

Проводимость полупроводников

Электроны и дырки

В [предыдущей лекции](#), в параграфе «Оптические переходы», мы говорили о внутреннем фотоэффекте, приводящем к переходу электронов из валентной зоны в зону проводимости, о неравновесной фотопроводимости, вклад в которую дают обе эти зоны. Однако и при отсутствии светового потока, как мы знаем, при повышении температуры от абсолютного нуля некоторые электроны в полупроводнике переходят из валентной зоны в зону проводимости из-за тепловых процессов.¹ Это – равновесная ситуация, и именно о ней пойдет речь в следующих параграфах, но в любом случае в зоне проводимости появляются электроны, а в валентной зоне – пустые места (уровни энергии)².

Незаполненные электронами места в валентной зоне называются *дырками*.

Дырки

Сразу отметим, что поскольку отсутствие отрицательно заряда $-e$ (электрона) эквивалентно дополнительному заряду $+e$, то можно считать, что *дырка* заряжена положительно. Это подтверждается также тем, что при ускорении электрона электрическим полем, ускорение освобождающегося места – дырки – направлено, очевидно, в противоположную сторону.

Концентрацию электронов зоны проводимости обычно обозначают буквой n , а концентрацию дырок в валентной зоне – буквой p . (Хорошо, что нам больше практически не придется говорить об импульсе!) Эти же буквы используют как индексы, указывающие, о чем идет речь – об электронах или дырках.³

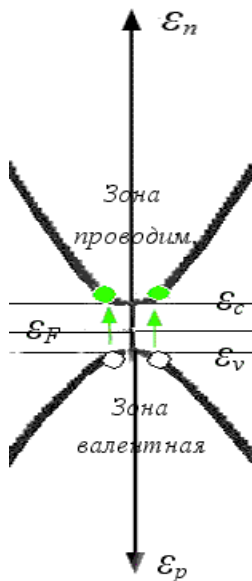
Оказывается удобным связать проводимость валентной зоны с дрейфом дырок, а не электронов по следующим соображениям.

¹ Тепловой заброс электронов в зону проводимости происходит обычно многофоновым поглощением с возможным участием примесных уровней энергии, расположенных в запрещенной зоне.

² Поскольку на одном невырожденном уровне энергии могут находиться 2 электрона, то пустой уровень – два пустых «места». Поэтому выражение «пустое место», наверно, предпочтительней.

³ С букв n и p начинаются слова *negative* (отрицательный) и *positive* (положительный)

Прежде всего, как мы видели, проводимость валентной зоны пропорциональна именно концентрации пустых мест, т.е. дырок. Теперь посмотрим на рисунок (по оси абсцисс, как и раньше,



отложена проекция волнового вектора k или квазиимпульса электрона $\hbar k$).

Энергия электрона в зоне проводимости минимальна в минимуме изображающей ее параболы («на дне зоны», при $k = 0$), а закон дисперсии имеет вид $\varepsilon_n = \varepsilon_c + \frac{\hbar^2 k^2}{2m_n}$, где ε_n – энергия электрона, ε_c – граница зоны проводимости, $m_n > 0$ – эффективная масса электрона.

В максимуме валентной зоны (при $k = 0$) энергия электрона максимальна и уменьшается, когда он «спускается» по параболе (т.е. при возрастании k). Но такое движение электрона эквивалентно движению дырки (освобождающегося места) к вершине зоны,¹ поэтому энергия дырки также уменьшается по мере ее приближения к вершине валентной зоны, т.е. энергия дырки минимальна в вершине валентной зоны. Но тогда ось энергий для дырок ε_p надо пустить в обратную сторону, вниз, как это показано на рисунке. Переверните мысленно рисунок и убедитесь, что эффективная масса дырок положительна, в отличие от эффективной массы электрона валентной зоны («перевернутая» парабола). Формально можно сделать так. Закон дисперсии электрона в валентной зоне (ε_v – ее граница) есть $\varepsilon = \varepsilon_v + \frac{\hbar^2 k^2}{2m^*}$ с эффективной массой электрона $m^* < 0$. Припишем дырке эффективную массу $m_p = -m^* > 0$ и направим ось энергий для дырок ε_p так, как показано на рисунке. Тогда закон дисперсии для дырок принимает вид

$$\varepsilon_p = \varepsilon_v + \frac{\hbar^2 k^2}{2m_p}.$$

Только на моем плохом рисунке $m_n = m_p$, в действительности эти массы могут сильно отличаться друг от друга (обычно $m_n < m_p$). Дырка – условный объект, поэтому ее иногда называют

Рис.1
Валентная зона и зона проводимости

¹ Вообще, динамика дырок полностью определяется динамикой электрона, так что абсолютные значения заряда, эффективной массы и т.п. электронов и дырок одинаковы. Одинаковы даже времена свободного пробега (в той же, конечно, зоне).

квазичастицей (как бы частицей). Дырки и электроны суть носители заряда, но их часто (как, впрочем, и ионы в электролитах) еще называют *носителями тока*. Приведем определение дырки, даваемое ГОСТом. «Дырка – квазичастица, носитель положительного заряда, равного элементарному заряду в полупроводниках». Не очень здорово, правда?

Для дырок (с учетом их положительного заряда) легко получить некоторые формулы, которые мы уже выводили для электронов. Дрейфовая скорость дырок в электрическом поле \mathbf{E} есть

$$\mathbf{u}_p = \frac{e\tau_p}{m_p} \mathbf{E}, \quad (1)$$

где τ_p – среднее время свободного пробега дырки (фактически, среднее время свободного пробега электрона в валентной зоне по отношению к любому рассеянию, кроме рассеяния на других электронах). Коэффициент пропорциональности

$$\mu_p = \frac{e\tau_p}{m_p} \quad (2) \text{ Подвижность дырок}$$

называется *подвижностью* дырок, так что

$$\mathbf{u}_p = \mu_p \mathbf{E}.$$

Аналогично, подвижностью электронов зоны проводимости называется величина

$$\mu_n = \frac{e\tau_n}{m_n}, \quad (3) \text{ Подвижность электронов}$$

так что дрейфовую скорость электронов в зоне проводимости можно записать так:

$$\mathbf{u}_n = -\mu_n \mathbf{E}. \quad (3')$$

Плотность дырочного тока есть (напомню, что p – концентрация дырок)

$$\mathbf{j}_p = e p \mathbf{u}_p = e p \mu_p \mathbf{E} = \frac{e^2 \tau_p p}{m_p} \mathbf{E} = \sigma_p \mathbf{E} \quad (4) \text{ Плотность дырочного тока}$$

В полный ток входит как электронный, так и дырочный ток, поэтому для полной плотности тока получаем

$$\mathbf{j} = \mathbf{j}_n + \mathbf{j}_p = \sigma \mathbf{E}, \quad \text{Закон Ома}$$

где полная удельная проводимость σ равна сумме электронной и дырочной проводимостей

$$\sigma = \sigma_n + \sigma_p = \frac{e^2 \tau_n n}{m_n} + \frac{e^2 \tau_p p}{m_p}. \quad (5) \text{ Полная удельная проводимость}$$

Изучая температурную зависимость [проводимости металлов](#), мы все внимание сконцентрировали на времени свободного

пробега. В полупроводниках от температуры слегка зависит даже эффективная масса, однако решающее значение имеет температурная зависимость концентрации электронов и дырок. Сейчас мы получим две очень важные формулы для этих концентраций, которые верны для *любых невырожденных¹ полупроводников* в условии равновесия, независимо от того, играют ли примеси какую-нибудь роль в их электропроводности.

Для определенности, рассмотрим газ электронов проводимости. Этот газ можно считать невырожденным, если в распределении Ферми – Дирака $f(\varepsilon) = 1/(e^{\frac{\varepsilon - \varepsilon_F}{kT}} + 1)$ для всех энергетических уровней выполняется условие

$$\varepsilon - \varepsilon_F \gg kT.$$

Сразу отметим, что аналогичное соотношение можно написать и для дырок, и это означает, что энергия Ферми, как мы уже знаем, находящаяся внутри запрещенной зоны, должна достаточно далеко отстоять как от границы валентной зоны, так и от границы зоны проводимости:

$$\varepsilon_c - \varepsilon_F \gg kT, \quad \varepsilon_F - \varepsilon_v \gg kT.$$

Будем считать эти условия выполненными, тогда²

$$f(\varepsilon) \approx e^{\varepsilon_F/kT} e^{-\varepsilon/kT},$$

как мы это уже видели в лекции «[Проводимость металлов](#)», (5).

Теперь воспользуемся формулой (9) из этой же лекции:

$$n = \int_{\varepsilon_c}^{\infty} g(\varepsilon - \varepsilon_c) e^{\varepsilon_F/kT} e^{-\varepsilon/kT} d\varepsilon.$$

Мы заменили верхнюю границу зоны проводимости на бесконечность. Так можно делать, если энергия kT значительно меньше ширины зоны проводимости, по которой и берется интеграл, поскольку в этом случае вклад добавленной области интегрирования в результат не существен (величина $e^{-\varepsilon/kT}$ становится очень малой). Кроме того, мы написали $g(\varepsilon - \varepsilon_c)$, понимая, что именно при $\varepsilon = \varepsilon_c$ плотность числа состояний обращается в ноль.

Сделав замену переменной интегрирования $\varepsilon = \varepsilon_c + \varepsilon'$ (так что энергия ε' отсчитывается от дна зоны проводимости), получаем

$$n = N_c e^{-(\varepsilon_c - \varepsilon_F)/kT}, \tag{6}!$$

где величина

¹ Полупроводник называется невырожденным, если в нем и электронный, и дырочный газ можно считать невырожденным (классическим).

² Во всех формулах kT после косой дроби есть неделимое целое. С прямой дробью получается очень мелко, а со скобками (kT) – некрасиво.

Условия невырожденности полупроводника.

$$N_c = \int_0^\infty g(\varepsilon') e^{-\frac{\varepsilon'}{kT}} d\varepsilon'$$

не зависит от энергии Ферми и называется *эффективной плотностью (числа) состояний* в зоне проводимости.¹

Если в качестве плотности числа состояний использовать функцию

$$g(\varepsilon) = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{\varepsilon},$$

полученную в лекции «[Проводимость металлов](#)» (11) для простейшей модели, то

$$N_c = \frac{1}{4} \left(\frac{2m_n kT}{\pi \hbar^2}\right)^{3/2} \quad (7)$$

Аналогично получается формула для концентрации дырок в валентной зоне

$$p = N_v e^{-(\varepsilon_F - \varepsilon_v)/kT}, \quad (8)!$$

где

$$N_v = \frac{1}{4} \left(\frac{2m_p kT}{\pi \hbar^2}\right)^{3/2} - \quad (9)$$

эффективная плотность (числа) состояний в валентной зоне.

Формулы (6) и (8) чрезвычайно важны для дальнейших рассуждений.

Собственная проводимость полупроводников

Это проводимость, обусловленная только *тепловыми* переходами электронов *из валентной зоны в зону проводимости*. Такая проводимость есть всегда при $T > 0$, но, будучи обычно очень малой, непосредственно наблюдается только в чистых полупроводниках (без донорных и акцепторных примесей, см. далее).

Чтобы получить формулу для удельной проводимости, надо найти концентрации электронов (n) и дырок (p). Как показывают формулы (6) и (8), для этого необходимо определить энергию Ферми ε_F .

Поскольку электроны и дырки в чистом полупроводнике рождаются парами (каждый переход электрона в зону про-

Эффективная плотность (числа) состояний в зоне проводимости

Собственная проводимость полупроводников

¹ Это название связано с тем, что написанный интеграл представляет собой число состояний в зоне проводимости ($\int_0^\infty g(\varepsilon') d\varepsilon'$), взятых с «универсальной» вероятностью их заполнения $e^{-\frac{\varepsilon'}{kT}}$, т.е. как бы с их «эффективностью».

димости приводит к образованию дырки в валентной зоне), то для собственной проводимости выполняется условие

$$n = p.$$

Это равенство представляет собой частный случай условия электронейтральности: концентрация отрицательных зарядов (электронов) равна концентрации положительных зарядов (дырок). Оно позволяет сразу определить энергию Ферми, используя формулы (6) и (8):

$$N_c e^{-(\varepsilon_c - \varepsilon_F)/kT} = N_v e^{-(\varepsilon_F - \varepsilon_v)/kT}$$

или

$$e^{\frac{2\varepsilon_F}{kT}} = \frac{N_v}{N_c} e^{\frac{\varepsilon_c + \varepsilon_v}{kT}}.$$

Логарифмируя и учитывая, что согласно формулам (7) и (9)

$$N_v/N_c = (m_p/m_n)^{3/2},$$

получаем

$$\varepsilon_F = \frac{1}{2}(\varepsilon_c + \varepsilon_v) + \frac{3}{4}kT \cdot \ln(m_p/m_n).$$

Таким образом, при $T = 0$ энергия Ферми располагается строго в середине запрещенной зоны, а при повышении температуры несколько смещается к границе зоны, носители заряда которой имеют меньшую эффективную массу (обычно это граница зоны проводимости).

Если теперь подставить энергию Ферми (11) в формулы (6) и (8), то получатся выражения для искомым концентраций (это общий метод), но в данном случае можно поступить проще. Из равенства (10,6,8) следует, что

$$n^2 = p^2 = np = N_c N_v e^{-(\varepsilon_c - \varepsilon_v)/kT}$$

или (ширина запрещенной зоны $\varepsilon_g = \varepsilon_c - \varepsilon_v$)

$$n = p = \sqrt{N_c N_v} e^{-\frac{\varepsilon_g}{2kT}}.$$

Подставляя эти концентрации в формулу (5) для удельной проводимости, получаем

$$\sigma = \sigma_0 e^{-\frac{\varepsilon_g}{2kT}},$$

где $\sigma_0 = \sqrt{N_c N_v} \cdot \left(\frac{e^2 \tau_n}{m_n} + \frac{e^2 \tau_p}{m_p} \right)$.

Коэффициент σ_0 зависит от температуры, но очень слабо по сравнению с экспоненциальной зависимостью в формуле (12), так что при оценках (в частности, в задачах) его можно считать независимым от температуры.

(10) Условие электронейтральности в чистом п/п

(11) Энергия Ферми в чистом п/п

Концентрации носителей в чистом п/п

(12) Формула для задач по температурной зависимости $\sigma(T)$

Мы видели, что проводимость металлов *падает* при повышении температуры из-за уменьшения среднего времени свободного пробега τ . Конечно, при росте температуры это время уменьшается и в полупроводниках, но решающее значение в них имеет очень быстрое нарастание концентрации носителей заряда. Поэтому *проводимость полупроводников быстро растет с ростом температуры*.

Весь анализ зонной структуры мы проводили на основе закона дисперсии $\varepsilon(\hbar k)$ (см. [рис.1](#)). Часто для иллюстрации ис-



пользуют упрощенную картинку зонной структуры (рис.2). В частности, на этом рисунке показано положение энергии Ферми в чистом полупроводнике и оптический переход из валентной зоны в зону проводимости. По оси ординат откладывается, конечно,

энергия, но почему это не показано явно? И еще один вопрос. Что откладывается по оси абсцисс?

Рис.2
Упрощенная схема зон

Примесная проводимость

Примеси в полупроводниках могут быть самых различных типов. Существуют неактивные примеси, которые никак не влияют на электронные процессы; таких примесей может быть очень много даже в «сверхчистых» материалах. Примеси-ловушки помогают корректировать подвижность носителей тока, рекомбинационные центры поставляют уровни, с помощью которых происходит рекомбинация электронов с дырками (обратный процесс для рождения пары электрон-дырка, см. лекцию «[Энергетические зоны](#)») и т.д.¹ Энергетические уровни примесных атомов могут

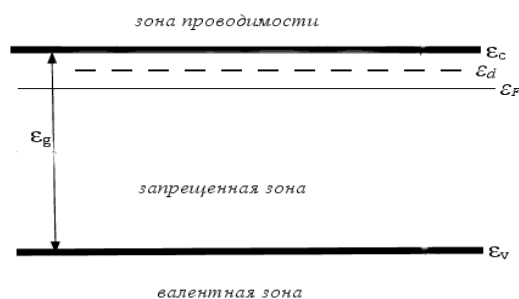
¹ Перечисленные примеси, как и примеси, о которых речь пойдет дальше, относятся к *атомарным примесям*, т.е. представляют собой отдельные атомы «чужого» вещества, замещающие атомы основной решетки. Если концентрация примесей велика, то их уровни могут образовывать энергетические зоны.

находиться как в запрещенной зоне, так и в зонах, содержащих электронные уровни основного кристалла.

Нас сейчас будут интересовать два особых вида примесей, коренным образом влияющих на концентрации носителей тока. Если атом-примесь имеет *больше* валентных электронов, чем их требуется для межатомных связей основной решетки, то эти электроны легко отрываются (имеют малую *энергию активации*) и переходят в зону проводимости. При этом нейтральный первоначально атом становится положительно заряженным ионом. Такие примеси называются *донорами*.

Мы будем рассматривать только *однозарядные* доноры, отдающие только один электрон. В кристаллах германия и кремния, относящихся к четвертой группе периодической системы, однозарядными донорами служат атомы из пятой группы (мышьяк, фосфор).

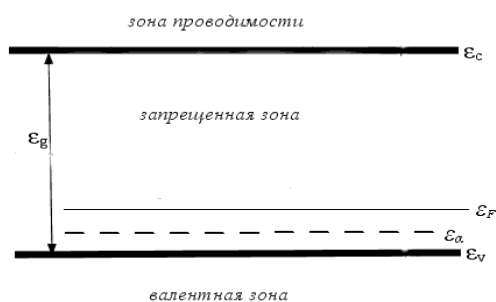
С точки зрения зонной теории, легко отрываемый электрон сидит на энергетическом уровне ε_d в запрещенной зоне вблизи зоны проводимости. Энергия *активации* (или *ионизации*) есть $\varepsilon_{id} = \varepsilon_c - \varepsilon_d$: именно такая энергия необходима для перевода электрона с донорного уровня в зону проводимости (см. рис.3).



Таким образом, доноры – поставщики электронов. При температуре $kT \gg \varepsilon_{id}$ большинство доноров оказываются ионизованными, а концентрация дополнительных электронов в зоне проводимости $n \approx n_d$. Энергия Ферми при этом располагается ниже донорных уровней.

Рис.3
Полупроводник *n*-
типа

Если атом-примесь имеет *меньше* валентных электронов, чем их требуется для межатомных связей основной решетки, то



эти примеси легко захватывают электроны из валентной зоны с образованием там подвижной дырки. При этом нейтральный первоначально атом становится отрицательно заряженным ионом. Такие примеси служат

Рис.4
Полупроводник *p*-
типа

поставщиками дырок и называются *акцепторами*, их уровни энергии ε_a лежат вблизи границы валентной зоны. *Энергией активации акцептора* называется энергия $\varepsilon_{ia} = \varepsilon_a - \varepsilon_v$: именно такая энергия необходима для перевода электрона из валентной зоны на акцепторный уровень (см. рис.4, энергия Ферми показана для ситуации, когда основная часть акцепторов ионизирована). Для германия и кремния однозарядными акцепторами служат элементы третьей группы (например, бор и алюминий).

Основными носителями тока называются те носители, которых больше в данном полупроводнике.

Если основные носители – электроны, то полупроводник относится к *n*-типу (при этом обычно $n \gg p$), если же дырки служат основными носителями, то это полупроводник *p*-типа (и тогда, как правило, $p \gg n$). Полупроводники *n*-типа получают легированием чистого полупроводника донорными примесями, а полупроводники *p*-типа – акцепторными. Всегда, в любом проводнике, есть, хотя бы в небольшом количестве, *неосновные* носители, т.е. дырки (и небольшое количество акцепторов) в полупроводнике *n*-типа и электроны (и небольшое количество доноров) в полупроводнике *p*-типа.

Формулы (6) и (8) позволяют определить тип полупроводника по положению энергии Ферми. Если считать, что $N_c \sim N_v$, то для того, чтобы концентрация электронов была существенно больше концентрации дырок нужно, чтобы выполнялось условие $\varepsilon_c - \varepsilon_F \ll \varepsilon_F - \varepsilon_v$ (см. рис. 3 и 4), т.е. чтобы энергия Ферми была заметно ближе к зоне проводимости, чем к валентной зоне. В полупроводнике *p*-типа, наоборот, энергия Ферми ближе к валентной зоне.

Обозначим N_d полную концентрацию¹ доноров (нейтральных и ионизованных), а N_a – концентрацию акцепторов. Тогда в полупроводнике *n*-типа $N_d \gg N_a$, а в полупроводнике *p*-типа – наоборот. Обозначим еще N_a^- и N_d^+ соответственно концентрацию акцепторов, захвативших электрон из валентной области, и доноров, отдавших электрон в зону проводимости. Для определения энергии Ферми вновь надо решить условие электронейтральности

Основные носители тока
Полупроводники *n*-типа и *p*-типа

Неосновные носители

Определение типа *n/p* по положению энергии Ферми.

¹ По сложившейся традиции концентрации примесей чаще обозначают заглавными буквами.

Условие электронейтральности в общем случае

$$n + N_a^- = p + N_d^+,$$

как уравнение относительно ε_F .

Концентрации n и p даются формулами (6) и (8).

Концентрация N_a^- определяется распределением Ферми – Дирака

$$N_a^- = N_a / (e^{\frac{\varepsilon_a - \varepsilon_F}{kT}} + 1).$$

Поскольку уровень ионизированного донора свободен, то доля таких доноров есть

$$1 - f(\varepsilon_d) = 1 - \frac{1}{e^{\frac{\varepsilon_d - \varepsilon_F}{kT}} + 1} = 1 / (e^{\frac{\varepsilon_F - \varepsilon_d}{kT}} + 1),$$

так что

$$N_d^+ = N_d / (e^{\frac{\varepsilon_F - \varepsilon_d}{kT}} + 1).$$

Решение получающегося уравнения возможно только в некоторых приближениях или численно. Мы опишем коротко поведение энергии Ферми и концентраций электронов и дырок в зависимости от температуры в полупроводнике *n*-типа. Энергию активации доноров будем здесь обозначать ε_i .

Сразу отметим, что практически *все* акцепторы (как мы уже говорили, их небольшое количество присутствует в полупроводнике *n*-типа), всегда ионизованы (захватили электрон), поскольку их уровни лежат существенно ниже энергии Ферми. При температурах, близких к абсолютному нулю, захваченные ими электроны оказываются отобранными у доноров, так что $N_d^+ \approx N_a^-$, концентрации электронов и дырок практически равны нулю, уровень Ферми лежит между донорными уровнями и нижней границей зоны проводимости.¹

При превышении некоторой температуры² T_1 начинается ионизация доноров и растет концентрация электронов. В этой области температур энергия Ферми смещается от середины между ε_d и ε_c вниз, к энергии донорных уровней ε_d :

$$\varepsilon_F \approx \frac{\varepsilon_d + \varepsilon_c}{2} - \frac{1}{2} kT \cdot \ln(N_c / N_d),$$

Температурная зависимость энергии Ферми и концентрации электронов (и тем самым проводимости) в п/п *n*-типа

¹ Можно показать, что для невырожденности полупроводника *n*-типа необходимо выполнение условия $N_d \ll N_c$, а для полупроводника *p*-типа – условия $N_a \ll N_v$.

² Эта температура T_1 определяется условием $kT_1 = \varepsilon_i / \ln(N_c / N_a)$. Ниже этой температуры $\varepsilon_F \approx \varepsilon_d + kT \cdot \ln(N_d / N_a)$, $n \approx \frac{N_d N_c}{N_a} e^{-\varepsilon_i / kT}$, $p \approx \frac{N_v N_a}{N_d} e^{-(\varepsilon_d - \varepsilon_v) / kT}$.

а концентрация электронов ведет себя подобно концентрации носителей при собственной проводимости:¹

$$n \approx \sqrt{N_c N_d} e^{-\varepsilon_i/2kT}.$$

Так продолжается до некоторой температуры T_2^2 , а при более высоких температурах наступает насыщение примесной проводимости, поскольку практически все доноры оказываются ионизованными:

$$\varepsilon_F \approx \varepsilon_c - kT \cdot \ln(N_c/N_d), \quad n \approx N_d, \quad p \approx \frac{N_c N_v}{N_d} e^{-\varepsilon_g/kT}.$$

Задание. Изобразите схематично зависимость удельной проводимости полупроводника n -типа от температуры $\sigma(T)$, считая, что в формуле

$$\sigma(T) = \frac{e^2 \tau_n n}{m_n}$$

от нее зависит только концентрация электронов n .

Электронно-дырочный переход

Электронно-дырочный переход (который часто называют *p-n-переходом*) – сравнительно узкая область вблизи поверхности контакта проводников с разными типами проводимости, обладающая нелинейной вольт-амперной характеристикой со своеобразными свойствами. Обычно под *p-n-переходом* понимают *гомпереход*, образуемый контактом двух одинаковых полупроводников с проводимостями n - и p -типа. Именно такой переход мы и рассмотрим здесь, однако используются также *гетеропереходы* между разными полупроводниками (иногда даже с одинаковым типом проводимости).

Итак, возьмем два одинаковых полупроводника с различными типами проводимости и приведем их в плотный контакт друг с другом. Возникает неравновесная ситуация, поскольку концентрации электронов и дырок различны теперь уже в единой системе соединенных полупроводников. Выравнивание концентраций, как мы знаем, происходит в процессе диффузии, которая в нашем случае осложнена тем, что диффундируют *заря-*

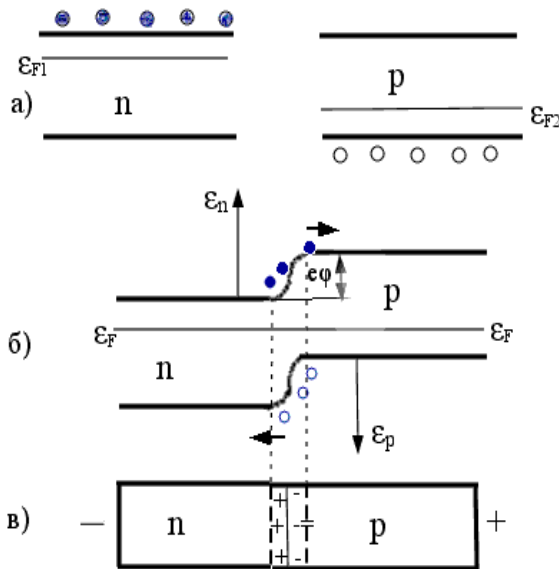
¹ Именно эта область температур обычно имеется в виду в задачах на семинарах и т.п.

² Эта температура определяется из соотношения $kT/\varepsilon_i = 1/\ln(N_c/N_d)$.

женные частицы, которые, к тому же, могут рекомбинировать друг с другом (см. «Рекомбинация» в [предыдущей лекции](#)).

Займемся сначала электронами. Они диффундируют из n -области в p -область, но там их ждут большие неприятности. Во-первых, они теряют статус основных носителей и переходят в разряд неосновных. Но главное, что их там поджидают дырки (много дырок!) и, стало быть, быстрая рекомбинация, так что

Диффузия электронов



далеко проникнуть им во владения дырок не удастся. Возникает сравнительно узкая область с особыми свойствами, называемая областью p - n -перехода. Уход электронов из n -области заряжает ее приконтактную часть положительным зарядом, а их приход в p -область приносит туда

Рис.5 p - n -переход

отрицательный заряд (см. рис.5-в). С обеих сторон от контакта число носителей сильно уменьшается, поэтому область перехода имеет значительно большее удельное сопротивление¹, чем остальная часть кристалла, и часто называется *обедненной* (носителями).

Где же сосредоточен избыточный заряд в области перехода, если она бедна носителями? В n -области доноры отдали свои электроны и зарядились положительно; электроны сбежали, а заряд на донорах остался. В p -области акцепторы, захватив электроны зоны проводимости, зарядились отрицательно и породили дырки; дырки прорекомбинировали с электронами, а заряд на акцепторах остался. Таким образом, возникший заряд (рис. 5-в) в основном *сконцентрирован на неподвижных донорах и акцепторах*.

Где расположен избыточный заряд?

Теперь обратим внимание, что возникающее распределение зарядов *порождает электрическое поле, препятствующее диффузии электронов*. Пока электрон находится на своей территории

Электрическое поле из-за диффузии электронов

¹ Именно поэтому область перехода определяет электрические свойства всей структуры.

в области p - n -перехода и носит гордое имя основного носителя, это поле *мешает* ему перейти через границу с p -областью, но как только эта граница преодолена (если, конечно, это получится и рекомбинация его не погубит), электрон становится неосновным носителем и возникшее электрическое поле *выталкивает* его домой. Запомним это.

Теперь обратимся к дыркам. Они диффундируют из p -области в n -область, переносят *противоположный* по сравнению с электронами заряд в *противоположном* направлении, так что «работают» в ту же сторону. Они тоже рекомбинируют с электронами в «чужой» n -области, также способствуют обеднению области перехода носителями, т.е., одним словом, усугубляют ситуацию. Возникающее электрическое поле препятствует проникновению дырок через границу (пока они являются основными носителями) и выталкивает их назад, когда таки проникнув в n -область, они становятся неосновными носителями.

Диффузия дырок

Связанное с диффузией движение электронов в пределах n -области и дырок в пределах p -области, образует единый непрерывный ток основных носителей, называемый *диффузионным*. Пусть положительным считается ток, текущий по направлению из p -области в n -область. Тогда *диффузионный ток* (основных носителей, обозначим его $I_{осн}$) *положителен*.

Ток основных носителей

Кроме диффузионного тока основных носителей в области перехода есть ток неосновных носителей заряда, в частности беглецов-электронов и беглецов-дырок.¹ Этот ток ($I_{неосн}$) направлен против тока основных носителей, т.е. отрицателен, поскольку знак заряда неосновных носителей противоположен знаку основных. Возникшее электрическое поле в области p - n -перехода способствует этому току.²

Ток неосновных носителей

По мере того, как происходят только что описанные процессы, ток основных носителей уменьшается за счет увеличения препятствующего ему поля, а ток неосновных носителей изменяется мало, так как лимитируется их малой концентрацией.³

¹ Речь идет не только о «беглецах». Мы помним, что в любом полупроводнике есть некоторое количество неосновных зарядов.

² Часто говорят, что неосновные носители участвуют в дрейфовом токе.

³ Электрическое поле p - n -перехода «выталкивает» практически все появляющиеся неосновные носители заряда.

Равновесие наступает тогда, когда полный ток становится равным нулю:

$$I_{\text{осн}} + I_{\text{неосн}} = 0.$$

Теперь посмотрим, как все это выглядит с точки зрения зонной структуры. На [рис.5-б](#) показана зонная структура приведенных в контакт полупроводников из рис.5-а после наступления равновесия (14).). Вспомним, что аналогично тому, как равновесие по отношению к процессам теплопередачи характеризуется единой температурой системы, равновесие по концентрациям приводит к единому значению химического потенциала, который, как мы знаем, в твердых телах называют (не очень хорошо) энергией Ферми. Таким образом, в состоянии равновесия энергия Ферми должна быть одинаковой во всех частях составного кристалла, что и показано на [рис.5-б](#).

Мы рассматриваем «идеальный p - n -переход»,¹ в частности границу раздела двух полупроводников считаем плоской и такой, что область p - n -перехода можно считать чем-то вроде распределенного плоского конденсатора, так что вне ее электрического поля нет. В таком случае нет и искажений зонной структуры, т.е. все выглядит так, как на [рис.5-а](#), только со смещением границы зон из-за смещения энергий Ферми. Но границы энергетических зон должны быть непрерывными, что достигается соединением их в области p - n -перехода корявой линией на [рис.5-б](#).

Теперь видно, что для электронов (синенькие такие) в роли основных носителей возник потенциальный барьер высотой $e\varphi = |\varepsilon_{F_2} - \varepsilon_{F_1}|$ (величину $e\varphi$ будем считать положительной). В точности такой же барьер возникает для основных носителей-дырок (беленькие кружочки, движущиеся на рис.5-б справа налево), ведь ось энергии для дырок направлена сверху вниз, переверните мысленно рисунок.

Долю электронов (и дырок), которые могут преодолеть барьер, можно оценить величиной $e^{-e\varphi/kT}$, и для тока основных носителей написать²

$$I_{\text{осн}} = I_d e^{-e\varphi/kT},$$

(14) Условие равновесия в p - n -переходе

Идеальный p - n -переход

¹ Свойства реального p - n -перехода могут существенно отличаться от описываемых здесь.

² Если не учитывать плотность числа состояний, то эта доля есть $\int_{e\varphi}^{\infty} e^{-\varepsilon/kT} d\varepsilon / \int_0^{\infty} e^{-\varepsilon/kT} d\varepsilon = e^{-e\varphi/kT}$.

где I_d – трудно считаваемый диффузионный ток без барьера (мы обойдемся и без его явного выражения).¹ Условие равновесия (14) принимает вид

$$I_d e^{-e\phi/kT} + I_{\text{неосн}} = 0.$$

(15) Условие равновесия в p - n -переходе

Посмотрим, что произойдет, если к концам полупроводников подвести напряжение U . Поскольку сопротивление области p - n -перехода существенно больше сопротивления остальной части полупроводников, то можно считать, что и при протекании тока напряжение U будет целиком приложено к этой области.

Будем говорить, что $U > 0$, если *плюс подается на полупроводник p -типа* (см. [рис.5-б](#)), что согласуется с [выбранным положительным направлением](#) тока через p - n -переход. Положительное напряжение называют *прямым*. При таком напряжении внешнее электрическое поле направлено против поля в p - n -переходе, потенциальный барьер уменьшается на величину энергии eU носителя заряда во внешнем поле, так что ток основных носителей увеличится:

$$I'_{\text{осн}} = I_d e^{-e(\phi-U)/kT}.$$

Прямое напряжение

Напряжение с плюсом на n -области называется *обратным*. При таком напряжении внутреннее поле складывается с внешним, барьер увеличивается, ток основных носителей уменьшается. Формула (16) остается верной.

(16) Обратное напряжение

Как при прямом, так и при обратном напряжении, можно считать, что ток неосновных носителей практически не меняется. Полный ток, как и раньше, представляет собой сумму токов основных и неосновных носителей

$$I = I_d e^{-e(\phi-U)/kT} + I_{\text{неосн}} = -I_{\text{неосн}} e^{eU/kT} + I_{\text{неосн}}.$$

Здесь мы сначала воспользовались формулой (16), а затем соотношением (15). Наконец, последний штрих. Мы знаем, что ток $I_{\text{неосн}} < 0$, это неудобно для окончательной формулы, поэтому положим $I_0 = -I_{\text{неосн}}$, что позволяет написать *вольт-амперную характеристику p - n -перехода* в следующем виде:

$$I = I_0 (e^{eU/kT} - 1).$$

(17) Вольт-амперная характеристика идеального p - n -перехода

¹ Написанная формула верна, но вывод ее, строго говоря, должен быть чуть-чуть другим. Дело в том, что электроны выступают в роли основных носителей только до границы n -области, где высота барьера равна $e\phi/2$ так что вероятность того, что электрон дает вклад в ток основных носителей, равна $\sim e^{-e\phi/2kT}$. Эту вероятность нужно умножить на такую же вероятность того, что дырка дойдет до границы с другой стороны. Результат получается тот же самый.

Таким образом, ток через p - n -переход положителен при прямом напряжении $U > 0$ и быстро нарастает при увеличении напряжения, если $eU/kT > 1$, обращается в ноль при $U = 0$ и становится отрицательным при обратном напряжении $U < 0$. При отрицательном напряжении таком, что $e|U|/kT \gg 1$,

$$I = -I_0 = I_{\text{неосн}},$$

т.е. предельный обратный ток целиком определяется током неосновных носителей. Этот ток существенно меньше прямого тока, что дает возможность использовать p - n -переход в диодах, в частности, для выпрямления переменного тока. Однако обращают на себя внимание два обстоятельства. Во-первых, зависимость $I(U)$ для прямого тока в идеальном случае линейна только в узкой области $eU/kT \ll 1$, а во-вторых, свойства p - n -перехода сильно зависят от температуры, что делает, к сожалению, использующие его приборы весьма нестабильными.

Задание. Нарисуйте вольт-амперную характеристику p - n -перехода (17).

Сравните рисунки 5-б и 5-в. Что же все-таки откладывается по оси абсцисс в упрощенной схеме энергетических зон?

Эффект Холла

С эффектом Холла мы уже встречались, изучая проводимость металлов. Однако особенно он интересен для полупроводников, потому что позволяет отличить электронный тип проводимости от дырочного.¹ Напомню, что суть его заключается в



возникновении
 речной разности
 потенциалов при
 щении образца с то-
 ком в магнитное по-
 ле. Посмотрим сна-
 чала на полупровод-
 ник n -типа, пред-
 ставляющий собой
 прямоугольный па-

Рис.6
 Эффект Холла
 в полупро-
 воднике n -
 типа

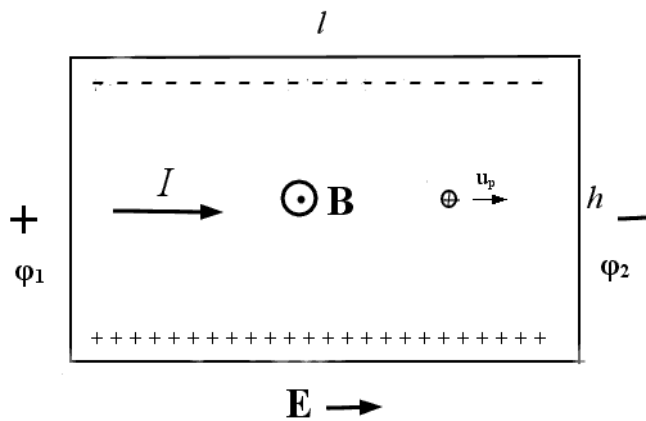
¹ с некоторыми оговорками, см. (20).

раллелепипед (см. рис.6). На этом рисунке показано направление внешнего электрического поля, вызванного им *первичного* тока I и направление дрейфовой скорости электронов \mathbf{u}_n . Видно также, что магнитное поле \mathbf{B} перпендикулярно плотности тока. На движущиеся электроны действует сила Лоренца $\mathbf{F}_{L,n} = -e[\mathbf{u}_n, \mathbf{B}]$. Эта сила направлена *вниз* и вызывает разделение зарядов как показано на рисунке. Поле этой силы есть

$$\mathbf{E}_{L,n} = [\mathbf{u}_n, \mathbf{B}] = -\mu_n[\mathbf{E}, \mathbf{B}]$$

(см. [формулу \(3'\)](#)).

Теперь посмотрим на полупроводник *p*-типа. На движущиеся со скоростью



с скоростью

\mathbf{u}_p дырки действует

сила Лоренца

$$\mathbf{F}_{L,p} = e[\mathbf{u}_p, \mathbf{B}],$$

направленная на рис.

7 тоже *вниз*,¹ но она

вызывает противоположное

разделение

зарядов. Таким обра-

зом, по знаку разно-

сти потенциалов между верхней и нижней гранью полупроводника можно определить его тип проводимости. Поле лоренцевой силы по отношению к дыркам есть

$$\mathbf{E}_{L,p} = [\mathbf{u}_p, \mathbf{B}] = \mu_p[\mathbf{E}, \mathbf{B}].$$

Разделение зарядов приводит к появлению поперечного электростатического *холловского* поля \mathbf{E}_H . Если же замкнуть нижнюю и верхнюю грани образца на небольшое сопротивление, то образуется *вторичная* цепь, по которой потечет ток. При этом сила Лоренца играет роль сторонней силы.

В полупроводниках с чисто электрической или чисто дырочной проводимостью найти поле \mathbf{E}_H (без вторичной цепи) очень легко, потребовав, чтобы в стационарной ситуации результирующая сила, действующая на носители заряда, обратилась в ноль. В случае полупроводника со смешанной проводимостью дело обстоит чуть-чуть сложнее. Условие равновесия сводится к обращению в ноль поперечного тока. Поперечный ток

Рис.7
Эффект Холла
в полупро-
воднике *p*-
типа

¹ Знак у силы Лоренца меняется дважды: из-за заряда и из-за направления дрейфовой скорости.

вызывается двумя полями – полем силы Лоренца ($\mathbf{E}_{L,n}$ для электронов и $\mathbf{E}_{L,p}$ для дырок) и электростатическим полем \mathbf{E}_H . По закону Ома в дифференциальной форме для плотности электронного поперечного тока пишем

$$\mathbf{j}_{H,n} = \sigma_n(\mathbf{E}_{L,n} + \mathbf{E}_H) = e\mu_n n(-\mu_n[\mathbf{E}, \mathbf{B}] + \mathbf{E}_H),$$

а для дырочного

$$\mathbf{j}_{H,p} = \sigma_p(\mathbf{E}_{L,p} + \mathbf{E}_H) = e\mu_p p(\mu_p[\mathbf{E}, \mathbf{B}] + \mathbf{E}_H).$$

В равновесии должно быть

$$\mathbf{j}_{H,n} + \mathbf{j}_{H,p} = 0,$$

т.е.

$$(e\mu_n n + e\mu_p p)\mathbf{E}_H = (e\mu_n^2 n - e\mu_p^2 p)[\mathbf{E}, \mathbf{B}],$$

или

$$\mathbf{E}_H = \frac{\mu_n^2 n - \mu_p^2 p}{\mu_n n + \mu_p p} [\mathbf{E}, \mathbf{B}].$$

(20) Поле Холла

Видно, что направление холловского поля определяется не только концентрациями носителей заряда, но и их подвижностями, поэтому, строго говоря, не всегда можно определить тип проводимости только по знаку эффекта Холла.

Задание. Проверьте, что направление \mathbf{E}_H в формуле (20) согласуется с рис. 6 и 7.

Если образец имеет форму прямоугольного параллелепипеда длиной l и шириной h (см. рис. 6 и 7), то легко получить и формулу для холловской разности потенциалов U_H :

$$U_H = hE_H = (h/l) \frac{\mu_n^2 n - \mu_p^2 p}{\mu_n n + \mu_p p} BU,$$

где $U = \varphi_1 - \varphi_2$ – разность потенциалов в первичной цепи.