

в ряде случаев оказалось возможным решить обратную задачу и на основании экспериментальных данных о ферми-поверхностях оценить значения OPW формфакторов.

Заметим, что форма поверхности, которую мы получили в двух-волновом OPW приближении, не зависит от знака w . Однако если учесть три или более OPW , то отдельные детали ферми-поверхности становятся чувствительными к знаку матричных элементов. Таким образом, анализируя экспериментальные данные, можно установить и знак псевдопотенциала. Однако в любом случае общий вид псевдопотенциала, изображенный на фиг. 33, хорошо известен, поэтому, зная величину двух или трех матричных элементов, обычно легко догадаться и об их знаках.

Когда мы перейдем к расчетам других свойств простых металлов, мы увидим, что, как правило, наиболее удобно начинать эти расчеты с матрицы гамильтониана и включать эффекты зонной структуры тогда, когда рассматривается само свойство. Знание истинной зонной структуры часто не очень нужно для расчета данного свойства. Обычно бывает трудно начать расчет и получить разумные результаты, когда стол завален огромной массой протабулированной информации. Исключениями, вероятно, являются такие специальные случаи, как расчет самой ферми-поверхности.

§ 6. ЗОННАЯ СТРУКТУРА ПОЛУПРОВОДНИКОВ И ПОЛУМЕТАЛЛОВ ¹⁾

В п. 6 § 4 мы отмечали, что для полупроводников характерно наличие полностью заполненных зон, называемых *валентными зонами*, и совершенно пустых зон, называемых *зонами проводимости*. Такое полное заполнение существует лишь при нулевой температуре, когда система находится в основном состоянии. При конечной же температуре некоторые из состояний зоны проводимости окажутся занятыми и соответственно некоторые из состояний валентной зоны — пустыми. Поэтому при конечной температуре некоторые электроны смогут принять участие в процессе проводимости — отсюда и название *полупроводник*. Мы увидим, каким образом примеси приводят к появлению электрической проводимости.

В рамках представленного здесь изложения мы можем рассматривать полуметаллы просто как полупроводники, край зоны проводимости которых лежит несколько ниже края валентной зоны. Поэтому некоторые из состояний зоны проводимости окажутся занятыми, а некоторые состояния валентной зоны — пустыми даже

¹⁾ Дополнительно с изложением этого вопроса на том же уровне можно ознакомиться в книге [5].

при абсолютном нуле температуры. В п. 3 настоящего параграфа мы вернемся специально к обсуждению полуметаллов.

Разница между полупроводниками и изоляторами чисто количественная и определяется величиной энергетической щели ¹⁾, отделяющей заполненную зону от незаполненной. В полупроводниках эта щель достаточно мала (порядка электронвольта) и они в обычных условиях становятся сравнительно хорошими проводниками. Таким свойством обладают германий и кремний и различные соединения элементов третьей и пятой групп, такие, как антимонид индия и арсенид галлия. В изоляторах щель достаточно велика (несколько электронвольт) и они не проводят ток. Сюда относятся многие ионные соединения, такие, как хлористый натрий.

Каким образом это количественное отличие приводит к качественной разнице, можно увидеть, если учесть, что вероятность заполнения состояния задается больцмановским множителем $A \exp(-E/KT)$, где энергия E порядка величины щели, а KT при комнатной температуре составляет 0,025 эВ. В полупроводниках больцмановский множитель может быть величиной порядка 10^{-10} , тогда как в изоляторах он может упасть до 10^{-30} или 10^{-40} . Поскольку в зоне проводимости имеется только 10^{23} состояний ²⁾, можно ожидать, что в зоне проводимости изолятора электронов не окажется. Когда мы будем рассматривать изоляторы, мы увидим, что кроме отсутствия носителей проводимость затрудняется еще одним обстоятельством, связанным с большой шириной запрещенной зоны. Дело в том, что, если носители и созданы какими-либо иными способами, они обладают малой подвижностью.

К сожалению, не существует сколько-нибудь простого приближенного метода вычисления запрещенной зоны полупроводников. Для получения достаточно полного и обоснованного описания зонной структуры необходимо проводить полный расчет. Если попытаться найти энергетические зоны путем внесения поправок к закону дисперсии свободных электронов, как это мы делали для простых металлов, то выяснится, что матричные элементы псевдопотенциала не просто деформируют ферми-поверхность, а полностью ее уничтожают. Мы не сможем найти электронные состояния полупроводника, используя одну или две ОРВ. Для этого всегда необходимо рассматривать их несколько, что опять отбрасывает нас, по сути дела, к зонным расчетам.

Все это не означает, что псевдопотенциалы неприменимы при рассмотрении полупроводников — фактически метод псевдопотенциала был развит Филлипсом и Клейнменом [36] именно в применении к полупроводникам. Оказалось, что псевдопотенциальные форм-

¹⁾ В отечественной литературе термин «энергетическая щель» (energy gap) при рассмотрении полупроводников часто заменяется термином «запрещенная зона». При переводе мы используем оба термина.— *Прим. ред.*

²⁾ На 1 см^3 .— *Прим. ред.*

факторы, возникающие при рассмотрении полупроводников, немногим больше формфакторов для простых металлов. Все трудности появляются из-за того, что необходимо учитывать много зон. В структуре алмаза, например, на одну примитивную ячейку приходится 2 атома, а на каждый атом — 4 электрона. Таким образом, на ячейку приходится 8 электронов — число, достаточное, чтобы заполнить 4 зоны. Конструирование ферми-поверхности свободных электронов привело бы к появлению большого количества очень мелких кусочков, которые, если мы для описания используем псевдопотенциал, исчезнут в плоскостях брэгговских отражений точно так же, как исчезли самые тонкие сегменты ферми-поверхности алюминия при включении конечного псевдопотенциала. Для больших волновых векторов, из которых образована ферми-поверхность 8 электронов, нет ни одного состояния, которое можно было бы адекватно описать без учета в разложении псевдоволновой функции нескольких плоских волн. Таким образом, мы возвращаемся к задаче о решении секулярного уравнения довольно высокого порядка. Тем не менее оказывается возможным в этой большой матрице гамильтониана выделить те наиболее важные матричные элементы, которые необходимы, чтобы понять самые существенные черты зонной структуры. Это важно для понимания оптических свойств полупроводников и природы их тетраэдрической (или ковалентной) связи. Мы вернемся к этому вопросу в п. 3 § 6 гл. IV, где будут обсуждаться атомные свойства.

К счастью, в теории полупроводников, как правило, не возникает необходимости рассматривать все эти сложные зоны. Так же как и в металлах, здесь электронные свойства будут определяться теми состояниями, которые лежат вблизи энергии Ферми. В металлах это состояния на ферми-поверхности, в полупроводниках это высшие состояния валентных зон и низшие состояния зоны проводимости.

Нас в действительности будут интересовать те состояния, которые лежат в пределах энергий порядка KT , отсчитанных от краев зон. Опять-таки, поскольку KT составляет всего лишь $1/40$ эВ, а ширина запрещенной зоны порядка 1 эВ, достаточно рассмотреть экстремальные состояния каждой зоны и вычислить малые поправки к энергии, появляющиеся при отклонении от этих состояний. Это делается с помощью $k \cdot p$ метода.

1. $k \cdot p$ метод и метод эффективной массы

Мы опять должны решить уравнение для собственных значений энергии:

$$\frac{p^2}{2m} \psi_k + V(r) \psi_k = E_k \psi_k,$$

где импульс

$$\frac{\hbar \nabla}{i} = \mathbf{p}.$$

Написав решение в блоховской форме

$$\psi_{\mathbf{k}} = u_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}},$$

получим

$$\frac{1}{2m} (\mathbf{p} + \hbar \mathbf{k})^2 u_{\mathbf{k}} + V(\mathbf{r}) u_{\mathbf{k}} = E_{\mathbf{k}} u_{\mathbf{k}}. \quad (2.32)$$

Рассмотрим состояния с волновым вектором, равным нулю. Это обычно (хотя, как мы увидим, не всегда) экстремальная точка $E_{\mathbf{k}}$. Описываемый метод можно непосредственно обобщить и для рассмотрения состояний, соседствующих с другими точками зоны Бриллюэна, однако случай окрестности точки $\mathbf{k} = 0$ наиболее прост.

В нулевом порядке по \mathbf{k} уравнение на собственные значения упрощается:

$$\frac{1}{2m} p^2 u_{\mathbf{k}} + V(\mathbf{r}) u_{\mathbf{k}} = E_{\mathbf{k}} u_{\mathbf{k}}.$$

Его решениями служат блоховские состояния при $\mathbf{k} = 0$, соответствующие различным зонам. Существует бесконечное число таких зон, и мы обозначим отвечающие им собственные функции через $u_0^{(1)}, u_0^{(2)}, \dots$, а соответствующие собственные энергии через $E_0^{(1)}, E_0^{(2)}, \dots$. Функции $\psi_{\mathbf{k}}^{(n)}$ при всех \mathbf{k} образуют полную систему для разложения любой функции, подчиняющейся периодическим граничным условиям в кристалле, а функции $\psi_{\mathbf{k}=0}^{(n)} = u_0^{(n)}$ — полную систему для разложения любой функции, имеющей трансляционную периодичность решетки и, следовательно, для разложения $u_{\mathbf{k}}^{(n)}$ при $\mathbf{k} \neq 0$. Мы можем, таким образом, использовать функции $u_0^{(n)}$ как базис для разложения $u_{\mathbf{k}}^{(n)}$ в ряд теории возмущений.

Рассмотрим теперь состояния с \mathbf{k} , не равным нулю. Слагаемое кинетической энергии в уравнении (2.32)

$$H_1 = \frac{2\hbar}{2m} \mathbf{k} \cdot \mathbf{p}$$

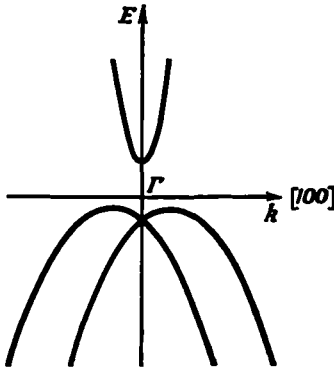
можно рассматривать как возмущение первого порядка по \mathbf{k} . Квадратичное же по \mathbf{k} возмущение есть просто

$$H_2 = \frac{\hbar^2 k^2}{2m}.$$

Теперь будем действовать в соответствии с обычной теорией возмущений. Поправка первого порядка имеет вид

$$\frac{\hbar}{m} \mathbf{k} \cdot \langle u_0^{(n)} | \mathbf{p} | u_0^{(n)} \rangle.$$

Если кристалл обладает центром инверсии, функции $u_0^{(n)}$ классифицируются по четности относительно этого центра. Отсюда следует, что написанный выше матричный элемент, а значит, и поправка первого порядка для такого кристалла обращаются в нуль. С другой стороны, в антимониде индия, например, центра инверсии нет и мы



Ф и г. 51. Схема энергетических зон в антимониде индия.

Показано расщепление максимума валентной зоны и минимум зоны проводимости, соответствующий малой массе. (Полное описание зонной структуры InSb можно найти в работе Кейяна [37].)

получаем поправку первого порядка. Из этого вытекает, что в подобных случаях экстремум энергии может сдвигаться из точки $k = 0$. Подобный сдвиг максимума энергии валентной зоны антимонида индия показан на фиг. 51.

При любой зонной структуре мы можем найти экстремум энергии и производить разложение по отклонениям от соответствующего волнового вектора. Тогда поправка первого порядка обращается в нуль по определению. Здесь же мы ограничимся случаем экстремума в точке $k = 0$, т. е. случаем, которому отвечает отсутствие поправки первого порядка.

Во втором порядке теории возмущения для энергии получаем следующее выражение:

$$E_k^{(n)} = E_0^{(n)} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2} \sum_m \frac{|k \cdot \langle u_0^{(m)} | \mathbf{p} | u_0^{(n)} \rangle|^2}{E_0^{(n)} - E_0^{(m)}}. \quad (2.33)$$

Можно было бы, конечно, найти энергию в любом нужном нам порядке, но этого выражения уже вполне достаточно для того, чтобы определить кривизну вблизи экстремума зоны.

Видно, что для нахождения энергии необходимо провести суммирование по всем зонам при $k = 0$. Однако рост энергетических знаменателей уменьшает роль высоколежащих зон, и часто оказывается, что существенно всего лишь несколько наиболее близких зон. К тому же часто можно значительно уменьшить число слагаемых, приняв во внимание симметрию состояний, отвечающих отдельным зонам. Учитывая, по какому из неприводимых представлений преобразуется каждое из состояний, равно как и опера-

тор импульса, можно получить *произведение представлений*, откуда мы сможем определить, какие из матричных элементов обращаются в нуль вследствие симметрии. Этот метод представляет собой одно из приложений теории групп, не обсуждавшееся в гл. I, но описание которого можно найти в любом обычном учебнике теории групп. В качестве конкретного примера укажем на то, что из-за нечетности оператора импульса он не может связывать два четных состояния ¹⁾.

Видно, что наибольший вклад в энергию дают те связанные друг с другом зоны, которые не слишком сильно отличаются друг от друга по энергиям. Можно ввести эффективную массу, представив энергию (2.33) в виде

$$E_k = E_0 + \frac{\hbar^2}{2m} \mathbf{k} \cdot \left(\frac{m}{m^*} \right) \cdot \mathbf{k}, \quad (2.34)$$

причем элементы тензора эффективной массы задаются выражением

$$\frac{m}{m^*} \Big|_{ij}^{(n)} = \delta_{ij} + \frac{2}{m} \sum_l \frac{\langle u_0^{(n)} | p_i | u_0^{(l)} \rangle \langle u_0^{(l)} | p_j | u_0^{(n)} \rangle}{E_0^{(n)} - E_0^{(l)}}. \quad (2.35)$$

Это выражение известно под названием *правила f-сумм*. Оно важно при рассмотрении оптических свойств, так как мы убедимся в том, что силы осциллятора, отвечающие межзонным переходам, определяются с теми же матричными элементами p .

Отметим, что если две зоны связываются оператором импульса, то матричный элемент $\langle u_0^{(n)} | p_i | u_0^{(l)} \rangle$ есть величина порядка \hbar/a , где a — межатомное расстояние. Следовательно,

$$\frac{m}{m^*} \approx 1 + \frac{2\hbar^2}{ma^2\Delta E},$$

причем \hbar^2/ma^2 порядка 10 эВ. Если $\Delta E \approx 0,2$ эВ, то мы получаем эффективную массу порядка 0,01 массы свободного электрона. Таким образом, не следует удивляться тому, что в полупроводниках электроны во многих случаях ведут себя подобно свободным частицам с массой, много меньшей истинной массы электрона.

Этот результат применим, конечно, и к простым металлам. Можно рассматривать, например, минимум зоны проводимости лития при $k = 0$. Единственная зона, лежащая ниже, — это 1s-зона внутренней оболочки, которая не связана с первым матричным элементом от p . Таким образом, все зоны, с которыми связывается зона проводимости, лежат выше ее минимума. Отсюда вытекает, что $1/m^*$ меньше $1/m$, т. е. m^* больше m . При переходе к более тяжелым щелочным металлам все с большим и большим числом зон, лежащих

¹⁾ Равно, впрочем, как и два нечетных. Точнее, оператор импульса не может связывать два состояния одинаковой четности. — *Прим. ред.*

ниже зоны проводимости, последние начинают доминировать. В таких металлах поэтому эффективные массы меньше истинной массы электрона.

Для любой зонной структуры можно провести вычисление эффективных масс, если мы только знаем или приближенно вычислим волновые функции других зон. Часто можно даже просто оценить величину эффективной массы, зная лишь порядок величины матричных элементов и воспользовавшись экспериментально измеренными значениями запрещенной зоны.

Мы получили метод определения состояний вблизи края зоны, очень похожий на метод псевдопотенциала в теории простых металлов. Метод эффективной массы намного старше метода псевдопотенциала. Ограничиваясь рассмотрением состояний вблизи $k = 0$, мы нашли зависимость энергии от k в виде (2.34). Это выражение можно использовать для построения эффективного гамильтониана $H(\hbar k)$, приводящего к уравнению, подобному уравнению Шредингера:

$$-\frac{\hbar^2}{2m} \sum_{ij} \frac{\partial}{\partial x_i} \left(\frac{m}{m^*} \right)_{ij} \frac{\partial}{\partial x_j} \psi + E_0 \psi = i\hbar \frac{\partial \psi}{\partial t}.$$

Так же как и уравнение с псевдопотенциалом, это уравнение, будучи точно решенным, дает правильные (при малых k) собственные значения энергии. Подобно уравнению с псевдопотенциалом, оно приводит к довольно плавной псевдоволновой функции ψ . Когда решение имеет вид плоской волны, истинная волновая функция получается путем умножения ψ на функцию Блоха $u_k^{(n)}$ (или на ее разложение, включающее первый порядок по k), в то время как в методе псевдопотенциала проводится ее ортогонализация к волновым функциям сердцевины. Если ψ нормирована на объем кристалла, $u_k^{(n)}$ должна быть нормирована на объем элементарной ячейки Ω_0 :

$$\int u_k^{(n)*} u_k^{(n)} d\tau = \Omega_0.$$

Имеется еще одно обобщение метода, которое также позволяет получить решения, отвечающие линейным комбинациям вырожденных состояний зоны. Тогда истинная волновая функция получается из ψ заменой k на $-iV$ в выражении для $u_k^{(n)}$.

К сожалению, это приближение не приводит непосредственно к псевдогамильтониану в присутствии возмущения, так как его обоснование базируется на условии трансляционной периодичности. Его можно обобщить лишь на тот случай, когда возмущение кристалла очень медленно меняется в пространстве.

В § 2 мы уже отмечали, что динамическое поведение электрона в идеальном кристалле под воздействием приложенных сил подобно поведению частицы с гамильтонианом $H(\hbar k)$. Это было продемон-

стрировано с помощью конструирования волновых пакетов и, стало быть, справедливо, если приложенная сила медленно меняется в пространстве. В частности, любой приложенный потенциал должен слабо изменяться на межатомном расстоянии. Для такого медленно меняющегося потенциала $V(\mathbf{r})$ в кристалле динамика электрона описывается уравнением

$$-\frac{\hbar^2}{2m} \sum_{i,j} \frac{\partial}{\partial x_i} \left(\frac{m}{m^*} \right)_{ij} \frac{\partial}{\partial x_j} \varphi + (E_0 + V(\mathbf{r})) \varphi = i\hbar \frac{\partial \varphi}{\partial t}, \quad (2.36)$$

которое интуитивно кажется вполне обоснованным. Медленно меняющийся потенциал просто-напросто поднимает или опускает зоны в разных частях кристалла. Мы воспользуемся этим уравнением [заменяв $i\hbar (\partial/\partial t)$ на E] в п. 2 § 8 при рассмотрении примесных состояний в полупроводниках. Замена истинного уравнения Шредингера уравнением (2.36) называется *методом эффективной массы*. Мы видим, что он совершенно аналогичен теории псевдопотенциала.

Интересно рассмотреть влияние медленно меняющихся деформаций кристалла, например длинноволновых колебаний решетки. Обобщение на этот случай производится не столь непосредственно. Можно выделить три эффекта, которые должны быть учтены.

1. Смещение с деформацией края зоны. Если экстремум приходится на точку $\mathbf{k} = 0$, то линейный по деформации сдвиг может вызвать только дилатация. (Из соображений симметрии коэффициент, связывающий смещение края зоны со сдвиговой деформацией, должен обратиться в нуль.) Если же экстремум расположен при $\mathbf{k} \neq 0$, то линейное по деформации смещение может возникать и от сдвиговых деформаций. Такое смещение, очевидно, подобно приложенному потенциалу и его можно учесть, полагая энергию E_0 из уравнения (2.36) функцией координат.

2. Деформация может изменять тензор эффективной массы. Это обстоятельство учитывается тем, что отношение m/m^* в уравнении (2.36) следует считать зависящим от координат. Отметим, что этот множитель в уравнении (2.36) расположен таким образом, чтобы гамильтониан остался эрмитовым, когда m/m^* становится функцией координат. [Эрмитовость можно проверить, умножив уравнение (2.36) слева на φ'^* и проинтегрировав. Затем следует провести два интегрирования по частям, используя при этом периодические граничные условия.]

3. Представим себе, что необходимо сшить волновые функции, вычисленные в двух областях со слегка отличающимися деформациями. Сшивать следует не псевдоволновые, а истинные волновые функции. Блоховская функция $u_{\mathbf{k}}(\mathbf{r})$ между атомами (или в частном случае — на границах ячейки) может зависеть от деформации и, следовательно, будет испытывать разрыв между двумя этими областями. Чтобы обеспечить непрерывность волновой функции, необходимо ввести разрыв в φ . Обозначим через β некоторое среднее

значение $u_k(\mathbf{r})$ на границах ячейки; этот параметр в случае переменной деформации окажется функцией координат. В уравнении (2.36) это обстоятельство можно учесть таким образом, чтобы величина $\beta\varphi$ была непрерывной, что приводит к эрмитовому псевдогамильтониану

$$-\beta^* \frac{\hbar^2}{2m} \sum_{i,j} \frac{\partial}{\partial x_i} \left(\frac{m}{\beta^* m^*} \right)_{ij} \frac{\partial}{\partial x_j} \beta\varphi + \beta (E_0(\mathbf{r}) + V(\mathbf{r})) \varphi = i\hbar \frac{\partial}{\partial t} \beta\varphi. \quad (2.37)$$

Эрмитовость необходима для выполнения уравнения непрерывности, гарантирующего сохранение заряда, и для обеспечения действительных собственных значений энергии. Такую довольно сложную форму уравнения можно приближенно получить с помощью метода ячеек [38], и ее можно с полным основанием распространить и на случай деформаций, быстро меняющихся в пространстве. Она применима даже при рассмотрении электронных состояний твердых растворов.

Если пренебречь теперь изменениями m/m^* и β с координатой и положить массу m^* изотропной, мы можем прокоммутировать $\partial/\partial x_i$ с β и m/m^* и получить

$$-\frac{\hbar^2}{2m^*} \nabla^2 \varphi + [E_0(\mathbf{r}) + V(\mathbf{r})] \varphi = i\hbar \frac{\partial \varphi}{\partial t}.$$

Эта простая форма уравнения будет использована для описания взаимодействия электронов с продольными колебаниями решетки в полупроводниках.

Из уравнения (2.37) видно, что описание полупроводников с помощью псевдодолновой функции намного сложнее, чем соответствующее описание простых металлов. Усложнения проистекают, во-первых, из-за необходимости рассматривать зоны с величиной $1/m^*$, сильно отличающейся от значения для свободных электронов, — использование только первого слагаемого в выражении (2.35) было бы совершенно бессмысленным для большинства полупроводников. Вторая трудность, тесно связанная с первой, состоит в том, что блоховскую функцию $u_k(\mathbf{r})$ нельзя представить константой в пространстве вне внутренних оболочек [к чему приводит оператор $(1 - P)$ в теории псевдопотенциала]. В кристаллах со структурой алмаза она близка к нулю в пространстве между атомами. В этом причина усложнения уравнения (2.37), связанного с появлением множителя β , и трудностей при переходе от псевдодолновой к истинной волновой функции. Наконец, это не позволяет дать ясную постановку задачи при неупорядоченном расположении атомов. Описание эффектов, связанных с изменениями $u_k(\mathbf{r})$, с помощью параметра β представляет собой, конечно, грубую аппроксимацию.

Тем не менее метод эффективной массы, сформулированный в форме уравнения (2.37), есть наилучший способ исследования многих

свойств полупроводников и именно на его основе мы будем строить наше изучение полупроводников.

В следующем разделе мы обсудим динамику электрона в полупроводниках. Для решения этой задачи мы рассмотрим идеальный кристалл и будем конструировать волновые пакеты. Это вернет нас к квазиклассической теории и не потребует уравнения (2.37) во всей его общности.

2. Динамика электронов и дырок в полупроводниках

Как мы уже отмечали, в чистом полупроводнике имеется электронов ровно столько, сколько нужно, чтобы заполнить валентную зону, а для того, чтобы заполнить состояние зоны проводимости, их уже не остается. Однако при конечной температуре (или, как мы увидим позже, в присутствии примесей) некоторые из состояний в зоне проводимости оказываются занятыми, а в валентной зоне — пустыми. Возбуждения первого типа называют *электронами*, а второго — *дырками*. Переход одного электрона через запрещенную зону приводит к появлению одного электрона и одной дырки. Сейчас мы займемся изучением динамики этих возбуждений.

Рассмотрим сначала электрон в зоне проводимости. Нас будут интересовать низкоэнергетические возбуждения, находящиеся вблизи минимума зоны. Здесь мы можем воспользоваться $k \cdot p$ методом и разложить энергию в ряд по k . Обозначим волновой вектор, отсчитанный от точки, отвечающей минимуму, через κ . В общем случае результат можно представить в виде квадратичной по волновому вектору форме, задаваемой соотношениями (2.34) и (2.35). Однако специальным выбором трех ортогональных осей эту форму можно диагонализировать:

$$E_{\kappa} = E_0 + \frac{\hbar^2}{2m} \sum_i \left(\frac{m}{m^*} \right)_{ii} \kappa_i^2.$$

В большинстве полупроводников по крайней мере две из трех эффективных масс совпадают и энергию можно обычно записать в виде

$$E_{\kappa} = E_0 + \frac{\hbar^2}{2m_t} \kappa_t^2 + \frac{\hbar^2}{2m_l} \kappa_l^2,$$

где m_t и m_l называют *поперечной* и *продольной эффективными массами*, κ_l — компонента κ вдоль оси симметрии, а κ_t означает поперечную компоненту.

Для любого закона дисперсии справедливы уравнения

$$\hbar \dot{\mathbf{k}} = \hbar \dot{\boldsymbol{\kappa}} = \mathbf{F} = -e \left(\mathcal{E} + \frac{\mathbf{v} \times \mathbf{H}}{c} \right),$$

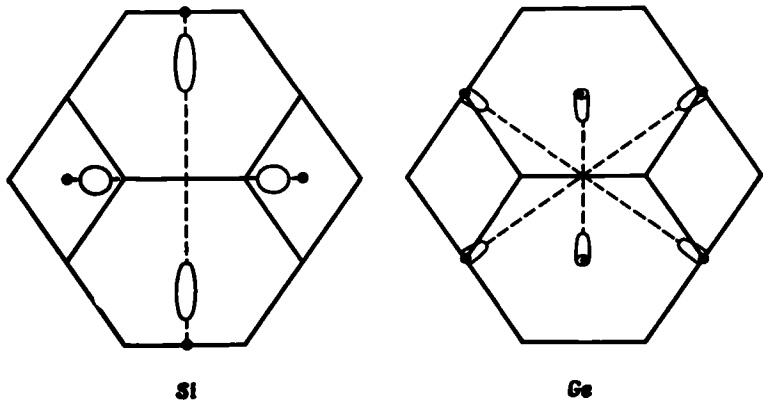
$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E = \frac{1}{\hbar} \nabla_{\boldsymbol{\kappa}} E.$$

Таким образом, мы можем определить импульс электрона $\mathbf{p} = \hbar \mathbf{k}$, подчиняющийся уравнению Ньютона $\dot{\mathbf{p}} = \mathbf{F}$, а связь скорости с импульсом задается выражением

$$\mathbf{v} = \left(\frac{1}{m^*} \right) \mathbf{p},$$

где мы ввели тензор эффективной массы (2.35). Динамическое поведение электронов такое же, как и свободной частицы с анизотропной массой. Отметим, что в общем случае направление скорости не совпадает с направлением импульса.

В зоне проводимости кремния имеется 6 минимумов с совпадающей энергией, расположенных на 6 различных осях [100], как это



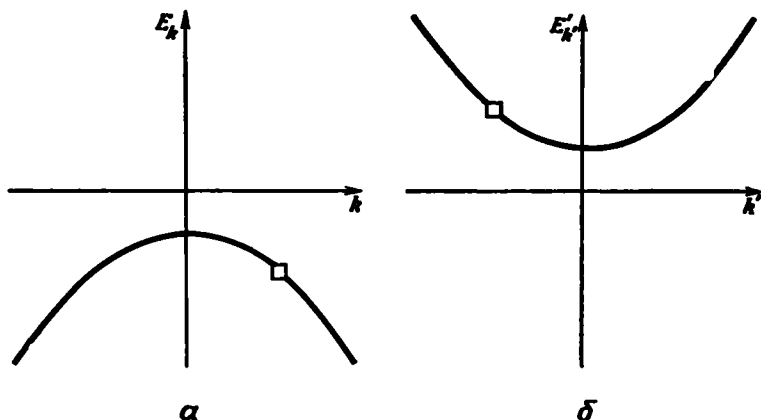
Ф и г. 52. Расположение минимумов зоны проводимости кремния и германия в зоне Бриллюэна.

Эллипсоиды представляют собой изоэнергетические поверхности в зоне проводимости. Зона Бриллюэна показана в направлении [110].

показано на фиг. 52. Эффективные массы следующие: $m_t = 0,19 m$ и $m_l = 0,98 m$. Для многих целей бывает достаточно заменить эти шесть групп носителей с анизотропными эффективными массами одной группой с должным образом выбранной средней изотропной эффективной массой. В зоне проводимости германия минимумы расположены в точках пересечения осей [111] с гранями зоны, т. е. в точке L . Это также показано на фиг. 52. Состояния на любых двух противоположных гранях зоны Бриллюэна можно объединить так, чтобы образовать одну долину. Таким путем мы получаем четыре долины с тензором эффективной массы, задаваемым массами $m_t = 0,082 m$ и $m_l = 1,64 m$. Как в кремнии, так и в германии электроны дают больший вклад в ток, чем дырки, и поэтому играют главную роль в тех случаях, когда число носителей обоих типов совпадает.

Понять динамическое поведение дырок в полупроводниках сложнее. Хорошо известно, и это ясно интуитивно, что незанятые состояния валентной зоны ведут себя подобно положительно заряженным частицам. Однако не вполне очевидно, что они должны вести себя подобно положительным частицам с положительной массой. Поэтому мы рассмотрим такие системы более внимательно.

Валентная зона как германия, так и кремния имеет максимум в точке $k = 0$ и вырождена. В приводимом ниже анализе мы пренебрегаем этим вырождением и рассматриваем поведение дырок в каждой зоне отдельно. Удалим один электрон из такого состояния



Ф и г. 53. *а* — верх валентной зоны с одним незанятым состоянием, обозначенным квадратиком; *б* — та же зона и то же состояние, представленные с помощью измененных энергии и волнового вектора.

Последнее представление удобно для описания динамики дырок.

валентной зоны, волновой вектор которого k лежит вблизи максимума (фиг. 53, *а*). Рассмотрим теперь поведение всей системы, обозначив опять через e абсолютную величину заряда электрона.

Ток системы, равный нулю до удаления электрона, после его удаления оказывается равным

$$j = 0 - (-e) \frac{1}{\hbar} \nabla_k E_k,$$

так как ток, отвечающий рассматриваемому состоянию, если бы оно было занято, равнялся бы $(-e) (1/\hbar) \nabla_k E_k$.

Если поместить эту систему во внешние поля, то все электроны двигались бы в пространстве волновых векторов в соответствии с приведенным ранее уравнением:

$$\hbar \dot{k} = (-e) \left(\mathcal{E} + \frac{v}{c} \times H \right).$$

Незанятое состояние описывается, конечно, тем же уравнением движения. Аналогичным образом скорость пакета, сконструированного из таких состояний, задается формулой

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E_{\mathbf{k}},$$

и пакет, связанный с незанятым состоянием, движется с такой же скоростью.

Отметим, наконец, что можно определить энергию удаления этого электрона как энергию, необходимую для переноса электрона из состояния \mathbf{k} в какой-то резервуар с нулевой энергией. (Мы могли бы подобным же образом определить энергию рождения электронного состояния как энергию, необходимую для переноса электрона из того же резервуара с нулевой энергией в соответствующее состояние зоны проводимости.) Полное изменение энергии, требующееся для рождения дырки, есть

$$\Delta E_{\text{tot}} = -E_{\mathbf{k}}.$$

Эта формула, полученная в соответствии с поведением дырки, совпадает с аналогичной формулой, основывающейся на поведении электрона, с точностью до изменения знака тока и энергии, необходимой для рождения возбуждения.

Используя эти формулы, можно было бы получить правильный закон движения дырок во внешних полях. Однако удобнее переписать их в таком виде, который делает поведение дырок интуитивно более понятным. Определим для этого волновой вектор $\mathbf{k}' = -\mathbf{k}$ и энергию возбуждения $E_{\mathbf{k}'} = -E_{\mathbf{k}}$. С помощью этих параметров, которые точно задают состояние системы, можно переписать импульс \mathbf{p}' , поток и закон изменения \mathbf{p}' :

$$\mathbf{p}' = \hbar \mathbf{k}'.$$

Тогда

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E = \frac{1}{\hbar} \nabla_{\mathbf{k}'} E',$$

$$\mathbf{j} = e\mathbf{v},$$

$$\dot{\mathbf{p}}' = -\hbar \dot{\mathbf{k}}' = e \left(\mathcal{E} + \frac{\mathbf{v}}{c} \times \mathbf{H} \right).$$

Таким образом, $E_{\mathbf{k}'}$ есть энергия распространяющегося и переносимого ток возбуждения, динамика которого совпадает с динамикой положительно заряженной частицы с волновым вектором \mathbf{k}' и импульсом \mathbf{p}' . В соответствии с этим мы можем просто изменить систему координат и описывать такие состояния так, как это показано на фиг. 53, б.

Заметим, что магнитное поле отклоняет дырочный волновой пакет вправо, как должно быть для положительно заряженной частицы. Этого не получилось бы, если бы мы сконструировали

пакет из незанятых состояний в зоне с положительной кривизной, такой, как зона проводимости. Можно также отметить, что дырочный пакет будет создавать в системе поле, аналогичное полю положительно заряженной частицы, хотя в конечном счете это поле создается зарядом ионов, компенсирующих заряд электронного газа. Заметим, наконец, что для изотропной зоны энергию E'_k вблизи края зоны можно разложить в ряд по волновым векторам:

$$E'_k = \frac{\hbar^2}{2m^*} k'^2,$$

где эффективная масса m^* положительна.

Мы показали, что возбуждения в полупроводниках можно разделить на два различных сорта — электроны и дырки — и что поведение каждого из них можно представлять себе как поведение электронов и позитронов. Следует заметить, что когда электрон из зоны проводимости падает обратно в валентную зону, то такому процессу соответствует аннигиляция электрон-дырочной пары с выделением энергии, равной сумме введенных выше энергий возбуждений. При рассмотрении явлений переноса в полупроводниках мы во всей полноте будем использовать это простое и наглядное описание системы.

Несмотря на сложную зонную структуру полупроводников, для многих целей бывает достаточно ограничиться рассмотрением окрестности края зон, где справедлива описанная выше простая и ясная картина. Что же касается свойств, для которых необходимо знать зонную структуру вдали от экстремумов, то для их изучения требуются результаты зонных расчетов или приближенное описание, которое дается в п. 3 § 6 гл. IV.

3. Полуметаллы

Как мы уже указывали в начале § 6, многое из того, что говорилось о зонах полупроводников, относится и к полуметаллам. Здесь также нельзя использовать приближение почти свободных электронов, так как почти вся ферми-поверхность, построенная для свободных электронов, поглощается гранями зоны Бриллюэна и для описания одного состояния необходимо несколько ортогонализированных плоских волн.

Многие электронные свойства, однако, полностью определяются состояниями, находящимися вблизи ферми-поверхности, а в полуметаллах уровень Ферми расположен очень близко к краю валентной зоны и зоны проводимости (для висмута, например, разность между ними составляет 0,02 эВ). Поэтому можно воспользоваться разложением около экстремальных точек и применять метод эффективной массы, хотя отклонения от параболичности здесь несколько большие, чем для многих полупроводников.

При низких температурах все состояния зоны проводимости, лежащие ниже уровня Ферми, заняты и существует четко определенная ферми-поверхность. В приближении эффективной массы эта поверхность представляет собой эллипсоид, и, как показано, такое приближение хорошо описывает форму наблюдаемой поверхности. Подобным же образом эллипсоидальную форму имеет и дырочная поверхность, причем полный объем внутри ее равен объему, заключенному внутри электронной поверхности. Полуметалл ведет себя подобно металлу с ферми-поверхностью в виде тонкой щепочки, и для ее изучения применимы те экспериментальные методы, которые описаны в п. 6 § 5. Большое количество экспериментальных результатов было получено, в частности, для висмута [24].

§ 7. ЗОННАЯ СТРУКТУРА ИЗОЛЯТОРОВ

Как было отмечено в предыдущем параграфе, изоляторы отличаются от полупроводников большими запрещенными зонами. Можно ожидать, что число термически возбужденных электронов или дырок здесь пренебрежимо мало, хотя их и можно возбудить путем освещения изолятора светом.

Прототипом изолятора служит ионный кристалл, такой, как хлористый натрий. Можно ожидать, что он окажется изолятором, так как имеет заполненные зоны, и хотя электроны ускоряются приложенным полем, как это было показано выше, все состояния остаются занятыми, так что ток в системе не идет. Используя терминологию теории простых металлов, мы можем сказать, что электроны ускоряются полем, но в результате дифракции полное ускорение обращается в нуль.

Однако сейчас стало ясно, что такое описание изоляторов приводит к неправильным результатам. Зонные расчеты для ионных кристаллов, молекулярных кристаллов и кристаллов инертных газов (которые в общем рассматриваются как молекулярные) приводят к очень узким валентным зонам, заполненным в основном состоянии. Если удалить электрон из такой узкой зоны, потребуется очень малая энергия для «реорганизации» оставшихся состояний таким образом, чтобы уменьшить электростатическую энергию дырки. Поэтому для валентных электронов зонная картина может быть несправедливой и более подходящим окажется описание, основывающееся на представлении о состояниях, локализованных на одном ионе.

Анализ свойств [39], при описании которых интуитивные теории приводят к различному поведению в зависимости от того, предполагаются ли состояния локализованными или зонными, позволяет довольно резко разделить материалы на два типа. Соединения можно расположить по степени «ионности», т. е. разнице зарядов