

обозначать через E_v , а нижнюю границу (дно) зоны проводимости — через E_c .

Пример, когда заведомо реализуется случай а), дают нам кристаллы щелочных металлов, содержащих по одному валентному s -электрону на атом. Менее тривиальный пример представляют кристаллы щелочноземельных элементов, содержащих по два валентных s -электрона на атом. Металлические свойства здесь обусловлены перекрытием зон, возникающих в результате размытия s - и p -уровней электронов в атомах (реализуется случай б)).

Наконец, в кристаллах германия и кремния реализуется случай в). Валентные оболочки свободных атомов этих элементов обладают электронной конфигурацией s^2p^2 (четыре валентных электрона — по два в s - и p -состояниях).

Разности энергий s - и p -уровней сравнительно невелики; поэтому при образовании кристаллов происходит гибридизация соответствующих состояний (ср. § II.2 и § 5). В результате волновые функции валентной зоны (в приближении сильной связи) строятся из одной s - и трех p -функций. Такая зона содержит по четыре состояния на атом — ровно столько, сколько там есть валентных электронов. При $T=0$ она полностью заполнена. Следующая разрешенная зона отделена от валентной запрещенным участком. Так же обстоит дело и в алмазе.

Заметим, что наличие запрещенной зоны в этом и в других случаях обусловлено не геометрическими (как форма зоны Бриллюэна), а динамическими особенностями данного кристалла — конкретным видом и величиной потенциальной энергии электрона. В принципе можно, изменяя ее путем, например, изменения межатомных расстояний (под давлением), добиться перекрытия валентной зоны и зоны проводимости. При этом рассматриваемое вещество превратится из полупроводника в металл. Возможен и обратный случай — возникновение запрещенной зоны и переход металла в полупроводник при достаточно сильном изменении давления на образец.

Из сказанного выше ясно, что электроны заполненных зон не принимают непосредственного участия в переносе постоянного тока, а также в переносе энергий. Они образуют фон, на котором движутся электроны проводимости и дырки в валентной зоне. В дальнейшем мы будем называть электроны проводимости просто электронами (если это не поведет к путанице).

§ 8. Эффективная масса

Закон дисперсии дает довольно полную информацию о поведении электрона в данной решетке. Однако его трудно определять из опыта, ибо в общем случае речь идет здесь о нахождении явного вида функции трех практически непрерывно изменяющихся аргу-

ментов — компонент квазимульса. К счастью, в большинстве задач физики полупроводников играет роль лишь сравнительно малая часть зоны Бриллюэна, отвечающая энергиям у дна или у потолка той или иной энергетической зоны. Как мы сейчас увидим, для описания закона дисперсии вблизи экстремумов зоны нет необходимости определять всю функцию $E_l(\mathbf{p})$ — достаточно задать лишь несколько констант.

Пусть экстремум l -той зоны лежит в точке $\mathbf{p} = \mathbf{p}_0$ и зона в ней не вырождена. Тогда вблизи точки экстремума функцию $E_l(\mathbf{p})$ можно разложить в ряд Тэйлора. Принимая обычное соглашение о суммировании по дважды повторяющимся векторным индексам, мы имеем (опуская для краткости индекс l при \mathbf{p}_0)

$$E_l(\mathbf{p}) = E_l(\mathbf{p}_0) + \frac{1}{2} m_{\alpha\beta}^{-1} (\rho_\alpha - \rho_{\alpha,0}) (\rho_\beta - \rho_{\beta,0}) + \dots \quad (8.1)$$

Здесь

$$m_{\alpha\beta}^{-1} = \left. \frac{\partial^2 E_l(\mathbf{p})}{\partial \rho_\alpha \partial \rho_\beta} \right|_{\mathbf{p}=\mathbf{p}_0} \quad (8.2)$$

— величины размерности обратной массы, а многоточием обозначены члены более высокого порядка, которыми в окрестности экстремума можно пренебречь. Слагаемые, линейные по $\mathbf{p} - \mathbf{p}_0$, отсутствуют в правой части (8.1) по определению точки \mathbf{p}_0 .

Так как значение второй производной не зависит от порядка дифференцирования, величины $m_{\alpha\beta}^{-1}$ образуют симметричный тензор второго ранга. Значения $m_{\alpha\beta}^{-1}$ зависят от выбора системы координат в \mathbf{p} -пространстве. В частности, можно выбрать ее так [М5], чтобы компоненты $m_{\alpha\beta}^{-1}$ при $\alpha \neq \beta$ обратились в нуль. Соответствующие оси координат называются *главными*, а компоненты $m_{\alpha\beta}^{-1}$ в этой системе координат — *главными значениями* данного тензора. Для краткости мы будем писать при них лишь один значок α ; следует, однако, помнить, что это — компоненты тензора, а не вектора.

Равенство (8.1), будучи отнесено к главным осям, принимает вид

$$E_l(\mathbf{p}) = E_l(\mathbf{p}_0) + \frac{1}{2} \sum_{\alpha=x,y,z} m_\alpha^{-1} (\rho_\alpha - \rho_{\alpha,0})^2. \quad (8.1')$$

При этом $m_\alpha > 0$ или $m_\alpha < 0$, если рассматриваемый экстремум есть, соответственно, минимум или максимум.

Поскольку величина $E_l(\mathbf{p}_0)$ при заданном индексе l есть константа, уравнение (6.1) в этом случае принимает вид

$$\frac{(\rho_x - \rho_{x,0})^2}{2m_x} + \frac{(\rho_y - \rho_{y,0})^2}{2m_y} + \frac{(\rho_z - \rho_{z,0})^2}{2m_z} = \text{const.} \quad (8.3)$$

(Здесь и в дальнейшем мы опускаем для краткости зонный индекс l .) Таким образом, изоэнергетические поверхности вблизи невырож-

денного экстремума представляют собой эллипсоиды с полуосями, пропорциональными $\sqrt{|m_x|}$, $\sqrt{|m_y|}$ и $\sqrt{|m_z|}$ *), и с центром в точке $\mathbf{p} = \mathbf{p}_0$. Выбранные нами оси координат суть главные оси этого эллипсоида.

Константы m_x , m_y и m_z не всегда независимы: они могут быть связаны условиями симметрии изоэнергетической поверхности. Рассмотрим, например, кристалл кубической системы. Тогда, как указано в § 6, изоэнергетическая поверхность должна обладать симметрией куба. В частности, она должна переходить сама в себя при поворотах на 90° вокруг каждой из взаимно перпендикулярных осей, проходящих через центр куба и центры его граней; сами эти оси физически эквивалентны.

Здесь могут представиться два случая:

1) Экстремум функции $E(\mathbf{p})$ лежит в центре зоны Бриллюэна: $\mathbf{p}_0 = 0$. Таковы, например, изоэнергетические поверхности в зоне проводимости InSb, GaSb, GaAs и в некоторых других соединениях: дно этой зоны отвечает там точке $\mathbf{p} = 0$. Из соображений симметрии ясно, что главные оси тензора $m_{\alpha\beta}^{-1}$ в данном случае суть указанные выше оси куба. При этом $m_x = m_y = m_z \equiv m$ и изоэнергетические поверхности у дна зоны представляют собой сферы с центром в точке $\mathbf{p} = 0$:

$$E(\mathbf{p}) = E_c + \frac{\mathbf{p}^2}{2m}. \quad (8.4)$$

Закон дисперсии (8.4) называют параболическим изотропным.

Для частного случая одномерной цепочки в приближении сильной связи равенство (8.4) уже было получено в § 5 (формула (5.19)). Как и там, величину m в (8.4) называют эффективной массой. В дальнейшем (§ IV.1) мы увидим, что это название имеет не только формальный смысл: величина m определяет и среднее ускорение электрона во внешнем электрическом или магнитном поле.

Величины m_x , m_y , m_z , даже когда они неодинаковы, также называются эффективными массами, а совокупность величин $m_{\alpha\beta}^{-1}$ — тензором обратной эффективной массы.

2) Экстремум лежит в точке $\mathbf{p}_0 \neq 0$, при этом вектор \mathbf{p}_0 , соединяющий данный экстремум с центром зоны Бриллюэна, направлен вдоль одной из осей симметрии кристалла. Из свойств симметрии изоэнергетической поверхности следует, что в этом случае должно существовать несколько физически эквивалентных экстремумов, расположенных в соответствующих точках на эквивалентных осях симметрии. Соответственно надо говорить о нескольких векторах \mathbf{p}_0 и нескольких эллипсоидах вида (8.3); мы будем нумеровать их верхним латинским индексом i . Изоэнергетическая поверхность

*) В случае максимума, когда $m_\alpha < 0$, константа в правой части (8.3) отрицательна: это есть энергия, отсчитанная от своего максимума.

в данном случае представляет собой совокупность всех этих эллипсоидов. Таковы, например, изоэнергетические поверхности в зонах проводимости германия и кремния. В первом из указанных материалов дну зоны проводимости отвечают точки, лежащие на объемных диагоналях куба (на оси $[111]$ и ей эквивалентных в обратной решетке); во втором — точки, расположенные на оси $[100]$ и ей эквивалентных.

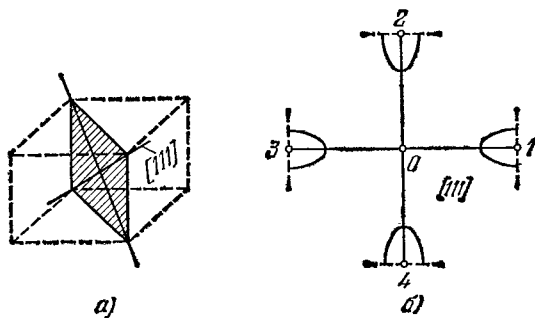


Рис. 3.6. а) Две (из четырех) пространственные диагонали куба $[111]$. б) Положение изоэнергетических поверхностей электронов в зоне Бриллюэна (Ge). Сечение плоскостью (110) . Пунктир — границы зоны Бриллюэна.

На рис. 3.6 и 3.7 схематически изображены сечения электронных изоэнергетических поверхностей в германии и кремнии плоскостями $p_x = 0$. Вблизи каждого минимума изоэнергетическая поверхность здесь имеет вид эллипсоида (8.3). Заметим, что это не противоречит кубической симметрии кристалла. Действительно, при поворотах на 90° вокруг трех взаимно перпендикулярных осей, проходящих через какую-нибудь из точек \mathbf{p}'_0 , вся изоэнергетическая поверхность отнюдь не обязана переходить сама в себя. Иначе говоря, симметрия вблизи каждой такой точки ниже кубической. Из рис. 3.6 и 3.7 видно, однако, что в рассматриваемых случаях физически эквивалентными оказываются любые две взаимно перпендикулярные оси, лежащие в плоскостях, перпендикулярных

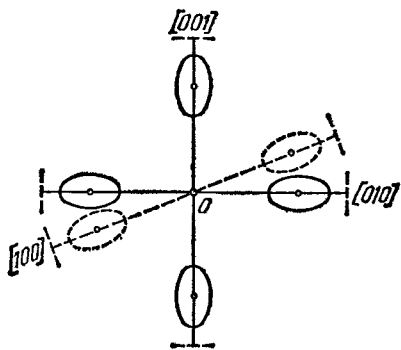


Рис. 3.7. Положение изоэнергетических поверхностей электронов в зоне Бриллюэна (Si). Пунктир — границы зоны.

векторам \mathbf{p}_0^i . Очевидно, эти оси вместе с \mathbf{p}_0^i и будут главными осями тензора $m_{\alpha\beta}^i$ для данного эллипсоида. Примем ось, вдоль которой направлен вектор \mathbf{p}_0^i , за ось z , а остальные две — за оси x, y . В соответствии со сказанным выше

$$m_x = m_y \equiv m_{\perp}, \quad m_z \equiv m_{\parallel} \neq m_{\perp}.$$

Таким образом, в рассматриваемом случае изоэнергетическая поверхность вблизи каждого экстремума имеет вид эллипсоида вращения:

$$E(\mathbf{p}) = E_c + \frac{(p_x - p_x^i, 0)^2}{2m_{\perp}} + \frac{(p_y - p_y^i, 0)^2}{2m_{\perp}} + \frac{(p_z - p_z^i, 0)^2}{2m_{\parallel}}. \quad (8.5)$$

Подчеркнем, что выражение (8.5) имеет смысл лишь вблизи каждого отдельного экстремума. Закон дисперсии вида (8.5) (или, вообще, (8.1') при не совпадающих друг с другом эффективных массах) называют параболическим анизотропным.

Области энергии вблизи каждого минимума иногда называют долинами, а полупроводники с несколькими эквивалентными минимумами энергии — многодолинными. В этих материалах имеется несколько разных групп носителей заряда. Все такие группы имеют одинаковые минимумы энергии, и в этом смысле они эквивалентны. В условиях термодинамического равновесия носители заряда распределяются поровну между разными долинами. Однако эффективные массы, а следовательно, и подвижности носителей заряда при данном направлении движения, вообще говоря, различны для разных долин. Наблюдаемая на опыте подвижность (и другие подобные величины) есть некоторая результирующая величина. В кристаллах кубической симметрии она оказывается изотропной вследствие кубической симметрии расположения эллипсоидов и равного числа электронов вблизи каждого из них.

Эквивалентность различных долин может быть нарушена при некоторых внешних воздействиях. Это имеет место, например, при одностороннем сжатии, когда энергия носителей заряда в одних долинах увеличивается, а в других — уменьшается. В результате изменяется распределение носителей между долинами, что приводит к изменению результирующей подвижности и проявляется в зависимости электрического сопротивления от давления (пьезосопротивление). Эквивалентность различных долин нарушается также при воздействии сильных внешних электрических и магнитных полей, что приводит к ряду интересных явлений. Некоторые из них будут рассмотрены ниже.

Экстремумы функции $E(\mathbf{p})$ могут располагаться как внутри зоны Бриллюэна, так и на ее границах. В германии, видимо, реализуется вторая возможность: дну зоны проводимости отвечают

точки L (рис. 3.1). При этом точки 1 и 3, 2 и 4 на рис. 3.6 отвечают одному и тому же состоянию, и соответствующие им минимумы надо рассматривать как один: в каждом из эллипсоидов, сечения которых изображены на рис. 3.6, сохранена только половина, лежащая в пределах первой зоны Бриллюэна (ср. ниже рис. 3.9). Таким образом, фактически на рис. 3.6 изображены два, а не четыре эквивалентных экстремума. На трехмерной картине, отвечающей кристаллу германия, имеются четыре эквивалентных экстремума (по числу диагоналей куба), а не восемь. С другой стороны, в кремнии, видимо, реализуется первая возможность: дну зоны проводимости отвечают точки Δ . Соответственно на рис. 3.7 изображены четыре эквивалентных экстремума. На трехмерной картине, отвечающей кристаллу кремния, их имеется шесть.

Итак, вид изоэнергетической поверхности вблизи невырожденной точки экстремума определяется не более чем тремя константами. В частных случаях, когда две или все три эффективные массы оказываются одинаковыми в силу свойств симметрии изоэнергетической поверхности, число констант соответственно уменьшается.

Если точка экстремума зоны одновременно является и точкой вырождения, то положение несколько усложняется. Изоэнергетические поверхности тогда уже не являются эллипсоидами; форма их зависит, в частности, от кратности вырождения и от свойств симметрии решетки. Нетрудно написать, однако, общее выражение для функции $E(\mathbf{p})$ вблизи экстремума в практически важном случае, когда вырождение двукратное и экстремум энергетической зоны лежит в центре зоны Бриллюэна. Действительно, искомое выражение должно:

- а) обладать двумя ветвями, которые смыкаются в точке $\mathbf{p} = 0$;
- б) иметь нулевые первые производные в точке $\mathbf{p} = 0$;
- в) содержать $|\mathbf{p}|$ в степени не выше второй.

В кристалле кубической системы единственную непрерывную функцию, удовлетворяющую поставленным условиям и обладающую должными свойствами симметрии, можно записать в виде

$$E(\mathbf{p}) = E(0) + \frac{1}{2m_0} \{Ap^2 \pm \sqrt{B^2p^4 + C^2(p_x^2p_y^2 + p_y^2p_z^2 + p_z^2p_x^2)}\}, \quad (8.6)$$

где A, B, C — безразмерные скаляры *); при этом A и B вещественны, а C может быть и мнимым (но $C^2 > -3B^2$).

Тензор обратной эффективной массы $m_{\alpha\beta}^{-1}$ можно ввести и в случае закона дисперсии (8.6). Однако теперь его компоненты уже не постоянны, а зависят от направления вектора \mathbf{p} . Двум ветвям функ-

*) Условие того, чтобы точка $\mathbf{p} = 0$ была экстремальной, а не седловой, сводится к неравенству $A^2 > B^2 + \frac{1}{3}C^2$ (если $C^2 > 0$) или $A^2 > B^2$ (если $C^2 < 0$).

ции (8.6) отвечают разные значения компонент $m_{\alpha\beta}^{-1}$. Соответственно говорят о зонах «легких» и «тяжелых» носителей заряда (обычно дырок — см. ниже § 9). Изоэнергетические поверхности, изображаемые формулой (8.6), называются *гофрированными сферами*. На рис. 3.8 изображено их сечение плоскостью (100).

Приводить тензор $m_{\alpha\beta}^{-1}$ к главным осям в данном случае не имеет смысла: поскольку компоненты его зависят от \mathbf{p} , такое приведение

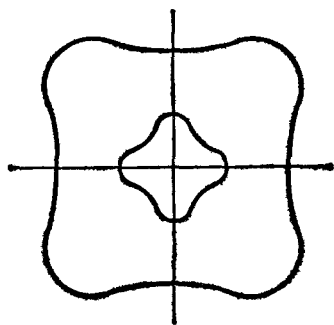


Рис. 3.8. Сечение гофрированной сферы (8.6) плоскостью (100). Внешний контур соответствует «тяжелым» носителям заряда, внутренний — «легким».

можно сделать, вообще говоря, лишь в нескольких точках зоны Бриллюэна, переходящих друг в друга при преобразованиях симметрии.

В ряде случаев отклонение поверхности (8.6) от сферической оказывается не очень большим. Тогда правую часть (8.6) можно приближенно заменить средним ее значением на сфере радиуса $|\mathbf{p}|$. При этом мы получаем две параболические зоны с изотропными законами дисперсии и разными эффективными массами m_1 и m_2 .

Вырождение зон встречается довольно часто. Как указывалось в § 5, его можно ожидать, если зоны получаются в результате размытия вырожденных энергетических уровней электронов в атомах. Согласно § 7 такая картина может иметь место, например, в валентных зонах германия и кремния.

По определению, выражения вида (8.1) и т. п. справедливы лишь в окрестности данной точки экстремума. При удалении от нее начинают играть роль и более высокие члены разложения функции $E(\mathbf{p})$ по степеням $p_\alpha - p_{\alpha,0}$. Об этом говорят как о непараболичности закона дисперсии.

Для оценки условий, при которых закон дисперсии начинает заметно отличаться от параболического, следует решить уравнение (4.1). Оказывается, что непараболичность особенно заметна в полупроводниках с достаточно узкой запрещенной зоной. Законы дисперсии для некоторых таких материалов выписаны в § 9.

§ 9. Зонная структура некоторых полупроводников

Структуру энергетических зон в полупроводнике можно было бы рассчитать с помощью уравнения (4.1). Фактически, однако, при этом возникают весьма серьезные вычислительные трудности, для преодоления которых необходимы самые мощные современные электронно-вычислительные машины. Практически чаще используется комбинированный подход: такие параметры, как ширина