электропроводности. Ранее было получено выражение для концентрации свободных электронов в интервале энергий от ε до $\varepsilon + d\varepsilon$:

$$dn = \frac{4\pi(2m^*)^{\frac{3}{2}} \varepsilon^{\frac{1}{2}} d\varepsilon}{h^3 \left(1 + \exp\left(\frac{\varepsilon - \mu_{\vartheta}}{kT}\right)\right)}. \quad (3.13.8)$$

Применим его как для квазиэлектронов зоны проводимости, так и для дырок валентной зоны, приняв всюду эффективную массу постоянной величиной. Тогда

$$dn_e = \frac{4\pi(2m_e^*)^{\frac{3}{2}} \varepsilon^{\frac{1}{2}} d\varepsilon}{h^3 \left(1 + \exp\left(\frac{\varepsilon - \mu_{\vartheta}}{kT}\right)\right)}. \quad (3.13.9)$$

Так как концентрация квазиэлектронов достаточно мала, то для них распределение Ферми-Дирака совпадает с распределением Максвелла-Больцмана:

$$f_{\Phi} = \left(1 + \exp\left(\frac{\varepsilon - \mu_{\vartheta}}{kT}\right)\right)^{-1} \approx e^{-\frac{\varepsilon - \mu_{\vartheta}}{kT}}. \quad (3.13.10)$$

Тогда с учетом (3.13.10) уравнение (3.13.9) можно записать в следующем виде:

$$n_e = \frac{4\pi(2m_e^*)^{\frac{3}{2}}}{h^3} e^{\frac{\mu_{\vartheta}}{kT}} \int_0^{\infty} \varepsilon^{\frac{1}{2}} e^{-\frac{\varepsilon}{kT}} d\varepsilon = \left(\frac{2\pi m_e^* kT}{h^2}\right)^{\frac{3}{2}} e^{\frac{\mu_{\vartheta}}{kT}}. \quad (3.13.11)$$

Так как

$$\mu_{\vartheta} = -\frac{\Delta\varepsilon_0}{2},$$

то

$$n_e = \left(\frac{2\pi m_e^* kT}{h^2}\right)^{\frac{3}{2}} e^{-\frac{\Delta\varepsilon_0}{2kT}}. \quad (3.13.12)$$

Если подставить найденное выражение (3.13.12) для n_e в формулу (3.13.7), то получим выражение для удельной электропроводности собственного полупроводника в виде

$$\sigma = e \left(\frac{2\pi m_e^* kT}{h^2}\right)^{\frac{3}{2}} (b_e + b_g) e^{-\frac{\Delta\varepsilon_0}{2kT}}. \quad (3.13.13)$$

Величины подвижностей квазиэлектронов и дырок с изменением температуры меняются незначительно по сравнению с экспоненци-

альным множителем. Тогда с учетом этого приближения можем записать:

$$\sigma = \sigma_0 e^{-\frac{\Delta \varepsilon_0}{2kT}}. \quad (3.13.14)$$

Отсюда температурная зависимость удельного электрического сопротивления

$$\rho = \rho_0 e^{\frac{\Delta \varepsilon_0}{2kT}}. \quad (3.13.15)$$

Формулы (3.13.14) и (3.13.15) позволяют по известной экспериментальной зависимости вычислить энергию активации, т.е. ширину запрещенной зоны $\Delta \varepsilon_0$.

13.2. Примесные полупроводники

Первоначально рассмотрим особенности электропроводности у полупроводников n -типа. У них концентрация носителей заряда определяется концентрацией носителей в собственном полупроводнике и концентрацией примесей, соответственно так же определяется и плотность тока:

$$n = n_c + n_n, j = j_c + j_n, \quad (3.13.16)$$

где n_c , n_n – концентрации носителей заряда собственного полупроводника и атомов примеси, соответственно; j – плотность тока проводимости, созданная квазиэлектронами и дырками соответственно в зонах проводимости и валентной.

$$j_c = \sigma_c E_0, \quad j_n = \sigma_n E_0, \quad (3.13.17)$$

$$\sigma_c = en_e(b_e + b_g), \sigma_n = en_n b_e; j = \sigma E_0, \sigma = \sigma_c + \sigma_n. \quad (3.13.18)$$

Проведя рассуждения, аналогичные разделу 3.13.1, получим:

$$\sigma = \sigma_{0c} e^{-\frac{\Delta \varepsilon_0}{2kT}} + \sigma_{0n} e^{-\frac{\Delta \varepsilon_{дон}}{2kT}}. \quad (3.13.19)$$

Если же полупроводник p -типа, то на основании предыдущего можем записать

$$\sigma = \sigma_{0c} e^{-\frac{\Delta \varepsilon_0}{2kT}} + \sigma_{0p} e^{-\frac{\Delta \varepsilon_{акц}}{2kT}}. \quad (3.13.20)$$

Но поскольку $\Delta \varepsilon_0 \gg \Delta \varepsilon_{дон}$, $\Delta \varepsilon_0 \gg \Delta \varepsilon_{акц}$, то при достаточно низкой температуре электропроводность примесных полупроводников в основном определяется величиной концентрации примесей. Однако с ростом температуры примесные уровни будут постепенно истощаться,

и при высоких температурах электропроводность будет определяться концентрацией носителей заряда в собственном полупроводнике.

Формулы (3.13.19), (3.13.20) позволяют по результатам опыта вычислить энергию активации в полупроводниках n - и p -типа при комнатной температуре или близкой к ней. В этом случае $\Delta\varepsilon_0 \gg \Delta\varepsilon_{\text{дон}}$, $\Delta\varepsilon_0 \gg \Delta\varepsilon_{\text{акц}}$. Тогда в уравнениях (3.13.19), (3.13.20) первыми слагаемыми можно пренебречь, записав

$$\sigma_n = \sigma_{n0} e^{-\frac{\Delta\varepsilon_{\text{дон}}}{2kT}}, \quad \sigma_p = \sigma_{p0} e^{-\frac{\Delta\varepsilon_{\text{акц}}}{2kT}}. \quad (3.13.21)$$

3.14. Контакт двух полупроводников

Рассмотрим контакт n - и p -полупроводников, причем для простоты дальнейших рассуждений предположим, что у этих полупроводников собственный полупроводник один и тот же. Опыт показывает, что при введении в контакт таких полупроводников в области контакта возникает особый слой, получивший название p - n перехода. p - n переход можно получить химическим, электролитическим способом, либо путем внедрения атомов примеси методом диффузии или ионной бомбардировкой. Механическим способом p - n переход получить не удастся из-за малости толщины контактного слоя по сравнению с неровностями поверхности образцов. Обычно толщина контактного слоя менее 10 мкм. Поскольку собственный полупроводник один и тот же, то $\Delta\varepsilon_0$ и размеры зон у них одинаковы, хотя электрохимические потенциалы разные, в нашем случае $\mu_{\text{эп}} < \mu_{\text{эн}}$ (рис. 3.14.1).

В процессе введения в контакт происходит выравнивание уровней электрохимических потенциалов и постепенное установление термодинамического равновесия между носителями заряда в исходных полупроводниках. Этот процесс идет в результате диффузии квазиэлектронов из n - в p -полупроводник и дырок из p - в n -полупроводник. Наглядно все это можно представить следующим образом (рис. 3.14.1).

При соприкосновении полупроводников электроны с донорных уровней преимущественно переходят не в зону проводимости, а на расположенные вблизи валентной зоны акцепторные уровни. На этих уровнях они локализуются вблизи атомов примеси и в образовании тока проводимости участвовать не могут. Такие переходы эквивалентны обратному переходу дырок с примесного акцепторного уровня на примесный донорный уровень. Возникшие на примесном уровне

дырки также перемещаться не могут, они локализованы вблизи атомов примеси.

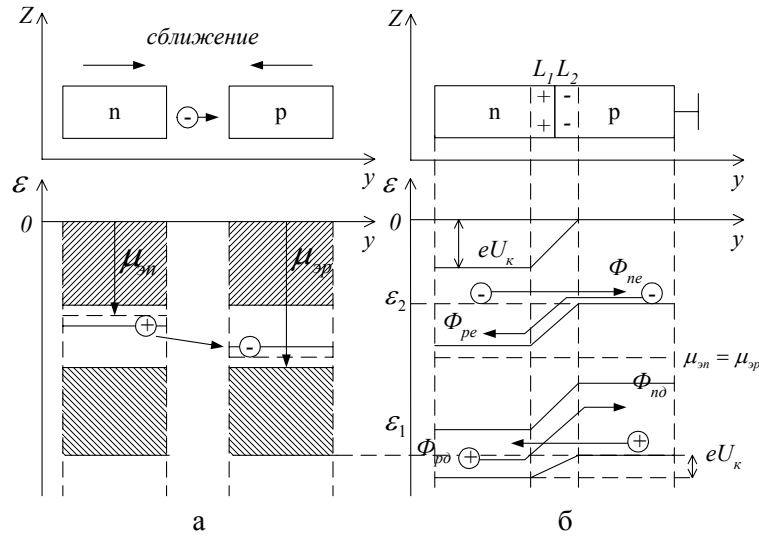


Рис. 3.14.1. Контактный слой: а – диффузия частиц в контактном слое, б – энергии носителей в контактном слое

Итак, после контакта происходит процесс диффузии носителей заряда, причем в процессе диффузии уровень электрохимического потенциала $\mu_{эп}$ постепенно опускается вниз и после выравнивания электрохимических потенциалов $\mu_{эп}$ и $\mu_{эп}$ дальнейшего переноса избыточного количества зарядов через контактный слой происходить не будет. Однако в результате такого процесса нарушается электрическая нейтральность исходных полупроводников. С одной стороны, электроны и дырки, переходя через границу раздела, оставляют после себя неподвижные ионы донорных и акцепторных атомов (рис. 3.14.1, а), а с другой стороны, увеличивается концентрация электронов вблизи контакта в области p -полупроводника и дырок – в области n -полупроводника, контактная область p -полупроводника оказывается заряженной отрицательно, а контактная область n -полупроводника оказывается заряженной положительно. На участке $L = L_1 + L_2$ возникает электрическое поле, созданное двумя слоями объемных зарядов.

Таким образом, вблизи границы раздела электронного и дырочного полупроводников возникают области положительных и отрица-

тельных объемных зарядов. За пределами контактного слоя в каждом из полупроводников объемных зарядов не существует. Но в результате таких переходов в пределах контактного слоя искривляются зона проводимости и валентная зона. Как следствие этого, возникает контактная разность потенциалов, которая одновременно является и внутренней и внешней. Внутренней она является потому, что в области контакта энергия зон изменяется на величину eU_k , а внешней – потому, что между точками, расположенными над поверхностями контактирующих полупроводников, возникает разность потенциалов. Эта контактная разность потенциалов не имеет специального названия.

Почему же возникает, на первый взгляд, такое странное обстоятельство, когда внешняя и внутренняя контактные разности потенциалов равны между собой? Внешняя и внутренняя контактные разности потенциалов имеют разное физическое происхождение: внешняя определяется разностью работ выхода электронов, внутренняя – взаимным расположением уровней электрохимических потенциалов. А так как в данном случае предполагается, что примесные полупроводники созданы на основе одного и того же собственного полупроводника, то внешняя и внутренняя контактные разности потенциалов равны между собой. Если бы это было не так, то внешняя контактная разность потенциалов отличалась бы от внутренней.

С учетом вышеизложенного, внешняя контактная разность потенциалов

$$U_k = \frac{1}{e}(A_p - A_n) = \frac{1}{e}(\mu_{sp} - \mu_{sn}), \quad (3.14.1)$$

A_p, A_n – работа выхода электрона из p - и n -полупроводников, $e > 0$.

После установления термодинамического равновесия в области контакта имеют место следующие потоки квазичастиц (рис. 3.14.1, б). Пусть квазиэлектрон с энергией $\varepsilon < \varepsilon_2$ движется из n - в p -полупроводник. В области контакта он попадает в тормозящее для него электрическое поле, постепенно теряет энергию и возвращается обратно в n -полупроводник. С другой стороны, любая дырка с энергией $\varepsilon < \varepsilon_1$ при движении в сторону n -полупроводника также попадает в тормозящее электрическое поле и возвращается обратно. Но если энергия квазиэлектронов в n -полупроводнике $\varepsilon > \varepsilon_2$, а энергия дырок в p -полупроводнике $\varepsilon > \varepsilon_1$, то квазиэлектроны и дырки преодолеют контактный слой и соответственно попадут в p - и n -полупроводники. У примесных полупроводников всегда имеется некоторое число таких квазиэлектронов и дырок. Поэтому там возникают различные диффузионные потоки квазичастиц.

Из постановки задачи следует, что через контактный слой должны протекать четыре потока квазичастиц (рис. 3.14.1, б). Рассмотрим их отдельно.

1. Диффузионные потоки $\Phi_{не}$ и $\Phi_{рд}$ возникают из-за изменения концентраций в пределах образца как у квазиэлектронов, так и у дырок. Их направления указаны на рис. 3.14.1, б.

2. В p -полупроводнике в зоне проводимости неосновными носителями заряда являются квазиэлектроны, в n -полупроводнике в валентной зоне – дырки.

Благодаря хаотическому движению некоторое число квазичастиц попадают в контактный слой, ускоряются электрическим полем контактного слоя и переносятся, соответственно: квазиэлектроны – в n -полупроводник, дырки – в p -полупроводник. Возникают потоки $\Phi_{не}$ и $\Phi_{рд}$.

Уравнение баланса потоков этих квазичастиц:

$$\Phi_{не} + \Phi_{рд} = \Phi_{ре} + \Phi_{нд}.$$

А так как поток квазичастиц $\Phi = I / q^*$, q^* – эффективный заряд квазичастицы, I – электрический ток квазичастиц, то

$$I_{нд} + I_{ре} = I_{не} + I_{рд}. \quad (3.14.2)$$

В отсутствие действия внешнего источника тока напряженность электрического поля $E = 0$, ток проводимости в образце отсутствует, диффузионный ток равен току проводимости:

$$I_{диф} = I_{пров}, \quad (3.14.3)$$

где $I_{диф} = I_{не} + I_{рд}$, $I_{пров} = I_{нд} + I_{ре}$.

3.15. Физические особенности контактного слоя в p - n полупроводнике

Как было установлено в разделе 3.12, в процессе контакта n - и p -полупроводников в пространстве между ними возникает слой конечной толщины, обедненный носителями заряда. В этом контактном слое возникают объемные заряды, локализованные вблизи атомов примесей, поэтому эти заряды под действием поля в пространстве перемещаться не могут, следовательно, не могут участвовать в образовании тока проводимости.

Оценим теперь толщину контактного слоя и выясним, от каких параметров вещества зависит его толщина. Для дальнейшего анализа

первоначально предположим, что к полупроводнику внешний источник тока не подключен (рис. 3.15.1).

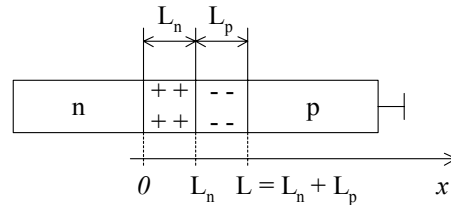


Рис. 3.15.1. Объемный заряд в области p - n перехода

Температура полупроводника такова, что в результате перехода электронов с примесных уровней в зону проводимости все примесные уровни n -полупроводника свободны, а в результате перехода электронов из валентной зоны p -полупроводника на примесные уровни все примесные уровни оказываются занятыми электронами. Это условие будет выполняться либо при достаточно высокой температуре образца, либо когда донорный уровень расположен в непосредственной близости к зоне проводимости n -полупроводника, а акцепторный уровень расположен в непосредственной близости к валентной зоне.

Все дальнейшие рассуждения проведем только для области контактного слоя $L = L_n + L_p$ (рис. 3.15.1). В общем случае объемная плотность зарядов в области контактного слоя имеет разный знак и разную пространственную протяженность. Выберем начало координат на левой границе положительного заряда n -полупроводника (рис. 3.15.1). Введем следующие обозначения для объемных плотностей зарядов в n - и p -полупроводниках:

$$0 < x < L_n, \rho_n = -en_{\text{дон}}, L_n < x < L, \rho_p = en_a, e < 0.$$

Здесь $n_{\text{дон}}$ – концентрация атомов примеси в n -полупроводнике (концентрация донорных уровней), n_a – концентрация атомов примеси в p -полупроводнике (концентрация акцепторных уровней). Предполагаем, что в области контактных слоев L_n и L_p концентрация объемных зарядов при различных значениях x сохраняется неизменной.

Так как p -полупроводник заземлен, то n -полупроводник оказывается положительно заряженным относительно p -полупроводника. Следовательно, в области контакта происходит постепенное изменение потенциала. Для определения изменения потенциала в области контактного слоя L воспользуемся уравнением Пуассона. Тогда в контактном слое толщиной L_n уравнение Пуассона

$$\frac{d^2\varphi_n}{dx^2} = -\frac{e}{\varepsilon_0 \varepsilon} n_{\text{дон}} \quad (3.15.1)$$

при следующих граничных условиях:

$$\varphi_n|_{x=L_n} = U_\kappa > 0, \quad \frac{d\varphi_n}{dx}|_{x=0} = 0, \quad eU_\kappa < 0.$$

Решением уравнения (1) является

$$\varphi_0 - U_\kappa = \frac{e}{2\varepsilon_0 \varepsilon} n_{\text{дон}} L_n^2, \quad (3.15.2)$$

φ_0 – потенциал на границе раздела n - и p -полупроводников.

В контактном слое толщиной L_p уравнение Пуассона

$$\frac{d^2\varphi_p}{dx^2} = \frac{e}{\varepsilon_0 \varepsilon} n_a \quad (3.15.3)$$

при граничных условиях

$$\varphi_p|_{x=L} = 0, \quad \frac{d\varphi_p}{dx}|_{x=L} = 0.$$

Решением уравнения (3.15.3) является

$$\varphi_0 = \frac{e}{2\varepsilon_0 \varepsilon} n_a L_p^2. \quad (3.15.4)$$

Подставим найденное из (3.15.4) значение φ_0 в уравнение (3.15.2):

$$\frac{e}{2\varepsilon_0 \varepsilon} n_a L_p^2 - U_\kappa = -\frac{e}{2\varepsilon_0 \varepsilon} n_{\text{дон}} L_n^2.$$

Отсюда внутренняя контактная разность потенциалов

$$U_\kappa = \frac{e}{\varepsilon_0 \varepsilon} (n_{\text{дон}} L_n^2 + n_a L_p^2). \quad (3.15.5)$$

Как следует из (3.15.1) и (3.15.3) с учетом граничных условий,

$$\left. \begin{aligned} \frac{d\varphi_n}{dx} &= -\frac{e}{\varepsilon_0 \varepsilon} n_{\text{дон}} \cdot x, \\ \frac{d\varphi_p}{dx} &= \frac{e}{\varepsilon_0 \varepsilon} n_a (x - L_n - L_p). \end{aligned} \right\} \quad (3.15.6)$$

Так как на границе раздела градиент потенциала должен быть непрерывным, то

$$\left(\frac{d\varphi_n}{dx} = \frac{d\varphi_p}{dx} \right)_{x=L_n}. \quad (3.15.7)$$

После подстановки значений (3.15.6) в (3.15.7) получаем

$$n_{\text{дон}} L_n = n_a L_p$$

или

$$L_p^2 = \left(\frac{n_{\text{дон}}}{n_a} \right)^2 L_n^2. \quad (3.15.8)$$

Найденное значение L_p^2 подставляем в (3.15.5):

$$U_{\kappa} = \frac{e}{2\varepsilon_0 \varepsilon} \frac{n_{\text{дон}}}{n_a} (n_{\text{дон}} + n_a) L_n^2.$$

Отсюда

$$L_n = \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_a}{n_{\text{дон}}} \frac{U_{\kappa}}{n_{\text{дон}} + n_a} \right)^{\frac{1}{2}}. \quad (3.15.9)$$

Значение L_n из (3.15.9) подставим в (3.15.8):

$$L_p = \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_{\text{дон}}}{n_a} \frac{U_{\kappa}}{n_{\text{дон}} + n_a} \right)^{1/2}. \quad (3.15.10)$$

Толщина контактного слоя $L = L_n + L_p$, тогда, как следует из (3.15.9) и (3.15.10),

$$L = \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_{\text{дон}} + n_a}{n_{\text{дон}} + n_a} U_{\kappa} \right)^{1/2}. \quad (3.15.11)$$

Если к концам полупроводника приложена внешняя разность потенциалов U , то в полупроводнике будет протекать электрический ток. Но контактный слой обеднен свободными носителями заряда, значит, он обладает электрическим сопротивлением, значительно превышающим электрическое сопротивление остального полупроводника, при этом основное падение напряжения будет происходить в контактом слое, а внешнее напряжение U будет суммироваться с внутренней контактной разностью потенциалов U_{κ} . Тогда формулы (3.15.9)–(3.15.11) можно видоизменить:

$$\left. \begin{aligned} L_n &= \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_a}{n_{\text{дон}}} \frac{U_{\kappa} + U}{n_{\text{дон}} + n_a} \right)^{1/2}, \\ L_n &= \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_{\text{дон}}}{n_a} \frac{U_{\kappa} + U}{n_{\text{дон}} + n_a} \right)^{1/2}, \\ L &= \left(\frac{2\varepsilon_0 \varepsilon}{e} \frac{n_{\text{дон}} + n_a}{n_{\text{дон}} \cdot n_a} (U_{\kappa} + U) \right)^{1/2}. \end{aligned} \right\} \quad (3.15.12)$$

причем внешняя разность потенциалов U может быть как положительной, так и отрицательной.

Из (3.15.12) следует, что ширина контактного слоя L будет меняться в зависимости от величины и знака разности потенциалов U и, в общем случае, $L_n \neq L_p$. Если же $U_k + U = 0$, то в области контактного слоя возникает пробой.

В табл. 3.15.1 приводятся результаты вычислений по формулам (3.15.9)–(3.15.11) толщины контактных слоев L_n , L_p , $L = L_n + L_p$ для Si. Вычисления выполнены для $\epsilon = 11,7$ эВ, $U_k = 1$ В. Из табл. 3.15.1 видно, что толщина контактного слоя сильно зависит от концентрации примесей, наибольшей величины она достигает у полупроводников с малой концентрацией. При высокой концентрации толщина контактного слоя очень мала и с увеличением концентрации стремится к нулю.

Так как электрическое сопротивление контактного слоя велико, то любое его изменение должно сильно влиять на электропроводность полупроводников. Согласно (3.15.12), толщина контактного слоя L зависит от величины и знака приложенной внешней разности потенциалов U , она определенным образом влияет на величину протекающего через контакт тока. В результате электрическое сопротивление контактного слоя для $+U$ и $-U$ оказывается различным по величине, и это различие может быть весьма значительным. На этом основано выпрямляющее действие p - n перехода, поскольку с изменением знака разности потенциалов U будет меняться как толщина контактного слоя L , так и его электрическое сопротивление.

$n_{\text{дон}}, \text{м}^{-3}$	$n_{\text{а}}, \text{м}^{-3}$	$L_n, \text{мкм}$	$L_p, \text{мкм}$	L_n / L_p	$L, \text{мкм}$
$1 \cdot 10^{21}$	$1 \cdot 10^{22}$	1,08	0,11	10	1,19
$1 \cdot 10^{21}$	$1 \cdot 10^{21}$	0,80	0,80	1	1,60
$1 \cdot 10^{22}$	$1 \cdot 10^{21}$	11	1,08	0,1	1,19
$1 \cdot 10^{22}$	$1 \cdot 10^{23}$	0,34	0,034	10	0,37
$1 \cdot 10^{22}$	$1 \cdot 10^{22}$	0,25	0,25	1	0,50
$1 \cdot 10^{23}$	$1 \cdot 10^{22}$	0,034	0,34	0,1	0,37

Таблица 3.15.1. Толщины контактных слоев L_n , L_p , $L = L_n + L_p$ для Si