

Фиг. 28.9. Типичные сигналы циклотронного резонанса в германии (а) и кремнии (б). (Из работы [6].)

Поле лежит в плоскости (110) и составляет с осью [001] углы  $60^\circ$  (Ge) и  $30^\circ$  (Si).

При таких условиях электропроводность полупроводника столь мала (в отличие от ситуации в металлах, см. т. 1, стр. 278), что возбуждающее электромагнитное поле может проникнуть в глубь образца достаточно далеко и вызвать резонанс; при этом не возникает никаких трудностей, связанных с глубиной скин-слоя. С другой стороны, при таких условиях (низкие температуры, чистые образцы) число носителей, которые при тепловом равновесии способны участвовать в резонансе, может оказаться столь малым, что носители должны создаваться другим путем, например с помощью фотовозбуждения. Некоторые типичные данные, получающиеся при изучении циклотронного резонанса, представлены на фиг. 28.9.

### ЧИСЛО НОСИТЕЛЕЙ ТОКА ПРИ ТЕРМОДИНАМИЧЕСКОМ РАВНОВЕСИИ

Самой важной характеристикой любого полупроводника при температуре  $T$  является число  $n_c$  электронов в зоне проводимости, приходящееся на единицу объема, и число дырок  $p_v$ <sup>1)</sup> на единицу объема в валентной зоне. Определение зависимости этих величин от температуры представляет собой весьма простое по существу, но иногда алгебраически весьма громоздкое упражнение по применению статистики Ферми — Дирака к соответствующей системе одноэлектронных уровней.

Значения  $n_c(T)$  и  $p_v(T)$ , как мы потом увидим, очень сильно зависят от наличия примесей. Однако существуют некоторые общие соотношения, которые выполняются независимо от чистоты образца; мы рассмотрим их в первую очередь. Предположим, что плотность уровней (см. т. 1, стр. 156) есть  $g_c(\mathcal{E})$  в зоне проводимости и  $g_v(\mathcal{E})$  в валентной зоне. Из-за наличия примесей, как мы увидим ниже, возникают дополнительные уровни, лежащие между потолком валентной зоны  $\mathcal{E}_v$  и дном зоны проводимости  $\mathcal{E}_c$ , тогда как вид  $g_c(\mathcal{E})$  и  $g_v(\mathcal{E})$  не меняется заметным образом. Поскольку проводимость полностью обусловлена электронами на уровнях, лежащих в зоне проводимости, или же дырками на уровнях

<sup>1)</sup> Концентрация дырок обычно обозначается буквой  $p$  (positive — положительный). Широко распространенное обозначение для концентрации электронов использует совпадение первых букв слов number (число) и negative (отрицательный).

в валентной зоне, то независимо от концентрации примесей число носителей, имеющих при данной температуре  $T$ , описывается выражениями

$$n_c(T) = \int_{\mathcal{E}_c}^{\infty} d\mathcal{E} g_c(\mathcal{E}) \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1},$$

$$p_v(T) = \int_{-\infty}^{\mathcal{E}_v} d\mathcal{E} g_v(\mathcal{E}) \left(1 - \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1}\right) = \int_{-\infty}^{\mathcal{E}_v} d\mathcal{E} g_v(\mathcal{E}) \frac{1}{e^{(\mu-\mathcal{E})/k_B T} + 1}. \quad (28.9)$$

При определении  $n_c$  и  $p_v$  влияние примесей сказывается только через величину химического потенциала  $\mu$ <sup>1)</sup>, который подставляется в (28.9). Чтобы найти  $\mu$ , нужно иметь некоторую информацию о примесных уровнях. Однако из выражений (28.9) можно получить кое-какие полезные сведения, не зависящие от точного значения химического потенциала, если только оно удовлетворяет условиям

$$\begin{aligned} \mathcal{E}_c - \mu &\gg k_B T, \\ \mu - \mathcal{E}_v &\gg k_B T. \end{aligned} \quad (28.10)$$

Существует область значений  $\mu$ , для которой условия (28.10) выполняются даже при весьма малой ширине запрещенной зоны  $E_g = \mathcal{E}_c - \mathcal{E}_v$ , составляющей всего несколько десятых электронвольта, и при температурах, близких к комнатной. Дальнейшее рассмотрение будет заключаться в следующем: допустив справедливость условий (28.10), упростим с их помощью выражения (28.9); затем, исходя из полученных таким образом значений  $n_c$  и  $p_v$  и имеющейся информации о возможных примесных уровнях, вычислим реальное значение химического потенциала и проверим, действительно ли оно попадает в область, определяемую условиями (28.10). Если химический потенциал лежит в указанном интервале, полупроводник называется «невыврожденным» и проведенное рассмотрение справедливо. Если это условие не выполняется, мы имеем дело с «вырожденным полупроводником» и должны работать непосредственно с выражениями (28.9), не производя упрощений, основанных на использовании условий (28.10).

Пусть условия (28.10) выполняются; тогда благодаря тому, что каждый уровень в зоне проводимости лежит выше  $\mathcal{E}_c$ , а каждый уровень валентной зоны — ниже  $\mathcal{E}_v$ , мы можем упростить выражение для функций распределения в (28.9):

$$\begin{aligned} \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1} &\approx e^{-(\mathcal{E}-\mu)/k_B T}, \quad \mathcal{E} > \mathcal{E}_c, \\ \frac{1}{e^{(\mu-\mathcal{E})/k_B T} + 1} &\approx e^{-(\mu-\mathcal{E})/k_B T}, \quad \mathcal{E} < \mathcal{E}_v. \end{aligned} \quad (28.11)$$

<sup>1)</sup> Очень часто химический потенциал полупроводника называют «уровнем Ферми», однако это — довольно неудачная терминология. Поскольку химический потенциал почти всегда лежит в запрещенной зоне, ни один одноэлектронный уровень не совпадает с уровнем Ферми (в отличие от ситуации, существующей в металлах). Поэтому обычное определение уровня Ферми (как той энергии, ниже которой в металле в основном состоянии одноэлектронные уровни заполнены, а выше не заполнены) в полупроводнике оказывается неоднозначным: любая энергия в пределах запрещенной зоны отделяет занятые уровни от незанятых при  $T = 0$ . Когда речь идет о полупроводниках, термин «уровень Ферми» следует рассматривать не более как синоним термина «химический потенциал».

Таким образом, выражения (28.9) принимают вид

$$\boxed{\begin{aligned} n_c(T) &= N_c(T) e^{-(\mathcal{E}_c - \mu)/k_B T}, \\ p_v(T) &= P_v(T) e^{-(\mu - \mathcal{E}_v)/k_B T}, \end{aligned}} \quad (28.12)$$

где

$$\begin{aligned} N_c(T) &= \int_{\mathcal{E}_c}^{\infty} d\mathcal{E} g_c(\mathcal{E}) e^{-(\mathcal{E} - \mathcal{E}_c)/k_B T}, \\ P_v(T) &= \int_{-\infty}^{\mathcal{E}_v} d\mathcal{E} g_v(\mathcal{E}) e^{-(\mathcal{E}_v - \mathcal{E})/k_B T}. \end{aligned} \quad (28.13)$$

Поскольку области интегрирования в (28.13) включают точки, где показатель экспонент обращается в нуль,  $N_c(T)$  и  $P_v(T)$  представляют собой относительно медленно меняющиеся функции температуры по сравнению с теми экспонентами, на которые они умножаются в выражениях (28.12). Это их наиболее важное свойство. Обычно, впрочем, функции  $N_c$  и  $P_v$  можно вычислить в явном виде. Из-за наличия экспоненциальных множителей в подынтегральных выражениях (28.13) существенный вклад вносят только энергии из областей шириной  $k_B T$  вблизи краев зоны; в этих областях квадратичное приближение (28.2) или (28.3) обычно оказывается превосходным. Плотности уровней тогда можно представить в следующей форме (задача 3):

$$g_{c,v} = \sqrt{2 |\mathcal{E} - \mathcal{E}_{c,v}|} \frac{m_{c,v}^{3/2}}{\hbar^3 \pi^2}, \quad (28.14)$$

и вычисление интегралов (28.13) дает

$$\begin{aligned} N_c(T) &= \frac{1}{4} \left( \frac{2m_c k_B T}{\pi \hbar^2} \right)^{3/2}, \\ P_v(T) &= \frac{1}{4} \left( \frac{2m_v k_B T}{\pi \hbar^2} \right)^{3/2}. \end{aligned} \quad (28.15)$$

Здесь  $m_c^3$  — произведение главных значений тензора эффективной массы для зоны проводимости (т. е. его детерминант)<sup>1</sup>, величина  $m_v^3$  определяется аналогичным образом.

Формулы (28.15) могут быть приведены к виду, удобному для численных расчетов:

$$\boxed{\begin{aligned} N_c(T) &= 2,5 \left( \frac{m_c}{m} \right)^{3/2} \left( \frac{T}{300\text{K}} \right)^{3/2} \cdot 10^{19} \text{ см}^{-3}, \\ P_v(T) &= 2,5 \left( \frac{m_v}{m} \right)^{3/2} \left( \frac{T}{300\text{K}} \right)^{3/2} \cdot 10^{19} \text{ см}^{-3}, \end{aligned}} \quad (28.16)$$

где температура  $T$  должна быть выражена в кельвинах. Поскольку экспоненты в (28.12) меньше единицы по крайней мере на порядок, а  $m_c/m$  и  $m_v/m$  обычно близки к единице, формулы (28.16) показывают, что абсолютный верхний предел концентрации носителей в невырожденном полупроводнике составляет  $10^{18}$  или  $10^{19}$  см<sup>-3</sup>.

<sup>1</sup> Если в зоне проводимости имеется больше одного минимума, то необходимо взять сумму членов типа (28.14) и (28.15), отвечающих каждому минимуму. Эти суммы будут иметь такой же вид, как и (28.14) и (28.15), если определение  $m_c$  заменить следующим:  $m_c^{3/2} \rightarrow \Sigma m_c^{3/2}$ .

Пока нам неизвестно значение химического потенциала  $\mu$ , мы еще не можем получить  $n_c(T)$  и  $p_v(T)$  из (28.12). Однако произведение этих двух концентраций не зависит от  $\mu$ :

$$n_c p_v = N_c P_v e^{-(\mathcal{E}_c - \mathcal{E}_v)/k_B T} = N_c P_v e^{-E_g/k_B T}. \quad (28.17)$$

Этот результат (иногда называемый «законом действующих масс»<sup>1)</sup>) означает, что при данной температуре достаточно знать концентрацию носителей одного типа, чтобы найти концентрацию носителей другого типа. Каким образом это сделать, зависит от того, насколько существенный вклад в концентрацию носителей вносят примеси.

### СЛУЧАЙ СОБСТВЕННЫХ ПОЛУПРОВОДНИКОВ

Если кристалл является настолько чистым, что примеси вносят пренебрежимо малый вклад в концентрацию носителей, то мы имеем дело с «собственным полупроводником». В собственном полупроводнике электроны могут попасть в зону проводимости, только покинув заполненные ранее уровни в валентной зоне, оставив там вместо себя дырки. Таким образом, число электронов в зоне проводимости равно числу дырок в валентной зоне:

$$n_c(T) = p_v(T) \equiv n_i(T). \quad (28.18)$$

Поскольку  $n_c = p_v$ , величину  $n_i$ , равную им обеим, можно записать как  $(n_c p_v)^{1/2}$ . Тогда из формулы (28.17) получаем

$$n_i(T) = [N_c(T) P_v(T)]^{1/2} e^{-E_g/2k_B T}, \quad (28.19)$$

или же, используя (28.15) и (28.16),

$$\begin{aligned} n_i(T) &= \frac{1}{4} \left( \frac{2k_B T}{\pi \hbar^2} \right)^{3/2} (m_c m_v)^{3/4} e^{-E_g/2k_B T} = \\ &= 2,5 \left( \frac{m_c}{m} \right)^{3/4} \left( \frac{m_v}{m} \right)^{3/4} \left( \frac{T}{300\text{K}} \right)^{3/2} e^{-E_g/2k_B T} \cdot 10^{19} \text{ см}^{-3}. \end{aligned} \quad (28.20)$$

Теперь мы можем получить для собственных полупроводников условие применимости предположения (28.10), на котором основывалось наше рассмотрение. Обозначив через  $\mu_i$  значение химического потенциала собственного полупроводника, находим, что значения  $n_c$  и  $p_v$ , найденные по формулам (28.12), будут равны  $n_i$  [см. (28.19)], если

$$\mu = \mu_i = \mathcal{E}_v + \frac{1}{2} E_g + \frac{1}{2} k_B T \ln \left( \frac{P_v}{N_c} \right), \quad (28.21)$$

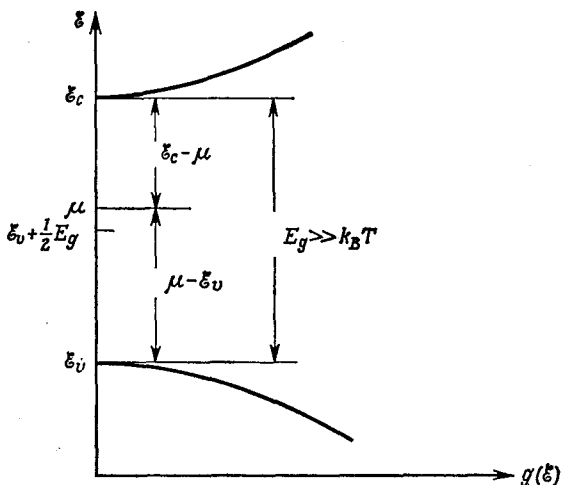
или, как следует из (28.15),

$$\mu_i = \mathcal{E}_v + \frac{1}{2} E_g + \frac{3}{4} k_B T \ln \left( \frac{m_v}{m_c} \right). \quad (28.22)$$

Это показывает, что при  $T \rightarrow 0$  химический потенциал  $\mu_i$  лежит точно посредине запрещенной зоны. Более того, поскольку  $\ln(m_v/m_c)$  есть величина порядка единицы,  $\mu_i$  не может отстоять от середины запрещенной зоны на расстояние,

<sup>1)</sup> Тут имеется полная аналогия с химическими реакциями: носители тока возникают при диссоциации электрона и дырки, связанных между собой.

существенно превышающее  $k_B T$  (фиг. 28.10). Следовательно, при значениях  $k_B T$ , малых по сравнению с  $E_g$ , химический потенциал находится далеко (в единицах  $k_B T$ ) от границ запрещенной зоны  $\mathcal{E}_c$  и  $\mathcal{E}_v$  и условие невырожденности (28.10) выполняется. Таким образом, формула (28.10) дает правильную оценку



Фиг. 28.10. В собственном полупроводнике, у которого ширина запрещенной зоны  $E_g$  велика по сравнению с  $k_B T$ , химический потенциал  $\mu$  лежит в области шириной порядка  $k_B T$  вблизи середины запрещенной зоны, а следовательно, далеко (по сравнению с  $k_B T$ ) от обеих ее границ  $\mathcal{E}_c$  и  $\mathcal{E}_v$ .

равных между собой величин  $n_c$  и  $p_v$  для собственного полупроводника, если только ширина запрещенной зоны  $E_g$  велика по сравнению с  $k_B T$ , а это условие выполняется почти во всех полупроводниках при температурах порядка комнатной и ниже.

### СЛУЧАЙ НЕСОБСТВЕННЫХ ПОЛУПРОВОДНИКОВ. НЕКОТОРЫЕ ОСНОВНЫЕ ЧЕРТЫ

Если примеси поставляют существенную часть электронов зоны проводимости и (или) дырок валентной зоны, то мы имеем дело с несобственным полупроводником. Из-за наличия таких добавочных источников носителей концентрация электронов в зоне проводимости уже не обязательно должна быть равна концентрации дырок в валентной зоне, т. е.

$$n_c - p_v = \Delta n \neq 0. \quad (28.23)$$

Поскольку закон действующих масс (28.17) выполняется независимо от роли примесей, мы можем использовать определение (28.19) величины  $n_i(T)$  и записать в общем случае

$$n_c p_v = n_i^2. \quad (28.24)$$

Соотношения (28.24) и (28.23) дают возможность выразить концентрации носителей в несобственном полупроводнике через их значения для собственного

полупроводника и через отклонение  $\Delta n$  от этих значений:

$$\left\{ \begin{array}{l} n_c \\ p_v \end{array} \right\} = \frac{1}{2} [(\Delta n)^2 + 4n_i^2]^{1/2} \pm \frac{1}{2} \Delta n. \quad (28.25)$$

Величину  $\Delta n/n_i$ , которая является мерой роли примесей как источника носителей, можно представить в особенно простой форме, рассматривая ее как функцию химического потенциала  $\mu$ . Поскольку формула (28.12) имеет вид <sup>1)</sup>

$$n_c = e^{\beta(\mu - \mu_i)} n_i, \quad p_v = e^{-\beta(\mu - \mu_i)} n_i, \quad (28.26)$$

получаем

$$\frac{\Delta n}{n_i} = 2 \operatorname{sh} \beta (\mu - \mu_i). \quad (28.27)$$

Мы отмечали, что если ширина запрещенной зоны  $E_g$  велика по сравнению с  $k_B T$ , то химический потенциал собственного полупроводника  $\mu_i$  будет удовлетворять условию отсутствия вырождения (28.10). Но из соотношения (28.27) следует, что когда разность между  $\mu_i$  и  $\mathcal{E}_c$  или  $\mathcal{E}_v$  велика по сравнению с  $k_B T$ , то столь же велика и разность между этими величинами и  $\mu$ , если только  $\Delta n$  не превышает на много порядков концентрацию носителей собственного полупроводника  $n_i$ . Таким образом, в области, где примеси не играют существенной роли, предположение о невырожденности, лежащее в основе вывода формулы (28.27), справедливо при  $E_g \gg k_B T$ .

Отметим также, что если отклонение  $\Delta n$  велико по сравнению с  $n_i$ , то, согласно формуле (28.25), концентрация носителей одного типа фактически равна  $\Delta n$ , а концентрация носителей другого типа меньше примерно в  $(\Delta n/n_i)^2$  раз. Поэтому если основным источником носителей тока являются примеси, то один из двух типов носителей будет доминирующим. Несобственный полупроводник называется полупроводником  $n$ -типа или  $p$ -типа в зависимости от того, какой тип носителей преобладает — электроны или дырки.

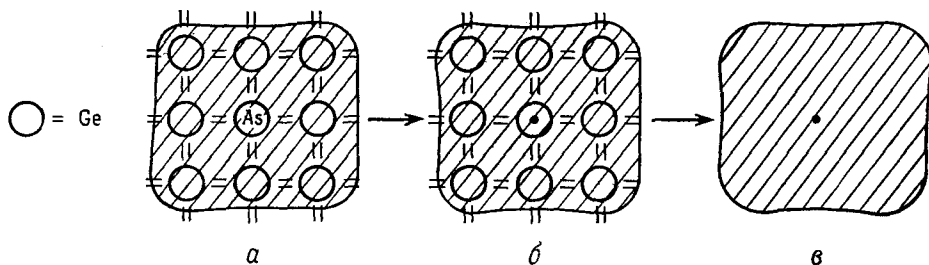
Чтобы полностью описать концентрации носителей в несобственных полупроводниках, необходимо определить  $\Delta n$  или  $\mu$ . Для этого нужно исследовать природу электронных уровней, обусловленных примесями, и статистическую механику заполнения этих уровней при термодинамическом равновесии.

## ПРИМЕСНЫЕ УРОВНИ

Примеси, которые вносят вклад в концентрацию носителей тока, называются *донорами*, если они поставляют дополнительные электроны в зону проводимости, и *акцепторами*, когда они поставляют дополнительные дырки в валентную зону (т. е. захватывают оттуда электроны). Донорные примеси — это атомы с более высокой валентностью, чем атомы, образующие чистое вещество (вещество-матрицу), а акцепторы — атомы с более низкой валентностью.

Рассмотрим, например, примеси замещения в полупроводнике IV группы. Допустим, что мы взяли кристалл чистого германия и заменили произвольный атом германия его правым соседом по периодической системе элементов — мышьяком (фиг. 28.11). Ион германия отдает четыре валентных электрона

<sup>1)</sup> Чтобы убедиться в правильности этих соотношений, не обязательно подставлять явные выражения для  $n_i$  и  $\mu_i$ ; достаточно заметить, что  $n_c$  и  $p_v$  пропорциональны соответственно  $\exp(\beta\mu)$  и  $\exp(-\beta\mu)$  и что при  $\mu = \mu_i$  обе эти концентрации равны  $n_i$ .



**Фиг. 28.11.** *a* — схематическое изображение донорной примеси (атом мышьяка с валентностью 5), замещающей в кристалле атом германия (валентность 4).  
*б* — атом мышьяка (As) можно представить как атом германия *плюс* добавочный единичный положительный заряд, закрепленный в узле решетки (точка в кружке).  
*в* — в полуклассическом приближении, в котором чистый полупроводник рассматривается как однородная среда, примесный атом мышьяка заменяется фиксированным точечным зарядом  $+e$  (точка).

и имеет заряд  $4e$ , а ион мышьяка отдает пять валентных электронов и имеет заряд  $5e$ . Если в первом приближении пренебречь различием структуры ионных остатков мышьяка и германия, то можно представить замещение атома германия атомом мышьяка как менее существенную модификацию и считать, что атом германия остается на месте, но в тот же узел кристаллической решетки помещается дополнительный неподвижный положительный заряд  $e$ , а также еще один электрон.

Рассмотрим общую модель полупроводника, легированного донорными примесями. В единице объема чистого идеального кристалла хаотически распределены <sup>1)</sup>  $N_D$  фиксированных притягивающих центров с зарядами  $+e$  и такое же число дополнительных электронов с зарядами  $-e$ . Предполагается, что каждый центр с зарядом  $+e$  может образовать связанное состояние с одним из дополнительных электронов, обладающих зарядом  $-e$  <sup>2)</sup>. Если бы примесь находилась не внутри полупроводника, а в пустом пространстве, то энергия связи электрона была бы просто равна первому ионизационному потенциалу атома примеси, равному 9,81 эВ для мышьяка. Однако (и *это имеет решающее значение для теории полупроводников*) благодаря тому, что примесь находится в полупроводнике, энергия связи значительно уменьшается (до 0,013 эВ для мышьяка в германии). Это происходит по следующим причинам.

1. Поле заряда, моделирующего примесь, должно быть уменьшено в  $\epsilon$  раз ( $\epsilon$  — статическая диэлектрическая проницаемость полупроводника) <sup>3)</sup>. Значения  $\epsilon$  могут быть весьма велики ( $\epsilon \approx 16$  в германии); они обычно лежат в пределах примерно между 10 и 20, но в ряде случаев могут достигать 100 и выше. Большие значения диэлектрических проницаемостей обусловлены малой шириной запрещенной зоны. Если бы запрещенная зона вообще отсутствовала, то кристалл был бы не полупроводником, а металлом, и статическая диэлектрическая

<sup>1)</sup> При весьма специальных условиях может оказаться, что и сами примеси образуют пространственно-упорядоченную структуру. Мы не будем здесь рассматривать эту возможность.

<sup>2)</sup> Как мы увидим, связь очень слаба, и электроны, связанные с центром, легко высвобождаются в результате теплового возбуждения.

<sup>3)</sup> Использование макроскопической электродинамики при рассмотрении энергии связи отдельного электрона оправдывается тем обстоятельством, что (как показано ниже) волновая функция связанного электрона спадает на характерном расстоянии в несколько сотен ангстрем.

проницаемость была бы бесконечно большой, поскольку в этом случае статическое электрическое поле вызывало бы ток и электроны могли бы уйти сколь угодно далеко от своего первоначального положения. Если ширина запрещенной зоны мала, но не равна нулю, то диэлектрическая проницаемость не будет бесконечной, но может быть весьма велика, поскольку пространственное распределение электронов сравнительно легко поддается изменению<sup>1)</sup>.

2. Электрон, движущийся в полупроводнике, должен описываться не соотношением между энергией и импульсом, характерным для свободного пространства, а полуклассическим соотношением (см. гл. 12)  $\mathcal{E}(\mathbf{k}) = \mathcal{E}_c(\mathbf{k})$ , где  $\hbar\mathbf{k}$  — квазиимпульс электрона, а  $\mathcal{E}_c(\mathbf{k})$  — зависимость энергии электрона от импульса в зоне проводимости. Иначе говоря, можно считать, что добавочный электрон, внесенный примесью, находится в состоянии, которое описывается суперпозицией уровней зоны проводимости чистого вещества, соответствующим образом измененных из-за наличия дополнительного локализованного заряда  $+e$ , моделирующего примесь. Чтобы энергия электрона была минимальной, он должен занимать только уровни вблизи дна зоны проводимости, для которых применимо квадратичное приближение (28.2). Если бы минимум зоны проводимости располагался в точке с кубической симметрией, то электрон вел бы себя почти как свободный, но обладал бы эффективной массой, отличной от массы свободного электрона  $m$ . В более общей ситуации зависимость энергии от волнового вектора будет некоторой анизотропной квадратичной функцией  $k$ . В любом случае, однако, мы можем в первом приближении считать, что электрон движется в вакууме, но имеет соответствующим образом определенную эффективную массу  $m^*$ , а не массу свободного электрона. Эта масса, вообще говоря, меньше массы свободного электрона во многих случаях в 10 и более раз.

В силу указанных двух обстоятельств электрон, находящийся в полупроводнике в присутствии донорной примеси с зарядом  $e$ , можно рассматривать как частицу с зарядом  $-e$  и массой  $m^*$ , движущуюся в вакууме в поле притягивающего центра с зарядом  $e/\epsilon$ . Эта задача в точности совпадает с задачей об атоме водорода, если заменить произведение зарядов ядра и электрона  $-e^2$  на  $-e^2/\epsilon$ , а массу свободного электрона  $m$  — на  $m^*$ . Таким образом, радиус первой бортовой орбиты  $a_0 = \hbar^2/m\epsilon^2$  окажется равным

$$r_0 = \frac{m}{m^*} \epsilon a_0, \quad (28.28)$$

а энергия связи в основном состоянии  $me^4/\hbar^2 = 13,6$  эВ имеет теперь величину

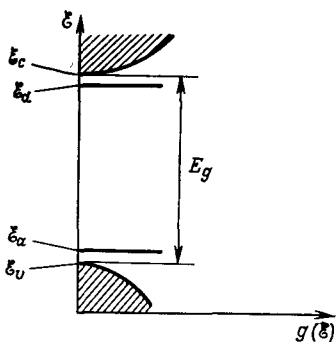
$$\mathcal{E} = \frac{m^*}{m} \frac{1}{\epsilon^2} \cdot 13,6 \text{ эВ}. \quad (28.29)$$

Для разумных значений  $m^*/m$  и  $\epsilon$  радиус  $r_0$  может достигать  $100 \text{ \AA}$  и более. Это очень важно для согласованности всего рассмотрения в целом, поскольку использование как полуклассической модели, так и макроскопической диэлектрической проницаемости было основано на предположении, что описываемые поля мало меняются на расстояниях порядка постоянной решетки.

<sup>1)</sup> Связь между малой шириной запрещенной зоны и большим значением диэлектрической проницаемости можно понять также с точки зрения теории возмущений: величина диэлектрической проницаемости является мерой степени искажения волновой функции электрона слабым электрическим полем. Малая ширина запрещенной зоны означает, что малы энергетические знаменатели и соответственно велики поправки первого порядка к волновым функциям.



Кроме того, при типичных значениях  $m^*/m$  и  $\epsilon$  энергия связи может уменьшиться в тысячу и более раз по сравнению со значением, равным 13,6 эВ. Действительно, малой ширине запрещенной зоны соответствует обычно большая диэлектрическая проницаемость, поэтому почти всегда оказывается, что энергия связи электрона с донорной примесью мала по сравнению с шириной запрещенной зоны полупроводника. Как уже говорилось, энергия связи отсчитывается от энергии уровней зоны проводимости, из которых формируется связанный примесный уровень. Таким образом, можно сделать вывод, что донорные примеси приводят к образованию дополнительных электронных уровней, энергия  $\mathcal{E}_d$  которых



**Фиг. 28.12.** Плотность уровней в полупроводнике, содержащем как донорные, так и акцепторные примеси.

Донорные уровни  $\mathcal{E}_d$  обычно находятся на малом по сравнению с  $E_g$  расстоянии от дна зоны проводимости  $\mathcal{E}_c$ , а акцепторные уровни  $\mathcal{E}_a$  обычно близки к потолку валентной зоны  $\mathcal{E}_v$ .

ривать зонную структуру, то такой связанной дырке будет соответствовать дополнительный электронный уровень с энергией  $\mathcal{E}_a$ , лежащий немного выше потолка валентной зоны (см. фиг. 28.12). Когда этот уровень пуст, дырка находится в связанном состоянии. Энергия связи дырки как раз равна энергии  $\mathcal{E}_a - \mathcal{E}_v$ , необходимой для того, чтобы перебросить электрон с потолка валентной зоны на акцепторный уровень, заполнив таким образом дырку около акцептора и создав свободную дырку в валентной зоне.

Самая важная информация о донорных и акцепторных уровнях заключается в том, что они лежат вблизи границ запрещенной зоны<sup>2)</sup>. При тепловом возбуждении гораздо легче вызвать переход электрона в зону проводимости с донорного уровня или дырки в валентную зону с акцепторного уровня, чем переход через всю запрещенную зону из валентной зоны в зону проводимости. Поэтому донорные и акцепторные примеси, если их концентрация не слишком мала, играют гораздо более важную роль в создании носителей, чем переброс электронов через всю запрещенную зону, характерный для собственных полупроводников.

<sup>1)</sup> По тем же причинам, что и в случае донорных примесей, энергия связи дырки весьма мала; иначе говоря, электроны валентной зоны легко переходят на акцепторный уровень при тепловом возбуждении.

<sup>2)</sup> Некоторые экспериментальные данные об известных донорных и акцепторных уровнях приведены в табл. 28.2.

меньше энергии  $\mathcal{E}_c$ , отвечающей дну зоны проводимости, причем разность между ними мала по сравнению с шириной запрещенной зоны  $E_g$  (фиг. 28.12).

Аналогичные рассуждения применимы к акцепторным примесям, валентность которых на единицу меньше валентности основного вещества (например, галлий в германии). Чтобы описать такую примесь, можно считать, что фиксированный заряд  $-e$  расположен в том же месте, где и атом основного вещества, и при этом один электрон в кристалле отсутствует. Отсутствующий электрон можно рассматривать как дырку, связанную с избыточным отрицательным зарядом, моделирующим примесь, причем энергия связи, как и в предыдущем случае, мала по сравнению с шириной запрещенной зоны  $E_g$ <sup>1)</sup>. Если рассмат-

Таблица 28.2

Уровни примесных атомов V группы (доноры) и III группы (акцепторы) в германии и кремнии <sup>a)</sup>

## III группа, акцепторы

	$\mathcal{E}_a - \mathcal{E}_v$ , эВ				
	B	Al	Ga	In	Tl
Si	0,046	0,057	0,065	0,16	0,26
Ge	0,0104	0,0102	0,0108	0,0112	0,01

## V группа, доноры

	$\mathcal{E}_c - \mathcal{E}_d$ , эВ			
	P	As	Sb	Bi
Si	0,044	0,049	0,039	0,069
Ge	0,0120	0,0127	0,0096	—

Ширина запрещенной зоны  $E_g = \mathcal{E}_c - \mathcal{E}_v$  при комнатной температуре

	$E_g$ , эВ
Si	1,12
Ge	0,67

<sup>a)</sup> Данные взяты из книги [7].НАСЕЛЕННОСТЬ ПРИМЕСНЫХ УРОВНЕЙ  
ПРИ ТЕРМОДИНАМИЧЕСКОМ РАВНОВЕСИИ

Чтобы определить, какое число носителей тока может быть переброшено с примесных уровней путем теплового возбуждения, нужно вычислить среднее число электронов на этих уровнях при заданных температуре и химическом потенциале. Концентрацию примесей будем считать достаточно малой, чтобы можно было пренебречь взаимодействием между электронами и дырками, локализованными на различных примесях. Мы можем тогда найти концентрацию электронов  $n_d$  (или дырок  $p_a$ ), связанных с донорными (или акцепторными) примесями, просто умножив концентрацию доноров  $N_d$  (или акцепторов  $N_a$ ) на среднее число электронов, локализованных на отдельной примеси. Будем предполагать для простоты, что примесь создает только один одноэлектронный уровень <sup>1)</sup>, и вычислим его среднюю населенность.

**Донорный уровень.** Если мы пренебрегаем электрон-электронными взаимодействиями, то данный уровень может быть пустым, может быть занят одним электроном с произвольно направленным спином или же занят двумя электро-

<sup>1)</sup> Это требование не вытекает из каких-либо общих соображений, и мы делаем предположение о наличии только одного связанного уровня лишь для простоты рассмотрения. Однако наши качественные выводы являются достаточно общими (см. задачу 4, п. «в»).

нами с противоположными спинами. Однако в последнем случае кулоновское отталкивание двух локализованных электронов столь сильно увеличивает энергию уровня, что нахождение двух электронов на одном уровне практически запрещено. Общее выражение для среднего числа электронов в системе при термодинамическом равновесии имеет вид

$$\langle n \rangle = \frac{\sum N_j e^{-\beta(E_j - \mu N_j)}}{\sum e^{-\beta(E_j - \mu N_j)}}, \quad (28.30)$$

где суммирование проводится по всем состояниям системы,  $E_j$  и  $N_j$  — энергия и число электронов в  $j$ -м состоянии,  $\mu$  — химический потенциал. В данном случае система представляет собой отдельную примесь с тремя состояниями: одно состояние, в котором отсутствуют электроны и которое не вносит вклада в энергию, а два состояния с одним электроном, имеющие энергию  $\mathcal{E}_d$ . Поэтому из (28.30) получаем

$$\langle n \rangle = \frac{2e^{-\beta(\mathcal{E}_d - \mu)}}{1 + 2e^{-\beta(\mathcal{E}_d - \mu)}} = \frac{1}{\frac{1}{2}e^{\beta(\mathcal{E}_d - \mu)} + 1}, \quad (28.31)$$

откуда <sup>1)</sup>

$$\boxed{n_d = \frac{N_d}{\frac{1}{2}e^{\beta(\mathcal{E}_d - \mu)} + 1}}. \quad (28.32)$$

**Акцепторный уровень.** В отличие от донорного уровня акцепторный, если его рассматривать как электронный уровень, может быть однократно или двукратно заполненным, но не может быть пустым. Это легко увидеть, исходя из дырочной картины. Акцепторную примесь [можно представить как фиксированный отрицательно заряженный притягивающий центр, наложенный на оставшийся неизменным атом основного вещества. Этот дополнительный заряд  $-e$  может образовать слабо связанное состояние с одной дыркой (отвечающей одному электрону на акцепторном уровне). Энергия связи дырки есть  $\mathcal{E}_a - \mathcal{E}_v$ ; когда происходит «ионизация», на акцепторный уровень приходит добавочный электрон. Однако конфигурация, в которой на акцепторном уровне отсутствуют электроны, отвечает двум локализованным дыркам вблизи акцепторной примеси и имеет очень большую энергию из-за взаимного кулоновского отталкивания двух дырок <sup>2)</sup>.

Имея это в виду, мы можем, используя (28.30), вычислить среднее число электронов на акцепторном уровне, отметив, что состояние без электронов теперь является запрещенным, а энергия двухэлектронного состояния превышает энергию двух одноэлектронных состояний на величину  $\mathcal{E}_a$ .

Итак,

$$\langle n \rangle = \frac{2e^{\beta\mu} + 2e^{-\beta(\mathcal{E}_a - 2\mu)}}{2e^{\beta\mu} + e^{-\beta(\mathcal{E}_a - 2\mu)}} = \frac{e^{\beta(\mu - \mathcal{E}_a)} + 1}{\frac{1}{2}e^{\beta(\mu - \mathcal{E}_a)} + 1}. \quad (28.33)$$

<sup>1)</sup> Некоторое представление о причинах появления необычного множителя  $1/2$ , отличающего выражение (28.32) от более привычной функции распределения Ферми — Дирака, можно составить, если исследовать, что происходит при уменьшении энергии двукратно занятого уровня от  $+\infty$  до  $2\mathcal{E}_d$ ; см. задачу 4.

<sup>2)</sup> Описывая акцепторный уровень как электронный, обычно пренебрегают электроном, который *должен находится* на этом уровне, и рассматривают только наличие или отсутствие второго электрона. Уровень называют заполненным или пустым в зависимости от того, имеется ли на нем второй электрон или нет.

Среднее число дырок на акцепторном уровне равно разности между максимальным числом электронов (два), которые могут находиться на этом уровне, и средним числом  $\langle n \rangle$  электронов на нем, т. е.  $\langle p \rangle = 2 - \langle n \rangle$ ; следовательно, величина  $p_a = N_a \langle p \rangle$  определяется выражением

$$p_a = \frac{N_a}{1/2 e^{\beta(\mu - \mathcal{E}_a)} + 1}. \quad (28.34)$$

## РАВНОВЕСНАЯ КОНЦЕНТРАЦИЯ НОСИТЕЛЕЙ В ПРИМЕСНОМ ПОЛУПРОВОДНИКЕ

Рассмотрим полупроводник, в котором на единицу объема приходится  $N_d$  донорных и  $N_a$  акцепторных примесей. Чтобы определить концентрации носителей, мы должны обобщить условие  $n_c = p_v$  [см. (28.18)], которое позволило нам найти эти концентрации в случае собственного (чистого) полупроводника. Мы можем сделать это, рассмотрев для начала электронную конфигурацию при  $T = 0$ . Пусть  $N_d \geq N_a$  [случай  $N_d < N_a$  рассматривается аналогично и ведет к тому же результату (28.35)]. Тогда в единице объема полупроводника  $N_a$  из  $N_d$  электронов, отданных донорными примесями, могут перейти на акцепторные уровни<sup>1)</sup>. Это приводит к электронной конфигурации основного состояния, в которой заполнены акцепторные уровни, уровни валентной зоны и  $N_d - N_a$  донорных уровней, а зона проводимости пуста. В случае термодинамического равновесия при температуре  $T$  электроны перераспределятся между уровнями таким образом, что число электронов в зоне проводимости и на донорных уровнях станет больше, чем при  $T = 0$ . Поскольку общее число электронов не меняется, увеличение их числа в зоне проводимости и на донорных уровнях в точности равно числу пустых уровней (т. е. дырок),  $p_v + p_a$  в валентной зоне и среди акцепторных уровней:

$$n_c + n_d = N_d - N_a + p_v + p_a. \quad (28.35)$$

Используя это уравнение в сочетании с полученными нами выше явными выражениями для  $n_c$ ,  $p_v$ ,  $n_d$  и  $n_a$  как функций от  $\mu$  и  $T$ , можно найти  $\mu$  как функцию от  $T$  и таким образом получить равновесные значения концентрации носителей при любой температуре. Общее рассмотрение весьма сложно, и мы здесь ограничимся поэтому только особенно простым и важным частным случаем.

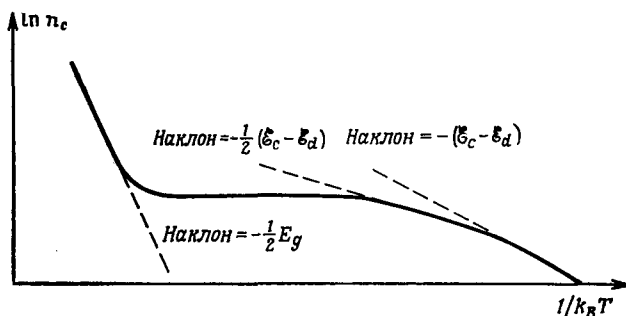
Пусть

$$\begin{aligned} \mathcal{E}_d - \mu &\gg k_B T, \\ \mu - \mathcal{E}_a &\gg k_B T. \end{aligned} \quad (28.36)$$

Поскольку  $\mathcal{E}_d$  и  $\mathcal{E}_a$  лежат вблизи краев запрещенной зоны, это ограничение ненамного сильнее условия невырожденности системы (28.10). Условие (28.36) и выражения (28.32), (28.34) для  $n_d$  и  $p_a$  обеспечивают полную «ионизацию» примесей за счет теплового возбуждения: только ничтожно малое число примесей сохраняет связанные с ними электроны или дырки:  $n_d \ll N_d$ ,  $p_a \ll N_a$ . При этом (28.35) принимает вид

$$\Delta n = n_c - p_v = N_d - N_a. \quad (28.37)$$

<sup>1)</sup> Поскольку  $\mathcal{E}_d$  лежит немного ниже минимума зоны проводимости  $\mathcal{E}_c$ , а  $\mathcal{E}_a$  немного выше максимума валентной зоны  $\mathcal{E}_v$ , мы имеем  $\mathcal{E}_d > \mathcal{E}_a$  (см. фиг. 28.12).



Фиг. 28.13. Температурная зависимость концентрации основных носителей тока (для случая  $N_d > N_a$ ).

Два высокотемпературных режима обсуждаются в тексте, поведение при очень низких температурах описано в задаче 6.

Теперь можно, исходя из формул (28.25) и (28.27), представить концентрации носителей и химический потенциал в виде явных функций только температуры:

$$\left\{ \begin{matrix} n_c \\ p_v \end{matrix} \right\} = \frac{1}{2} [(N_d - N_a)^2 + 4n_i^2]^{1/2} \pm \frac{1}{2} [N_d - N_a], \quad (28.38)$$

$$\frac{N_d - N_a}{n_i} = 2 \operatorname{sh} \beta (\mu - \mu_i). \quad (28.39)$$

Если ширина запрещенной зоны велика по сравнению с  $k_B T$ , то предположение (28.36), с которого мы начинали, остается справедливым, пока разность между  $\mu$  и  $\mu_i$  не станет достаточно большой по сравнению с  $k_B T$ . Согласно выражению (28.39), это произойдет, если величина  $|N_d - N_a|$  будет на несколько порядков больше собственной концентрации носителей  $n_i$ . Поэтому выражение (28.38) правильно описывает переход от полупроводника с почти собственной проводимостью ( $n_i \gg |N_d - N_a|$ ) к полупроводнику с преимущественно примесной проводимостью ( $n_i \ll |N_d - N_a|$ ). Разлагая выражение (28.38), мы находим, что при малых количествах примесей поправки к концентрациям носителей в собственном полупроводнике можно записать следующим образом:

$$\left\{ \begin{matrix} n_c \\ p_v \end{matrix} \right\} \approx n_i \pm \frac{1}{2} (N_d - N_a), \quad (28.40)$$

тогда как для широкого интервала концентраций носителей в несобственном полупроводнике имеем

$$\left. \begin{matrix} n_c \approx N_d - N_a \\ p_v \approx \frac{n_i^2}{N_d - N_a} \end{matrix} \right\} N_d > N_a; \quad (28.41)$$

$$\left. \begin{matrix} n_c \approx \frac{n_i^2}{N_a - N_d} \\ p_v \approx N_a - N_d \end{matrix} \right\} N_a > N_d.$$

Равенства (28.41) весьма важны для теории полупроводниковых приборов (см. гл. 29). Они показывают, что имеющийся за счет примесей избыток электронов (или дырок)  $N_d - N_a$  почти полностью передается в зону проводимости (или валентную зону); концентрация носителей в другой зоне имеет значительно меньшую величину, равную  $n_i^2 / |N_d - N_a|$  в соответствии с законом действующих масс (28.24).

При очень низкой температуре (или очень большой концентрации носителей) условие (28.36) в конце концов нарушается и какая-либо из величин  $n_d/N_d$  или  $p_a/N_a$  (но не обе вместе) уже не будет пренебрежимо малой. Это означает, что тепловое возбуждение теперь уже не может привести к полной ионизации примесей одного из типов. В результате концентрация основных носителей тока падает с понижением температуры (фиг. 28.13) <sup>1)</sup>.

## ПРОВОДИМОСТЬ ЗА СЧЕТ ПРИМЕСНОЙ ЗОНЫ

По мере приближения температуры к нулю то же самое происходит с долей ионизованных примесей и, следовательно, с концентрацией носителей в зоне проводимости или в валентной зоне. Тем не менее некоторая небольшая остаточная проводимость наблюдается даже при самых низких температурах. Это происходит потому, что волновая функция электрона, связанного с примесью, имеет значительную протяженность в пространстве, поэтому перекрытие волновых функций различных примесных центров возможно даже при довольно низких концентрациях. (То же самое относится и к дыркам.) Когда этим перекрытием пренебречь нельзя, электрон имеет возможность туннелировать с центра на центр. Происходящий в результате перенос заряда называется «проводимостью за счет примесной зоны».

Использование термина «зона» основывается на аналогии с методом сильной связи (гл. 10), согласно которому набор атомных уровней с одинаковой энергией при учете перекрытия волновых функций размывается в энергетическую зону. Однако примеси обычно расположены не в узлах решетки Браве, поэтому необходимо соблюдать осторожность, приписывая примесным «зонам» черты, присущие электронным зонам в *периодических* потенциалах <sup>2)</sup>.

## ТЕОРИЯ ЯВЛЕНИЙ ПЕРЕНОСА В НЕВЫРОЖДЕННЫХ ПОЛУПРОВОДНИКАХ

Из статистики Ферми — Дирака и условия невырожденности (28.10) непосредственно следует (см. задачу 7), что равновесное распределение скоростей электронов вблизи данного минимума зоны проводимости (или дырок вблизи данного максимума валентной зоны) имеет вид

$$f(\mathbf{v}) = n \frac{|\det \mathbf{M}|^{1/2}}{(2\pi k_B T)^{3/2}} \exp \left\{ -\frac{\beta}{2} \sum_{\mu\nu} v_\mu \mathbf{M}_{\mu\nu} v_\nu \right\}, \quad (28.42)$$

где  $n$  — вклад этих носителей в полную концентрацию.

Именно такую форму принимает при термодинамическом равновесии распределение скоростей молекул классического газа. Однако здесь имеются два отличия.

1. В классическом газе концентрация молекул  $n$  задана; в полупроводнике же величина  $n$  чрезвычайно чувствительна к изменениям температуры.

2. В классическом газе тензор массы  $\mathbf{M}$  диагонален.

В результате теория явлений переноса в невырожденном полупроводнике аналогична соответствующей теории для классического газа с несколькими типами

<sup>1)</sup> Более количественное рассмотрение такого поведения проводится в задаче 6.

<sup>2)</sup> Проблема поведения электронов в непериодическом потенциале (которая возникает не только в связи с примесными зонами, но также, например, в случае неупорядоченных сплавов) еще не вышла из детского возраста и является одним из наиболее оживленных направлений современных исследований по физике твердого тела.