

ЛЕКЦИИ

Дисциплина

«Методы и средства контроля параметров материалов электроники и наноэлектроники» (8 ч)

Названия ЛЕКЦИЙ (указанные в программе для очников !!!)

1. Механика микроскопических систем: принципы, законы и математический аппарат.
2. Методы описания локализованных и делокализованных состояний в твердых телах.
3. Методы изучения электротранспортных свойств материалов.
4. Эффекты в материалах со свободными носителями заряда во внешних полях и их использование для контроля свойств этих материалов.
5. Электронный парамагнитный резонанс как метод изучения структуры материалов.
6. Оптическая спектроскопия как метод изучения зонной структуры материалов.
7. Кристаллическая структура материалов по данным рентгеноструктурного анализа.
8. Метод термо-ЭДС: физические принципы и применения для исследования проводящих материалов.

Лекция 1

1. Механика микроскопических систем: принципы, законы и математический аппарат.

Наиболее важные исторические факты и математические объекты:

1923 г. де Бройль предложил равенство:

$$\lambda = \frac{h}{p}, \quad (p = \frac{h}{\lambda} = hk) \quad (1)$$

(отсюда возник принцип корпускулярно-волнового дуализма)

1927 г. Принцип неопределенности Гейзенберга

$$\Delta x \cdot \Delta p_x \geq \frac{h}{2\pi}, \quad \Delta y \cdot \Delta p_y \geq \frac{h}{2\pi}, \quad \Delta z \cdot \Delta p_z \geq \frac{h}{2\pi} \quad (2)$$

(отсюда следовал вывод – не все физические величины измеримы одновременно)

Согласно квантовой механике любая физическая величина f в стационарных условиях характеризуется спектром *собственных значений*, f_n . Каждому f_n соответствует свое состояние, описываемое некоторой определенной функцией координат и времени, $\Psi_n(q)$. Функцию $\Psi_n(q)$ называют **волновой функцией** квантовой системы, находящейся в данном n -ом состоянии.

$$\text{Условие нормировки: } \int_{-\infty}^{\infty} \Psi^*(q) \Psi(q) dq = 1. \quad (3)$$

Наиболее важные принципы:

Принцип суперпозиции утверждает следующее.

1) Если в состоянии с волновой функцией $\Psi_1(q)$ некоторое измерение дает с достоверностью результат 1, а в состоянии с $\Psi_2(q)$ – результат 2, то линейная комбинация этих функций, $\Psi(q) = c_1 \Psi_1(q) + c_2 \Psi_2(q)$ (где c_1 и c_2 – константы), будет описывать состояние, в котором такое же измерение даст либо результат 1, либо результат 2.

2) Кроме того, если нам известна зависимость состояний от времени, $\Psi_1(q, t)$ и $\Psi_2(q, t)$, то любая их линейная комбинация тоже даст возможную зависимость некоторого квантово-механического состояния от времени.

3) Если квантовая система состоит из двух подсистем, находящихся в состояниях с волновыми функциями $\Psi_1(q)$ и $\Psi_2(q)$ и не взаимодействующих друг с другом, тогда состояние всей системы будет характеризоваться волновой функцией

$$\Psi_{12}(q) = \Psi_1(q) \cdot \Psi_2(q). \quad (4)$$

Последнее равенство будет справедливо не только для конкретного момента времени, но и во все будущие моменты времени: $\Psi_{12}(q, t) = \Psi_1(q, t) \cdot \Psi_2(q, t)$.

Волновая функция и операторы

Каждой физической величине f в квантовой механике ставится в соответствие линейный оператор \hat{f} , действие которого на волновую функцию квантовой системы в данном (n -ом) состоянии сводится к умножению этой функции на определенное значение данной физической величины, f_n :

$$\hat{f}\Psi_n = f_n \Psi_n.$$

Это значение f_n называется собственным значением величины f для n -ого состояния. Для волновых функций можно записать:

$$\int \Psi_n^* \Psi_m dq = \delta_{nm}. \quad (5)$$

Линейность оператора \hat{f} означает выполнение равенства:

$$\hat{f}(c_1 \Psi_1 + c_2 \Psi_2 + \dots + c_k \Psi_k) = c_1 \hat{f}\Psi_1 + c_2 \hat{f}\Psi_2 + \dots + c_k \hat{f}\Psi_k.$$

Возможно разложение произвольной волновой функции Ψ по полной системе собственных функций,

$$\Psi = \sum_{k=1}^n a_k \Psi_k.$$

Также можно записать:

$$\Psi^* = \sum_{k=1}^n a_k^* \Psi_k^*.$$

В таком случае линейность оператора \hat{f} означает выполнение равенства:

$$\hat{f}\Psi = \sum_{k=1}^n a_k f_k \Psi_k. \quad (6)$$

Из $\int \Psi^* \Psi dq = 1$ получаем:

$$\int (a_1^* \Psi_1^* + a_2^* \Psi_2^* + \dots + a_n^* \Psi_n^*) (a_1 \Psi_1 + a_2 \Psi_2 + \dots + a_n \Psi_n) dq = \sum_{k=1}^n |a_k|^2 = 1.$$

Рассмотрим интеграл

$$\int \Psi^* \hat{f} \Psi dq .$$

$$\int \Psi^* \hat{f} \Psi dq =$$

$$\int (a_1^* \Psi_1^* + a_2^* \Psi_2^* + \dots + a_n^* \Psi_n^*) (a_1 \hat{f} \Psi_1 + a_2 \hat{f} \Psi_2 + \dots + a_n \hat{f} \Psi_n) dq = \sum_{k=1}^n |a_k|^2 f_k .$$

Очевидно, что рассматриваемый интеграл дает:

$$\bar{f} = \sum_n f_n |a_n|^2 = \int \Psi^* \hat{f} \Psi dq .$$

Вернемся к уравнению:

$$\hat{f} \Psi_n = f_n \Psi_n .$$

Видно, что если функцией Ψ является одна из собственных функций квантовой системы (так, что все a_k кроме одного равны нулю), то в результате действия на нее оператора \hat{f} эта функция оказывается умноженной на соответствующее собственное значение f_n .

Отсюда следует, что собственные функции Ψ_n являются решениями уравнения:

$$\hat{f} \Psi = f \Psi .$$

Физические величины вещественны.

Поэтому собственные значения оператора физической величины должны быть вещественными и должно выполняться равенство

$$\int \Psi^* (\hat{f} \Psi) dq = \int \Psi (\hat{f}^* \Psi^*) dq ,$$

где оператор \hat{f}^* является комплексно сопряженным с оператором \hat{f} (то есть, если выполняется равенство $\hat{f} \Psi = \Phi$, то должно выполняться также и равенство $\hat{f}^* \Psi^* = \Phi^*$).

Из сказанного ясно, что операторы физических величин должны быть ЭРМИТОВЫМИ.

Сложение операторов

Если \hat{f} и \hat{g} - операторы, отвечающие двум физическим величинам f и g , то сумме $(f + g)$ отвечает оператор $(\hat{f} + \hat{g})$.

Смысл сложения различных физических величин в квантовой механике, однако, существенно различен в зависимости от того, измеримы ли эти величины одновременно или нет.

Если величины f и g одновременно измеримы, то операторы \hat{f} и \hat{g} имеют совместные собственные функции, которые являются в то же время и собственными функциями оператора $(\hat{f} + \hat{g})$, а собственные значения последнего оператора равны суммам.

Если же величины f и g **не могут иметь** одновременно определенных значений, то смысл их суммы $(f + g)$ более ограничен. Можно лишь утверждать, что среднее значение этой величины в произвольном состоянии равно сумме средних значений каждого из слагаемых в отдельности:

$$\overline{f + g} = \bar{f} + \bar{g}.$$

Что же касается собственных значений и функций оператора $(\hat{f} + \hat{g})$, то здесь они, вообще говоря, не будут иметь никакого отношения к собственным значениям и функциям величин **f и g** .

Пусть теперь снова **f и g** – **одновременно измеримые величины**.

Наряду с их суммой можно ввести понятие и об их произведении как о величине, собственные значения которой равны произведениям собственных значений величин f и g . Легко видеть, что такой величине соответствует оператор, действие которого состоит в последовательном действии на функцию сначала одного, а затем другого оператора.

Такой оператор изображается математически как произведение операторов \hat{f} и \hat{g} .

Действительно, если Ψ_n – общие собственные функции операторов \hat{f} и \hat{g} , то

Имеем

$$\hat{f}\hat{g}\Psi_n = \hat{f}g_n\Psi_n = f_n g_n \Psi_n.$$

Точно также

$$\hat{g}\hat{f}\Psi_n = \hat{g}f_n\Psi_n = g_n f_n \Psi_n = f_n g_n \Psi_n.$$

Следовательно

$$\hat{f}\hat{g} - \hat{g}\hat{f} = 0,$$

или

$$[\hat{f}, \hat{g}] = 0.$$

Таким образом, мы приходим к важному результату: **если две величины f и g могут иметь одновременно определенные значения, то их операторы коммутируют друг с другом.**

Может быть доказана и обратная теорема: если операторы f и g коммутативны, то у них все собственные функции можно выбрать общими, что физически означает одновременную измеримость соответствующих физических величин.

Таким образом, коммутативность операторов является необходимым и достаточным условием одновременной измеримости физических величин.

Если же величины f и g не измеримы одновременно, то понятие их произведения не имеет указанного выше прямого смысла. Это проявляется уже в том, что оператор $\hat{f}\hat{g}$ в этом случае не будет эрмитовым, а поэтому не может соответствовать вещественной физической величине.

Заметим, что если $[\hat{f}, \hat{h}] = 0$ и $[\hat{g}, \hat{h}] = 0$, то отсюда, вообще говоря, отнюдь не следует, что и \hat{f} и \hat{g} коммутативны.

Гамильтониан

Волновая функция Ψ полностью определяет состояние физической системы в квантовой механике.

Это означает, что задание этой функции в некоторый момент времени не только описывает все свойства системы в этот момент, но определяет ее поведение также и во все будущие моменты времени (конечно, лишь с той степенью полноты, которая вообще допускается квантовой механикой).

Математически это обстоятельство выражается тем, что значение производной $\partial\Psi/\partial t$ от волновой функции по времени в каждый данный момент времени должно определяться значением самой функции Ψ в тот же момент, причем зависимость эта должна быть, согласно принципу суперпозиции, линейной.

В наиболее общем виде волновое уравнение будет иметь вид

$$i\hbar \frac{\partial\Psi}{\partial t} = \hat{H}\Psi, \quad (7)$$

где \hat{H} - линейный оператор, соответствующий энергии системы (гамильтониан).

Дифференцирование операторов по времени

Понятие о производной физической величины по времени не может быть определено в квантовой механике в том смысле, какой оно имеет в классической механике.

Действительно, определение производной в классической механике связано с рассмотрением значений величины в два близких, но различных момента времени.

Но в квантовой механике величина, имеющая в некоторый момент времени определенное значение, не имеет в следующие моменты вообще никакого определенного значения.

Поэтому понятие производной по времени должно быть определено в квантовой механике иным образом.

Естественно определить производную df/dt от величины f как величину, среднее значение которой равно производной по времени от среднего значения \bar{f} .

Таким образом, имеем, по определению,

$$\frac{\overline{df}}{dt} = \frac{d(\bar{f})}{dt}.$$

В результате имеем:

$$\frac{\overline{df}}{dt} = \frac{d(\bar{f})}{dt} = \frac{d}{dt} \left(\int \Psi^* \hat{f} \Psi dq \right) = \int \Psi^* \frac{\partial \hat{f}}{\partial t} \Psi dq + \int \frac{\partial \Psi^*}{\partial t} \hat{f} \Psi dq + \int \Psi^* \hat{f} \frac{\partial \Psi}{\partial t} dq,$$

где предполагается, что оператор \hat{f} может зависеть от времени как от параметра.

Из (8.1) следует, что

$$\frac{\partial \Psi}{\partial t} = -\frac{i}{\hbar} \hat{H} \Psi, \quad \frac{\partial \Psi^*}{\partial t} = -\frac{i}{\hbar} \hat{H}^* \Psi^*.$$

Поэтому

$$\frac{\overline{df}}{dt} = \int \Psi^* \frac{\partial \hat{f}}{\partial t} \Psi dq + \frac{i}{\hbar} \int (\hat{H}^* \Psi^*) \hat{f} \Psi dq - \frac{i}{\hbar} \int \Psi^* \hat{f} (\hat{H} \Psi) dq.$$

Поскольку оператор \hat{H} - эрмитов, то

$$\int (\hat{H}^* \Psi^*) (\hat{f} \Psi) dq = \int (\Psi^* \hat{H}) (\hat{f} \Psi) dq.$$

Поэтому

$$\frac{\overline{df}}{dt} = \int \Psi^* \left(\frac{\partial \hat{f}}{\partial t} + \frac{i}{\hbar} \hat{H} \hat{f} - \frac{i}{\hbar} \hat{f} \hat{H} \right) \Psi dq.$$

Но поскольку по определению

$$\frac{\overline{df}}{dt} = \int \Psi^* \left(\frac{\hat{df}}{dt} \right) \Psi dq,$$

то получаем

$$\left(\frac{\hat{df}}{dt} \right) = \frac{\partial \hat{f}}{\partial t} + \frac{i}{\hbar} (\hat{H} \hat{f} - \hat{f} \hat{H}).$$

Если оператор \hat{f} не зависит от времени явно, то

$$\left(\frac{\hat{d}f}{dt}\right) = \frac{i}{\hbar}(\hat{H}\hat{f} - \hat{f}\hat{H}).$$

Очень важной категорией физических величин являются те, операторы которых не зависят явно от времени и, кроме того, коммутативны с гамильтонианом,

$$\left(\frac{\hat{d}f}{dt}\right) = 0.$$

Такие величины называют **сохраняющимися**.

Для них среднее значение величины остается постоянным во времени.

Стационарные состояния

Как известно, сохраняющаяся функция Гамильтона называется энергией.

Смысл закона сохранения энергии в квантовой механике состоит в том, что если в данном состоянии энергия имеет определенное значение, то это значение остается постоянным во времени.

Состояния, в которых энергия имеет определенные значения, называются *стационарными состояниями* системы.

Они описываются волновыми функциями Ψ_n , являющимися собственными функциями оператора Гамильтона:

$$i\hbar \frac{\partial \Psi_n}{\partial t} = \hat{H}\Psi_n = E_n \Psi_n.$$

Из уравнения

$$i\hbar \frac{\partial \Psi_n}{\partial t} = E_n \Psi_n.$$

находится вид собственной функции оператора \hat{H} :

$$\Psi_n = \exp\left(-\frac{i}{\hbar} E_n t\right) \cdot \phi_n(q),$$

где $\psi(q)$ – функция, зависящая только от координат.

Разложение произвольной волновой функции Ψ по волновым функциям стационарных состояний имеет вид

$$\Psi = \sum_n a_n \exp\left(-\frac{i}{\hbar} E_n t\right) \cdot \phi_n(q).$$

Теория возмущений

1. Возмущения, не зависящие от времени (случай дискретного спектра и невырожденных уровней энергии)

$$H = H_0 + V,$$

известны точные решения уравнения

$$H \cdot \psi^{(0)} = E^{(0)} \psi^{(0)}.$$

Требуется найти приближенные решения уравнения

$$H \cdot \psi = (H_0 + V) \cdot \psi = E \psi.$$

Разложим

$$\psi = \sum_m c_m \psi_m^{(0)}.$$

Получим

$$\sum_m c_m (E_m^{(0)} + V \cdot \psi_m^{(0)}) = \sum_m c_m E \psi_m^{(0)}.$$

Умножив слева на $\psi_k^{(0)*}$ и проинтегрировав обе стороны по всему пространству, получим

$$(E - E_k^{(0)})c_k = \sum_m V_{km} c_m, \quad (8)$$

где

$$V_{km} = \int \psi_k^{(0)*} V \psi_m^{(0)} dq.$$

Величины коэффициентов c_m и энергии E предполагаются разложенными в ряд по степени малости:

$$c_m = c_m^{(0)} + c_m^{(1)} + c_m^{(2)} + \dots; \quad E = E^{(0)} + E^{(1)} + E^{(2)} + \dots,$$

где $c_m^{(1)}$ и $E^{(1)}$ - величины того же порядка малости, что и возмущение V .

Определим поправки первого порядка малости к n -му собственному значению и собственной функции.

Положим $c_n^{(0)} = 1$; $c_m^{(0)} = 0$; $n \neq m$.

В выражение (8) подставим $c_k = c_k^{(0)} + c_k^{(1)}$ и $E = E_n^{(0)} + E_n^{(1)}$,

сохранив только члены первого порядка малости.

При $k = n$ получаем: $E_n^{(1)} = V_{nn} = \int \psi_n^{(0)*} V \psi_n^{(0)} dq$.

При $k \neq n$ получаем: $c_k^{(1)} = \frac{V_{kn}}{E_n^{(0)} - E_k^{(0)}} \quad (k \neq n)$.

Положим $c_n^{(1)} = 0$.

Тогда:

$$\psi_n^{(1)} = \sum_m' \frac{V_{mn}}{E_n^{(0)} - E_m^{(0)}}$$

$$E_n^{(2)} = \sum_m' \frac{|V_{mn}|^2}{E_n^{(0)} - E_m^{(0)}}$$

2. Возмущения, не зависящие от времени (случай дискретного спектра и вырожденных уровней энергии)

Пусть оператор H_0 имеет вырожденные собственные значения.

$\psi_n^{(0)}, \psi_{n'}^{(0)}, \dots$ собственные функции, относящиеся к одному и тому же собственному значению энергии $E_n^{(0)}$

Правильные функции нулевого приближения –

$$c_n^{(0)}\psi_n^{(0)} + c_{n'}^{(0)}\psi_{n'}^{(0)} + \dots$$

Выпишем уравнения

$$(E - E_k^{(0)})c_k = \sum_m V_{km}c_m \quad (9)$$

с $k = n, n', \dots$, подставив в них в первом приближении

$$E = E^{(0)} + E^{(1)}.$$

Для величин c_k достаточно ограничиться нулевыми значениями $c_n = c_n^{(0)}, c_{n'} = c_{n'}^{(0)}, \dots$, $c_m = 0$. Тогда получим

$$E^{(1)}c_n^{(0)} = \sum_{n'} V_{nn'}c_{n'}^{(0)}$$

или

$$\sum_{n'} (V_{nn'} - E^{(1)}\delta_{nn'})c_{n'}^{(0)} = 0. \quad (10)$$

где числа n, n' пробегает все значения, нумерующие состояния, относящиеся к данному невозмущенному собственному значению $E_n^{(0)}$. Эта система однородных линейных уравнений для величин $c_n^{(0)}$ имеет отличные от нуля решения при условии обращения в нуль определителя, составленного из коэффициентов при неизвестных.

Таким образом, получаем уравнение

$$|V_{nn'} - E^{(1)}\delta_{nn'}| = 0. \quad (11)$$

Это уравнение — s -й степени по $E^{(1)}$.

Уравнение (11) называют секулярным (или вековым) уравнением.

Собственные значения секулярного уравнения могут быть найдены методом теории возмущений. В первом порядке теории возмущений поправки к собственному значению нулевого приближения равны диагональным матричным элементам матрицы оператора V . Поправки второго порядка находятся с учетом недиагональных матричных элементов V_{nm} . Находим их следующим образом.

В уравнении (8) с $k = n$ в левой стороне равенства полагаем $E = E^{(0)} + E^{(1)}$ (сохраняем обозначение $E^{(1)}$ для поправки к энергии в рассматриваемом приближении), а вместо c_n пишем $c_n^{(0)}$.

Имея в виду, что $c_m^{(0)}$ для всех $m \neq n, n', \dots$, имеем

$$E^{(1)}c_n^{(0)} = \sum_m V_{nm} c_m^{(1)} + \sum_{n'} V_{nn'} c_{n'}^{(0)}, \quad (12)$$

откуда

$$c_m^{(1)} = \sum_{n'} \frac{V_{mn'}}{E_n^{(0)} - E_m^{(0)}} c_{n'}^{(0)} \quad (m \neq n).$$

Подставив это в (12), находим

$$E^{(1)}c_n^{(0)} = \sum_{n'} c_{n'}^{(0)} \left(V_{nn'} + \sum_m \frac{V_{nm}V_{mn'}}{E_n^{(0)} - E_m^{(0)}} \right).$$

Эта система уравнений заменяет теперь систему (11); условие их совместности снова приводит к секулярному уравнению, отличающемуся от (10) заменой

$$E^{(1)}c_n^{(0)} = V_{nn'} \rightarrow V_{nn'} + \sum_m \frac{V_{nm}V_{mn'}}{E_n^{(0)} - E_m^{(0)}}. \quad (13)$$

Теперь найдем формулы для поправок первого приближения к собственным функциям и второго приближения для собственных значений. Будем считать, что в качестве функций $\psi_n^{(0)}$ выбраны правильные функции нулевого приближения. Определенная с их помощью матрица $V_{nn'}$ диагональна по индексам n, n' (относящимся к одной и той же группе функций вырожденного уровня), причем диагональные элементы $V_{nn}, V_{n'n'}, \dots$ равны соответствующим поправкам первого приближения $E_n^{(1)}, E_{n'}^{(1)}, \dots$. Рассматриваем возмущение собственной функции $\psi_n^{(0)}$, так что в нулевом приближении $E = E_n^{(0)}$, $c_n^{(0)} = 1$ и $c_m^{(0)} = 0$ при $n \neq m$. В первом приближении $E = E_n^{(0)} + V_{nn}$, $c_n = 1 + c_n^{(1)}$ и $c_m = c_m^{(1)}$. Выпишем из общей системы

(9) уравнение с $k \neq n, n', \dots$, сохраняя в нем члены первого порядка:

$$(E_n^{(0)} - E_k^{(0)})c_k^{(1)} = V_{kn}c_n^{(0)} = V_{kn},$$

откуда

$$c_k^{(1)} = \frac{V_{kn}}{E_n^{(0)} - E_k^{(0)}} \quad (k \neq n, n', \dots).$$

Находим, что при сохранении членов второго порядка и при $k = n$

$$E^{(1)}c_n^{(1)} = V_{n'n'}c_{n'}^{(1)} + \sum_m V_{n'm}c_m^{(1)}.$$

$$c_{n'}^{(1)} = \frac{1}{(V_{nn} - V_{n'n'})} \sum_m \frac{V_{n'm}V_{mn}}{E_n^{(0)} - E_m^{(0)}}. \quad (2)$$

Лекция 2. Методы описания локализованных и делокализованных состояний в твердых телах.

1. Локализованные состояния электронов.

Для атома водорода или водородоподобного иона с единственным электроном на орбите гамильтониан имеет вид:

$$H = -\frac{\hbar^2}{2m} \Delta - \frac{Ze^2}{r}.$$

Второй член в этом операторе представляет потенциальную энергию единственного электрона и называется кулоновским потенциалом.

Особенностью кулоновского потенциала

$$-\frac{Ze^2}{r}$$

является то, что энергия электрона зависит только от значения квантового числа n .

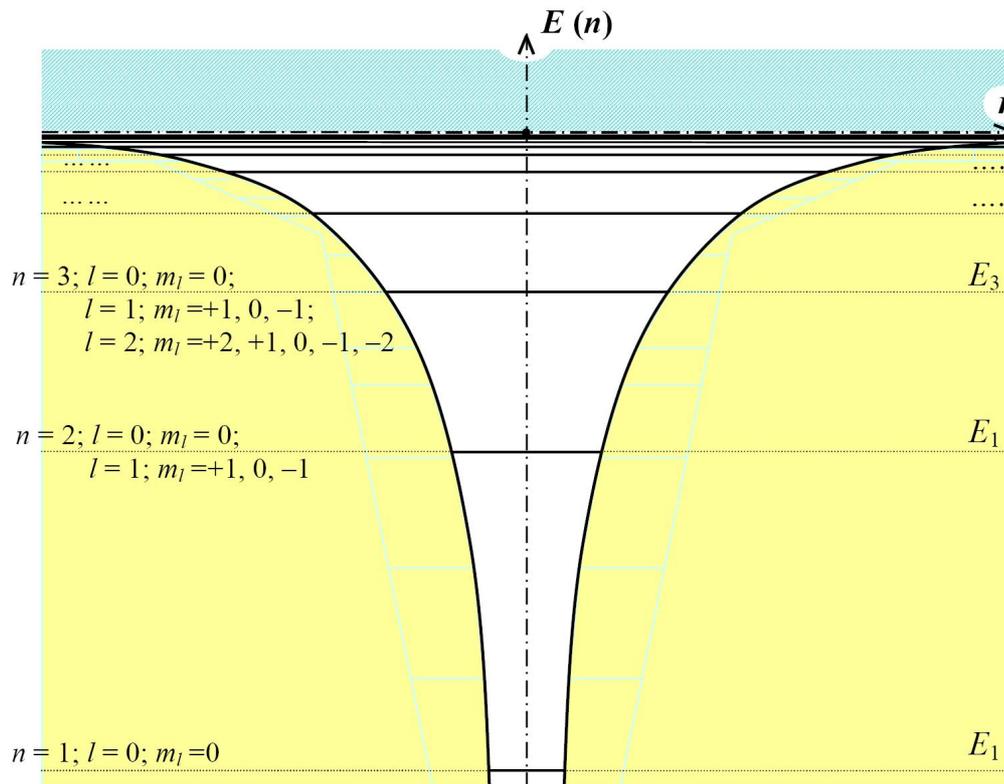


Рис. Уровни энергии электрона в атоме водорода

В случае многоэлектронных атомов или ионов

гамильтониан, включающий наиболее сильные взаимодействия, имеет вид:

$$H = \sum_{i=1}^N \left(\frac{p_i^2}{2m} - \frac{Ze^2}{r_i} \right) + \sum_{j>k=1}^N \frac{e^2}{r_{jk}} .$$

С учетом того, что

$$p = i\hbar\nabla \text{ и } (i \cdot \nabla)^2 = -\Delta = -\left(\frac{\partial}{\partial^2 x} + \frac{\partial}{\partial^2 y} + \frac{\partial}{\partial^2 z} \right),$$

это уравнение запишется как:

$$H = \sum_{i=1}^N \left(-\frac{\hbar^2}{2m} \Delta - \frac{Ze^2}{r_i} \right) + \sum_{j>k=1}^N \frac{e^2}{r_{jk}},$$

где

$$r_{ik} = |r_i - r_k|$$

Учет оператора межэлектронного взаимодействия,

$$\sum_{j>k=1}^N \frac{e^2}{r_{jk}},$$

не позволяет получить точное решение гамильтониана многоэлектронного атома.

Взаимодействие между электронами достаточно велико, поэтому рассматривать его в целом как возмущение нельзя. Однако полное взаимодействие можно разделить на две части таким образом, что для одной из них, главной, оказывается возможным относительно просто найти волновые функции и энергии, а оставшуюся часть можно рассматривать как возмущение.

Основная часть энергии:

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2}{2m} \Delta + U(r_i) \right].$$

Возмущение:

$$V = \sum_{j>k=1}^N \frac{e^2}{r_{jk}} .$$

Этот метод в общем случае называется методом самосогласованного поля (ССП) или методом Хартри-Фока. В рамках такого подхода удобно представить главную часть взаимодействия в виде суммы взаимодействий каждого из электронов (например, i -го электрона) с некоторым сферически-симметричным потенциалом $U(r_i)$. Предполагается, что этот потенциал создается ядром и всеми электронами, за исключением i -го электрона.

Рассмотрим основную часть:

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2}{2m} \Delta + U(r_i) \right]. \quad (2)$$

Предположим, что наш атом не взаимодействует со своим окружением. Рассматривая его как стационарную систему, для определения возможных энергий системы применим стационарное уравнение Шредингера:

$$H' \cdot \Psi = E' \Psi$$

или в развернутом виде

$$\left\{ \sum_{i=1}^N \left[-\frac{\hbar^2}{2m} \Delta + U(r_i) \right] \right\} \cdot \Psi = E' \Psi . \quad (3)$$

Из (3) видно, что при пренебрежении нецентральной частью взаимодействия электронов друг с другом мы достигли полного разделения переменных. Обозначим набор $\{n, l, m_l\}$ квантовых чисел для i -го электрона посредством k_i . Тогда многоэлектронная волновая функция Ψ для системы «ядро+ N -электронов» в данном приближении может быть представлена в виде произведения волновых функций, найденных для каждого электрона,

$$\Psi = \prod_{i=1}^N \Psi_i(k_i), \quad (4)$$

а энергия N -электронов, окружающих ядро, будет определяться суммой энергий каждого из этих электронов,

$$E' = \sum E'_i(k_i). \quad (5)$$

В дальнейшем мы опустим индекс i в обозначении набора квантовых чисел i -го электрона. Отличие многоэлектронной волновой функции от одноэлектронной (определяющей состояния рассматриваемого электрона) будет состоять в том, что в первом случае мы будем писать Ψ , а во втором случае - $\Psi(k)$.

Итак, используя приближение, в котором нецентральная часть электрон-электронного взаимодействия не рассматривается, в качестве решений уравнения Шредингера (3) мы получаем волновые функции, совпадающие с волновыми функциями водородоподобного атома:

$$\Psi(k) = r^{-1} R_{nl}(r) Y_{lm_l}(\vartheta, \varphi). \quad (6)$$

В выражении (6) использованы обозначения:

$R_{nl}(r)$ – радиальная функция, зависящая от вида центрального потенциала $U(r)$;

$Y_{lm_l}(\vartheta, \varphi)$ - сферическая функция, описывающая угловую зависимость волновой функции

$\Psi(k)$.

Примерные графики некоторых радиальных функций ns , np и nd показаны на рис. 1.

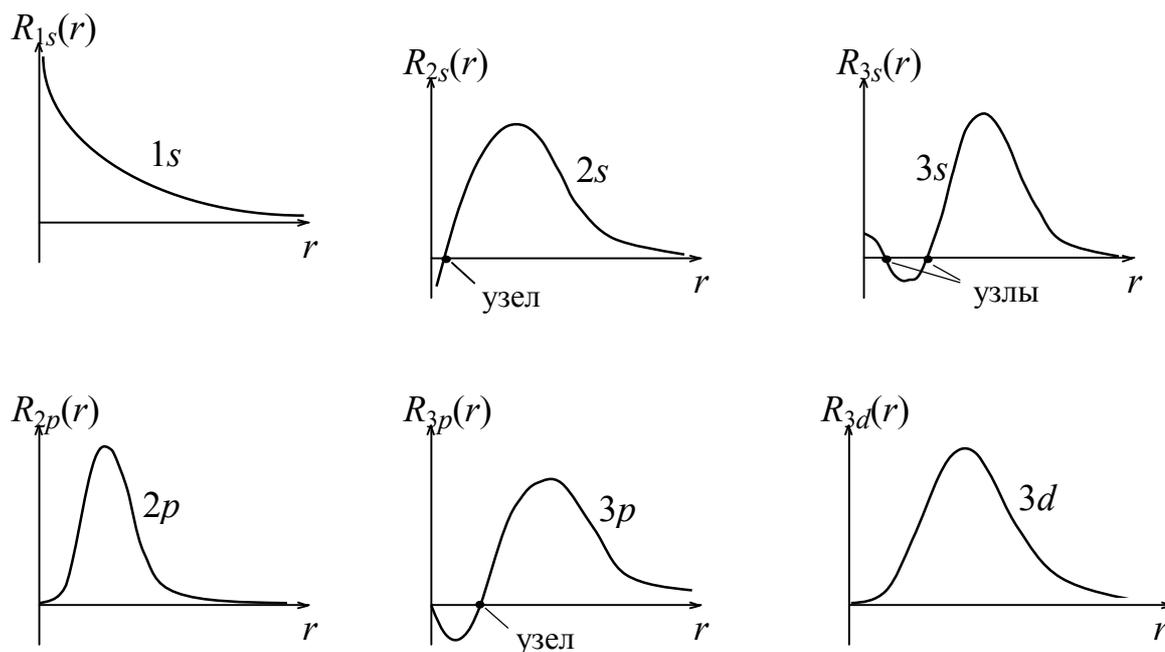


Рис. 1.

Известно, что сферические функции определяются равенством

$$Y_{lm_l}(\vartheta, \varphi) = (-1)^m \left\{ \frac{(2l+1)(l-m_l)!}{4\pi(l+m_l)!} \right\}^{\frac{1}{2}} P_l^{m_l}(\cos \vartheta) e^{im_l\varphi}, \quad (7)$$

где $P_l^{m_l}(\cos \vartheta)$ - полином Лежандра, определяемый равенством

$$P_l^{m_l}(x) = \frac{(1-x^2)^{(l/2)m_l}}{2^l l!} \cdot \frac{d^{(l+m_l)}}{dx^{(l+m_l)}} (x^2-1)^l. \quad (8)$$

В случае многоэлектронного атома собственными волновыми функциями являются слеттеровские детерминанты построенные на следующих функциях

$Y_{lm_l}(\vartheta, \varphi)$:

$$Y_{00} = |00\rangle = \left(\frac{1}{4\pi} \right)^{\frac{1}{2}},$$

$$Y_{10} = |10\rangle = \left(\frac{3}{4\pi} \right)^{\frac{1}{2}} \cdot \text{Cos}(\vartheta),$$

$$Y_{1\pm 1} = |1\pm 1\rangle = \mp \left(\frac{3}{8\pi} \right)^{\frac{1}{2}} \cdot \text{Sin}(\vartheta) \cdot e^{\pm i\varphi},$$

$$Y_{20} = |20\rangle = \left(\frac{5}{4\pi} \right)^{\frac{1}{2}} \left(\frac{1}{2} \right) \cdot (3\text{Cos}^2(\vartheta) - 1),$$

$$Y_{2\pm 1} = |2 \pm 1\rangle = \mp \left(\frac{5}{4\pi}\right)^{\frac{1}{2}} \left(\frac{3}{2}\right)^{\frac{1}{2}} \cdot \sin(\vartheta) \cos(\vartheta) \cdot e^{\pm i\varphi},$$

$$Y_{2\pm 2} = |2 \pm 2\rangle = \left(\frac{5}{4\pi}\right)^{\frac{1}{2}} \left(\frac{3}{8}\right)^{\frac{1}{2}} \cdot \sin^2(\vartheta) \cdot e^{\pm 2i\varphi}.$$

Для графического представления обычно используют действительные комбинации сферических функций:

$$|s\rangle = |00\rangle = \frac{1}{4\pi},$$

$$|p_z\rangle = |10\rangle = \sqrt{\frac{3}{4\pi}} \cos(\vartheta) = \sqrt{\frac{3}{4\pi}} \frac{1}{r} z,$$

$$|p_x\rangle = \frac{1}{\sqrt{2}} (|1-1\rangle - |1+1\rangle) = \sqrt{\frac{3}{4\pi}} \sin(\vartheta) \cos(\varphi) = \sqrt{\frac{3}{4\pi}} \frac{1}{r} x,$$

$$|p_y\rangle = \frac{i}{\sqrt{2}} (|1-1\rangle + |1+1\rangle) = \sqrt{\frac{3}{4\pi}} \sin(\vartheta) \sin(\varphi) = \sqrt{\frac{3}{4\pi}} \frac{1}{r} y,$$

$$|d_{z^2}\rangle = |20\rangle = \sqrt{\frac{5}{16\pi}} \cdot (3\cos^2(\vartheta) - 1) = \sqrt{\frac{5}{16\pi}} \frac{1}{r^2} (3z^2 - r^2),$$

$$|d_{x^2-y^2}\rangle = \frac{1}{\sqrt{2}} (|2+2\rangle + |2-2\rangle) = \sqrt{\frac{15}{16\pi}} \sin^2(\vartheta) \cos(2\varphi) = \sqrt{\frac{15}{16\pi}} \frac{1}{r^2} (x^2 - y^2),$$

$$|d_{xy}\rangle = -\frac{i}{\sqrt{2}} (|2+2\rangle - |2-2\rangle) = \sqrt{\frac{15}{16\pi}} \sin^2(\vartheta) \sin(2\varphi) = \sqrt{\frac{15}{16\pi}} \frac{1}{r^2} xy,$$

$$|d_{xz}\rangle = -\frac{1}{\sqrt{2}} (|2+1\rangle - |2-1\rangle) = \sqrt{\frac{15}{4\pi}} \sin(\vartheta) \cos(\vartheta) \cos(\varphi) = \sqrt{\frac{15}{4\pi}} \frac{1}{r^2} xz,$$

$$|d_{yz}\rangle = \frac{i}{\sqrt{2}} (|2+1\rangle + |2-1\rangle) = \sqrt{\frac{15}{4\pi}} \sin(\vartheta) \cos(\vartheta) \sin(\varphi) = \sqrt{\frac{15}{4\pi}} \frac{1}{r^2} yz.$$

Радиальные части ВФ ($R_{nl}(r)$) в расчетах задаются в виде гауссовских функций:

$$R_{00}(r) = \sum_i C_i e^{-\beta_i r^2},$$

$$R_{1m_l}(r) = \sum_i C_i r e^{-\alpha_i r^2},$$

$$R_{2m_l}(r) = \sum_i C_i r^2 e^{-\alpha_i r^2},$$

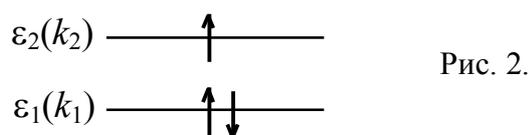
$$R_{3m_l}(r) = \sum_i C_i r^3 e^{-\alpha_i r^2}.$$

В волновых функциях спин электрона можно учесть, умножая орбитальную функцию $\Psi(k)$ на спиновую функцию, α или β .

В процессе развития теории оказалось, что волновая функция

$$\Psi(k) = r^{-1} R_{nl}(r) Y_{lm_l}(\vartheta, \varphi)$$

не удовлетворяет принципу Паули. Поэтому используют симметризованную функцию (детерминант Слеттера). Для примера рассмотрим случай трехэлектронного атома, где три электрона заселяют два энергетических уровня $\varepsilon_1(k_1)$ и $\varepsilon_2(k_2)$, характеризуемых наборами квантовых чисел k_1 и k_2 (см. рис. 2).



Волновая функция такого трехэлектронного атома будет представляться в виде следующего слеттеровского детерминанта:

$$\Psi' = (3!)^{-1/2} \begin{vmatrix} \Psi_1(k_1) \cdot \alpha_1 & \Psi_1(k_1) \cdot \beta_1 & \Psi_1(k_2) \cdot \alpha_1 \\ \Psi_2(k_1) \cdot \alpha_2 & \Psi_2(k_1) \cdot \beta_2 & \Psi_2(k_2) \cdot \alpha_2 \\ \Psi_3(k_1) \cdot \alpha_3 & \Psi_3(k_1) \cdot \beta_3 & \Psi_3(k_2) \cdot \alpha_3 \end{vmatrix}. \quad (9)$$

Для упрощения этот слеттеровский детерминант может быть записан в следующем виде:

$$\Psi' = \left\{ \begin{matrix} + & - & + \\ k_1; k_1; k_2 \end{matrix} \right\}, \quad (10)$$

где фигурные скобки подразумевают слеттеровский детерминант.

Одно из свойств таких детерминантов может быть представлено равенствами:

$$\left\{ \begin{matrix} + & - & + \\ k_1; k_1; k_2 \end{matrix} \right\} = - \left\{ \begin{matrix} - & + & + \\ k_1; k_1; k_2 \end{matrix} \right\} = \left\{ \begin{matrix} - & + & + \\ k_1; k_2; k_1 \end{matrix} \right\}. \quad (11)$$

Иногда наборы квантовых чисел одноэлектронных волновых функций обозначаются без надстрочных символов «+» или «-». Например, вместо

$$\bar{k}_j$$

записывают

$$K_j = \bar{k}_j,$$

подразумевая при этом набор из четырех квантовых чисел, $\{n, l, m_l, m_s\}$.

Обобщенный слеттеровский детерминант может быть записан в следующем раскрытом виде:

$$\{K_1; K_2; \dots; K_N\} = (N!)^{-1/2} \sum_{\underline{P}} (-1)^{\overline{P}} [\Psi_1(K_1) \Psi_2(K_2) \Psi_3(K_3) \dots \Psi_N(K_N)], \quad (12)$$

где символ \underline{P} обозначает перестановку координат электронов среди состояний, определяемых наборами квантовых чисел K_1, K_2, \dots, K_N , а \overline{P} определяет четность перестановки – если перестановка четная, то $\overline{P} = +1$, в противном случае $\overline{P} = -1$.

Решения уравнения

$$\left\{ \sum_{i=1}^N \left[-\frac{\hbar^2}{2m} \Delta + U(r_i) \right] \right\} \cdot \Psi = E \Psi \quad (13)$$

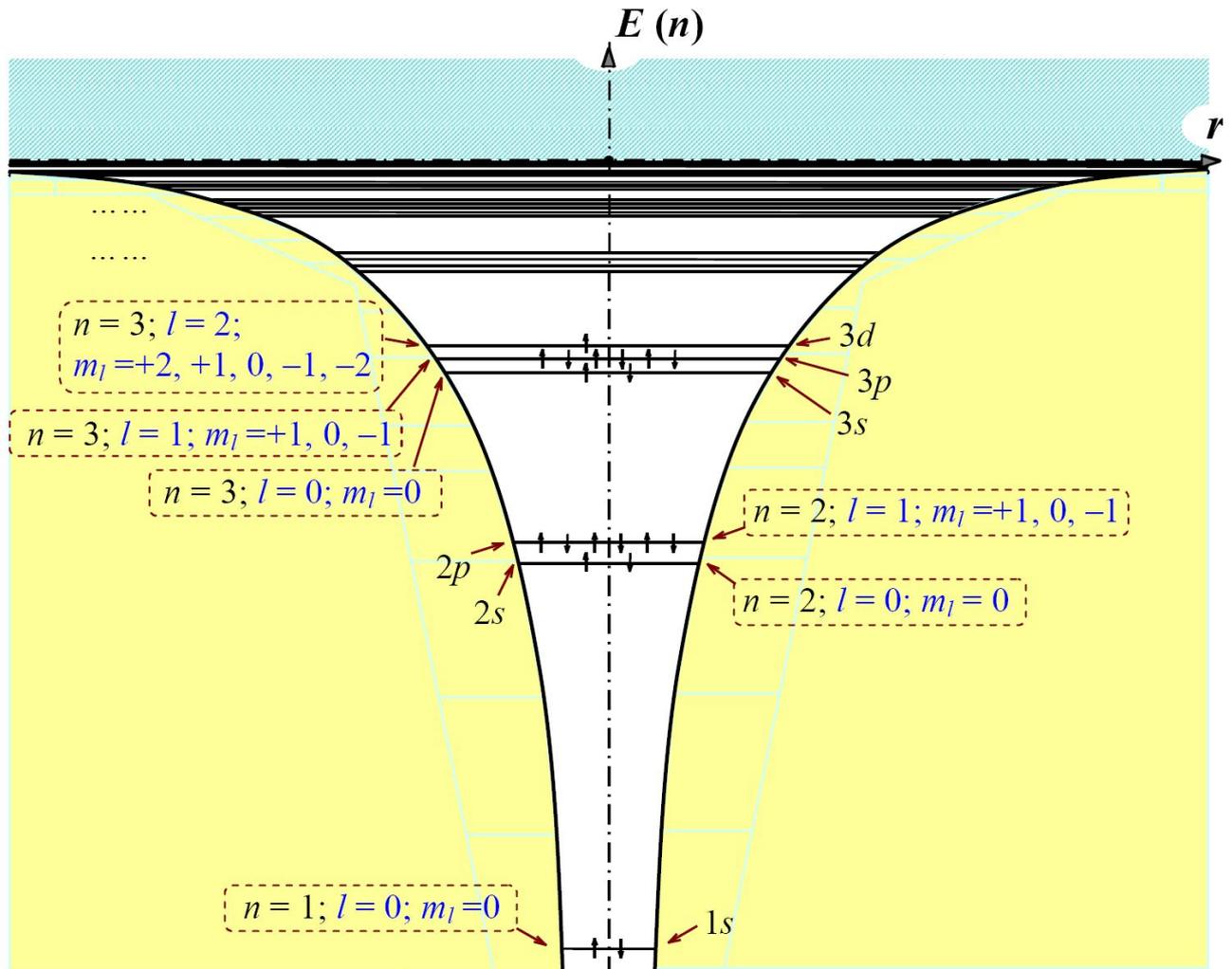
вырождены по m_l и m_s , но зависят от квантовых чисел n и l . Поэтому энергии каждого из N электронов можно обозначать с помощью символов n и l с соответствующими для данного электрона значениями. Эти два квантовых числа обозначают электронную оболочку, к которой принадлежит данный электрон. Совокупностью N пар символов $\{(n_1 l_1), (n_2 l_2), \dots, (n_N l_N)\}$ определяется электронная конфигурация рассматриваемого N -электронного атома. Если значения символов n и l в каких то парах указанной совокупности совпадают (например в k парах), то выражение для электронной конфигурации упрощается путем обозначения таких пар символов выражением $(n l)^k$. Обычно в обозначении электронной конфигурации значения $l = 0, 1, 2, 3, 4, 5, \dots$ заменяют на соответствующие им буквы s, p, d, f, g, h, \dots . Максимальное число пар с одинаковыми символами n и l равно $2(2l+1)$. В таком случае электронная оболочка $(n l)^{2(2l+1)}$ является замкнутой. В замкнутой оболочке $(n l)^{2(2l+1)}$ присутствуют все наборы $K = (n, l, m_l, m_s)$, возможные при данных n и l . Наличие всех азимутальных чисел в замкнутой оболочке придает ей инертный характер. В частности, для нее полные орбитальный и спиновый моменты равны нулю. Дополнительные возмущения конечно могут привести к поляризации замкнутой оболочки, но если эти возмущения малы, эффектами поляризации можно пренебречь. Поэтому в подавляющем большинстве случаев квантово механических расчетов влияние малых возмущений рассматривают только в отношении незаполненных электронных оболочек.

Если электроны вне замкнутых оболочек эквивалентны (обладают одинаковыми n и l), то в слеттеровском детерминанте эти числа можно не повторять. Например, электронная конфигурация иона Nd^{3+} записывается как $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2 5p^6 4f^3$. У этого иона в незамкнутой оболочке находятся три $4f$ -электрона. Поэтому исходные собственные функции можно писать в следующем виде:

$$\left\{ \begin{matrix} + & + & - \\ 3 & 2 & 1 \end{matrix} \right\}, \left\{ \begin{matrix} + & + & - \\ 3 & 1 & -2 \end{matrix} \right\}, \left\{ \begin{matrix} - & - & + \\ 0 & -1 & -3 \end{matrix} \right\} \text{ и т.д.}$$

Здесь цифрами заданы величины m_l , а символы + и – определяют величину m_s (соответственно, $m_s = +1/2$ или $m_s = -1/2$). Величины n и l здесь не указаны, а только подразумеваются (в данном примере $n = 4$, а l соответствует f).

В описанном выше приближении уровни электронов атома или иона будут иметь энергии:



Здесь представлена энергетическая диаграмма иона с электронной конфигурацией $3d^1$.

Далее рассматривают возмущение;

$$V = \sum_{j>k=1}^N \frac{e^2}{r_{jk}}$$

Теперь уже **электроны теряют свою индивидуальность**.

Поэтому переходят к интегральному описанию энергии атома.

В результате мы **приходим к электронным термам**:

2. Делокализованные состояния электронов.

Квантовая статистика Ферми Дирака

Свободные электроны, образующие в металлах электронный газ, по своим свойствам отличны от молекул газа. Газ молекул подчиняется статистике Максвелла-Больцмана, а электронный газ – статистике Ферми-Дирака. Рассмотрим различия между этими статистиками.

1. Статистике Максвелла-Больцмана подчиняются частицы, движение которых согласуется с законами классической механики. Состояние любой такой частицы однозначно определяется ее координатами x, y, z и составляющими ее импульса p_x, p_y, p_z . Значения этих шести переменных могут меняться непрерывно. Считаются различными состояниями, отличающиеся бесконечно мало величинами соответствующих координатных компонент и компонент импульса.

Электроны обладают волновыми свойствами. Поэтому энергия и другие характеристики их движения оказываются квантованными, будучи решениями волнового уравнения Шредингера. Они квантованы и соответствуют определенному квантовому состоянию электрона. Здесь работает принцип неопределенности Гейзенберга, согласно которому должно выполняться соотношение:

$$dx \cdot dy \cdot dz \cdot dp_x \cdot dp_y \cdot dp_z \geq h^3, \quad (1)$$

где h – постоянная Планка. Произведение $dx \cdot dy \cdot dz \cdot dp_x \cdot dp_y \cdot dp_z$ представляет элемент шестимерного фазового пространства $d\tau$. Соотношение (1) говорит о том, что различным элементам шестимерного фазового пространства $d\tau$ будут соответствовать различные квантовые состояния электрона лишь в том случае, если размер этих элементов не меньше h^3 . Поэтому в квантовой статистике за элементарную ячейку 6-мерного фазового пространства принимается объем, равный

$$d\tau = h^3. \quad (2)$$

При рассмотрении свободных электронов предполагается, что их потенциальная энергия одинакова во всех точках металла, вследствие чего их распределение в объеме V является равномерным. В этом случае вместо 6-мерного фазового пространства X, Y, Z, P_x, P_y, P_z , пользуются 3-мерным пространством импульсов P_x, P_y, P_z и разбивают его на элементарные ячейки размером

$$d\tau = \frac{h^3}{V}. \quad (3)$$

Каждой такой ячейке соответствует отдельное квантовое состояние, отличное от других состояний.

Таким образом, первое отличие квантовой статистики от статистики Максвелла-Больцмана состоит в методе деления фазового пространства на элементарные ячейки. Классическая статистика не делает никаких ограничений на величину этих ячеек, полагая, что они могут быть сколь угодно малыми. Квантовая же статистика считает, что физический смысл имеют лишь ячейки размером не менее h^3 для 6-мерного фазового пространства и h^3/V – для 3-мерного пространства импульсов.

2. Электроны подчиняются принципу Паули. Согласно этому принципу в твердом теле в каждом квантовом состоянии с энергией E может находиться не более двух электронов, отличающихся друг от друга направлением спинов. Это означает, что каждая ячейка h^3/V пространства импульсов, соответствующая данному квантовому состоянию, может вместить не более 2 электронов.

3. Статистика Максвелла–Больцмана индивидуализирует молекулы газа, считая, что их можно отличать друг от друга. Поэтому перестановка местами двух молекул, находящихся в разных состояниях, приводит, согласно классической статистике, к новому микросостоянию системы. Квантовая же статистика Ферми–Дирака считает все частицы неотличимыми друг от друга. Поэтому перестановка их местами не приводит к появлению нового микросостояния.

Функция распределения Ферми–Дирака.

Основная задача статистики Ферми–Дирака, как и любой другой физической статистики, состоит в отыскании функции распределения f_F , отвечающей равновесному (наивероятнейшему) состоянию системы и имеющей следующий смысл.

Представим себе кусок металла объемом V , в котором находится N свободных электронов, образующих электронный газ. Построим пространство импульсов с осями координат P_x, P_y, P_z . Разобьем это пространство на элементарные ячейки объемом h^3/V . Каждой такой ячейке соответствует определенное квантовое состояние с энергией E . Функция распределения выражает вероятность заполнения этих ячеек электронами, или, что то же самое, среднюю степень заполнения ячеек электронами. Так, ячейки, для которых $f_F = 1/2$ заполнены в среднем наполовину, ячейки, для которых $f_F = 1$, заполнены целиком и т. д.

Статистический расчет, основанный на учете указанных выше свойств частиц электронного газа, приводит к следующему выражению для f_F :

$$f_F = \frac{1}{\exp\left(\frac{E - \mu}{kT}\right) + 1}. \quad (4)$$

Здесь E – энергия, соответствующая заполняемой ячейке, k — постоянная Больцмана, T – абсолютная температура, μ – химический потенциал, отнесенный к отдельной частице и равный:

$$\mu = \frac{U - TS + pV}{N},$$

где U – внутренняя энергия системы, S – энтропия, V – объем системы, p – давление, N – число частиц в системе. Химический потенциал выражает работу, которая затрачивается при данных условиях на увеличение числа частиц системы на единицу. При равенстве химических потенциалов какого-либо вещества, находящегося в двух соприкасающихся фазах, устанавливается равновесие, при котором не происходит направленного переноса частиц из одной фазы в другую. В частности, при равенстве химических потенциалов двух соприкасающихся электронных проводников между ними устанавливается равновесие, при котором потоки электронов, идущие из одного проводника в другой, оказываются одинаковыми;

p называют также энергией Ферми. При $E = \mu$ имеем $f_F = 1/2$.

Следовательно, энергии Ферми μ соответствуют квантовые состояния, вероятность (средняя степень) заполнения которых равна половине.

Распределение электронов по составляющим импульса

Функция распределения позволяет определить число электронов, составляющие импульса которых лежат в интервалах от p_x до $p_x + dp_x$, от p_y до $p_y + dp_y$, от p_z до $p_z + dp_z$. С этой целью построим в пространстве импульсов элемент фазового объема $dp_x dp_y dp_z$ и разобьем его на ячейки h^3/V . Число таких ячеек будет равно:

$$Z = \frac{dp_x dp_y dp_z}{h^3 / V} = \frac{V}{h^3} dp_x dp_y dp_z. \quad (5)$$

Поделив правую и левую части этого выражения на объем металла V , получим число ячеек в элементе $dp_x dp_y dp_z$ для единицы объема металла:

$$z = \frac{dp_x dp_y dp_z}{h^3}. \quad (6)$$

Так как вероятность того, что любая из этих ячеек занята, равна f_F и так как в каждой из них может разместиться, согласно принципу Паули, 2 электрона, то число электронов в единице объема металла, составляющие импульса которых лежат в интервалах от p_x до $p_x + dp_x$, от p_y до $p_y + dp_y$, от p_z до $p_z + dp_z$, будет равно:

$$dn(p_x, p_y, p_z) = z \cdot 2f_F = \frac{2}{h^3} \frac{dp_x dp_y dp_z}{\exp\left(\frac{E - \mu}{kT}\right) + 1}. \quad (7)$$

Распределение электронов по энергиям.

Можно определить число электронов в единице объема металла, энергия которых заключена в интервале от E до $E+dE$. Для этого представим в пространстве импульсов две концентрические сферы с радиусами p и $p+dp$. Этим сферам соответствуют энергии E и $E+dE$. Объем шарового слоя между сферами равен $4\pi p^2 dp$. В нем разместится Z элементарных ячеек h^3/V :

$$Z = \frac{4\pi p^2 dp}{h^3 / V}. \quad (8)$$

Так как $p^2 = 2mE$ и $dp = (m/2E)^{1/2}$, то для Z получим следующее выражение:

$$Z = \frac{2\pi V(2m)^{3/2} E^{1/2} dE}{h^3}. \quad (9)$$

Поделив правую и левую части этого выражения на объем металла V , определим **число квантовых состояний (число ячеек) в единице объема металла, приходящееся на интервал энергий от E до $E+dE$:**

$$z = \frac{2\pi(2m)^{3/2} E^{1/2} dE}{h^3}. \quad (10)$$

На рис. 1 показана зависимость числа квантовых состояний (числа ячеек) z от энергии E . Площадь заштрихованной плоскости соответствует числу ячеек (числу квантовых состояний), приходящемуся на интервал dE . Из рисунка видно, что с увеличением E площадь полосок увеличивается. Это означает, что в равных по величине интервалах dE содержится тем больше ячеек z , чем выше энергия E , соответствующая этим ячейкам. Иначе говоря, энергетические уровни, соответствующие квантовым состояниям, располагаются тем ближе друг к другу, чем выше энергия, отвечающая этим уровням (рис. 2).

Найдем число электронов в единице объема металла, энергия которых лежит в интервале от E до $E+dE$, умножив число ячеек (10) на f_F и на 2:

$$dn(E) = 2f_{\phi} z = \frac{4\pi(2m)^{3/2}}{h^3} \frac{E^{1/2} dE}{\exp\left(\frac{E-\mu}{kT}\right) + 1}. \quad (11)$$

Эта формула выражает закон распределения электронов по энергиям.

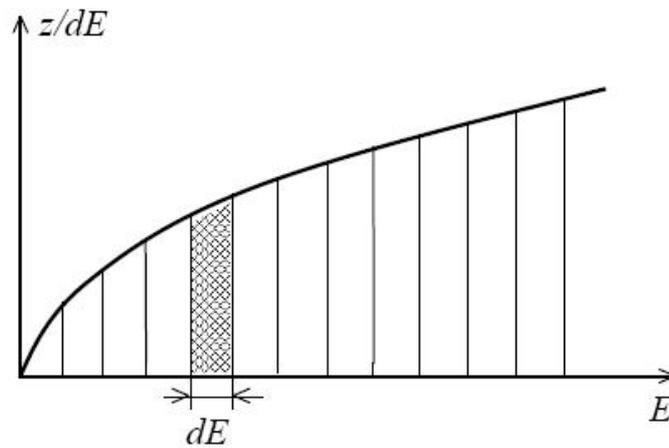


Рис. 1. Распределение квантовых состояний по энергиям

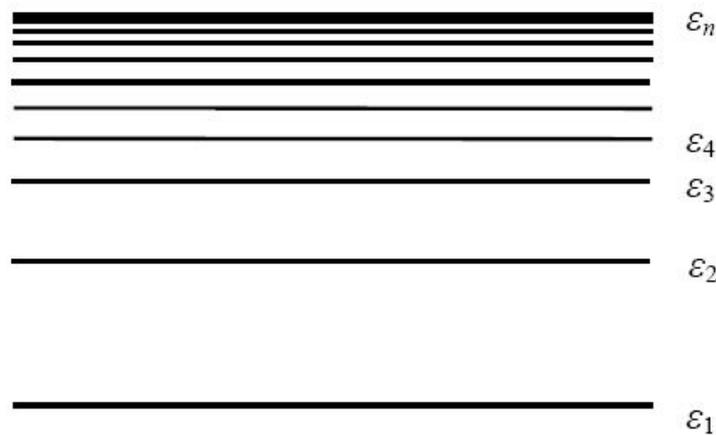


Рис. 2. Расположение уровней энергии электронов в металле

Распределение электронов по импульсам.

Если в формулу (11) вместо E и dE подставить соответственно $(p^2/2m)$ и (pdp/m) , то получится следующее выражение для $dn(p)$:

$$dn(p) = \frac{4\pi p^2}{h^3} \frac{dp}{\exp\left(\frac{E - \mu}{kT}\right) + 1}. \quad (12)$$

Соотношение (7.12) определяет число электронов в единице объема металла, полные импульсы которых заключены в интервале от p до $p+dp$.

3. Вырожденный и невырожденный электронный газ

Различие между молекулярным и электронным газом выступает особенно резко при температуре абсолютного нуля.

Распределение молекул идеального газа, не подверженного действию силового поля, описывается функцией Максвелла–Больцмана:

$$f = N \left(\frac{1}{2\pi m k T} \right)^{3/2} \exp \left(-\frac{E}{kT} \right). \quad (13)$$

Выше показано, что при абсолютном нуле f обращается в бесконечность. Это выражает тот факт, что при $T = 0^\circ\text{K}$ фазовые точки, изображающие состояние молекул в пространстве импульсов, стягиваются в начало координат и плотность их распределения становится бесконечно большой.

Распределение свободных электронов в металле описывается функцией Ферми-Дирака:

$$f_F = \frac{1}{\exp \left(\frac{E - \mu}{kT} \right) + 1}. \quad (14)$$

На рис. 3 показан график этой функции для $T = 0^\circ\text{K}$; μ_0 представляет собою энергию Ферми при $T = 0^\circ\text{K}$. При $E < \mu_0$ разность $(E - \mu_0) < 0$ и выражение (14) можно переписать в следующем виде:

$$f_F = \frac{1}{\left[\exp \left(\frac{\mu_0 - E}{kT} \right) \right]^{-1} + 1}.$$

Так как $(E - \mu_0) < 0$, то при $T \rightarrow 0$, $\exp[(\mu_0 - E)/(kT)] \rightarrow \infty$ и $f_F \rightarrow 1$. При $E > \mu_0$ разность $(E - \mu_0) > 0$, вследствие чего при $T \rightarrow 0$, $\exp[(E - \mu_0)/(kT)] \rightarrow \infty$ и $f_F \rightarrow 0$.

Таким образом, при абсолютном нуле температур в интервале энергий от 0 до μ_0 функция распределения Ферми-Дирака равна 1. При переходе через μ_0 она скачкообразно падает до нуля. Это означает, что при абсолютном нуле свободные электроны занимают все дозволенные энергетические уровни вплоть до энергии μ_0 (вероятность заполнения этих уровней равна 1). Уровни же, энергия которых выше μ_0 , остаются совершенно свободными (вероятность их заполнения равна 0). Следовательно, энергия Ферми μ_0 представляет собою максимальную энергию, которую могут иметь свободные электроны при абсолютном нуле. Она является константой, в значительной мере характеризующей электрические свойства металлов.

Если по оси абсцисс откладывать порядковый номер энергетических уровней, занятых электронами, то последним из них будет уровень $N/2$, где N – число свободных электронов в твердом теле. Этому уровню соответствует энергия μ_0 . Поэтому его называют уровнем Ферми.

На рис. 4 показана кривая распределения свободных электронов по энергиям при $T = 0^\circ\text{K}$, являющаяся графиком функции (11). По своему характеру эта кривая совпадает с кривой рис. 1, изображающей зависимость числа квантовых состояний $z(E)$ от энергии E . Такое совпадение понятно, так как при абсолютном нуле для всех состояний, для которых $E < \mu_0$, $f_F = 1$, вследствие чего $dn(E) = 2zf_F = 2z$. При переходе через μ_0 величина f_F скачкообразно падает до нуля, поэтому и $dn(E)$ в этой точке также скачкообразно падает до нуля

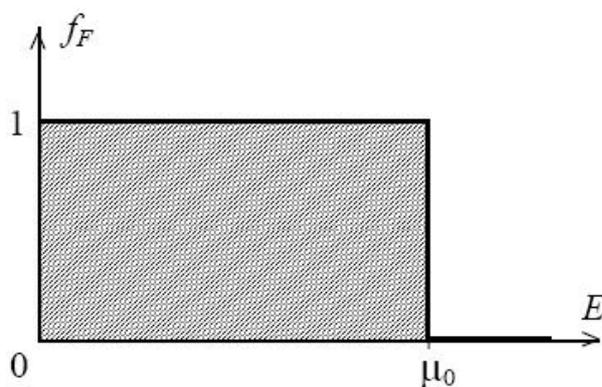


Рис. 3. График функции распределения Ферми–Дирака при $T = 0^\circ\text{K}$

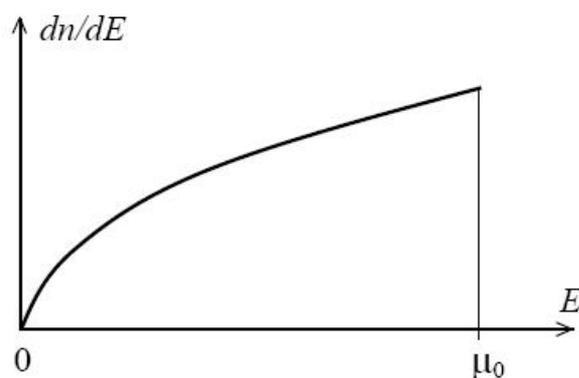


Рис. 4. Распределение электронов по энергиям при $T = 0^\circ\text{K}$

Итак, при абсолютном нуле распределение свободных электронов, описываемое функцией Ферми–Дирака, резко отличается от распределения молекул газа, подчиняющихся статистике Максвелла-Больцмана.

Согласно классической статистике при температуре абсолютного нуля прекращается тепловое движение молекул газа и их кинетическая энергия обращается в нуль. Кинетическая же энергия свободных электронов при абсолютном нуле не равна нулю. Как видно из рис. 4, она меняется от значений, близких к 0, до μ_0 ; μ_0 можно определить следующим образом. При абсолютном нуле в интервале энергий от 0 до μ_0 функция распределения $f_F = 1$ и

$$dn(E) = 2z = \frac{4\pi(2m)^{3/2} E^{1/2} dE}{h^3}.$$

Интегрируя это выражение по энергиям от 0 до μ_0 , получим общее число свободных электронов в единице объема металла:

$$n = \frac{4\pi(2m)^{3/2}}{h^3} \int_0^{\mu_0} E^{1/2} dE = \frac{4\pi(2m)^{3/2}}{h^3} \frac{2}{3} (\mu_0)^{3/2}.$$

Отсюда имеем:

$$\mu_0 = \frac{h^2}{2m} \left(\frac{3n}{8\pi} \right)^{2/3}. \quad (15)$$

Как показывает расчет, средняя энергия электрона равна $(3/5)\mu_0$:

$$\bar{E} = \frac{3}{5} \mu_0 = \frac{3}{5} \frac{h^2}{2m} \left(\frac{3n}{8\pi} \right)^{2/3}. \quad (16)$$

Подставляя в формулу (16) $h = 6,6 \cdot 10^{-34}$ Дж·сек, $m = 9 \cdot 10^{-31}$ кг, $n = 6 \cdot 10^{28}$ м⁻³, получим: $\bar{E} = 9 \cdot 10^{-19}$ Дж = 5,4 эВ. Такую энергию молекулы обычного газа имели бы при температуре в несколько десятков тысяч градусов.

Тот факт, что свойства электронного газа в металлах резко отличны от свойств обычного молекулярного газа и его распределение не подчиняется законам классической статистики, говорит, что электронный газ находится в вырожденном состоянии.

Если, однако, в функции распределения Ферми-Дирака

$$f_F = \frac{1}{\exp\left(\frac{E - \mu}{kT}\right) + 1}$$

экспоненциальный член, стоящий в знаменателе, будет значительно больше 1, то единицей в знаменателе можно пренебречь и выражение для функции распределения можно записать в следующем виде:

$$f_F \rightarrow f_B = \exp\left(\frac{\mu - E}{kT}\right) = A \cdot \exp\left(-\frac{E}{kT}\right). \quad (17)$$

Выражение (17) представляет собою функцию распределения Максвелла-Больцмана.

Таким образом, при $E - \mu \gg kT$ распределение Ферми-Дирака переходит в распределение Максвелла-Больцмана, вырождение у электронного газа снимается и он начинает следовать законам классической статистики как обычный газ. Очевидно, что это утверждение справедливо при

$$\exp\left(-\frac{\mu}{kT}\right) \gg 1. \quad (18)$$

Вычислим значение этой экспоненты.

На основании (7.11)

$$dn(E) = 2z = \frac{4\pi(2m)^{3/2} E^{1/2} dE}{h^3} \cdot f_F.$$

Если

$$\exp\left(\frac{E - \mu}{kT}\right) \gg 1,$$

то

$$f_F = \exp\left(\frac{\mu}{kT}\right) \exp\left(-\frac{E}{kT}\right), \quad dn(E) = \frac{4\pi(2m)^{3/2} E^{1/2} dE}{h^3} \exp\left(\frac{\mu}{kT}\right) \exp\left(-\frac{E}{kT}\right).$$

Интегрируя по энергии в пределах от 0 до ∞ , получим число электронов в единице объема:

$$n = 2 \exp\left(\frac{\mu}{kT}\right) \cdot \left(\frac{2\pi mkT}{h^2}\right)^{3/2},$$

Откуда

$$\exp\left(-\frac{\mu}{kT}\right) = \frac{2}{n} \left(\frac{2\pi mkT}{h^2}\right)^{3/2}.$$

Поэтому условие снятия вырождения (18) можно записать следующим образом:

$$\frac{1}{n} \left(\frac{2\pi mkT}{h^2}\right)^{3/2} \gg 1. \quad (19)$$

При выполнении обратного неравенства,

$$\frac{1}{n} \left(\frac{2\pi mkT}{h^2}\right)^{3/2} \ll 1, \quad (20)$$

электронный газ будет вырожденным и применять к нему статистику Максвелла — Больцмана нельзя.

Из выражения (20) следует, что вырождению газа электронов способствуют малая масса его частиц, высокая их концентрация и низкая температура. В качестве примера определим состояние электронного газа в металлах при температурах $T = 300^\circ\text{K}$ (комнатная

температура) и $T = 2000^\circ\text{K}$. Приняв $m = 9 \cdot 10^{-31}$ кг и $n = 6 \cdot 10^{28}$ м³ и подставив в соотношение (20) численные значения остальных констант, получим:

для $T = 300^\circ\text{K}$

$$\frac{1}{n} \left(\frac{2\pi mkT}{h^2} \right)^{3/2} \approx 2 \cdot 10^{-4},$$

а для $T = 2000^\circ\text{K}$

$$\frac{1}{n} \left(\frac{2\pi mkT}{h^2} \right)^{3/2} \approx 3 \cdot 10^{-3}.$$

В обоих случаях получилась величина, значительно меньшая 1. Поэтому во всех металлах электронный газ оказывается вырожденным даже при температурах в несколько тысяч градусов. Это вызвано малой величиной массы электрона и очень высокой концентрацией электронного газа в металлах. При тех же условиях протонный газ был бы газом невырожденным только потому, что масса протона примерно в 2000 раз больше массы электрона [для протонного газа левая часть соотношения (19) оказывается равной 2700 для $T = 2000^\circ\text{K}$ и 15 для $T = 300^\circ\text{K}$].

В полупроводниках концентрация электронного газа много меньше, чем в металлах, и составляет $10^{18} \div 10^{24}$ м⁻³. При комнатной температуре левая часть соотношения (19) оказывается в этом случае равной $10^{-6} \div 10^{-1}$. Отсюда видно, что в полупроводниках электронный газ вследствие своей низкой концентрации является в большинстве случаев (но не всегда) газом невырожденным и подчиняется классической статистике Максвелла–Больцмана.

Можно оценить ту температуру T_k , при которой вырождение электронного газа в металлах снимается. Грубую оценку T_k можно получить, приравняв единице левую часть (19). Тогда будем иметь:

$$T_k = \frac{n^{2/3} h^2}{2\pi mk}.$$

Подстановка в эту формулу n , h , k , m ($h = 6,6 \cdot 10^{-34}$ дж·сек, $m = 9 \cdot 10^{-31}$ кг, $n = 6 \cdot 10^{28}$ м³) дает для T_k величину порядка $5 \cdot 10^4$ град.

Следовательно, только выше нескольких десятков тысяч градусов электроны металла (если бы таковой мог существовать при этих температурах в конденсированном состоянии) удовлетворяли бы законам обычных газов и подчинялись бы статистике Максвелла–Больцмана.

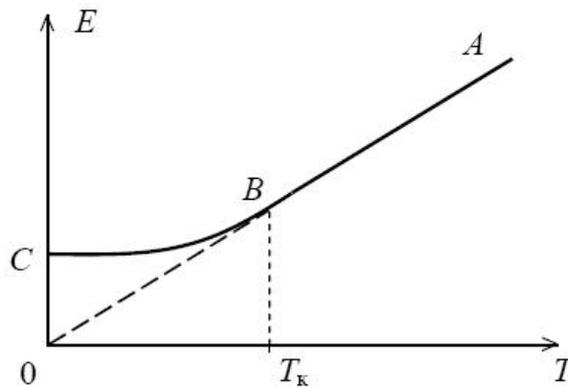


Рис. 5. Зависимость энергии электронного газа от температуры

На рис. 7.5 показана зависимость энергии электронного газа от температуры. При температурах выше T_k электронный газ является невырожденным и его энергия, как и энергия обычного газа, меняется пропорционально T (участок AB). Ниже T_k электронный газ становится вырожденным, и его энергия и скорость движения электронов перестают практически зависеть от температуры (участок BC).

4. Влияние температуры на распределение электронов, теплоемкость электронного газа

При температуре $T = 0$ К свободные электроны металла занимают все энергетические уровни вплоть до уровня Ферми μ_0 . Уровни, расположенные выше уровня Ферми, остаются совершенно свободными. Графиком функции распределения является ломаная линия (см. рис. 3), удовлетворяющая следующему уравнению:

$$f = \begin{cases} 1 & \text{при } E \leq \mu_0 \\ 0 & \text{при } E > \mu_0 \end{cases} \quad (21)$$

С повышением температуры электроны подвергаются тепловому возбуждению и переходят на более высокие энергетические уровни, вследствие чего меняется характер распределения их по состояниям. Однако в интервале температур, в котором энергия теплового движения kT остается значительно меньшей энергии Ферми μ_0 , т. е. в котором выполняется условие

$$kT \ll \mu_0, \quad (22)$$

тепловому возбуждению могут подвергаться электроны лишь узкой полосы порядка kT , непосредственно расположенной вблизи уровня Ферми (рис. 6). Электроны более глубоких

уровней остаются практически незатронутыми, так как энергия теплового движения kT недостаточна для их перевода выше уровня Ферми.

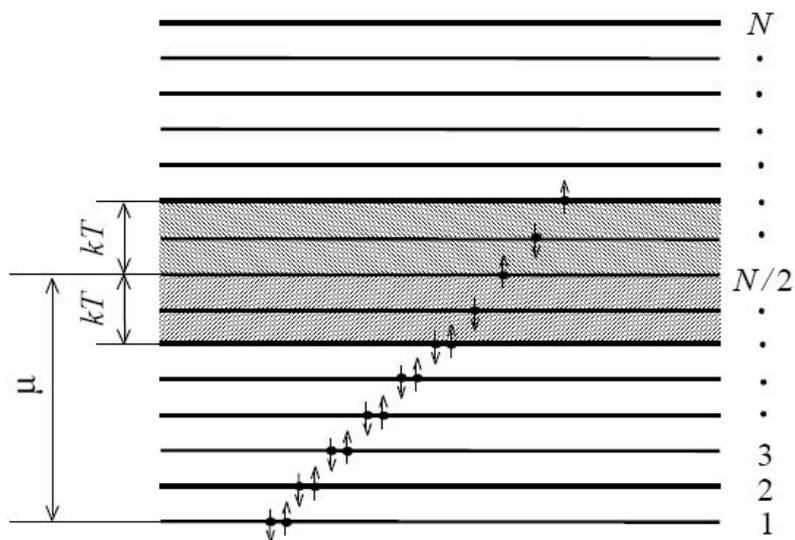


Рис. 6. Тепловое возбуждение электронов

Таким образом, в результате теплового возбуждения устанавливается новое распределение электронов по состояниям. На рис. 7 пунктиром показана кривая распределения электронов при $T = 0$ К, сплошной линией – кривая распределения при температуре, отличной от абсолютного нуля, но удовлетворяющей условию (22). Из рис. 7 видно, что эти кривые отличаются друг от друга лишь характером спада правых частей: в то время как при $T = 0$ К спад кривой в точке μ происходит скачкообразно по вертикальной прямой, при $T > 0$ этот спад протекает по плавной кривой ABC , асимптотически сливающейся с осью абсцисс.

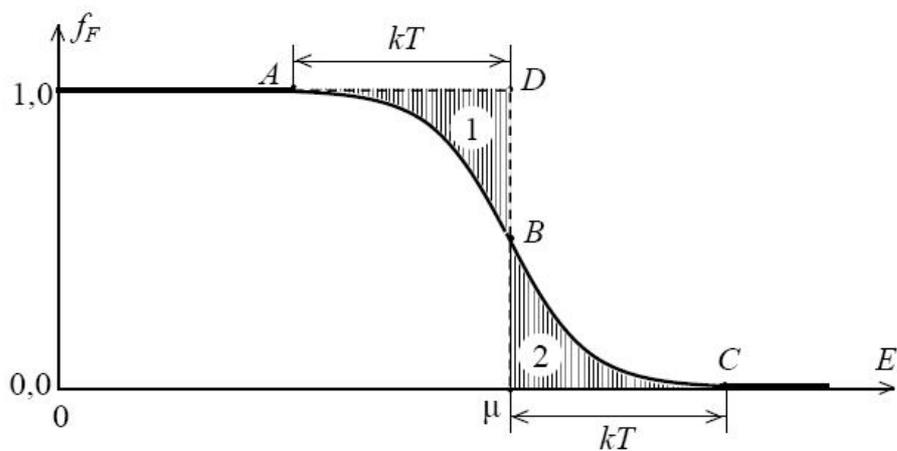


Рис. 7. График функции распределения Ферми-Дирака при $T > 0$ К

На рис. 7 видно, что при температурах, удовлетворяющих условию (7.22), тепловому возбуждению подвергаются электроны лишь уровней, тесно примыкающих к уровню Ферми. Распределение же электронов, располагающихся на более глубоких уровнях, остается таким же, как и при абсолютном нуле. Важно заметить, что плавная кривая ABC , которая в области низких энергий E сливается с горизонтальной прямой $f_F = 1$, в области высоких энергий переходит в кривую, соответствующую распределению Максвелла. На рис. 7 заштрихованные площадки пропорциональны числу электронов, покидающих состояния с энергией $E < \mu$ (площадка 1) и переходящему на уровни, расположенные выше уровня Ферми μ (площадка 2). По величине эти площадки равны друг другу, так как соответствуют одному и тому же числу электронов.

Произведем приближенную оценку этого числа, которое обозначим как ΔN . В интервале энергий от 0 до μ располагается $N/2$ энергетических уровней, где N – число свободных электронов в одном киломоле металла. Упрощая задачу, мы можем считать, что эти уровни отстоят друг от друга на одинаковом расстоянии, именно на расстоянии

$$\Delta E = \frac{\mu}{(N/2)}.$$

Термическому возбуждению подвергаются электроны энергетической полосы шириной kT , расположенной ниже уровня Ферми (рис. 7.6). В этой полосе размещается

$$\frac{kT}{\Delta E} = \frac{kT}{2\mu} N$$

уровней, на каждом из которых может находиться не более двух электронов. До теплового возбуждения все они занимали места ниже уровня Ферми. Полагая, что в результате теплового возбуждения выше уровня Ферми оказывается не более половины этих электронов, получим следующее приближенное соотношение для ΔN :

$$\Delta N \approx \frac{kT}{2\mu} N. \quad (23)$$

При комнатной температуре $kT \approx 0,025$ эВ, $\mu = 3 \div 10$ эВ, поэтому

$$\frac{\Delta N}{N} \cdot 100\% < 1\%.$$

Таким образом, во всем диапазоне температур, в котором электронный газ в металле является **вырожденным**, его распределение мало отличается от распределения при абсолютном нуле. Тепловому возбуждению подвергается лишь незначительная доля электронов, располагающихся у уровня Ферми. При комнатной температуре эта доля составляет меньше 1% от общего числа электронов проводимости. Так как при температурах, при которых возможно существование конденсированного состояния металла, электронный газ в

нем является всегда вырожденным, то рассмотренные закономерности распределения электронов в металлах остаются справедливыми практически во всех случаях.

То обстоятельство, что при обычных температурах термическому возбуждению подвергается лишь незначительная часть свободных электронов, делает понятным, почему теплоемкость электронного газа оказывается чрезвычайно малой по сравнению с теплоемкостью решетки. При термическом возбуждении каждый электрон поглощает энергию, по порядку величины равную kT . Энергия, поглощаемая всем электронным газом, равна произведению kT на число электронов ΔN , испытывающих термическое возбуждение:

$$\Delta E \approx \Delta N \cdot kT = N kT \frac{kT}{2\mu}. \quad (24)$$

Для одного киломоля металла, отдающего в электронный газ по одному электрону от каждого атома, N равно числу Авагадро, а произведение Nk равно универсальной газовой постоянной R . Поэтому энергия, поглощенная электронным газом в результате теплового возбуждения части электронов, равна:

$$\Delta E = RT \frac{kT}{2\mu}. \quad (25)$$

Теплоемкость электронного газа при постоянном объеме равна:

$$C_V = \frac{d(\Delta E)}{dT} = R \frac{kT}{\mu}. \quad (26)$$

Теплоемкость же невырожденного одноатомного газа, подчиняющегося классической статистике, равна:

$$C_{V(\text{кл})} = \frac{3}{2} R.$$

Поделив (26) на $C_{V(\text{кл})}$, получим:

$$\frac{C_V}{C_{V(\text{кл})}} = \frac{2}{3} \frac{kT}{\mu}. \quad (27)$$

Из соотношения (27) видно, что теплоемкость вырожденного электронного газа в металле во столько раз меньше теплоемкости невырожденного одноатомного газа, во сколько раз kT меньше μ . Для комнатных температур $kT/\mu \approx 0,01$, поэтому $C_V \approx 0,01 \cdot C_{V(\text{кл})}$. Это хорошо согласуется с опытом и снимает трудность, существовавшую в классической теории металлов при трактовке их теплоемкости.

Теперь рассмотрим кратко зависимость энергии Ферми от температуры. С повышением температуры μ , представляющий собою химический потенциал, отнесенный к одному электрону,

$$\mu = \frac{U - TS + pV}{N},$$

уменьшается, и уровень Ферми смещается влево от μ_0 (см. рис. 7). Количественное соотношение, выражающее зависимость μ от T , можно получить следующим образом. Интегрируя (11) по энергиям в пределах от 0 до ∞ , мы получим общее число свободных электронов в единице объема металла:

$$n = \frac{4\pi(2m)^{2/3}}{h^3} \int_0^{\infty} \frac{\sqrt{E}}{e^{(E-\mu)/kT} + 1} dE.$$

Интеграл, стоящий в правой части, в общем виде не берется. Приближенное вычисление его для области низких температур, в которой выполняется условие (22), приводит к следующему результату:

$$n \approx \frac{4\pi(2m)^{2/3}}{h^3} \cdot \frac{2}{3} \mu^{3/2} \left[1 + \frac{\pi^2}{8} \left(\frac{kT}{\mu} \right)^2 \right].$$

Полагая в малом слагаемом $\mu \approx \mu_0$ и решая уравнение относительно μ , получим:

$$\mu \approx \mu_0 \left[1 - \frac{\pi^2}{12} \left(\frac{kT}{\mu_0} \right)^2 \right]. \quad (28)$$

Так как вплоть до точки плавления металла kT остается значительно меньше μ_0 , то уменьшение μ с повышением T оказывается настолько малым, что во многих случаях им можно пренебречь и энергию Ферми при любой температуре считать равной μ_0 . Однако, в контактных явлениях, протекающих в месте соприкосновения разнородных проводников, зависимость (7.28) играет важную роль.

5. Квантовая модель свободных электронов

Первой удачной попыткой объяснения электрических и магнитных свойств твердых тел, в первую очередь металлов, явилась теория свободных электронов. Она исходит из представления, что металл содержит свободные электроны, способные перемещаться по всему объему. Положительный заряд, компенсирующий заряд электронов, распределен в металле с равномерной плотностью и создает поле с постоянным положительным потенциалом V_0 . Во многих задачах этот потенциал полагается равным нулю и свободные электроны наделяются лишь кинетической энергией подобно частицам идеального газа. У поверхности металла потенциал резко падает от V_0 до 0, а потенциальная энергия электрона увеличивается с $U_0 = -eV_0$ до 0 (рис. 8). Это позволяет рассматривать металл как потенциальную яму (потенциальный ящик), заполненную свободными электронами.

Решение задачи о движении микрочастицы в потенциальной яме (см. выше) могут быть использованы и для описания движения свободных электронов в металле. Именно, если внутренний потенциал металла положить равным нулю, что не нарушает общности решения задачи, то уравнение Шредингера для свободного электрона будет совпадать с уравнением:

$$\Delta\psi + k^2\psi = 0, \quad (29)$$

где

$$k^2 = \frac{8\pi^2 m E}{h^2}. \quad (30)$$

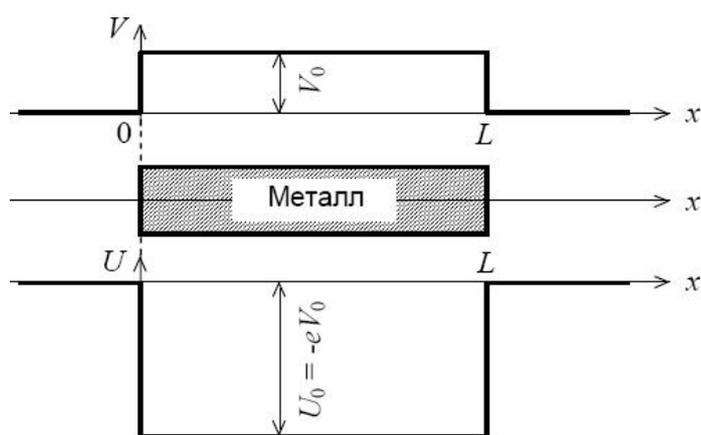


Рис. 8. Внутренний потенциал металла V_0 и потенциальная энергия электрона в металле U

В этих равенствах k – волновой вектор, E – энергия электрона. График зависимости E от k представляет собой параболу. Так как электрон не может свободно покинуть металл, то волновая функция на его границах (в точках $x = 0$ и $x = L$) должна обращаться в нуль. Поэтому краевыми условиями задачи являются:

$$\psi(0) = 0, \quad \psi(L) = 0. \quad (31)$$

Этим условиям удовлетворяют следующие решения уравнения (29):

$$\psi_n(x) = A \sin(k_n x), \quad (32)$$

где

$$A = \sqrt{\frac{2}{L}}, \quad k_n = n \frac{\pi}{L}, \quad n = 1, 2, 3, \dots \quad (33)$$

Энергетический спектр свободных электронов в металле определяется равенством:

$$E_n = n^2 \frac{h^2}{8mL^2}. \quad (34)$$

Из (34) видно, что этот спектр является дискретным.

В теории твердого тела вместо краевых условий (31) часто используются так называемые периодические краевые условия: твердое тело предполагается бесконечно большим, а его характеристики, рассматриваемые как функции координат, считаются периодическими с периодом, равным L . Иначе говоря, бесконечно большое твердое тело мысленно разбивается на кубики с ребром L и считается, что закон изменения ψ повторяется в каждом кубике. Для тела конечных размеров допускается, что в простейшем случае оно также представляет собою куб с ребром L , в котором ψ имеет трехмерную периодичность с периодом L ; при этом никаких условий, ограничивающих положение электрона внутри тела, не налагается.

Движение свободных электронов в бесконечно большом теле описывается бегущими волнами:

$$\psi = Ae^{ikx},$$

являющимися решением уравнения Шредингера для свободных микрочастиц. Из требования периодичности волновой функции следует, что

$$\psi(x+L) = \psi(x), \quad Ae^{ik(x+L)} = Ae^{ikx}, \quad e^{ikL} = 1. \quad (35)$$

Последнее равенство удовлетворяется лишь при следующих значениях k :

$$k_n = n \frac{2\pi}{L}, \quad n = 0, \pm 1, \pm 2, \pm 3, \dots \quad (36)$$

Уровни энергии определяются равенством:

$$E_n = n^2 \frac{h^2}{2mL^2}. \quad (37)$$

Таким образом, периодические краевые условия приводят также к дискретному энергетическому спектру свободных электронов в металле. Как было отмечено выше, распределение этих электронов по уровням описывается статистикой Ферми–Дирака, если электронный газ является вырожденным, и статистикой Максвелла–Больцмана, если газ является невырожденным.

Модель свободных электронов является весьма грубым приближением к реальной картине твердого тела. Она не учитывает структуры твердого тела, дискретного характера распределения положительных зарядов в нем, заменяя реальные силы взаимодействия электронов с узлами решетки идеальным полем с постоянным потенциалом. Несмотря на это, теория свободных электронов позволила выяснить основные свойства электронного газа в металле, установить температурную зависимость электропроводности и теплопроводности,

правильно оценить их абсолютную величину, описать широкий круг явлений, связанный с эмиссией электронов, термоэлектрическими явлениями, магнитными свойствами металлов и др. Некоторые приложения этой теории будут рассмотрены нами в последующих главах. Но теория свободных электронов оказалась бессильной при рассмотрении свойств твердых тел, зависящих от их внутренней структуры. Она не дает ответа даже на такой вопрос, почему одни тела являются проводниками, другие – изоляторами.

6. Квантовая модель связанных электронов (слабая и сильная связь)

Дальнейшим этапом в развитии электронной теории является зонная теория твердых тел. В зонной теории твердое тело рассматривается как тело кристаллическое, имеющее правильное внутреннее строение, и делается попытка описания поведения в таком теле электронов. Существует два приближенных метода такого описания.

1. Приближение, исходящее из связанных электронов (приближение сильной связи). В основе этого приближения лежит допущение, что энергия связи электронов с атомами твердого тела значительно больше их кинетической энергии перемещения сквозь решетку. Электроны связаны со своими атомами настолько сильно, что основное время находятся около них и лишь время от времени переходят от одного атома к другому (сильная связь). Поэтому в основу метода кладутся волновые функции изолированных атомов (связанных электронов) и из них строятся волновые функции электронов твердого тела.

2. Приближение, исходящее из свободных электронов (приближение слабой связи). Этот метод исходит из прямо противоположного предположения. Считается, что энергия взаимодействия электрона с решеткой мала по сравнению с его кинетической энергией (слабая связь). Это позволяет рассматривать электрон как практически свободный и пользоваться решением уравнения Шредингера для свободных электронов, исправив его учетом слабого возмущающего действия периодического поля решетки.

Рассмотрим кратко результаты, к которым приводят оба метода.

Приближение сильной связи.

Рассмотрим влияние электрического и магнитного полей на энергетические уровни свободных атомов. Энергетическое состояние атома удобно изображать посредством энергетических уровней (рис. 9).

Относительная энергия	Обозначение уровня	Кратность вырождения	
=====	6d	10	6
	7p	6	3
	7s	2	1
=====	4f	14	7
	5d	10	5
	6p	6	3
	6s	2	1
=====	4d	10	5
	5p	6	3
	5s	2	1
=====	3d	10	5
	4p	6	3
	4s	2	1
=====	3p	6	3
	3s	2	1
=====	2p	6	3
	2s	2	1
=====	1s	2	1

Рис. 9. Диаграмма уровней энергии атома

Так как в свободном атоме энергия электрона определяется лишь главным (n) и орбитальным (l) квантовыми числами и не зависит от двух других квантовых чисел (магнитного m_l и спинового s), то каждый энергетический уровень оказывается вырожденным: ему соответствует не одно, а $2(2l+1)$ различных электронных состояний, отличающихся друг от друга числами m_l или s . Поэтому кратность вырождения уровней равна $2(2l+1)$. В третьем столбце рис. 9 приведена кратность вырождения каждого уровня.

При помещении атома во внешнее электрическое или магнитное поле энергетические уровни претерпевают двоякое изменение. Они, во-первых, несколько смещаются относительно своих первоначальных положений (рис. 10) и, во-вторых, у них снимается вырождение: каждый уровень расщепляется на ряд подуровней в соответствии с кратностью своего вырождения. Так как последняя равна $2(2l + 1)$, то следовало бы ожидать, что каждый уровень расщепится на $2(2l + 1)$ подуровней. Однако вследствие того, что спиновое квантовое число оказывает весьма слабое влияние на энергию электрона, состояния,

различающиеся лишь направлением спина, обладают практически одной и той же энергией. Поэтому энергетические уровни расщепляются не на $2(2l + 1)$, а лишь на $(2l + 1)$ подуровней. На каждом таком подуровне может находиться одновременно 2 электрона с противоположно направленными спинами.

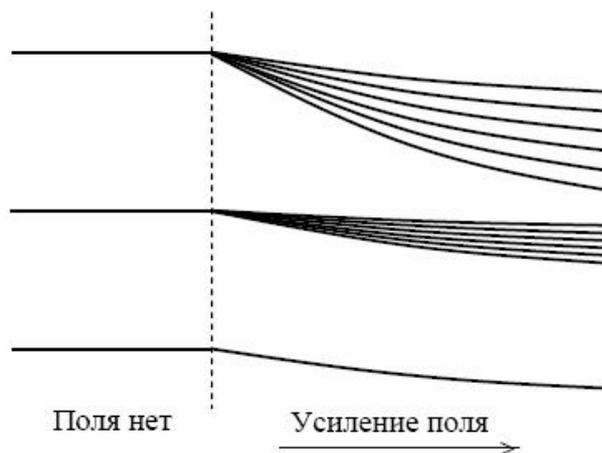


Рис. 10. Влияние внешнего поля на энергетические уровни атома

На различные энергетические уровни атома влияние внешнего поля не одинаково (рис. 10). Уровни внутренних электронов, сильно взаимодействующих с ядром, испытывают настолько слабое расщепление, что им можно пренебречь; по мере же перехода ко все более внешним электронам энергия взаимодействия их с ядром уменьшается и влияние внешнего поля увеличивается. Наиболее сильное изменение под влиянием поля претерпевают энергетические уровни внешних валентных электронов, сравнительно слабо связанных с ядром.

Теперь рассмотрим изменения в состояниях электронов при сближении атомов.

В твердом теле расстояния между атомами настолько малы, что каждый из них оказывается в достаточно сильном электромагнитном поле, созданном соседними атомами, с которыми он взаимодействует. Для того чтобы проследить, какое влияние оказывает это взаимодействие на энергетические уровни атомов, рассмотрим следующий идеализированный процесс. Расположим N атомов натрия в виде пространственной решетки, свойственной кристаллу натрия, но на столь больших расстояниях r друг от друга, что взаимодействием между ними можно пренебречь. Каждый такой атом можно рассматривать тогда как свободный и энергетическое состояние электронов в нем считать таким же, как и в отдельно взятом изолированном атоме. На рис. 11 внизу показаны энергетические схемы двух изолированных атомов натрия.

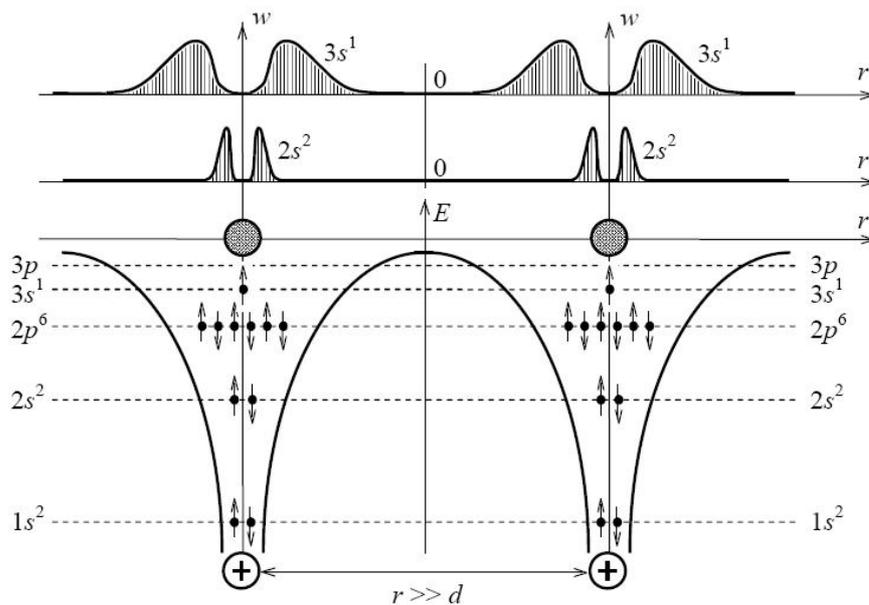


Рис. 11. Энергетическая схема двух изолированных атомов натрия

Каждый атом изображен в виде плоской веретенообразной потенциальной ямы, ограниченной потенциальными кривыми, выражающими зависимость потенциальной энергии электрона от расстояния его от ядра. Внутри ямы проведены энергетические уровни, на которых размещаются электроны: $1s$, $2s$, $2p$, $3s$ и т. д. На уровнях $1s$ и $2s$ располагается по два электрона, на уровне $2p$ – шесть электронов, на уровне $3s$ – один электрон; уровни, лежащие выше $3s$, являются свободными. Атомы отделены друг от друга потенциальными барьерами толщиной r . Высота барьеров для электронов, находящихся на разных уровнях, различна: она равна расстоянию от этих уровней до нулевого уровня. Потенциальный барьер препятствует свободному переходу электронов от одного атома к другому. В верхней части рис. 11 показана качественная картина распределения вероятности обнаружения электронов $3s$ и $2s$ на том или ином расстоянии от ядра. Максимумы этих кривых соответствуют примерно положению боровских орбит электронов $3s$ и $2s$.

Подвергнем теперь решетку медленному однородному сжатию, не нарушающему ее симметрии. По мере сближения атомов взаимодействие между ними растет и на расстояниях r , равных параметру решетки d , достигает нормальной своей величины. Из рис. 12 видно, что потенциальные кривые, разграничивающие соседние атомы, частично накладываются друг на друга и дают результирующую потенциальную кривую 1–2, проходящую ниже нулевого энергетического уровня. Это означает, что сближение атомов вызывает уменьшение не только ширины потенциального барьера, препятствующего переходу электронов из одного атома в другой, но и его высоты. Замечательным является то обстоятельство, что при $r = d$ высота этого барьера оказывается даже несколько ниже первоначального положения

энергетического уровня валентных электронов (уровня $3s$). Поэтому эти электроны приобретают способность практически беспрепятственно переходить от одного атома к другому и свободно странствовать по кристаллу. В верхней части рис. 12 показано изменение волновых функций валентных электронов при сближении атомов. При $r = d$ эти функции перекрываются настолько, что дают электронное облако практически равномерной плотности. Этому соответствует состояние полного обобществления валентных электронов, при котором вероятность обнаружения их в любом месте решетки является одной и той же.

Из рис. 12 видно, что электронные облака внутренних оболочек атома практически не перекрываются. Это свидетельствует о том, что состояние внутренних электронов в кристалле остается почти таким же, как и в изолированных атомах. Уменьшение потенциальной энергии электронов при сближении атомов приводит к смещению их энергетических уровней. На рис. 12 пунктиром показано положение энергетических уровней в изолированных атомах, сплошными линиями – положение этих уровней в кристалле. В соответствии с тем, что наибольшее изменение потенциальной энергии происходит у электронов верхних уровней, эти уровни смещаются наиболее сильно; уровни же внутренних электронов, потенциальная энергия которых почти не меняется при образовании кристаллов, остаются на своих местах.

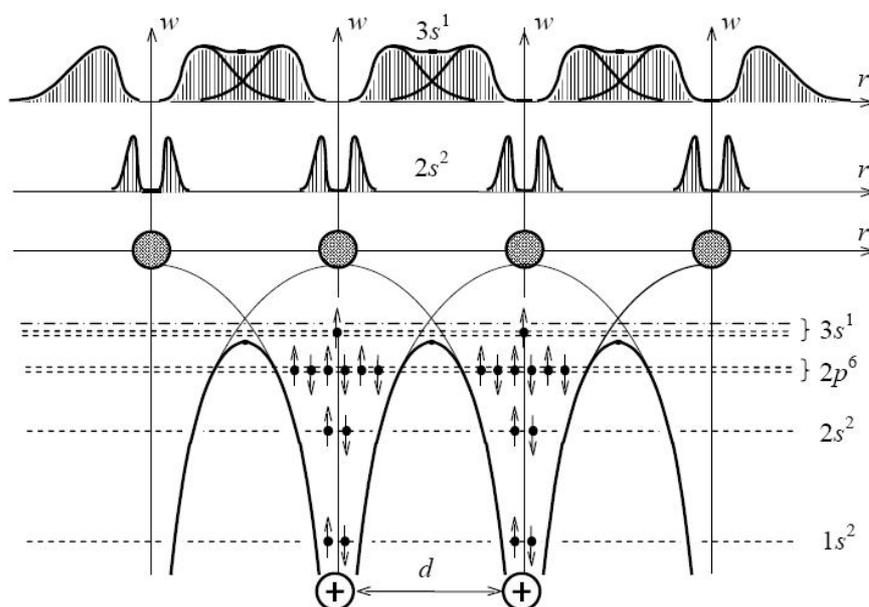


Рис. 12. Энергетическая схема двух атомов натрия, сближенных на расстояние, равное параметру решетки кристалла натрия

Рассмотрим, как энергетические уровни свободных атомов превращаются в энергетические зоны при образовании кристалла.

Взаимодействие атомов при образовании кристаллической решетки, приводит к еще одному важному результату – к расширению энергетических уровней атомов и превращению их в кристалле в энергетические зоны.

Подобное расширение является следствием волновых свойств электронов и непосредственно связано с уменьшением степени локализации их в кристалле по сравнению с изолированными атомами. В атомах время жизни электрона в возбужденном состоянии $\tau \sim 10^{-8}$ сек. В соответствии с принципом неопределенностей ширина энергетических уровней, отвечающая этим состояниям, равна:

$$\Delta E \approx \frac{h}{\tau} \approx 10^{-7} \text{ эВ.}$$

Она определяет естественную ширину спектральных линий, испускаемых атомом.

В кристалле все электроны приобретают способность переходить от одного атома к другому. Этот переход осуществляется путем туннельного просачивания сквозь потенциальный барьер, разделяющий атомы. Наличие таких переходов уменьшает степень локализации электронов на определенных атомах и приводит к большей или меньшей неопределенности в значении их энергий, т. е. к размытию уровней энергии и превращению их в полосы или зоны. Для определения порядка ширины этих зон заменим потенциальные барьеры, разделяющие атомы, простейшими барьерами прямоугольной формы (рис. 13).

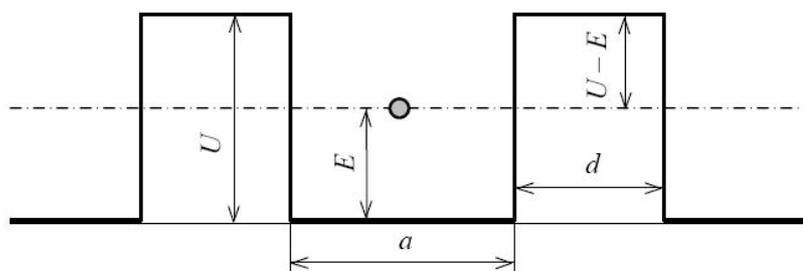


Рис. 13. Электрон в прямоугольном потенциальном барьере конечной толщины

Прозрачность такого барьера D определяется формулой:

$$D = D_0 \exp\left(-\frac{4\pi\sqrt{2m(U-E)}}{h}d\right),$$

где $(U - E)$ – высота барьера, d – его толщина, D_0 – множитель, по порядку величины близкий к единице. Если ширина потенциальной ямы, в которой находится электрон, равна

a , а скорость его движения равна v , то за 1 сек электрон подойдет к барьеру v/a раз. Умножая v/a на D , получим «частоту» перехода электрона на соседний атом:

$$f = \frac{v}{a} \exp\left(-\frac{4\pi\sqrt{2m(U-E)}d}{h}\right). \quad (38)$$

Величина τ , обратная f , выражает среднее время пребывания' электрона у определенного атома. Так как $a = 10^{-8}$ см, $v \approx 10^8$ см/сек, то $(v/a) \approx 10^{16}$ сек⁻¹. Для изолированных атомов d велико. Примем его равным 30 А. Это соответствует среднему расстоянию между молекулами газа при нормальных условиях. Положив $(U - E) = 10$ эВ (энергия ионизации атомов натрия), получим для τ величину $\approx 10^{20}$ лет. Следовательно, переход электрона от атома к атому может осуществляться в этом случаев среднем один раз за 10^{20} лет. Это указывает на то, что в системе изолированных атомов вероятность туннельного перехода даже для валентных электронов столь ничтожна, что они практически никогда не покидают своих атомов. Вместе с этим при столь огромных значениях % туннельный эффект не может оказать (в соответствии с принципом неопределенностей) никакого влияния на ширину энергетических уровней системы атомов; как и в отдельно взятом атоме она определяется средним временем жизни возбужденных состояний и по порядку величины равна 10^{-7} эВ.

В кристалле атомы сближены настолько, что электронные облака валентных электронов перекрываются (рис. 12). Примем толщину барьера для таких электронов $d \approx 1$ А. Подставив это значение d в формулу (38), получим $f \approx 10^{15}$ сек⁻¹, $\tau \approx 10^{-15}$ сек. При столь высокой частоте перехода валентных электронов от атома к атому теряет смысл говорить о принадлежности их определенным атомам. Электроны «обобществляются», образуя «электронный газ».

Таким образом, туннельный эффект в кристалле доводит среднее время пребывания валентного электрона в определенном узле решетки до величины $\tau \approx 10^{-15}$ сек. В соответствии с принципом Гейзенберга неопределенность в значении энергии таких электронов равна $\Delta E \approx 1$ эВ. Это означает, что энергетический уровень валентных электронов, имеющий в изолированном атоме ширину $\approx 10^{-7}$ эв, превращается в кристалле в энергетическую зону шириною порядка единиц эВ.

Иная картина имеет место для электронов внутренних оболочек атома. Вследствие резкой зависимости ширины и высоты потенциального барьера от расстояния (рис. 12) вероятность туннельного перехода внутренних электронов от атома к атому оказывается очень низкой. Так, для электронов 1s атома натрия ,высота барьера $(U - E) = 1000$ эВ, ширина барьера $d \approx 3$ А. Подстановка этих значений в формулу (38) приводит к следующим

величинам: $f \approx 10^{-27} \text{ сек}^{-1}$ и $\tau \approx 10^{20}$ лет. Следовательно, электроны $1s$ остаются в кристалле натрия столь же жестко локализованными на определенных узлах решетки, как и в изолированных атомах. Энергетические уровни этих электронов в кристалле такие же узкие, как и в отдельно взятом атоме. По мере перехода к более внешним оболочкам атома высота и ширина потенциального барьера уменьшаются, вероятность туннельного перехода электронов увеличивается, вследствие чего растет ширина энергетических зон. На рис. 14 показано изменение энергетических уровней атомов натрия по мере их сближения. Справа приведены уровни изолированного атома натрия, слева – образование зон, обусловленное расширением уровней при уменьшении расстояния r между атомами; d – межатомное расстояние в кристалле натрия. При $r = d$ уровни $1s$ и $2s$ не претерпевают еще заметного расширения, в то время как уровень $3s$ сильно расширился. Еще большему расширению подвергаются более высокие уровни, соответствующие возбужденным состояниям атома – уровни $3p$, $3d$, $4s$ и т. д. Уровни $3s$ и $3p$ расширяются настолько сильно, что перекрываются. Подобное явление перекрытия соседних энергетических зон наблюдается у многих твердых тел и играет большую роль в определении их электрических свойств.

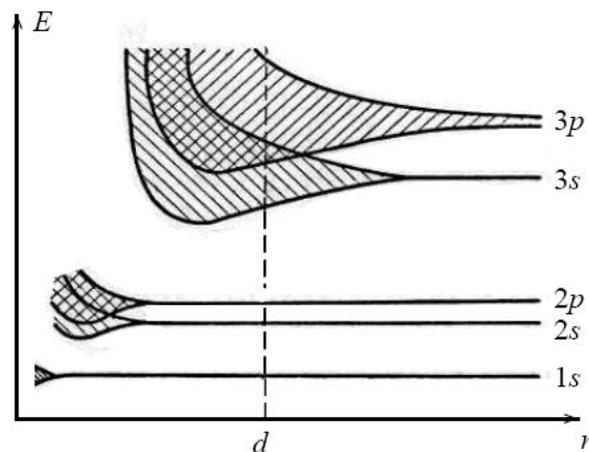


Рис. 14. Расширение энергетических уровней при сближении атомов натрия

Скажем несколько слов о внутренней структуре энергетических зон кристалла.

В системе, состоящей из N изолированных атомов, каждый уровень, не вырожденный в атоме, повторяется N раз. Иначе говоря, уровень, не вырожденный в атоме, N -кратно вырожден в системе, состоящей из N изолированных атомов. Такое вырождение называется перестановочным. Оно выражает тот факт, что электрон с данной энергией может находиться на любом из N атомов системы. При постепенном сближении атомов каждый из

них попадает во все более возрастающее электромагнитное поле своих соседей, с которыми он взаимодействует. Это взаимодействие приводит к расширению энергетических уровней и снятию вырождения. В результате каждая энергетическая зона, образовавшаяся из невырожденного атомного уровня, оказывается состоящей из N близко прилегающих друг к другу подуровней.

Следует указать, что расщепление уровня на подуровни является общим свойством любой связанной системы, состоящей из тождественных элементов. Так, если взять N одинаковых шариков, подвешенных на одинаковых пружинах, то, будучи несвязанными между собою, они колеблются с одной и той же частотой γ_0 . При установлении связи между шариками образуется связанная система, состоящая из N одинаковых элементов. Такая система имеет уже N собственных частот колебаний, сгруппированных около первоначальной частоты γ_0 . Таким образом происходит расщепление уровня, соответствующего частоте γ_0 , на N подуровней.

Если энергетический уровень в атоме имеет $(2l+1)$ -кратное вырождение, то соответствующая ему энергетическая зона в кристалле будет состоять из $(2l+1)N$ подуровней. Так, трехкратно вырожденный уровень p дает зону, состоящую из $3N$ подуровней, пятикратно вырожденный уровень d – зону, состоящую из $5N$ подуровней и т. д. Расстояние между подуровнями в зоне для кристаллов обычных размеров очень мало. В кристалле с объемом 1 см^3 содержится $N \approx 10^{22}$ атомов. При ширине зоны порядка 1 эВ расстояние между уровнями в ней равно 10^{-22} эВ. Это расстояние столь ничтожно, что зоны можно считать практически непрерывными. Однако тот факт, что число уровней в зоне является все-таки *конечным*, играет важную роль в определении характера распределения электронов по состояниям.

Из рис. 14 видно, что энергетические зоны отделены друг от друга зонами запрещенных энергий. Ширина этих зон измеряется также единицами электрон-вольт. Для перехода электрона из нижней энергетической зоны в ближайшую верхнюю зону он должен обладать энергией, достаточной для преодоления запрещенной зоны. Эту энергию называют энергией активации.

Лекция 3. Методы изучения электротранспортных свойств материалов.

1. Фундаментальная система уравнений, полученная на базе зонной теории твердых тел

Для различных материалов, используемых в твердотельной электронике, одним из основных электротехнических параметров является проводимость. Она пренебрежимо мала в диэлектриках, имеет средние значения для полупроводников и достаточно высока для металлов. При одинаковых внешних условиях, металлы и полупроводники отличаются зависимостью проводимости от температуры. В металлах эта зависимость определяется равенством:

$$\sigma(T) = \sigma(T_0)[1 - \alpha(T - T_0)], \quad (1)$$

где T_0 – некоторая стартовая температура, T – текущее значение температуры, получаемое в процессе измерения ($T > T_0$), а α – температурный коэффициент проводимости. В полупроводниках зависимость проводимости от температуры в общем случае может оказаться очень сложной функцией - все зависит от типа и количества различных примесных атомов в них. В наиболее простом случае бездефектных и беспримесных полупроводников температурная зависимость проводимости приближенно описывается экспоненциальной функцией:

$$\sigma(T) \approx \sigma_0 e^{-\frac{\Delta E}{kT}}, \quad (2)$$

где σ_0 – константа, температурной зависимостью которой пренебрегают, ΔE – величина, имеющая размерность энергии и называемая энергией температурной активации проводимости (ее обычно указывают в эВ), k – постоянная Больцмана, равная 8.614210^{-5} эВ·К⁻¹, T – абсолютная температура полупроводника (в градусах К). После логарифмирования левой и правой частей выражения (1.2) получается новое приближенное равенство,

$$\ln[\sigma(T)] \approx \sigma_0 - \frac{\Delta E}{kT} = \sigma_0 - \frac{\Delta E}{k} \frac{1}{T}, \quad (3)$$

которое представляет линейную зависимость величины $\ln[\sigma(T)]$ от $1/T$. График этой зависимости – прямая. На рис. 1 представлены графически экспериментальные зависимости подобного типа, определенные для трех твердых тел – медного проводника (Cu) и германиевого (Ge) и кремниевого (Si) полупроводников. Они позволят оценить значения проводимости этих материалов при различных температурах и определить величины их энергий активации. Поскольку наклоны прямых линий, представляющих графики

зависимостей (3) для германия и кремния, определяются значениями ΔE , то для их определения достаточно произвести простейшие вычисления.

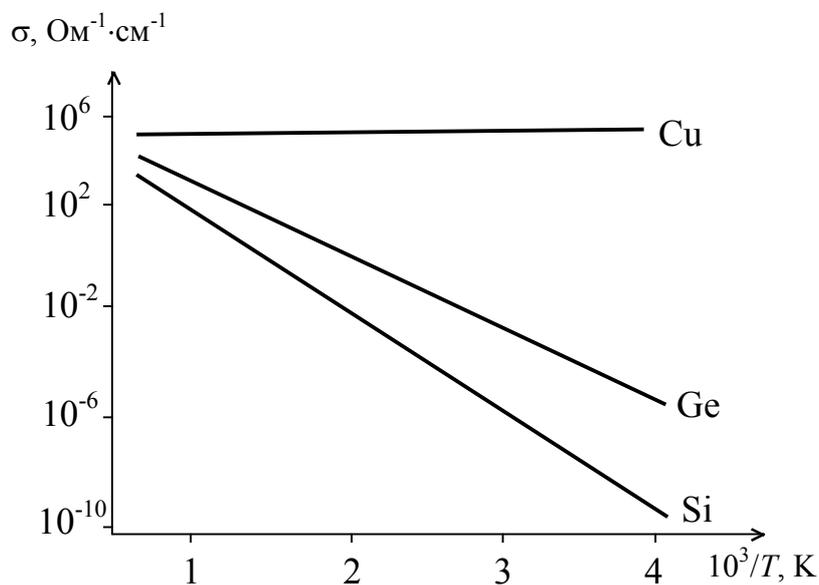


Рис. 1. Зависимость электропроводности чистых материалов от температуры

На рисунке 2 показана зависимость электропроводности беспримесного кремния от температуры при освещении его солнечным светом (график 2). Зависимость, представленная на рис. 2, демонстрирует сильное влияние электромагнитных излучений светового диапазона на свойства полупроводников.

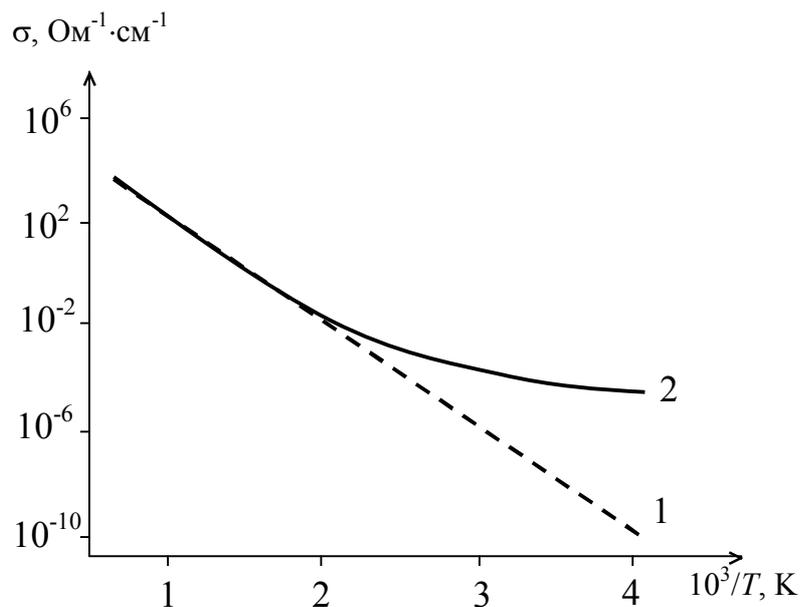


Рис. 2. Влияние освещения на температурную зависимость электропроводности кремния: зависимости, полученные в темноте (1) и на свету (2)

Расчет проводимости полупроводников обычно производят на основе зонной модели. Чтобы описать физические свойства кристаллических полупроводниковых материалов на

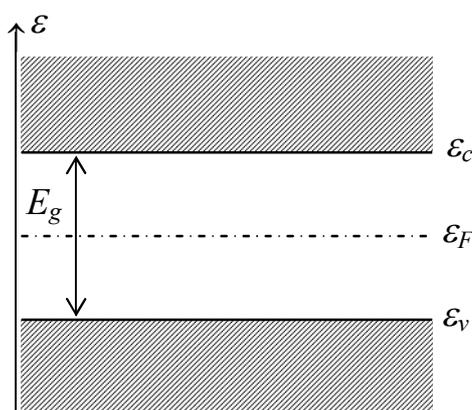


Рис. 3. Зонная диаграмма полупроводника

основе этой модели, нужно, прежде всего, определить структуру энергетических зон.

Схематическое изображение наиболее важных зон полупроводника представлено на рис. 3, где: ε_c , ε_v и ε_F – значения энергии, соответствующие дну зоны проводимости, потолку валентной зоны и уровню Ферми; E_g – ширина запрещенной зоны.

В результате теоретических расчетов было найдено, что равновесные концентрации свободных электронов n_0 и дырок p_0 в

полупроводнике определяются следующими равенствами:

$$n_0 = N_c \exp\left[-\frac{(\varepsilon_c - \varepsilon_F)}{k_B T}\right],$$

$$p_0 = N_v \exp\left[\frac{(\varepsilon_v - \varepsilon_F)}{k_B T}\right], \quad (4)$$

где: N_c и N_v – эффективные плотности состояний в зоне проводимости и в валентной зоне (соответственно), k_B – постоянная Больцмана, T – абсолютная температура полупроводника. Для кремния $N_c \approx N_v \approx 2 \cdot 10^{19} \text{ см}^{-3}$. Энергия Ферми (электрохимический потенциал) фактически равна работе, которую необходимо совершить для изменения числа свободных носителей заряда на единицу. В беспримесном и бездефектном полупроводнике (в собственном полупроводнике) энергия Ферми обозначается ε_i и она равна:

$$\varepsilon_i = \varepsilon_F = \frac{\varepsilon_c + \varepsilon_v}{2} + \frac{k_B T}{2} \ln\left(\frac{N_v}{N_c}\right). \quad (5)$$

где N_v/N_c приблизительно пропорционально отношению эффективных масс дырки и электрона.

Если за счет термического возбуждения электрон покидает валентную зону, то на его месте в этой зоне образуется дырка. Поэтому собственные концентрации свободных

носителей в идеальном беспримесном полупроводнике одинаковы и определяются равенством

$$n_i = p_i = \sqrt{N_c N_v} \exp\left(\frac{E_g}{2k_B T}\right), \quad (6)$$

где $N_v N_c \sim T^3$, а $E_G = E_{G0} - k_1 T$. Зависимость E_G от температуры приводит к тому, что при $T = 0$ К ширина запрещенной зоны, например, кремния $E_G = E_{G0} = 1,21$ эВ, в то время как при комнатной температуре ($T = 300$ К) $E_G = 1,1$ эВ.

Практические расчеты концентрации свободных носителей в бездефектном беспримесном полупроводнике для заданной температуры следует вести по формуле

$$n_i(T) = B T^{3/2} \exp\left(-\frac{E_{g0}}{2k_B T}\right), \quad (7)$$

где значения констант для кремния и германия следующие: 1) кремний: $B = 3,873 \cdot 10^{16} \text{ К}^{-3/2} \cdot \text{см}^{-3}$, $E_{G0} = 1,21$ эВ(1,17); 2) германий: $B = 1,760 \cdot 10^{16} \text{ К}^{-3/2} \cdot \text{см}^{-3}$, $E_{G0} = 0,785$ эВ(0,66). Используя эту формулу, для кремния при $T = 300$ К можно получить $n_i(\text{Si}) = 1,4 \cdot 10^{10} \text{ см}^{-3}$, а для германия $n_i(\text{Ge}) = 2,3 \cdot 10^{13} \text{ см}^{-3}$.

Для целенаправленного изменения проводимости и других физических свойств полупроводника его легируют донорными (фосфор P, мышьяк As, сурьма Sb) или акцепторными (алюминий Al, галий Ga, индий In) примесями.

В полупроводнике с донорными или акцепторными примесями произведение равновесных концентраций носителей заряда является некоторой постоянной величиной (которая, однако, зависит от температуры). Единственное условие, которое при этом налагается, состоит в том, чтобы интервалы между уровнем E_F и границами разрешенных зон должны были достаточно велики по сравнению с параметром $k_B T$.

За счет термического возбуждения постоянно происходит генерация свободных носителей как за счет переноса электронов из валентной зоны в зону проводимости, так и вследствие ионизации донорных и акцепторных атомов. Одновременно с этим происходят рекомбинационные процессы, приводящие к исчезновению свободных носителей. Генерация электронов и дырок и их рекомбинация происходят непрерывно. При любой температуре устанавливается такое равновесное состояние, что произведение np при неизменных значениях T и E_g оказывается постоянным:

$$np = n_i^2 = p_i^2 = \text{const}. \quad (8)$$

Равенство (8) называют *законом действующих масс*. В случае термодинамического равновесия этот закон справедлив как для собственных, так и для примесных

полупроводников. Если же равновесие нарушено и величины n и p в некотором конкретном приборе зависят от внешних факторов, то $np \neq n_i^2$.

Концентрации носителей заряда можно представить в виде функций, зависящих от n_i и E_i :

$$\begin{aligned} n &= n_i \exp\left[\frac{E_F - E_i}{k_B T}\right], \\ p &= n_i \exp\left[\frac{E_i - E_F}{k_B T}\right]. \end{aligned} \quad (9)$$

Эти формулы в равной мере относятся как к собственным, так и к примесным полупроводникам.

Равенства (2.7) и (2.8) позволяют получить формулу

$$np = n_i^2 = B^{1/2} T^3 \exp\left(-\frac{E_{G0}}{k_B T}\right). \quad (10)$$

Необходимо иметь в виду, что концентрация носителей заряда определяется положением уровня Ферми. Приближенные выражения, определяющие уровень E_F для примесного полупроводника с донорной примесью, можно найти, используя равенство:

$$N_d = N_c \exp[-(E_c - E_F)/(kT)].$$

В результате находим, что

$$E_F = E_c - kT \ln(N_c / N_d). \quad (11)$$

Для полупроводника р-типа используем равенство

$$N_a = N_v \exp[-(E_F - E_v)/(kT)],$$

откуда получаем

$$E_F = E_v + kT \ln(N_v / N_a). \quad (12)$$

Говорят, что полупроводник является *вырожденным*, если концентрация примесей настолько высока, что уровень E_F лежит на границах разрешенных зон или внутри той или другой из этих зон. В более строгом смысле вырожденный полупроводник характеризуется неравенством

$$E_c - E_F > 3kT.$$

Это условие приводит к тому, что формулы (2.11) и (2.12) становятся не совсем точными.

В примесных полупроводниках любого типа проводимости (n или p) одновременно присутствуют как свободные электроны, так и дырки. Тех из них, которые соответствуют типу проводимости полупроводника, называются *основными* носителями заряда. Носители

противоположного типа называются *неосновными*. Для определения их концентраций можно использовать следующие формулы:

$$\begin{aligned} n_n &= \frac{1}{2} N_d \left[1 + \left(1 + \frac{4n_i^2}{N_d^2} \right) \right], \\ p_n &= \frac{1}{2} N_d \left[-1 + \left(1 + \frac{4n_i^2}{N_d^2} \right) \right]. \end{aligned} \quad (13)$$

где n_n и p_n – соответственно основные и неосновные носители заряда в полупроводнике n -типа, p

В равновесных условиях полупроводник должен быть электронейтрален. Условие электронейтральности представляется равенством:

$$p + N_d - n - N_a = 0, \quad (14)$$

где N_d и N_a соответственно обозначают концентрации донорных и акцепторных атомов.

Часто в практически важных случаях (т.е. при комнатных температурах и выше) можно считать, что для полупроводника n -типа

$$n_n \approx N_d \quad \text{и} \quad p_n \approx n_i^2 / N_d.$$

Аналогично, для полупроводника p -типа

$$p_p \approx N_a \quad \text{и} \quad n_p \approx n_i^2 / N_a.$$

Однако при более высоких температурах генерация собственных носителей заряда становится настолько интенсивным, что определять тип проводимости по типу примесей становится невозможным – примесный полупроводник теряет свои индивидуальные свойства.

Поскольку основными носителями заряда в полупроводниках являются свободные электроны и дырки, то их электропроводность (σ) определяется электропроводностью электронов и дырок

$$\sigma = \sigma_n + \sigma_p = q\mu_n n + q\mu_p p = q(\mu_n n + \mu_p p). \quad (15)$$

Как видно из (15) электропроводность полупроводника зависит от концентрации носителей заряда и их подвижности, значения которых зависят как от технологии так и температуры. Примерная концентрация собственных носителей заряда в германии при $T=300\text{K}$ равна $n_i=2,5 \cdot 10^{19} \text{ м}^{-3}$.

Для чистого бездефектного кристалла с проводимостью близкой к собственной справедливо равенство $n_i = p_i$. Поэтому электропроводность беспримесного (собственного) полупроводника определяется равенством:

$$\sigma_i(T) = q[\mu_n(T) + \mu_p(T)] \sqrt{N_c(T)N_v(T)} e^{\frac{E_g}{2kT}} = \sigma_0(T) e^{\frac{E_g}{2kT}} \cong \sigma_0 e^{\frac{E_g}{2kT}}$$

В этом равенстве величина

$$\sigma_0(T) = q[\mu_n(T) + \mu_p(T)]\sqrt{N_c(T)N_v(T)} \cong \sigma_0$$

слабо зависит от температуры и в приближенных расчетах ее принимают равным постоянному значению σ_0 , соответствующему величине электропроводности полупроводника при $T \rightarrow \infty$.

Формула (16) хорошо описывает экспериментальную кривую электропроводности для чистых кристаллов с совершенной структурой. С ее помощью по экспериментальной зависимости $\sigma_i(T)$ можно определить такие характеристические параметры материала как E_g и σ_0 .

Для легированного (примесного) полупроводникового кристалла эффективная подвижность носителей заряда определяется рассеянием на колебаниях решетки и рассеянием на ионизованной примеси. Поэтому на экспериментальных графиках температурных зависимостей концентрации свободных носителей заряда и их подвижности можно выделить несколько областей. При этом в области, где доминирует примесная проводимость $n_i(T) \ll N_d$ или $n_i(T) \ll N_a$, помимо рассеяния на решетке на величину электропроводности может оказывать влияние и рассеяние на примесях. Особенно заметным влияние изменения подвижности становится в области истощения примеси, для которой концентрация основных носителей с хорошей точностью можно считать постоянной $n_n \approx N_d$ (или $p_p \approx N_a$), поскольку выполняется условие $n_i \ll N_d$, $n_i \ll N_a$. В этой области температурной зависимостью $n_i(T)$ можно пренебречь.

Как и в случае беспримесных полупроводниковых кристаллов, электропроводность легированных кристаллов складывается из электронной и дырочной проводимости:

$$\sigma(T) = \sigma_n(T) + \sigma_p(T)$$

Если $n \gg p$, то $\sigma_n \gg \sigma_p$, поэтому

$$\sigma \cong \sigma_n(T) = q\mu_n(T)n(T) = q\mu_n(T)N_c e^{-\frac{E_c - E_F(T)}{kT}}. \quad (17)$$

Если $p \gg n$, то $\sigma_p \gg \sigma_n$, поэтому

$$\sigma \cong \sigma_p(T) = q\mu_p(T)p(T) = q\mu_p(T)N_v e^{-\frac{E_F(T) - E_v}{kT}}. \quad (18)$$

Анализ соотношений (18) показывает, что изменение концентрации от температуры зависит экспоненциально от изменения положения уровня Ферми. Вообще уровень Ферми следует рассматривать как хороший индикатор процессов, происходящих с носителями

заряда. Если уровень Ферми приближается к зоне проводимости, то это значит, что возрастает концентрация свободных электронов n и электронная проводимость σ_n . В этом случае концентрация дырок p уменьшается и соответственно падает дырочная проводимость σ_p . Если уровень Ферми приближается к валентной зоне, то возрастает концентрация дырок p и увеличивается дырочная проводимость σ_p . В последнем случае концентрация свободных электронов n уменьшается и существенно падает электронная проводимость σ_n .

В области высоких температур, где доминируют межзонные переходы и собственная концентрация носителей больше примесной $n_i \gg n_{np}$, полупроводник ведет себя как собственный (область I). В области низких температур (область III, где примесь не ионизована) уровень Ферми должен находиться выше донорного уровня (вероятность заполнения электронами больше 1/2). По мере того, как температура повышается, доноры отдают электроны в зону проводимости и постепенно полностью ионизируются (область II). Область II принято называть областью истощения примеси, поскольку все атомы доноров отдали свои электроны, а концентрация собственных электронов все еще очень мала. Общая концентрация электронов в этой области остается постоянной и примерно равной концентрации примесных атомов. Именно эта температурная область и является основной областью работы значительной части полупроводниковых приборов с p - n -переходами. Поскольку в области II концентрация носителей изменяется незначительно, то величина электропроводности начинает зависеть от подвижности свободных носителей заряда. В результате с повышением температуры в пределах области II электропроводность будет падать в некотором интервале температур за счет доминирования процессов рассеяния свободных носителей на колебаниях решетки. Но при дальнейшем повышении температуры начнется переход к собственной проводимости - концентрация собственных свободных носителей заряда и электропроводность начинают возрастать экспоненциально с температурой.

Таким образом, электрическая проводимость полупроводников определяется концентрациями свободных носителей заряда в нем и их подвижностью. Причем факторами, влияющими на электропроводность легированных полупроводников, являются ширина запрещенной зоны E_g , концентрация и тип примесей, глубина залегания примесных уровней.

Во внешнем электрическом поле эти свободные носители заряда (электроны и дырки) начинают дрейфовать, вызывая появление тока. Кроме того, направленный перенос заряда возникает в форме диффузионного движения при неравновесных процессах, когда в объеме полупроводника появляются градиенты концентраций свободных носителей заряда.

Плотность дрейфового тока свободных электронов

$$J_n = qn v_n, \quad (19)$$

где q – заряд свободного носителя (электрона, $q = 1,602 \cdot 10^{-19}$ Кл), n – концентрация свободных электронов, v_n – их средняя дрейфовая скорость.

Плотность дрейфового тока свободных дырок

$$J_p = qp v_p, \quad (20)$$

где n – концентрация свободных дырок, v_p – их дрейфовая скорость.

Дрейфовые скорости электронов и дырок зависят от их подвижностей (μ_n и μ_p) и напряженности внешнего электрического поля (E):

$$\begin{aligned} \bar{v}_n &= \mu_n \bar{E}, \\ \bar{v}_p &= \mu_p \bar{E}. \end{aligned} \quad (21)$$

В относительно слабом электрическом поле подвижности свободных носителей не зависят от напряженности этого поля. В таком случае появляется возможность характеристики проводящих свойств полупроводника с помощью параметра σ , называемого удельной электропроводностью:

$$\begin{aligned} \sigma_n &= qn\mu_n, \\ \sigma_p &= qp\mu_p. \end{aligned} \quad (22)$$

Иногда используют параметр, называемый удельным сопротивлением ρ , причем $\rho_n = 1/\sigma_n$ и $\sigma_n = qn\mu_n$. Таким образом, для дрейфовых плотностей тока можно записать равенства:

$$\begin{aligned} \bar{J}_n &= \sigma_n \bar{E} = \frac{\bar{E}}{\rho_n}, \\ \bar{J}_p &= \sigma_p \bar{E} = \frac{\bar{E}}{\rho_p}. \end{aligned} \quad (23)$$

Однако, когда напряженность достигнет некоторого критического значения, подвижности носителей заряда начинают резко уменьшаться так, что дрейфовая скорость оказывается независимой от E . Достигнутое при этом значение называют дрейфовой скоростью насыщения и обозначают v_s . В кремнии и германии это значение примерно одинаково для электронов и дырок, $v_{ns} \approx v_{ps} \approx 10^7$ см/с.

Диффузионные токи вызываются внутренними электрическими полями, возникающими в тех местах полупроводника, где под влиянием какого то внешнего воздействия появляется градиент концентрации свободных носителей (электронов или дырок). Направление поля оказывается таким, что свободные носители будут перемещаться из области с более высокой концентрацией в область с меньшей концентрацией. При этом

плотности диффузионных токов свободных электронов и дырок определяются следующими векторными равенствами:

$$\begin{aligned}\bar{J}_{n.диф.} &= qD_n \overline{grad(n)}, \\ \bar{J}_{p.диф.} &= qD_p \overline{grad(p)},\end{aligned}\quad (24)$$

где коэффициенты диффузии свободных носителей заряда (D_n и D_p) связаны с их подвижностями,

$$\begin{aligned}D_n &= \mu_n \frac{k_B T}{e} = \mu_n \varphi_T, \\ D_p &= \mu_p \frac{k_B T}{e} = \mu_p \varphi_T,\end{aligned}\quad (25)$$

φ_T – тепловой потенциал.

Отметим, что градиент любой скалярной величины S определяется равенством

$$\overline{grad(s)} = \bar{i} \frac{\partial s}{\partial x} + \bar{j} \frac{\partial s}{\partial y} + \bar{k} \frac{\partial s}{\partial z},$$

где \bar{i} , \bar{j} и \bar{k} - единичные орты, соответственно направленные вдоль осей x , y и z .

Подвижности свободных носителей (μ_n и μ_p) зависят от температуры T и от суммарной концентрации донорных и акцепторных атомов $N=N_d+N_a$:

$$\mu(T, N) = \left[\mu_1 + \frac{\mu_2}{(N/N_{\mu 0})^a} \right] \left(\frac{T}{T_0} \right). \quad (26)$$

Таким образом, проводимость полупроводника зависит от концентрации свободных носителей заряда, которая, в свою очередь, зависит от типа и концентрации введенной в полупроводник примеси. Следует обратить внимание на тот факт, что при увеличении концентрации примеси на 9 порядков, проводимость образца возрастает на 8 порядков, т.е. существует почти линейная зависимость между проводимостью и концентрацией примеси, N . Исследования показывают, что в области высоких температур одинаково высока электропроводность как легированного кремния так и нелегированного. Но у сильно легированного полупроводника почти во всем диапазоне температур (меньших температуры плавления) проводимость меняется слабо. В области низких температур проводимость легированных полупроводников имеет слабо выраженный максимум, что напоминает температурную зависимость проводимости металлов.

Проводимость кристалла, содержащего несколько типов (k) подвижных носителей заряда

$$\sigma = \sum_{i=1}^k n_i \mu_i q_i, \quad (27)$$

где n_i - концентрация носителей заряда i -го вида, μ_i - подвижность носителей заряда i -го вида, q_i - величина заряда свободного носителя i -го вида.

Величина, обратная к проводимости, называется удельным сопротивлением. Для полупроводника с p -типом проводимости удельное сопротивление равно:

$$\rho_p = \frac{1}{\sigma_p} \approx \frac{1}{N_a e \mu_p}, \quad (28)$$

где N_a - концентрация акцепторных атомов, e - заряд электрона, μ_p - подвижность дырок.

Удельное сопротивление полупроводника с n -типом проводимости равно:

$$\rho_n = \frac{1}{\sigma_n} \approx \frac{1}{N_d e \mu_n}, \quad (29)$$

где N_d - концентрация донорных атомов, e - заряд электрона, μ_n - подвижность электронов.

Интегральные сопротивления полупроводниковых образцов с p - и n -типами проводимости можно найти по формулам

$$R_p = \frac{\rho_p l_p}{S} \quad \text{и} \quad R_n = \frac{\rho_n l_n}{S}, \quad (30)$$

где l_p и l_n - длины образцов, S - площади поперечного сечения, а ρ_p и ρ_n - их удельные сопротивления.

В отсутствие сильных магнитных полей физические свойства полупроводника, связанные с его электрической проводимостью, описываются системой из пяти основных уравнений, называемой *фундаментальной системой уравнений* полупроводника:

$$\bar{J}_n = qn\mu_n \bar{E} + qD_n \overline{\text{grad}(n)}, \quad (31)$$

$$\bar{J}_p = qp\mu_p \bar{E} - qD_p \overline{\text{grad}(p)}, \quad (32)$$

$$\frac{1}{q} \text{div}(\bar{J}_n) - R_n + G_n = \frac{\partial n}{\partial t}, \quad (33)$$

$$-\frac{1}{q} \text{div}(\bar{J}_p) - R_p + G_p = \frac{\partial p}{\partial t}, \quad (34)$$

$$\epsilon \epsilon_0 \text{div}(\bar{E}) = q(p - n + N_D - N_A), \quad (35)$$

где уравнения (31) и (32) являются уравнениями для плотностей электронного и дырочного токов, (33) и (34) - уравнения непрерывности, (35) - уравнение Пуассона. В этих уравнениях использованы следующие дополнительные обозначения: R_p и R_n - число рекомбинировавших носителей заряда (дырок и электронов, соответственно) в единицу времени, G_p и G_n - число появившихся носителей заряда (дырок и электронов, соответственно) в результате их тепловой генерации в единицу времени. Эти величины связаны с временами жизни свободных носителей:

$$\tau_p = 1/(G_p - R_p), \quad \tau_n = 1/(G_n - R_n). \quad (36)$$

2. Зондовые методы контроля электротранспортных свойств полупроводников.

При исследовании электрических свойств полупроводников и производстве полупроводниковых материалов, структур и приборов возникает необходимость измерения удельного электрического сопротивления или удельной электрической проводимости полупроводниковых материалов в виде монокристаллических слитков, образцов различной геометрической формы, пластин, диффузионных, эпитаксиальных и ионно-легированных слоев, составляющих часть полупроводниковых структур. Измерение удельного сопротивления осуществляется не только для установления его значения, но также для определения других важных параметров полупроводникового материала на основе теоретических расчетов или дополнительных экспериментальных данных.

Выбор метода измерения осуществляют с учетом получения требуемой информации, особенностей исследуемого материала, возможности изготовления электрических контактов, геометрической формы образца, метрологических характеристик метода измерения. В идеальном случае измерение характеристик материалов не должно приводить к разрушению образца и не должно требовать его специальной обработки.

Многие методы измерения удельного электрического сопротивления полупроводниковых материалов основаны на измерении разности электрических потенциалов на некотором участке образца, через который пропускается электрический ток. Исторически одним из первых методов измерения удельного сопротивления был *двухзондовый метод*, применявшийся для измерения удельного сопротивления металлов. Измерения с помощью методов, для которых необходима строго определенная геометрия образца, сопряжены со значительными затратами времени, главным образом на изготовление образцов и контактов к ним. Для быстрого измерения удельного сопротивления используют четырехзондовый метод, метод сопротивления растекания точечного контакта, а также метод Вандер-Пау. Указанные методы удобны, позволяют выполнить измерения не только на однородных образцах, но и на диффузионных, эпитаксиальных и ионно-легированных слоях, а также исследовать пространственное изменение удельного сопротивления. Для зондовых методов силовые линии напряженности электрического поля, непараллельны и распределены неоднородно по сечению образца, ввиду чего сначала необходимо теоретически рассчитывать распределение электрического

потенциала в образце, а затем находить взаимосвязь между удельным сопротивлением и измеряемыми величинами – током и напряжением.

При использовании зондовых методов нужен непосредственный контакт с полупроводниковым образцом. Поэтому эти методы приводят к разрушению образцов. Например, при измерениях четырехзондовым методом или методом сопротивления растекания металлические зонды разрушают поверхность образца; для применения же метода Ван-дер-Пау необходимо создание четырех контактов. Поэтому существует потребность в методах измерения, не требующих непосредственного контакта с образцом и получивших название *бесконтактных*. Имеется несколько подходов к осуществлению бесконтактных измерений. Они основаны на использовании отражения или поглощения электромагнитной волны, индуктивной или емкостной связи образца с измерительной схемой. В силу этого бесконтактные методы являются оптическими или высокочастотными.

Двухзондовый метод применяют для измерения удельного сопротивления образцов, имеющих правильную геометрическую форму и постоянное поперечное сечение. При этом на торцевых гранях образца, например в виде прямоугольной пластины, изготавливают омические контакты. Через эти контакты вдоль образца пропускают электрический ток. На одной из поверхностей образца вдоль линий тока устанавливают два контакта в виде металлической игловок – зондов, имеющих малую площадь соприкосновения с поверхностью и позволяющих измерить разность потенциалов. Если образец однороден, то его удельное сопротивление (Ом·см)

$$\rho = \frac{U_{12} \cdot S}{I \cdot s}.$$

где I – сила тока, протекающего через образец, А; U_{12} – разность потенциалов между измерительными или потенциальными зондами, В; s – расстояние между зонами, см; S – площадь поперечного сечения, см².

Ток через образец подается от регулируемого источника постоянного тока, гарантирующего достаточную его стабильность во времени. Сила тока измеряется миллиамперметром, разность потенциалов (см. рис. 1) – полуавтоматическим компенсирующим потенциометром или электронным цифровым вольтметром с высоким выходным сопротивлением.

Погрешность измерений напряжения и тока при использовании приборов с цифровой индикацией не превышает 0,1%. Однако систематическая погрешность измерения напряжения между потенциальными зондами зависит от сопротивления контакта зонда с образцом, которое может в $10^2 - 10^4$ раз превышать сопротивление образца.

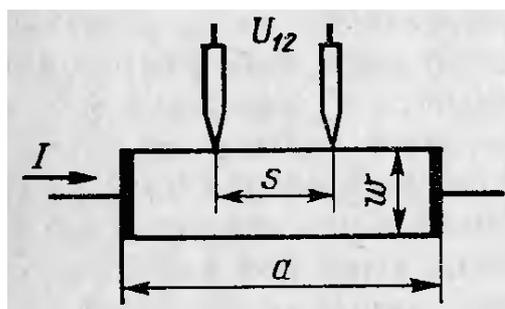


Рис. 1. Схема измерения удельного сопротивления двухзондовым методом

Чтобы устранить влияние сопротивлений контактов на результаты измерений, необходимо предельно уменьшать протекающий через них ток. Для этого используют вольтметры с высоким выходным сопротивлением (10^8 Ом и больше), причем сопротивление изоляции каждого из зондов относительно любого элемента измерительной установки должно быть много больше этого значения.

Неточное наблюдение геометрических размеров образца ведет к отклонению формы эквипотенциальных поверхностей от плоских поверхностей, параллельных торцевым граням образца, и возникновению систематической погрешности. Искажение формы эквипотенциальных поверхностей и одномерного пространственного распределения линий тока имеет место в образце правильной геометрической формы вблизи омических контактов, если они неоднородны по площади. В предельном случае использования точечных токовых контактов эту погрешность можно свести к значению, не превышающему 0,5%, при следующих соотношениях геометрических размеров образца:

$$a \gg 3b, \quad s \leq a/2, \quad b/2 \leq w \leq b,$$

где a , b , w – длина, ширина и толщина образца. При этом потенциальные зонды следует располагать на расстоянии от торцевых граней образца, большем $3d$.

Отклонение реального контакта зонда от теоретической модели точечного потенциального контакта также вносит систематическую погрешность в результат измерения.

Случайная составляющая относительной погрешности измерений, обусловленная невоспроизводимостью расстояния между зондами при независимой фиксации каждого из зондов со среднеквадратическим отклонением ΔS , определяется соотношением

$$\frac{\delta\rho}{\rho} = 1,41 \cdot \left(\frac{2\Delta S}{s} \right)$$

при доверительной вероятности 0,95.

На контакте металл – полупроводник при протекании электрического тока может возникать ряд физических эффектов и явлений, которые вносят существенную погрешность

в результаты измерений. Основными из них являются: инжекция носителей заряда контактом, влияющая на проводимость образца; эффект Пельтье, приводящий к возникновению градиента температуры на образце и соответствующей этому градиенту термо-ЭДС; нагрев образца электрическим током, протекающим через образец. Указанные явления необходимо учитывать не только при измерении удельной проводимости, но и во всех случаях, когда через измеряемый образец с металлическими контактами протекает электрический ток. Чтобы уменьшить погрешности, вызванные этими явлениями, ток через образец выбирается минимальным, но обеспечивающим необходимую точность измерения тока и напряжения. Торцевые грани образца шлифуются абразивным порошком, после чего электрически металлизируются. Абразивная обработка обеспечивает подавление инжекции носителей заряда.

Влияние неоднородности удельного сопротивления образца.

Практически любой полупроводниковый образец обладает некоторой неоднородностью в распределении удельного сопротивления. Измеренное на неоднородном образце значение удельного сопротивления характеризуется *эффективным удельным сопротивлением* и представляет собой значение, усредненное по некоторой области образца. Интерпретация результатов измерений эффективного удельного сопротивления на неоднородном образце, т.е. нахождения соотношения между эффективным и действительным или средним значением удельного сопротивления, предполагает знание конкретной функции $\rho(x, y, z)$ и требует выполнения соответствующих теоретических расчетов. Решение этой задачи в аналитической форме возможно для простых одномерных случаев: линейного, экспоненциального и периодического распределения удельного сопротивления. В ряде практически важных случаев для образцов с одномерной неоднородностью, а именно: для диффузионных, эпитаксиальных и ионно-легированных слоев, распределение удельного сопротивления по толщине слоя измеряют с помощью последовательного удаления слоев.

В объемных образцах наиболее часто встречаются локальные флуктуации удельного сопротивления, распределенные по объему образца и обусловленные флуктуациями концентрации и подвижности носителей заряда. Если линейный размер локальных неоднородностей превышает длину свободного пробега носителей заряда и дебаевскую длину экранирования, но мал по сравнению с геометрическими размерами образца, то эффективная удельная проводимость связана с ее средним значением σ соотношением

Второе слагаемое в квадратичной скобке представляет собой локальное среднеквадратическое отклонение удельной проводимости от ее среднего значения $\bar{\sigma}$. Из

сказанного следует, что измеряемые значения удельной проводимости отличается от ее среднего значения.

Четырехзондовый метод измерения.

Четырехзондовый метод измерения удельного сопротивления полупроводников является самым распространенным. Кроме высоких метрологических показателей преимущество четырехзондового метода состоит в том, что для его применения не требует создания омических контактов к образцу, возможно измерение удельного сопротивления объемных образцов самой разнообразной формы и размеров, а также удельного сопротивления слоев полупроводниковых сттруктур. Условием для его применения с точки зрения формы образца является наличие плоской поверхности, линейные размеры которой превосходит линейные размеры системы зондов.

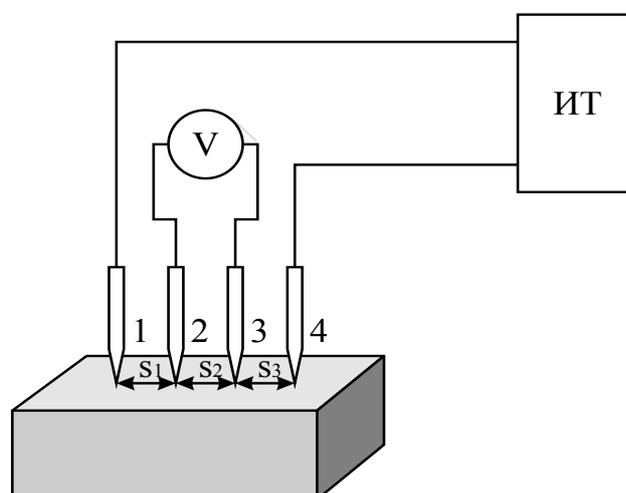


Рис.2. Схема измерения удельного сопротивления четырехзондовым методом

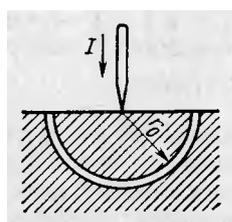


Рис.3. Модель зонда

Теория метода. Рассмотрим теоретические основы четырехзондового метода измерения удельного сопротивления применительно к образцу, представляющему собой полубесконечный объем, ограниченный плоской поверхностью.

На плоской поверхности образца вдоль площадью соприкосновения, расстояния между которыми S_1, S_2, S_3 . Через два внешних зондах 2 и 3 пропускают электрический ток I_{14} , на двух внутренних зондах 2 и 3 измеряют разность потенциалов U_{23} . По измеренным

значениями разности потенциалов между зондами 2 и 3 и тока, протекающего через зонды 1 и 4, можно определить удельное сопротивление образца.

Чтобы найти аналитическую связь между удельным сопротивлением ρ , током I_{14} и напряжением U_{23} , необходимо сначала решить более простую задачу, связанную с протеканием тока через отдельный точечный зонд, находящийся в контакте с плоской поверхностью полупроводникового образца полубесконечного объема.

Так как пространственное распределение электрического потенциала $U(r)$ в образце имеет сферическую симметрию, то для его определения достаточно решить уравнение Лапласа в сферической системе координат, в котором оставлен лишь член, зависящий от r ,

$$\Delta U(r) = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dU}{dr} \right) = 0 \quad (1)$$

при условии, что потенциал в точке $r=0$ положителен и стремится к нулю при очень больших r . Интегрирование этого уравнения с учетом указанных граничных условий позволяет получить следующее решение:

$$U(r) = -\frac{C}{r}$$

Константу интегрирования можно вычислить из условия для напряженности электрического поля $E(r)$ при некотором значении $r = r_0$:

$$E(r_0) = -\left. \frac{dU(r)}{dr} \right|_{r=r_0}.$$

Так как плотность тока, протекающего через полусферу радиусом r_0 ,

$$j = \frac{I}{2\pi r_0^2},$$

Учитывая закон Ома получим

$$U(r_0) = \frac{I \cdot \rho}{2\pi r_0}.$$

Окончательно имеем

$$U(r) = \frac{I \cdot \rho}{2\pi r}. \quad (2)$$

Очевидно, что распределение потенциала будет таким же, когда форма контакта зонда с поверхностью образца имеет вид полусферы конечного диаметра.

Пусть радиус контакта равен r_1 . Тогда электрическое напряжение на образце равно электрическому потенциалу зонда:

$$U(r_1) = \frac{I \cdot \rho}{2\pi r_1}. \quad (3)$$

Из уравнения напряжения на приконтактном слое толщиной $(r_2 - r_1)$ получим равенство

$$U(r_1) - U(r_2) = \frac{I \rho}{2\pi} \cdot \frac{r_2 - r_1}{r_2 r_1}$$

откуда следует, что основное изменение потенциала происходит вблизи зонда. Например, при $r_2 = 10r_1$ напряжение на образце превосходит напряжение на слое толщиной $(r_2 - r_1)$ всего лишь на 10%. Это означает, что значение протекающего через зонд тока определяется главным образом сопротивлением приконтактной области, протяженность которой тем меньше, чем меньше радиус контакта.

Рассмотрим линейное расположение зондов.

Сформулируем предположения, на которых основан четырехзондовый метод измерения удельного сопротивления: 1) зонды расположены на плоской поверхности однородного изотропного образца полубесконечного объема; 2) зонды имеют контакты с поверхностью образца в точках, которые расположены вдоль прямой линии; 3) инжекция носителей заряда в объем образца отсутствует.

По принципу суперпозиции электрический потенциал в любой точке образца равен сумме потенциалов, создаваемых в этой точке током каждого зонда. При этом потенциал имеет положительный знак для тока, вытекающего из образца (зонд 1), и отрицательный знак для тока, вытекающего из образца (зонд 4). Для системы зондов, расстояния между которыми s_1, s_2, s_3 , потенциалы измерительных зондов 2 и 3 будут:

$$U_2 = \frac{I_{14} \rho}{2\pi} \left(\frac{1}{s_1} - \frac{1}{s_2 + s_3} \right), \quad U_3 = \frac{I_{14} \rho}{2\pi} \left(\frac{1}{s_1 + s_2} - \frac{1}{s_3} \right).$$

Разность потенциалов

$$U_{23} = U_2 - U_3 = \frac{I_{14} \rho}{2\pi} \left(\frac{1}{s_1} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} + \frac{1}{s_3} \right). \quad (4)$$

Из последнего получаем, что удельное сопротивление образца равно,

$$\rho = 2\pi \left(\frac{U_{23}}{I_{14}} \right) / \left(\frac{1}{s_1} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} + \frac{1}{s_3} \right).$$

Если расстояние между зондами одинаковы, т.е. $s_1 = s_2 = s_3 = s$, то

$$\rho = 2\pi s \frac{U_{23}}{I_{14}}. \quad (5)$$

Используя другие комбинации включения токовых и потенциальных зондов, можно получить аналогичные выражения для удельного сопротивления, которые отличаются от (1.5) значениями числовых коэффициентов, приведенными в табл. 1.

Таблица 1.

№ п/п	Зонды	Ток	Зонды	Напряжение	Числовой коэффициент
1	1-4	I_{14}	2-3	U_{23}	
2	1-2	I_{12}	3-4	U_{34}	
3	1-3	I_{13}	2-4	U_{24}	
4	2-3	I_{23}	1-4	U_{14}	
5	2-4	I_{24}	1-3	U_{13}	
6	3-4	I_{34}	1-2	U_{12}	

Как следует из табл. 1, предпочтительны комбинации включения зондов 1 и 4, так как они обеспечивают максимальное регистрируемое напряжение.

Расположение зондов по вершинам квадрата. Линейное расположение зондов на поверхности полупроводникового образца не является единственно возможным. Можно, например, использовать систему четырех зондов, расположенных по вершинам квадрата. В этом случае ток пропускают через зонды, образующие одну из сторон квадрата, например через зонд 1 и 2, а напряжение измеряют на другой паре зондов 3 и 4. Используя соотношение (2), легко рассчитать, что при таком расположении зондов удельное сопротивление

$$\rho = \frac{2\pi s}{(2 - \sqrt{2})} \cdot \frac{U_{34}}{I_{12}}.$$

Расположение зондов по вершинам квадрата обеспечивает снижение случайных ошибок в два раза. Это достигается за счет выполнения измерений при пропускании тока последовательно через каждую пару соседних зондов, изменения полярности приложенного напряжения и следующего вычисления среднего значения удельного сопротивления по результатам восьми измерений. С помощью современной электронной измерительной аппаратуры такие измерения легко могут быть автоматизированы.

Электрическая схема и методика измерения.

Электрическая схема измерения удельного сопротивления четырехзондовым методом проста. Ток I от регулируемого источника постоянного напряжения U_0 пропускается через

зонды 1 и 4. Желательно, чтобы источник напряжения имел высокое выходное сопротивление, т.е. являлся генератором тока. Напряжение, возникающее при этом между зондами 2 и 3, регистрируется вольтметром V . Сила тока фиксируется миллиамперметром или находится путем измерения напряжения на эталонном резисторе, включенном последовательно в цепь зондов 1 и 4. Наименьший рабочий ток определяется возможностью измерения малых напряжений; наибольший рабочий ток ограничивается нагревом образца.

Система из четырех зондов конструктивно оформляется в виде *четырёхзондовой головки*. Расстояния между зондами строго фиксированы, угол заточки острия зонда составляет $45-150^\circ$. Четырёхзондовую головку крепят к манипулятору, с помощью которого зонды устанавливаются на поверхности образца. Зонды индивидуально прижимаются к поверхности с силой до 2 Н. в качестве материала зондов используют твердые металлы и сплавы.

При проведении измерений в интервале высоких температур вплоть до 950°C для ряда полупроводниковых материалов для изготовления зондов используют карбид вольфрама. Чтобы уменьшить разрушение поверхности, применяют зонды из жидких металлов – ртути и галлия (при температурах выше температуры плавления гелия $29,8^\circ\text{C}$). Материал зонда не должен химически реагировать с полупроводниковым материалом.

Чтобы контактные сопротивления потенциальных зондов не влияли на результаты измерений, разность потенциалов необходимо фиксировать в отсутствие тока через них. Поэтому измерения проводят компенсационным методом с помощью полуавтоматических потенциалов. Целесообразно также применение электронных цифровых вольтметров с высоким входным сопротивлением. В таком случае ток через измерительные зонды пренебрежимо мал, что позволяет отказываться от использования компенсационных методов измерений. Применение милливольтметра с входным сопротивлением порядка 10^8 Ом дает возможность измерять удельное сопротивление на слитках и пластинах кремния до 3000 Ом·см.

Погрешность измерения удельного сопротивления четырехзондовым методом определяется как составляющими погрешности входящих в формулу (5) величин, так и размером контактной площадки зонда, сопротивлением контактов, нестабильностью температуры образца, термо-ЭДС, освещенностью образца, инжекцией носителей заряда и др.

Различие в расстояниях между зондами ведет к погрешности измерения ρ . Если, например, каждый из зондов смещен относительно своего номинального положения на Δx_i , то относительная погрешность

$$\frac{\delta \rho}{\rho} = \frac{1}{4s} (3\Delta x_1 - 5\Delta x_2 + 5\Delta x_3 - 5\Delta x_4).$$

Это соотношение можно использовать для коррекции значения удельного сопротивления, если Δx_i известны.

Когда конструкция зондовой головки допускает независимое смещение каждого зонда относительно своего номинального положения, которое характеризуется среднеквадратическим отклонением Δs , то случайная погрешность измерения

$$\frac{\delta \rho}{\rho} = 2,06 \cdot \left(\frac{2\Delta s}{s} \right)$$

при доверительной вероятности 0,95.

Чтобы избежать погрешностей при измерениях тока и напряжения, которые могут возникнуть вследствие утечек тока и возникновения напряжения на контактных сопротивлениях, необходимо обеспечивать высокое сопротивление изоляции и использовать приборы для измерения напряжения с входным сопротивлением, превышающим сопротивление образца и контактов в 10^2 раз.

Источником погрешности могут служить фотопроводимость и фото-ЭДС, возникающие под действием освещения и особенно сильно проявляющиеся в образцах с высоким удельным сопротивлением.

Так как полупроводники имеют относительно высокий температурный коэффициент сопротивления, то при измерениях за счет протекания через образец тока может произойти не только локальный нагрев, но и повышение температуры всего образца. Например, повышение температуры кремния с удельным сопротивлением $10 \text{ Ом} \cdot \text{см}$ и 5°C приводит к изменению удельного сопротивления на 4,0%. Поэтому для уменьшения нагрева образца необходимо выбирать рабочий ток минимально возможным, а температуру образца поддерживать постоянной. Рабочий ток, однако, должен обеспечивать необходимую точность измерений равности потенциалов. Измерение разности потенциалов производят при двух направлениях тока и полученные значения усредняют, исключая таким способом продольную термо-ЭДС, возникающую на образце вследствие градиента температуры. Уменьшение рабочего тока одновременно снижает модуляцию проводимости образца, вызванную инжекцией носителей заряда при протекании тока.

Для уменьшения влияния инжекции и получения малых контактных сопротивлений металлических зондов поверхность образца, на которой производят измерения, механически обрабатывают (например, шлифуют). Однако подобная обработка, особенно для высокоомных образцов, иногда недостаточна. Например, механически обработанная поверхность германия позволяет получить малое сопротивление контакта, а поверхность

кремния с металлическим зондом, наоборот, - очень высокое сопротивление. Высокое сопротивление контакта не только ограничивает протекающий ток, но как правило, нестабильно во времени и в значительной степени затрудняет проведение измерений. Поэтому нередко контакты подвергают электрической формовке.

Следует отметить, что высокое контактное сопротивление между зондом и образцом является одной из основных причин, ограничивающих применение четырехзондового метода для измерения удельного сопротивления широкозонных полупроводников типа A^3B^5 и A^2B^6 . В технических условиях на монокристаллический кремний в слитках в соответствии с ГОСТ 19658–81 предусмотрена следующая методика измерения удельного сопротивления. Она распространяется на слитки монокристаллического кремния, получаемые по методу Чохральского и предназначенные для изготовления пластин-подложек, используемых в производстве эпитаксиальных структур и структур металл – диэлектрик - полупроводник. Методика предназначена для измерения удельного сопротивления на торцевой поверхности слитков кремния от 10^{-4} до 10^2 Ом·см. Измерения проводят на плоских поверхностях, имеющих шероховатость не более 2,5 мкм при фиксированной температуре ($23 \pm 2^\circ\text{C}$). Используют четырехзондовую измерительную головку типа С2080 с четырьмя линейно расположенными зондами из карбида вольфрама с межзондовым расстоянием ($1,3 \pm 0,01$) мм и максимальным линейным размером рабочей площадки зонда не более 60 мкм. Усилие прижима зонда к поверхности слитка составляет 0,5-2,0 Н. Измерительные приборы обеспечивают измерение силы электрического напряжения – с погрешностью не более 0,5%, а электрического напряжения – с погрешностью не более 1% при необходимом для правильного измерения входном сопротивлении. Предельные значения рабочих токов и измеряемых напряжений, а также входных сопротивлений в зависимости от удельного сопротивления кремния приведены в табл. 2.

Удельное сопротивление вычисляют как среднее арифметическое значений, полученных при двух измерениях, различающихся направлением тока. Измерение вычисляют с удельным сопротивлением более 200 Ом·см проводят при их затмении.

При выполнении всех требований к применяемым средствам измерения и соблюдении необходимых условий интервал, в котором находится случайная погрешность измерения удельного сопротивления, характеризующая сходимость результатов, равен $\pm 2\%$ при доверительной вероятности 0,95. Интервал, в котором находится погрешность измерения, определяющая воспроизводимость измерений при соблюдении требований стандарта, равен $\pm 5\%$ при доверительной вероятности 0,95.

Таблица 2.

Удельное сопротивление ρ , Ом·см	Рабочий ток I , А	Измеряемое напряжение U , В	Входное сопротивление R , Ом, не менее
10^{-5}	$1,0 \cdot 10^{-1}$	$1,2 \cdot 10^{-4}$	$1 \cdot 10^3$
10^{-2}	$1,0 \cdot 10^{-3}$	$1,2 \cdot 10^{-3}$	$1 \cdot 10^4$
10^{-1}	$1,0 \cdot 10^{-1}$	$1,2 \cdot 10^{-2}$	$1 \cdot 10^5$
1,0	$8,2 \cdot 10^{-2}$	$1,0 \cdot 10^{-1}$	$1 \cdot 10^6$
10^1	$8,2 \cdot 10^{-3}$	$1,0 \cdot 10^{-1}$	$1 \cdot 10^7$
10^6	$8,2 \cdot 10^{-4}$	$1,0 \cdot 10^{-1}$	$1 \cdot 10^8$
10^3	$8,2 \cdot 10^{-5}$	$1,0 \cdot 10^{-1}$	$1 \cdot 10^9$

3. Применение четырехзондового метода к образцам простой геометрической формы

Рассмотрим образец полубесконечного объема с проводящей или изолирующей границей.

Соотношение (5) пригодно для вычисления удельного сопротивления образца полубесконечного объема, т.е. для образца, линейные размеры которого много больше расстояния между зондами s . Этот критерий применимости формулы (5), по существу, является полуколичественным, так как не указывают точного соотношения между размерами образца и расстоянием s . Для получения более строгих критериев применимости (5) рассмотрим частные случаи измерения удельной проводимости образцов ограниченных размеров, которые имеют и самостоятельное значение, так как часто встречаются на практике.

Проанализируем следующий случай. Образец полубесконечного объема имеет плоскую проводящую границу; зонды расположены на линии, перпендикулярной этой границе. Наличие проводящей границы на одной из граней образца обуславливает его шунтирование и, следовательно, увеличение тока I , что приводит к уменьшению измеренного в соответствии с (5) значения ρ по сравнению с истинным. Практическим примером такой модели может служить образец, на одной из плоских граней которого создан омический контакт. Ввиду того, что сопротивление омического контакта очень мало, можно считать потенциал контакта постоянным.

Используем метод зеркальных изображений, который широко применяется в электростатике. Для этого на продолжении линии зондов симметрично относительно проводящей границы поместим два таких мнимых источника тока 5 и 6, чтобы

удовлетворять условию на проводящей границе $U=0$. Условие на проводящей границе будет выполнено, если считая образец бесконечным, в точку 5 поместить положительный источник тока I , а в точке 6 – отрицательный $-I$. В силу единственности решения задачи при данном граничном условии решение для системы источников тока 1,4,5 и 6 будет искомым. Вычисление потенциалов в точках 2 и 3 с учетом четырех источников тока позволяет определить удельное сопротивление образца:

$$\rho = 2\pi s f_1 \left(\frac{l}{s} \cdot \frac{U}{I} \right).$$

где

$$f_1 \left(\frac{l}{s} \right) = \left[1 + \frac{1}{2 + (2l/s)} - \frac{1}{5 + (2l/s)} - \frac{1}{1 + (2l/s)} + \frac{1}{4 + (2l/s)} \right]^{-1} \quad (6)$$

- поправочная функция, зависящая от отношения l/s . При $l=0$, т.е. в случае, когда зонд 4 установлен на проводящей границе образца, $f_1(l/s) \approx 2$; при $l \geq 3s$ функция $f_1(l/s)$ практически не отличается от 1. Таким образом, рассмотрим случае выражения (1.5) можно использовать, если выполняется соотношение $l \geq 3s$.

Поправочную функцию, подобную (6), можно легко вычислить и для случая, когда граница является не проводящей, а изолирующей. При этом, используя метод зеркальных изображений, необходимо учесть, что на изолирующей границе должно выполняться иное граничное условие: нормальная составляющая тока, а следовательно, и нормальная составляющая электрического поля на границе равны нулю. Это граничное условие будет выполнено, если знаки зеркальных источников тока совпадают со знаками реальных токов, протекающих источников зонды 1 и 4. В результате вычислений получим

$$f_2 \left(\frac{l}{s} \right) = \left[1 + \frac{1}{1 + (2l/s)} - \frac{1}{2 + (2l/s)} - \frac{1}{4 + (2l/s)} + \frac{1}{5 + (2l/s)} \right]^{-1}.$$

Для образца полубесконечного объема при параллельном расположении зондов относительно изолирующей границы

$$f_2 \left(\frac{l}{s} \right) = \left[1 + \frac{1}{\{1 + (2l/s)^2\}^{1/2}} - \frac{1}{\{1 + (l/s)^2\}^{1/2}} \right]^{-1}.$$

Относительно проводящей границы

$$f_2 \left(\frac{l}{s} \right) = \left[1 + \frac{2}{\{1 + (2l/s)^2\}^{1/2}} - \frac{1}{\{1 + (l/s)^2\}^{1/2}} \right]^{-1}.$$

Числовые значения функций $f_1(l/s)$, $f_2(l/s)$, $f_3(l/s)$ и $f_4(l/s)$ приведены в табл. 3.

Таблица 3.

(l/s)	$f_1(l/s)$	$f_2(l/s)$	$f_3(l/s)$	$f_4(l/s)$
0	1,82	0,69	0,5	∞
0,2	1,365	0,79	0,533	8,07
0,5	1,182	0,882	0,658	2,08
1,0	1,06	0,947	0,842	1,232
2,0	1,01	0,992	0,965	1,038
5,0	1,004	0,996	0,9974	1,003
10,0	1,0005	0,9995	0,9996	1,004

Анализ поправочных функций, полученных для образца полубесконечного объема с изолирующей или проводящей границей при параллельном и перпендикулярном расположении линии зондов относительно границы, показывает, что во всех случаях поправка пренебрежимо мала, если $l \geq 5s$.

Тонкая пластина.

Определение удельного сопротивления тонкой пластины, как и образец полубесконечного объема с границей, сводится к вычислению поправочной функции. Однако ее расчет для тонкой пластины более сложен, так как тонкая пластина определенной геометрической формы имеет большое число поверхностей и для каждой из них должно выполняться соответствующее граничное условие.

Рассмотрим простой случай тонкой пластины бесконечных размеров, нижняя граница которой является проводящей. Используя метод зеркальных изображений, расположим на расстоянии w ниже проводящей границы мнимые источники тока I и $-I$, что обеспечивает выполнение из нижней проводящей границе граничного условия $U=0$. Однако при этом нарушается требование равенства нулю нормальной составляющей тока на верхней поверхности пластины, введем на расстоянии выше пластины два мнимых источника тока I и $-I$. При этом граничное условие на верхней поверхности будет выполнено, но нарушается граничное условие на нижней проводящей границе. Чтобы удовлетворить условию на нижней границе, введем два мнимых источника I и $-I$ на расстоянии от нижней поверхности. Очевидно, введение мнимых источников тока для выполнения граничных условий нужно продолжить до бесконечности.

Значение потенциалов на измерительных зондах 2 и 3 можно вычислить путем суммирования потенциалов, создаваемых в данной точке каждым источником тока:

$$U_2 = \frac{I\rho}{2\pi} \sum_{n=-\infty}^{+\infty} (-1)^n \left| [s^2 + (2wn)^2]^{(-1/2)} \{ (2s)^2 + [s^2 + (2wn)^2]^2 n^2 \}^{(-1/2)} \right|$$

$$U_3 = \frac{I\rho}{2\pi} \sum_{n=-\infty}^{+\infty} (-1)^n \left| [(2s)^2 + (2wn)^2]^{(-1/2)} \{ (2s)^2 - [s^2 + (2wn)^2]^2 n^2 \}^{(-1/2)} \right|.$$

В результате удельное сопротивление пластины

$$\rho = 2\pi s g_1 \left(\frac{w}{s} \right) \cdot \left(\frac{U}{I} \right).$$

Функция поправок $g_1(w/s)$ зависит только от отношения толщины пластины w к расстоянию между зондами s :

$$g_1(w/s) = \left\{ 1 + 4 \sum_{n=1}^{\infty} (-1)^n \left[\left(1 + 4n^2 \frac{w^2}{s^2} \right)^{(-1/2)} - \left(4 + 4n^2 \frac{w^2}{s^2} \right)^{(-1/2)} \right] \right\}^{-1}.$$

Если $w \gg s$, то пластину можно считать образцом полубесконечного объема:

$$g_1(w/s) = 1.$$

Реальные пластины имеют боковые грани, которые влияют на распределение тока, и потому должны быть учтены соответствующими поправочными функциями. Поправочные функции в ряде случаев могут быть вычислены в результате решения уравнения Лапласа с соответствующими граничными условиями на боковых поверхностях пластин.

Лекция 4. Эффекты в материалах со свободными носителями заряда во внешних полях и их использование для контроля свойств этих материалов.

1. Эффект Холла и магниторезистивный эффект

Среди множества явлений, связанных с переносом носителей заряда, гальваномагнитные явления зарекомендовали себя как наиболее ценные для определения параметров полупроводников. К гальваномагнитным явлениям относятся эффект Холла и магниторезистивный эффект, возникающие при совместном действии на полупроводник электрического и магнитного полей.

Эффект Холла.

Обычно гальваномагнитные явления рассматривают для изотропного полупроводника в изотермических условиях. Предположим, что в прямоугольной системе координат электрическое и магнитное поле направлены вдоль осей x и z . Под действием электрического поля напряженностью E_x через полупроводниковый образец, имеющий форму прямоугольной пластины толщиной ω , протекает электрический ток, плотность которого j_x . На носители заряда, например электроны, дрейфующие со скоростью $\mu_n E_x$ в магнитном поле с индукцией B_z , в направлении y действует сила $F_y = e\mu_n E_x B_z$, причем ее направление одинаково для электронов и дырок. Если составляющая тока $j_y = 0$, то накопившийся у граней образца в направлении y объемный заряд создает электрическое поле напряженностью которое уравнивает ток, протекающий под действием силы F_y . Возникающее в образце электрическое поле напряженностью E_y и соответствующая ему разность потенциалов между гранями образца представляют собой эффект Холла. В общем случае напряженность электрического поля и ЭДС Холла U_x выражаются соотношениями

$$E_y = R_x j_x B_z; \quad U_x = R_x \frac{I_x B_z}{\omega} \quad (1)$$

где R_x — коэффициент Холла; I_x — сила тока, протекающего через образец.

Если процессы рассеяния носителей заряда описываются в приближении времени релаксации τ , то для зонной модели с одной параболической зоной в результате решения

кинетического уравнения Больцмана можно получить выражения для указанных гальвано-магнитных явлений при слабых и сильных магнитных полях.

Критерий слабого и сильного магнитного поля, как показывает теория, зависит не только от значения магнитной индукции, но и от значения подвижности носителей заряда.

Для слабого магнитного поля

$$\mu_n^2 B^2 \ll 1, \mu_p^2 B^2 \ll 1 \quad (2a)$$

для слабого

$$\mu_n^2 B^2 \gg 1, \mu_p^2 B^2 \gg 1 \quad (2b)$$

Для примесного полупроводника с одним типом носителей заряда коэффициент Холла связан с концентрацией носителей заряда n_0 , p_0 и факторами рассеяния r_n , r_p :

$$R_{x_n} = -r_n / (en_0) \quad (3)$$

$$R_p = r_p / (ep_0) \quad (4)$$

где

$$r_n = \frac{\langle \tau_n^2 \rangle}{\langle \tau_n \rangle^2} \geq 1; \quad r_p = \frac{\langle \tau_p^2 \rangle}{\langle \tau_p \rangle^2} \geq 1.$$

Усреднение времени релаксации производится в рамках соответствующей статистики носителей заряда. Коэффициент Холла имеет отрицательное значение для электронного полупроводника и положительное — для дырочного.

Для невырожденных полупроводников, для которых применимы классическая статистика Больцмана и степенной закон зависимости времени релаксации от энергии, средние величины, входящие в выражение для коэффициента r , вычислены для различных механизмов рассеяния носителей заряда. Значения этого коэффициента приведены в табл. 1.

Таблица 1.

Механизмы рассеяния	r	α	α_r
Ионизированные примеси	1,9	2,15	5,89
Нейтральные примеси	1	11	1
Акустические фононы	1,18	0,381	1,77

В сильных магнитных полях $r = 1$ независимо от механизма рассеяния, а отношение коэффициентов Холла в слабом и сильном магнитных полях позволяет найти значение γ , если условие сильного магнитного поля можно реализовать.

Если полупроводник имеет смешанную электропроводность, то в слабом магнитном поле коэффициент Холла

$$R_x = \frac{r}{p} \cdot \frac{\mu_p^2 p_0 - \mu_n^2 n_0}{(\mu_p p_0 + \mu_n n_0)^2} = \frac{r}{e} \cdot \frac{p_0 - n_0 b^2}{(p_0 + n_0 b)^2}, \quad (5)$$

где $b = \mu_n / \mu_p$.

При $n_0 \gg p_0$ выражение (5) переходит в (3), а при $p_0 \gg n_0$ переходит в (4).

Для собственного полупроводника $n_0 = p_0 = n_i$, где n_i — собственная концентрация носителей заряда. При этом

$$R_i = \frac{r(\mu_p - \mu_n)}{en_i(\mu_p + \mu_n)} = \frac{r}{en_i} \cdot \frac{1-b}{1+b}. \quad (6)$$

В выражениях (5) и (6) сделано допущение, что $r_n = r_p$. Это справедливо для любого механизма рассеяния при упругом рассеянии.

В сильном магнитном поле, когда выполняется условие (2), для полупроводника при наличии электронов и дырок коэффициент Холла

$$R_x = \frac{1}{e} \cdot \frac{p_0 - n_0 b^2 + b^2 \mu_p^2 B^2 (p_0 - n_0)}{(p_0 + n_0 b)^2 + b^2 \mu_p^2 B^2 (p_0 - n_0)} \quad (7)$$

При $B \rightarrow \infty$ коэффициент Холла

$$R_x = 1/[e(p_0 - n_0)]. \quad (8)$$

Для примесных полупроводников выражение (8) переходит в (3) и (4) при $r = 1$.

Как следует из равенств (3), (4) и (5) — (8), коэффициент Холла определяется такими параметрами полупроводникового материала, как концентрация и подвижность носителей заряда. Поэтому эффект Холла является эффективным методом определения этих параметров.

Если полупроводник имеет сложную зонную структуру, то выражение для коэффициента Холла усложняется. С учетом реальной структуры зоны проводимости кремния, имеющей шесть эквивалентных минимумов энергии с поверхностями постоянной энергии в виде эллипсоидов вращения, характеризующихся продольной и поперечной эффективными массами, в выражение для коэффициента Холла вводят множитель, зависящий от зонной структуры. Числовое значение этого множителя равно 0,866:

$$R_n = -0,868r/(en_0).$$

Валентная зона кремния состоит из зоны легких и тяжелых дырок; поверхности равной энергии для зоны тяжелых дырок являются гофрированными. Поэтому коэффициент Холла вычисляют с учетом концентрации легких и тяжелых дырок.

Магниторезистивный эффект. Этот эффект возникает вследствие искривления пути носителей заряда в магнитном поле и отклонения направления их движения от направления продольного электрического поля. В образце конечных размеров электрическая сила, создаваемая электрическим полем Холла, оказывает компенсирующее действие на силу Лоренца, в результате чего магниторезистивный эффект выражен слабее. Компенсирующее действие поля Холла тем меньше, чем меньше отношение длины образца к его ширине, что связано с шунтирующим действием токовых контактов на это поле.

Относительное магнитосопротивление, обусловленное искривлением пути носителей заряда в магнитном поле, при $j_x = const, j_y = 0$

$$\frac{\Delta p}{p} = \frac{E_x(B) - E_x(0)}{E_x(0)} = \frac{E_x(B)}{E_x(0)} - 1 \quad (9)$$

Для образца конечных размеров относительное магнитосопротивление измеряют при $E_x = const, E_y = 0$:

$$\frac{\Delta p_r}{p} = \frac{E_x(B) - E_x(0)}{E_x(0)} = \frac{j_x(B)}{j_x(0)} - 1 \quad (10)$$

В отличие от (9) магнитосопротивление, определяемое по (10), называют *геометрическим магнитосопротивлением*.

В обоих случаях для полупроводника с одним типом носителей заряда магнитосопротивление в области слабых магнитных полей пропорционально $(\mu_n B)^2$ или $(\mu_p B)^2$; оно выражается через коэффициенты α и α_r , зависящие от механизма рассеяния носителей заряда. Для электронного полупроводника

$$\Delta p / p = \alpha \mu_n^2 B^2; \quad \Delta p_r / p = \alpha_r \mu_n^2 B^2, \quad (11)$$

где

$$\alpha = \frac{\langle \tau^3 \rangle \langle \tau \rangle - \langle \tau^2 \rangle^2}{\langle \tau \rangle^4}; \quad \alpha_r = \frac{\langle \tau^3 \rangle}{\langle \tau \rangle^3}$$

Значения коэффициентов α и α_r приведены в табл. 1.

В области сильных магнитных полей геометрическое магнитосопротивление сохраняет квадратичную зависимость от $\mu_{np} B$, тогда как магнитосопротивление (9) насыщается до значения, не зависящего от подвижности носителей зарядов. Для

полупроводников с большими значениями подвижности носителей зарядов уже при небольших значениях магнитной индукции реализуется условие сильного магнитного поля, поэтому для таких полупроводников метод геометрического магнитосопротивления, как метод определения характеристик материала, предпочтительнее. Кроме того, магнитосопротивление (2.9) зависит от разности различным образом усредненных значений времени релаксации и может быть очень мало.

Сопутствующие явления. Рассмотрим ряд физических явлений, которые возникают совместно с эффектом Холла и, следовательно, могут вносить систематические и случайные погрешности при его измерении. К этим явлениям относятся: **поперечный термогальваномагнитный, термогальванический, термомагнитный эффекты.**

Поперечный термогальваномагнитный эффект.

Этот эффект возникает на помещенном в магнитном поле образце, по которому протекает электрический ток; он состоит в появлении градиента температуры в направлении, перпендикулярном магнитному полю и току. Если, как и в случае эффекта Холла, вектор магнитной индукции направлен вдоль оси z , а электрический ток — вдоль оси x , то вдоль оси y возникает градиент температуры, пропорциональный магнитной индукции и плотности электрического тока:

$$\frac{\partial T}{\partial y} = -k_{nm} B_z j_x, \quad (12)$$

где k_{nm} — коэффициент пропорциональности.

Физически поперечный термогальваномагнитный эффект обусловлен тем, что магнитное поле действует на «горячие» (быстрые) носители зарядов с большей силой, чем на «холодные» (медленные). Поэтому происходит пространственное разделение носителей зарядов по скоростям. В результате обмена энергией носителей зарядов с кристаллической решеткой возникает градиент температуры. Знак градиента температуры не зависит от типа носителей заряда, а зависит от направления магнитного поля и тока, так же как и полярность ЭДС Холла. Поэтому возникающая как следствие градиента температуры термо-ЭДС суммируется с ЭДС Холла и не может быть отделена от него при изменении направления тока или магнитного поля.

Термогальванический эффект.

Термогальванический эффект возникает при следующих условиях. Если образец, по которому в направлении x проходит тепловой поток, поместить в магнитное поле, направленное по оси z , то в направлении оси y возникает электрическое поле. При наличии градиента температуры потоку теплоты соответствует преимущественное движение «быстрых» носителей заряда от «горячего» конца образца к «холодному».

Напряженность электрического поля пропорциональна градиенту температуры и магнитной индукции:

$$E_y = k_{TT} B_z \frac{\partial T}{\partial x}, \quad (13)$$

где k_{TT} — коэффициент пропорциональности.

При этих же условиях возникает градиент температуры $\partial T/\partial y$, пропорциональный B_z и $\partial T/\partial x$:

$$\frac{\partial T}{\partial y} = k_{TM} B_x \frac{\partial T}{\partial x}, \quad (14)$$

где k_{TM} — коэффициент пропорциональности.

Этот эффект называют термомагнитным. Знак градиента температуры зависит от типа носителей заряда.

Таким образом, ЭДС термогальванического эффекта и термо-ЭДС, обусловленная термогальваномангнитным и термомагнитным эффектами, возникают совместно с ЭДС Холла и суммируются с ней. Градиент температуры $\partial T/\partial x$, обуславливающий возникновение этих эффектов, как правило, существует на образце при проведении холловских измерений в результате протекания тока через контакты к образцу и выделения на них некоторого количества теплоты. Знак градиента температуры, обусловленный эффектом Пельтье или нелинейностью ВАХ токовых контактов, зависит от направления тока через образец. Следовательно, знаки этих эффектов зависят от направления протекающего через образец тока.

Кроме того, существование градиента температуры $\partial T/\partial x$ может быть обусловлено неравномерным нагревом образца при проведении, например, температурных измерений или его неоднородностью.

Относительная оценка этих эффектов показывает, что наиболее важным является поперечный термогальваномангнитный эффект, для которого термо-ЭДС может составлять 10% от ЭДС Холла. Другие эффекты менее существенны и играют более важную роль в металлах.

Геометрия образцов и расположение контактов.

ЭДС Холла часто измеряют на пластинах прямоугольной формы. Для прямоугольного образца с размерами a , b и w (соответственно по направлениям x , y и z) ЭДС Холла определяют по (1). Если при измерении входящих в эту формулу величин использовать

единицы СИ (вольт, ампер, метр и тесла), то коэффициент Холла будет иметь размерность $\text{м}^3/\text{Ки}$:

$$R_x = \frac{wU_x}{I_x B_z}. \quad (15)$$

Таким образом, для определения коэффициента Холла R_x необходимо измерить холловскую разность потенциалов, ток, протекающий через образец, магнитную индукцию и геометрический размер образца w в направлении магнитного поля.

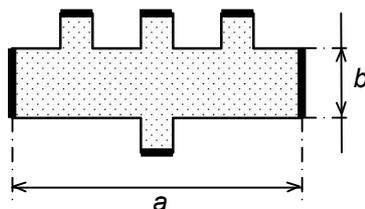


Рис 1. Модель образца для измерения ЭДС Холла

В соответствии с теорией эффекта Холла металлические контакты с образцом, предназначенные для измерения ЭДС Холла, должны иметь очень малую площадь, чтобы не искажать линии тока в образце. Практически в качестве контактов используют металлические зонды или сплавные контакты малых размеров. Применяют также боковые отростки, изготовленные с помощью ультразвуковой резки или химического травления (рис. 2.1). Образцу и боковым отросткам придают гантелеобразную форму. При этом технология изготовления контактов упрощается, облегчаются операции металлизации и сплавления. Такие контакты не искажают линии тока в образце и за счет большой площади имеют малое сопротивление контакта и более низкий уровень шума. Наличие нескольких боковых отростков позволяет одновременно с ЭДС Холла измерять удельное сопротивление образца. При измерении параметров диффузионных и эпитаксиальных слоев можно использовать стандартную технологию с применением фотолитографии.

Соотношение геометрических размеров образца и месторасположение контактов могут существенно влиять на измеряемую ЭДС Холла. Поскольку торцевые токовые контакты закорачивают ЭДС Холла, образец должен быть достаточно длинным. При симметричном расположении двух холловских контактов отношение $a/b \geq 4$, в противном случае необходимо вводить поправочную функцию, зависящую от этого отношения.

Для образца с двумя парами боковых гантелеобразных контактов во избежание закорачивания ими ЭДС Холла и искажения линий тока необходимо соблюсти ряд требований на соотношение длины и ширины образца, расположение и размеры боковых отростков: отношение ширины образца b к размеру отростка вдоль оси x должно быть не менее

3; расстояние от холловского контакта до токового контакта и расстояние между холловскими контактами— не менее $2b$ при длине образца

$1 \text{ см} \leq a \leq 1,5 \text{ см}$ и толщине $w \geq 0,1 \text{ см}$.

На пластинах произвольной геометрической формы ЭДС Холла измеряют аналогично измерению удельного сопротивления методом

Ван-дер-Пау. Электрический ток пропускают через контакты 1 и 3, а на контактах 2 и 4 измеряют разность потенциалов $U_{24}(0)$ в отсутствие и при наличии магнитного поля $U_{24}(B)$ ЭДС Холла

$$U_x = U_{24}(B) - U_{24}(0) = R_x \frac{I_{13} B}{w}. \quad (16)$$

Из свойств векторного поля следует, что отклонение носителей зарядов магнитным полем и возникновение объемных зарядов на боковой поверхности образца не изменяют линий тока в образце, т. е. линии тока остаются теми же, как и в отсутствие магнитного поля. ЭДС Холла образуется за счет возникновения электрического поля \vec{E}_n , перпендикулярного линиям плотности тока. Разность потенциалов между произвольными точками образца 2 и 4 в отсутствие магнитного поля равна разности потенциалов между точками 5 и 4, если точки 2 и 5 находятся на одной эквипотенциали. Эквипотенциальная поверхность показана на рис. 2 пунктиром.

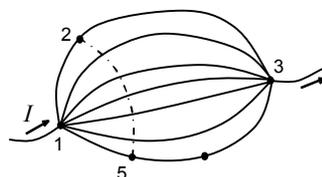


Рис. 2. Линии тока в пластине находящейся в поперечном магнитном поле

При наличии магнитного поля изменение разности потенциалов между точками 2 и 4 выражается интегралом от \vec{E}_n по пути, ортогональному линиям тока, между точками 2 и 5 и по пути вдоль линии тока между точками 5 к 4. Второй интеграл не изменяет разности потенциалов и, следовательно, равняется:

$$U_{25} = U_{24}(B) - U_{24}(0) = \int_2^4 \vec{E}_n d\vec{s} = R_x B \int_2^5 j ds = R_x \frac{I_{12} B}{w}. \quad (17)$$

Так как описанный метод измерения ЭДС Холла, так же как и аналогичный метод измерения удельного сопротивления, чувствителен к размерам контактов, то для уменьшения влияния контактов измерения проводят на образцах, имеющих форму,

показанную на рис. 1, а—в. Использование таких образцов снижает погрешности при измерении удельного сопротивления и ЭДС Холла во много раз.

Четырехзондовый метод измерения.

Для измерения ЭДС Холла применяют также метод, аналогичный четырехзондовому методу измерения удельного сопротивления с расположением зондов в вершинах квадрата. Если, например, образец — круглая пластина (рис. 3) и зонды размещены симметрично по ее краю, то этот случай аналогичен измерению ЭДС Холла на пластине произвольной формы. Следовательно, применимо соотношение (1).

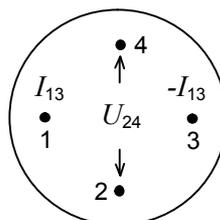


Рис. 3. Расположение зондов на пластине круглой формы

Однако по мере уменьшения расстояния между зондами при расположении системы зондов в центре пластины возникает закорачивание ЭДС Холла периферийной областью образца. Это явление называют вторичным эффектом Холла. Поэтому в формулу (2.12) необходимо вводить поправочную функцию $f(d/s)$, зависящую от отношения диаметра пластины к расстоянию между зондами. При увеличении отношения d/s образец приближается к пластине бесконечных размеров и, следовательно, ЭДС Холла уменьшается. Значения поправочной функции $f(d/s)$ приведены в табл. 2.

Таблица 2.

d/s	$f(d/s)$	d/s	$f(d/s)$	d/s	$f(d/s)$
$\sqrt{2}$	1	4,0	0,158	10,0	0,0255
2,0	0,59	5,0	0,102	20,0	0,004
3,0	0,279				

Поправочные функции рассчитаны также для пластин, представляющих собой половину или четверть круглой пластины, при расположении токовых зондов вдоль радиальной линии симметрии образца, причем потенциальные зонды лежат на линии, перпендикулярной линии симметрии и проходящей через ее середину.

Аналогичным образом применяют четырехзондовый метод измерения ЭДС Холла для пластин прямоугольной формы. Та же самая система зондов может использоваться для

измерения удельного сопротивления пластин с учетом соответствующих поправочных функций.

Четырехзондовый метод измерения ЭДС Холла целесообразно применять для низкоомных образцов, когда влияние контактных сопротивлений не препятствует осуществлению измерений, позволяя избежать трудоемких операций изготовления контактов.

Использование постоянного, переменного тока и магнитного поля.

Согласно (1), ЭДС Холла U_x , а следовательно, и коэффициент Холла можно измерить по крайней мере четырьмя различными способами, используя постоянный или переменный ток, а также постоянное или переменное магнитное поле.

Постоянное магнитное поле и постоянный ток. Этот метод измерения ЭДС Холла является самым простым и наиболее распространенным. От источника постоянного напряжения $ИН$ (рис. 4) через образец прямоугольной формы пропускают постоянный ток.

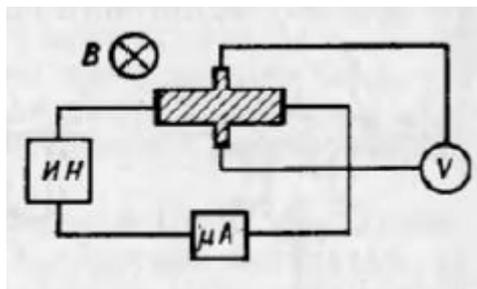


Рис. 4. Схема измерения ЭДС Холла при постоянном токе и магнитном поле

Образец помещают между полюсами постоянного магнита или электромагнита, создающего в рабочем зазоре магнитную индукцию до 1 Тл. Изменение магнитной индукции в рабочем зазоре магнита осуществляют, меняя расстояние между полюсами магнита, регулируя ток электромагнита, а также применяя магнитные концентраторы.

Чтобы исключить влияние контактных сопротивлений на результаты измерений, необходимо использовать компенсационный метод измерения ЭДС Холла или применять электронные вольтметры с высоким входным сопротивлением.

Как отмечалось, при измерении ЭДС Холла разность потенциалов на контактах складывается из четырех составляющих: ЭДС Холла U_x , термо-ЭДС поперечного термогальваномагнитного эффекта $U_{ПТ}$, ЭДС термогальванического эффекта $U_{ТГ}$ и термо-ЭДС термомагнитного эффекта $U_{ТМ}$. Существует еще одна составляющая, которая также влияет на измеряемую разность потенциалов,— напряжение U_0 , возникающее за счет

неточного расположения холловских контактов друг против друга, т. е. на разных эквипотенциальных поверхностях образца.

В общем случае разность потенциалов на контактах равна сумме пяти составляющих. Предполагая, что продольный градиент температуры $\partial T/\partial x$ вызван неравномерным нагревом образца за счет внешнего нагревателя и, следовательно, не зависит от направления тока через него, можно записать систему уравнений для разности потенциалов, которая соответствует всем возможным сочетаниям направлений тока и магнитного поля:

$$\text{при } +I, +B \quad U_1 = U_x + U_{IT} + U_{TI} + U_{TM} + U_0;$$

$$\text{при } -I, +B \quad U_2 = -U_x - U_{IT} + U_{TI} + U_{TM} - U_0;$$

$$\text{при } -I, -B \quad U_3 = U_x + U_{IT} - U_{TI} - U_{TM} - U_0;$$

$$\text{при } +I, -B \quad U_4 = -U_x - U_{IT} - U_{TI} - U_{TM} + U_0;$$

Проведя простые преобразования, получим

$$U_x + U_{IT} = (U_1 + U_2 + U_3 + U_4)/4. \quad (18)$$

Таким образом, выполнив четыре измерения при различных направлениях магнитного поля и тока через образец, из результатов измерений можно исключить ЭДС от всех сопутствующих эффектов, кроме термо-ЭДС поперечного термогальванемагнитного эффекта. Необходимость выполнения четырех измерений, требующих большого времени, является существенным недостатком этого метода измерения ЭДС Холла.

На практике не всегда можно избавиться от влияния неэквипотенциальности контактов, особенно для высокоомных образцов. Чтобы пропустить через высокоомный образец достаточный ток, увеличивают приложенное к образцу напряжение, что увеличивает разность потенциалов U_0 . Если эта разность потенциалов значительно превосходит ЭДС Холла, то точность измерения U_x может снизиться настолько, что сделает невозможным ее измерение. Поэтому стремятся не только уменьшить неэквипотенциальность контактов за счет их точной установки друг против друга, но и осуществить их электрическую балансировку.

Неэквипотенциальность холловских контактов можно устранить с помощью трехзондовой схемы измерения (рис. 5,а). Если потенциал зонда 1 меньше потенциала зонда 3, но больше потенциала зонда 2, то на потенциометре R_1 можно найти точку, потенциал которой равен потенциалу зонда 1.

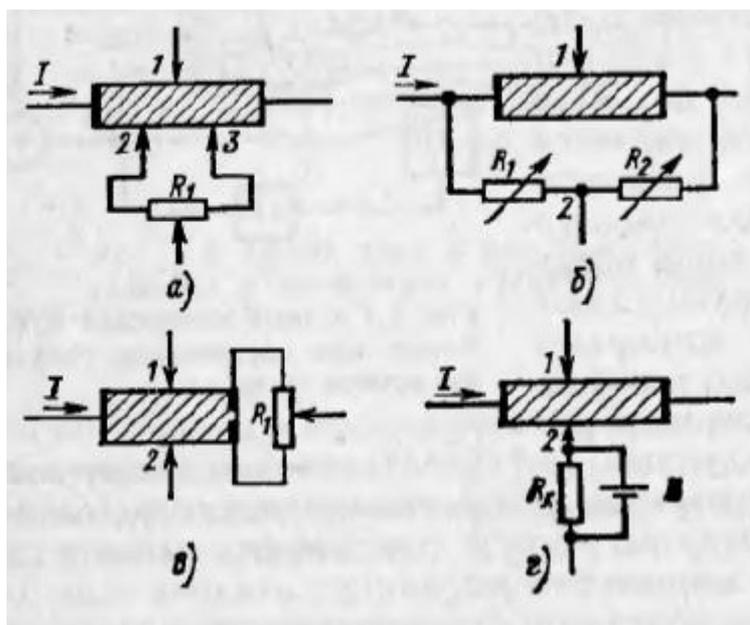


Рис. 5. Схемы устранения неэквипотенциальности холловских контактов

Чтобы потенциометр не шунтировал часть образца, заключенную между зондами 2 и 3, а ток через эти зонды был пренебрежимо мал, необходимо выполнить условие $R_1 \gg R_{23}$. Однако при этом следует учесть, что последовательное включение в измерительную цепь резистора с большим сопротивлением уменьшает чувствительность схемы измерения. Трехзондовую схему используют при измерении ЭДС Холла на низкоомных образцах.

На рис. 5.б представлена мостовая схема измерения холловской разности потенциалов. Части образца слева и справа от зонда 1 вместе с резисторами R_1 и R_2 составляют плечи моста. ЭДС Холла измеряют между зондом 1 и точкой 2. Потенциал точки 2 можно сделать равным потенциалу зонда 1, регулируя сопротивления R_1 и R_2 . К недостаткам схемы относится уменьшение чувствительности за счет измерения половины ЭДС Холла.

Схему, приведенную на рис. 5.в, называют схемой смещения эквипотенциалей. Как видно из рисунка, один из токовых контактов представляет собой, по существу, два отдельных контакта. С помощью потенциометра R_1 можно изменять протекающие через эти контакты токи и таким образом влиять на распределение электрического поля вдоль образца, добиваясь устранения неэквипотенциальности контактов.

Схема с отдельным источником компенсации (рис. 5.г) позволяет устранить неэквипотенциальность холловских контактов за счет компенсации U_0 разностью потенциалов на компенсирующем резисторе R_k при протекании через него тока от

внешнего источника питания. В этой схеме сопротивление R_K можно выбрать небольшим, чтобы не уменьшать чувствительности измерительной схемы.

Переменный ток или переменное магнитное поле. В этом случае возникает переменная ЭДС Холла с частотой тока или магнитного поля. При прохождении через образец переменного тока и использовании постоянного магнитного поля ЭДС Холла измеряют непосредственно с помощью селективного вольтметра.

Ввиду инерционности тепловых процессов ЭДС сопутствующих эффектов уменьшается уже при низких частотах (единицы герц), а усреднение измеряемой ЭДС Холла по двум направлениям магнитного поля устраняет влияние напряжения неэквипотенциальности холловских контактов. Если, кроме того, принять во внимание, что селективный усилитель не регистрирует постоянных составляющих напряжения, то постоянные во времени градиенты температуры не влияют на результаты измерений.

При прохождении через образец постоянного тока в переменном магнитном поле и усреднении измеренного напряжения по двум направлениям тока неустранимой остается лишь ЭДС термогальванического эффекта, обусловленная градиентом температуры вследствие эффекта Пельтье.

Переменный ток и переменное магнитное поле. Пусть магнитный поток и электрический ток в образце изменяются следующим образом:

$$\begin{aligned} B &= B_0 \cos(\omega_1 t + \theta_1); \\ I &= I_0 \cos(\omega_2 t + \theta_2); \end{aligned} \quad (19)$$

причем частоты $\omega_1 \neq \omega_2$, фазы $\theta_1 \neq \theta_2$. В этом случае ЭДС Холла

$$\begin{aligned} U_x &= R_x \frac{IB}{w} = R_x \frac{I_0 B_0}{w} \cos(\omega_1 t + \theta_1) \cos(\omega_2 t + \theta_2) = \\ &= R_x \frac{I_0 B_0}{2w} \{ \cos[(\omega_1 - \omega_2)t + (\theta_1 - \theta_2)] + \cos[(\omega_1 + \omega_2)t + (\theta_1 + \theta_2)] \} \end{aligned} \quad (20)$$

может быть измерена при суммарной или разностной частоте:

$$\begin{aligned} U_x [(\omega_1 - \omega_2)t] &= R_x \frac{I_0 B_0}{2w} \cos[(\omega_1 - \omega_2)t + (\theta_1 - \theta_2)]; \\ U_x [(\omega_1 + \omega_2)t] &= R_x \frac{I_0 B_0}{2w} \cos[(\omega_1 + \omega_2)t + (\theta_1 + \theta_2)]. \end{aligned} \quad (21)$$

При измерении среднеквадратичного значения ЭДС Холла I_0 — среднеквадратичное значение тока; B_0 — амплитудное значение магнитной индукции. При $\omega_1 = \omega_2 = \omega$ ЭДС

Холла складывается из постоянной составляющей ЭДС и переменной составляющей удвоенной частоты:

$$U_x = R_x \frac{I_0 B_0}{2w} \cos(\theta_1 + \theta_2) + R_x \frac{I_0 B_0}{2w} \cos[2\omega t + (\theta_1 + \theta_2)]. \quad (22)$$

Если проводить измерения постоянной составляющей ЭДС или переменной составляющей удвоенной частоты, то измеренная разность потенциалов будет суммой ЭДС Холла и ЭДС, возникающей за счет термо-ЭДС поперечного термогальваномагнитного эффекта. Остальные эффекты не будут оказывать влияния, так как создают ЭДС на частоте ω . Таким образом, для определения ЭДС Холла достаточно однократного измерения.

Большими достоинствами обладают схемы измерения ЭДС Холла, использующие электрический ток и магнитное поле, изменяющиеся с различными частотами. Применяя селективный вольтметр, регистрирующий, например, сигнал только на разностной частоте, можно отсеять все сопутствующие ЭДС. Однако при выборе рабочих частот следует иметь в виду возможность появления сигналов с частотой ЭДС Холла в результате возникновения гармоник первичных частот ω_1, ω_2 , а также гармоник частоты сети переменного тока и их комбинационных частот. Одним из основных источников появления таких сигналов являются нелинейность входных цепей измерителя ЭДС Холла и контактов самого образца, на которых могут смешиваться напряжения с частотами ω_1 и ω_2 , например напряжение неэквипотенциальности холловских контактов (ω_2) и напряжение электромагнитной наводки, создаваемое магнитным полем на частоте ω_1 .

Упрощенная схема одного из вариантов такой измерительной установки представлена на рис. 6. Схема имеет следующие основные элементы: низкочастотный генератор переменного тока G с низкоомным симметричным выходом; измеритель тока A образца, включенный в середину выходной обмотки трансформатора Tr , селективный усилитель- U с узкой полосой пропускания, фазочувствительный детектор $ФД$, двухкоординатный самописец $Сп$ и смеситель $См$. Для создания магнитного поля, в которое помещают образец, используют электромагнит, питаемый от промышленной сети переменного тока или от низкочастотного генератора. Источник магнитного поля на схеме не показан.

Измерительная схема работает следующим образом. Переменное напряжение частотой ω_2 от генератора G через трансформатор Tr подается на образец. ЭДС, возникающая на холловских контактах образца, один из которых заземлен, усиливается усилителем U , настроенным на разностную частоту $\omega_1 - \omega_2$.

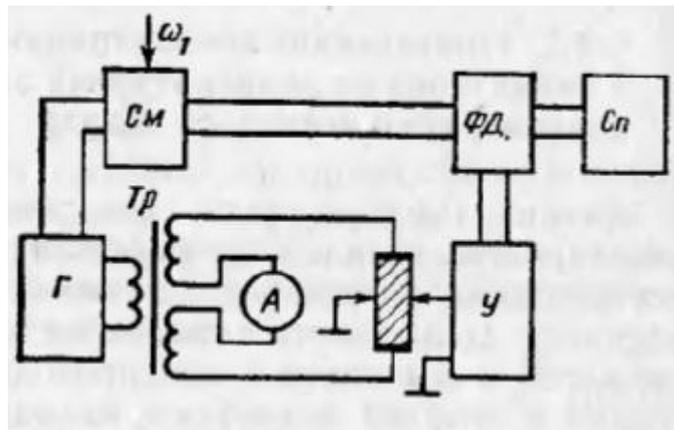


Рис. 6. Структурная схема измерения ЭДС Холла при переменном токе и переменном магнитном поле

Напряжение неэквивипотенциальности холловских контактов имеет частоту тока ω_2 и потому не регистрируется усилителем. Высокое входное сопротивление усилителя (10^6 Ом и более) достигается благодаря использованию электрометрического входного каскада и максимального уменьшения паразитных входных емкостей. Узкая полоса пропускания снижает, кроме того, уровень шумов усилителя. ЭДС Холла регистрируется самописцем *Sp*. Применение двухкоординатного самописца позволяет осуществить автоматическую регистрацию температурной зависимости ЭДС Холла, если, например, измерять температуру образца, находящегося в криостате, с помощью термопары.

Смеситель *Sm*, на который одновременно подаются напряжения с частотами ω_1 и ω_2 , служит для создания опорного напряжения с частотой $\omega_1 - \omega_2$ для *ФД*. В качестве смесителя частот используют датчики Холла из антимонида индия. Тип электропроводности образца определяют по знаку фазы ЭДС Холла относительно фазы опорного напряжения.

Перечислим факторы, влияющие на систематическую погрешность измерения ЭДС Холла: неэквивипотенциальность холловских контактов; термо-ЭДС, обусловленная внешним градиентом температуры и эффектом Пельтье; сопутствующие явления; соотношение между геометрическими размерами образца, расположение и размеры токовых и потенциальных контактов; электромагнитные наводки и ЭДС на комбинационных частотах.

Кроме того, на погрешность измерений влияют и следующие факторы, рассмотренные ранее применительно к методам измерения удельного сопротивления: фотопроводимость и фото-ЭДС, возникающая за счет освещения образца; инжекция носителей заряда через контакты при протекании тока через образец; поверхностные токи утечки как в измерительной цепи, так и по поверхности образца; увеличение температуры образца в процессе измерений за счет нагрева образца электрическим током.

Влияние многих источников погрешностей измерений можно уменьшить, используя минимальный ток через образец. Рекомендуется устанавливать такой ток, чтобы напряженность электрического поля в образце не превышала $10^2 \text{ В} \cdot \text{см}^{-1}$.

2. Определение концентрации доноров и акцепторов по температурной зависимости концентрации носителей заряда

Анализ температурной зависимости концентрации носителей заряда, определенной по результатам холловских измерений, является одним из основных способов изучения примесей в полупроводниках. Цель такого анализа состоит в определении типа и числа примесей, имеющих в полупроводниковом материале, их концентраций и энергий ионизации, вклада каждого из локальных уровней в общую концентрацию носителей заряда, спектра возбужденных состояний уровней и других характеристик. Основные причины, ограничивающие применение этого метода, связаны, например, с электропроводностью по примесным уровням в области низкой температуры и вырождением при больших концентрациях примесей.

В общем виде анализ температурной зависимости концентрации носителей заряда в полупроводнике представляет собой сложную задачу. Полупроводник может содержать несколько видов донорных и акцепторных примесей, что соответствует определенной степени его компенсации. Некоторые примеси могут отдавать или присоединять не один, а несколько электронов. Примесный атом может находиться не только в основном энергетическом состоянии, но и в возбужденном. Основное и возбужденное состояния характеризуются соответствующими факторами спинового вырождения. В некоторых полупроводниковых материалах валентная зона в центре зоны Бриллюэна оказывается вырожденной, что проявляется в увеличении фактора вырождения акцепторного состояния. Поэтому в каждом конкретном случае следует решить, чем можно пренебречь.

Предположим, что в запрещенной зоне полупроводникового материала имеются донорные и акцепторные примесные уровни с энергией E_d , E_a и концентрацией N_d , N_a , причем $N_d > N_a$. Концентрация отрицательных зарядов определяется суммарной концентрацией электронов n_0 в зоне проводимости и концентрацией электронов на уровнях акцепторной примеси. Будем считать, что каждый атом акцепторной примеси захватывает только один электрон. Концентрация положительных зарядов складывается из дырок валентной зоны p_0 и тех атомов донорной примеси, которые отдали свои электроны в зону проводимости. Концентрация электронов на уровнях донорной примеси,

т. е. концентрация нейтральных доноров,

$$n_d = \frac{N_d}{1 + g_d^{-1} \exp \frac{E_d - E_F}{kT}}, \quad (23)$$

а концентрация отрицательно заряженных ионов акцепторной примеси

$$N_{\bar{a}} = \frac{N_a}{1 + g_a \exp \frac{E_a - E_F}{kT}}, \quad (24)$$

где g_a , g_d — факторы спинового вырождения примесных уровней; k — постоянная Больцмана.

По условию электрической нейтральности, число положительных зарядов равно числу отрицательных зарядов; следовательно, уравнение электронейтральности имеет вид

$$n_0 + N_{\bar{a}} = (N_d - n_d) + p_0 \quad (25)$$

Уравнение (25) позволяет анализировать температурную зависимость концентрации носителей заряда.

Для невырожденного донорного полупроводника в температурной области примесной электропроводности концентрацией дырок в валентной зоне можно пренебречь по сравнению с концентрацией электронов ($n_0 \gg p_0$). Кроме того, можно считать, что все акцепторы ионизированы, т. е. $N_{\bar{a}} = N_a$. При этих предположениях уравнение электронейтральности приобретает следующий вид:

$$\frac{(n_0 + N_a)}{N_d - N_a - n_0} = g_d^{-1} \exp \left(\frac{E_d - E_F}{kT} \right) \quad (26)$$

Следует отметить, что уравнение (26) справедливо для любого закона дисперсии в разрешенных зонах и произвольного вырождения. Это уравнение описывает одну из наиболее простых моделей примесей в полупроводниках. Если экспериментальные данные не соответствуют этой модели, то нужно искать способы уточнения модели и расширения экспериментальных возможностей, хотя соответствие экспериментальных данных выбранной модели еще не означает, что модель выбрана правильно. Вопрос о применимости использованной модели является одним из основных вопросов при интерпретации экспериментальных результатов.

Если предположить, что полупроводник невырожден и что закон дисперсии в разрешенных зонах энергии квадратичен, то при

$$n_0 = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \quad (27)$$

и

$$\exp\left(\frac{E_d - E_F}{kT}\right) = \exp\left(-\frac{E_c - E_d}{kT}\right) \exp\left(\frac{E_c - E_F}{kT}\right) = \frac{N_c}{n_0} \exp\left(-\frac{\Delta E_d}{kT}\right)$$

уравнение электронейтральности примет вид

$$\frac{n_0(n_0 + N_a)}{N_d - N_a - n_0} = g_d^{-1} N_c \exp\left(-\frac{\Delta E_d}{kT}\right), \quad (28)$$

где $E_c - E_d = \Delta E_d$ — энергия ионизации доноров;

$$N_c = 2 \left(\frac{2\pi m_n kT}{h^2} \right)^{3/2} = 4,83 \cdot 10^{15} \left(\frac{m_n}{m_0} \right)^{3/2} T^{3/2} = AT^{3/2}$$

— эффективная плотность квантовых состояний в зоне проводимости;

m_n — эффективная масса электрона в зоне проводимости; m_0 — масса свободного электрона.

Замена функции Ферми в выражении для концентрации носителей заряда в зоне проводимости

$$h_0 = N_c F_{1/2}\left(\frac{E_F - E_c}{kT}\right)$$

на экспоненту (27) дает хорошее приближение, если

$$\frac{E_F - E_c}{kT} < -3.$$

Уравнение (28) можно использовать во всем интервале температур, где выполняется это неравенство. По таблицам функции Ферми можно определить положение уровня Ферми, если концентрация носителей заряда известна.

Для невырожденного акцепторного полупроводника соотношение имеет вид

$$\frac{p_0(p_0 + N_a)}{N_d - N_a - p_0} = g_a N_v \exp\left(-\frac{\Delta E_a}{kT}\right),$$

где N_v — эффективная плотность квантовых состояний в валентной зоне; $\Delta E_a = E_a - E_v$ — энергия ионизации акцепторов.

Если концентрация компенсирующей акцепторной примеси невелика, а температура полупроводника такова, что уровень донорной примеси далек от истощения, т. е.

$$N_d \gg n_0 \gg N_a, \quad (29)$$

то, пренебрегая N_a по сравнению с n_0 в числителе и N_a и n_0 по сравнению с N_d в знаменателе (28), получаем

$$n_0 = \left(g_d^{-1} N_d N_c \right)^{1/2} \exp\left(-\frac{\Delta E_d}{2kT} \right). \quad (30)$$

Преобразуем выражение (30):

$$-\frac{\Delta E_d}{T} = k \left[\ln g_d + 2 \ln \left(n_0 T^{-3/4} \right) - \ln \left(AN_d \right) \right]. \quad (31)$$

Продифференцировав (31) по $1/T$, определим

$$\Delta E_d = -2k \frac{d \left[\ln \left(n_0 T^{-3/4} \right) \right]}{d(1/T)}. \quad (32)$$

Экспериментальную зависимость $\ln(n_0)$ из соображений удобства масштаба изображают графически в виде функции от $1000/T$ или $100/T$. Поэтому при анализе выражение (32) целесообразно представить в более общем виде:

$$\Delta E_d = -0,43kT_0 \frac{d \left[\lg \left(n_0 T^{-3/4} \right) \right]}{d(T_0/T)}, \quad (33)$$

где T_0 — температура, которая выбирается с учетом удобства построения графика.

В форме конечных разностей выражение (33) приобретает вид

$$\Delta E_d = -3,968 \cdot 10^{-4} T_0 \frac{\lg \left[n_2 T_2^{-3/4} \left(n_0 T^{-3/4} \right) \right]}{T_0/T_2 - T_0/T_1},$$

где $k = 8,62 \cdot 10^{-5}$ эВ/К; ΔE_d — в эВ. Таким образом, энергия ионизации примеси может быть определена по тангенсу угла наклона прямолинейного участка зависимости $\lg(n_0 T)^{3/4}$ от T_0/T . После того как значение энергии ионизации примеси будет установлено,

значение $g_d^{-1} N_d$ можно вычислить по (30):

$$g_d^{-1} N_d = \frac{n_0^2}{N_c} \exp\left(\frac{\Delta E_d}{kT} \right).$$

При сильной компенсации концентрация компенсирующей примеси N_a сравнима с концентрацией основной примеси N_d , хотя $N_a < N_d$. Если в некотором интервале

температур концентрация электронов много меньше концентрации компенсирующей примеси и разности между основной и компенсирующей примесями, т. е.

$$n_0 \ll N_a, \quad (34)$$

$$n_0 \ll N_d - N_a \quad (35)$$

то согласно (28),

$$n_0 = g_d^{-1} \frac{N_d - N_a}{N_a} N_c \exp\left(-\frac{\Delta E_d}{kT}\right). \quad (36)$$

Отсюда

$$-\frac{\Delta E_d}{T} = kT \left[\ln g_d + \ln(n_0 T^{-3/4}) + \ln \frac{N_a}{A(N_d - N_a)} \right]. \quad (37)$$

Аналогично случаю слабой компенсации,

$$\Delta E_d = -0,43kT_0 \frac{d \lg(n_0 T^{-3/2})}{d(T_0/T)}. \quad (38)$$

В форме конечных разностей выражение (38) имеет вид

$$\Delta E_d = -0,43kT_0 \frac{d \lg \left[n_2 T_2^{-3/2} (n_1 T_1^{-3/2}) \right]}{T_0/T_2 - T_0/T_1},$$

или в электрон-вольтах

$$\Delta E_d = -1,984 \cdot 10^{-4} T_0 \frac{\lg \left[n_2 T_2^{-3/2} / (n_1 T_1^{-3/2}) \right]}{T_0/T_2 - T_0/T_1}.$$

Таким образом, для определения энергии ионизации примеси в данном случае необходимо построить зависимость $\lg(n_0 T^{-3/2})$ от T_0/T и по ее линейному участку вычислить тангенс угла наклона. В соответствии с (2.24) при установленном значении ΔE_d

$$g_d^{-1} \frac{N_d - N_a}{N_a} = \frac{n_0 T^{-3/2}}{A} \exp\left(\frac{\Delta E_d}{kT}\right) \quad (39)$$

Основное преимущество описанных графических способов определения энергии ионизации примеси по данным эффекта Холла заключается в простоте вычисления искомого

величин. Однако с их помощью нельзя найти остальные параметры локальных уровней и, что особенно важно, нельзя однозначно выбрать необходимое для расчета уравнение. Так как соотношение между величинами N_d , N_a и n_0 заранее не известно, то неизвестно, какое из уравнений (30) или (36) следует использовать.

Выбор способа расчета становится однозначным в том случае, когда в исследованном температурном интервале удается экспериментально наблюдать переход от зависимости (2.24) в интервале низких температур к зависимости (30) в интервале более высоких температур, что подтверждается уменьшением угла наклона экспериментальной зависимости $\lg n_0(1/T)$ примерно в два раза. Для этого, однако, необходимо, чтобы в широком температурном интервале действовал только один локальный уровень и условия (34), (35) и (29) выполнялись последовательно при повышении температуры. При дальнейшем росте температуры концентрация электронов стремится к насыщению: $n_0 = N_d - N_a$. На рис. 7 представлены зависимости $\lg n_0(10^2/T)$ для полупроводника, содержащего некоторое число доноров и акцепторов. Для наглядности предположим, что $N_d - N_a = 10^{16} \text{ см}^{-3}$ во всех рассматриваемых случаях. Донорные уровни расположены на 0,01 эВ ниже дна зоны проводимости, $g_d = 2$, $m_n = 0,25m_0$. Кривым на рис. 7 соответствуют различные степени компенсации.

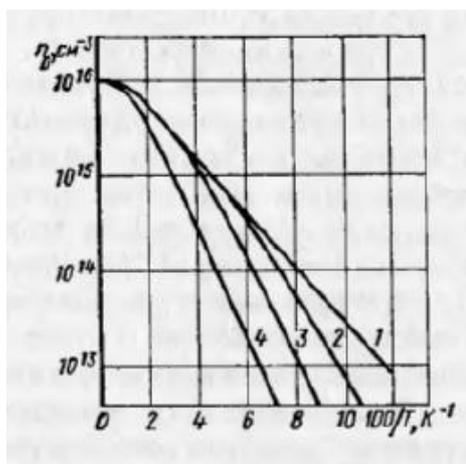


Рис. 7. Теоретически рассчитанные зависимости логарифма концентрации электронов по от $100/T$

Кривая 1 рассчитана при $N_a = 0$ (компенсация отсутствует), кривая 2 — при $N_a = 10^{14} \text{ см}^{-3}$ (степень компенсации около 1%), кривая 3 — при $N_a = 10^{15} \text{ см}^{-3}$ (степень компенсации около 9%); кривая 4 — при $N_a = 10^{16} \text{ см}^{-3}$ (степень компенсации 50%). Ход

кривых 2 и 3 свидетельствует о том, что влияние компенсации начинает проявляться лишь при достаточно низких температурах, когда концентрация электронов становится сравнимой с N_a . При этом угол наклона кривых возрастает, достигая значения $\Delta E_d/k$. При этих температурах справедливо соотношение (37). В сильно компенсированных материалах (кривая 4) почти во всем интервале температур применимо выражение

$$n_0 = \frac{N_d - N_a}{1 + \frac{N_a}{g_d^{-1} N_c} \exp\left(\frac{\Delta E_d}{kT}\right)}$$

Таким образом, температурная зависимость концентрации носителей заряда позволяет по предельному значению концентрации в области насыщения определить $N_d - N_a$; по углу наклона зависимости при низких температурах можно с достаточной точностью оценить энергию ионизации примеси, а по известным значениям $(N_d - N_a)$ и ΔE_d , согласно (39), найти N_a/g_d^{-1} . Если фактор спинового вырождения известен, то параметры N_d, N_a и ΔE_d определены (рис. 8).

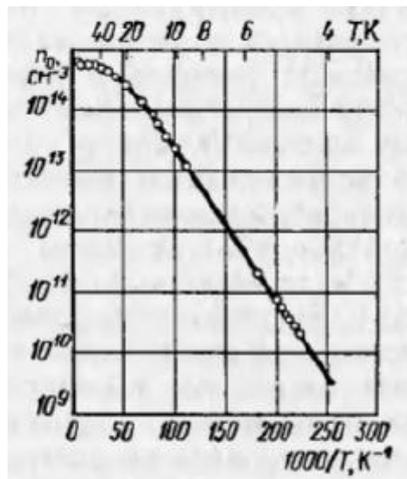


Рис. 8. Экспериментальная зависимость $\lg n_0$ от $1000/T$ для образца из арсенида

$$\text{галлия: } N_d = 7,76 \cdot 10^{14} \text{ см}^{-3}; N_a = 1,86 \cdot 10^{14} \text{ см}^{-3}; \Delta E_d = 4,28 \text{ мэВ}$$

В более общем случае, когда фактор спинового вырождения неизвестен, для определения параметров используют уравнение (28). Подбором параметров добиваются совпадения теоретической и экспериментальной зависимостей во всей области температур.

Если параметры подобраны правильно, то зависимость $\ln \left[T^{-3/2} \frac{g_d n_0 (n_0 + N_a)}{A(N_d - N_a - n_0)} \right]$

от $1/T$ будет линейной. Наилучше совпадения теоретической и экспериментальной зависимости добиваются при использовании метода наименьших квадратов.

При вычислении концентрации носителей заряда в зависимости от температуры по коэффициенту Холла согласно (3) или (4) следует принимать во внимание температурную зависимость холл-фактора $r(T)$. Незнание механизмов рассеяния, зонной структуры и закона дисперсии ведет к возникновению систематических погрешностей при нахождении концентрации носителей зарядов, примесей, их энергии ионизации и других параметров. Однако на практике определение механизмов рассеяния, т. е. учет холл-фактора, представляет собой трудную задачу, поэтому часто полагают $r(T) = 1$.

В настоящее время для анализа температурной зависимости концентрации носителей заряда используют ЭВМ. С помощью ЭВМ можно легко обрабатывать большое количество исходных данных, рассчитывать уравнения (26) любой степени сложности, анализировать сложные модели примесных уровней. В ряде работ, посвященных исследованию кремния, арсенида галлия, германия проанализированы экспериментальные данные с помощью модели, учитывающей возбужденные состояния и расщепление основного состояния примеси, с учетом влияния глубоких примесных уровней, многовалентных примесей и примесей двух различных типов.

При наличии возбужденных примесных состояний концентрация свободных носителей заряда уменьшается вследствие их распределения по возбужденным уровням. Основное состояние соответствует фиксированному значению энергии ниже дна зоны проводимости (для донорной примеси). Увеличение концентрации доноров приводит к уширению сильновозбужденных состояний, перекрытию волновых функций этих состояний и образованию зоны непрерывного спектра, соприкасающейся с зоной проводимости.

Следовательно, дно зоны проводимости опускается и энергия термической активации понижается. Для оставшихся дискретных возбужденных уровней разности энергии не изменяются. Структура возбужденных состояний водородоподобной модели описывается формулой $E_k = (1 - k^{-2})E_1$, где $k \geq 2$ — целое число; $E_1 = E_c - E_d$ фиксировано в пределе при малой концентрации примеси. Например, для арсенида галлия n -типа с концентрацией доноров более $8 \cdot 10^{14} \text{ см}^{-3}$ уширение уровней настолько велико, что все возбужденные состояния перекрываются и не влияют на функцию распределения. При концентрации доноров 10^{14} см^{-3} нужно учитывать лишь первое возбужденное состояние.

С учетом возбужденных состояний уравнение (28) имеет вид

$$\frac{n_0(n_0 + N_a)}{N_d - N_a - n_0} = \frac{g_d^{-1} N_c \exp\left(-\frac{\Delta E_d}{kT}\right)}{1 + F}.$$

Где

$$F = \sum_{k=2}^{\infty} \frac{g_1}{g_k} \exp\left(-\frac{\Delta E_k}{kT}\right);$$

g_1, g_k — факторы вырождения основного и возбужденных состояний.

Наличие нескольких примесей различного типа усложняет анализ экспериментальных данных. Однако он упрощается, когда соответствующие энергетические уровни хорошо разделены, т. е. в некотором температурном интервале только один из уровней изменяет свое заполнение. При определенных условиях ненаблюдаемыми могут оказаться все уровни одного и того же типа. Рассмотрим, например, компенсированный полупроводник p -типа с двумя акцепторными уровнями. Структура энергетических уровней такого полупроводника приведена на рис. 9,а.

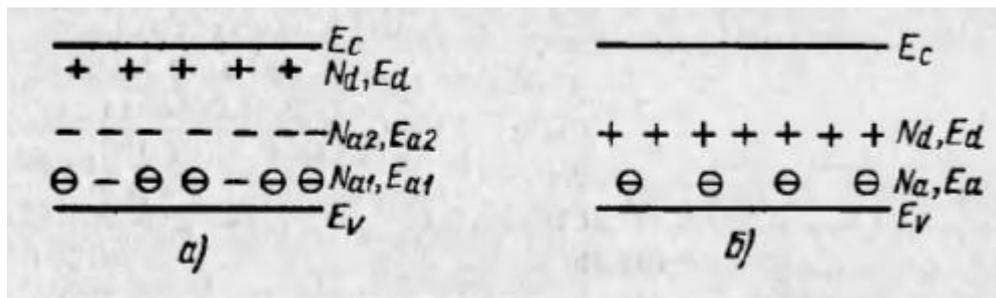


Рис. 9. Зонные диаграммы полупроводников p -типа:

$$a — N_d^+ - N_d; N_d < N_a; \quad б — N_a^- - N_a; N_d < N_a$$

Если концентрация компенсирующих доноров N_d меньше концентрации мелких акцепторов N_a , то оба акцепторных уровня проявляются в температурной зависимости $p_0(T)$. Если $N_d > N_{a1}$, то мелкие акцепторные уровни будут заполнены при любой температуре и можно наблюдать активационный переход только на второй акцепторный уровень E_{a2} . Таким образом, оба типа акцепторных уровней могут быть зарегистрированы лишь при $N_d < N_a$. В противоположном случае наличие ионизированных акцепторов может повлиять на подвижность дырок.

В полупроводнике p -типа с мелкими акцепторными и глубокими донорными уровнями, расположенными в нижней половине запрещенной зоны, глубокие доноры могут проявлять себя как акцепторы (рис. 9,б). Кроме того, если $N_d > N_a$, то при нулевой темпе-

ратуре мелкие акцепторные уровни заполнены электронами и их нельзя обнаружить с помощью эффекта Холла. Нужно отметить, что глубокий донорный уровень можно отличить от глубокого акцепторного уровня (рис. 9, *a*), поскольку при захвате электрона донор из состояния с положительным зарядом переходит в нейтральное состояние, тогда как акцептор из нейтрального состояния переходит в состояние с отрицательным зарядом. Хотя температурная зависимость ЭДС Холла позволяет выделить глубокие уровни по их влиянию на концентрацию носителей заряда, чувствительность этих измерений относительно глубоких примесных центров невелика. Она ограничена точностью измерения коэффициента Холла и обычно не превышает десятой доли концентрации носителей заряда, т. е. значения $0,1(N_d - N_a)$. Поэтому глубокие примесные уровни исследуют с помощью емкостных методов, обладающих большей чувствительностью.

Лекция 5. Электронный парамагнитный резонанс как метод изучения структуры материалов.

1. Электронный парамагнитный резонанс: основные принципы и определения

Электронный парамагнитный резонанс (ЭПР) – это явление резонансного поглощения или излучения электромагнитных волн радиочастотного диапазона (10^9 - 10^{12} Гц) в газах, веществах и твердых телах, содержащих частицы с не скомпенсированными электронными моментами (парамагнитные частицы). В качестве таких частиц могут выступать ионизированные атомы или молекулы, а также свободные (в газах) или захваченные дефектами (в твердых телах) электроны. ЭПР открыт в 1944 казанским ученым Евгением Константиновичем Завойским.

Энергетический спектр парамагнитных частиц является дискретным (т.е., квантовыми законами разрешены только отдельные значения энергии). В отсутствие внешних возмущений спектр представляет собой группы уровней с близкими или совпадающими значениями энергии. Причем энергетические интервалы между «центрами тяжести» указанными групп уровней намного превышают интервалы между уровнями в каждой группе. Такие выделенные группы уровней называются термами. Каждый терм характеризуется определенными значениями суммарного орбитального момента и суммарного спинового момента (L и S). Значения этих моментов присутствуют в обозначении каждого терма – ^{2S+1}L , где L заменяется буквами S, P, D, F (и т.д.) в случаях $L = 0, 1, 2, 3, \dots$ (соответственно). Численное значение левого верхнего индекса ($2S+1$) указывает на число спиновых уровней энергии, соответствующих каждому орбитальному состоянию данного терма. Так как число орбитальных состояний равно $2L+1$, общее число состояний данного терма равно $(2L+1) \cdot (2S+1)$. Расщепления уровней внутри терма обусловлены спин-орбитальным взаимодействием. Однако эти расщепления уровней не полные; некоторым значениям энергии соответствуют несколько квантовых состояний, характеризуемых величиной полного момента $J = L + S$.

В жидкостях и твердых телах (вследствие небольших расстояний между атомами или молекулами вещества) орбитальное и спиновое движения электронов парамагнитной частицы возмущаются полями, индуцируемыми ее соседями (лигандами). Под воздействием полей лигандов термы парамагнитной частицы расщепляются, поэтому степень вырождения уровней энергии существенно понижается. Поскольку магнитная составляющая поля лигандов пренебрежимо мала по сравнению с электрической составляющей, в основном происходит возмущение орбитального движения электронов парамагнитной частицы; поле

лигандов действует на спиновые моменты только через спин-орбитальное взаимодействие, которое обычно бывает намного слабее действия поля лигандов.

Но в некоторых случаях магнитные электроны (электроны с не скомпенсированными спиновыми моментами) находятся на внутренних незаполненных электронных оболочках, оказываясь экранированными от внешних полей внешней заполненной электронной оболочкой. В таких случаях влияние поля лигандов очень сильно ослабляется, поэтому спин-орбитальное взаимодействие может стать преобладающим [2].

В кристаллах с парамагнитными ионами число ближайших к ним лигандов в преобладающем большинстве случаев равно шести [14-16]. Реже это число равно четырем или восьми. Многогранники, вершины которых совпадают с центрами ближайших лигандов парамагнитного иона, называются координационными многогранниками. Если кристалл имеет решетку кубической группы симметрии, то координационными многогранниками парамагнитных ионов оказываются октаэдр (6 лигандов), тетраэдр (4 лиганда) или куб (8 лигандов). Вид координационных многогранников парамагнитных ионов является важной информацией, характеризующей их магнитные свойства. Именно лиганды, чьи ядра являются реперами для построения координационного многогранника, в основном определяют топологию поля лигандов и именно они в основном определяют статические магнитные свойства кристалла в целом. Обычно эти свойства описываются с помощью кластерной модели, в которой рассматривают только небольшую часть кристалла, включающую в себя один парамагнитный ион и ближайшие к нему лиганды. Эту небольшую совокупность ионов называют парамагнитным центром или комплексом [16].

В кристаллах кубической группы симметрия координационных многогранников во многих случаях также является кубической, поскольку содержат все элементы симметрии, характерные для групп O_h или T_d . Однако, в случаях парамагнитных ионов с вырожденными основными орбитальными состояниями координационные многогранники деформируются так, что симметрия поля лигандов понижается до тетрагональной, тригональной или ромбической. Понижение симметрии координационного многогранника происходит также в случаях локальной компенсации заряда не изовалентной примеси. В таких случаях спектры ЭПР позволяют установить факт понижения симметрии поля лигандов и описать искажения симметрии координационного многогранника количественно.

Выше мы отметили, что для теоретического описания спектра энергий парамагнитных ионов в твердых телах важно соотношение между интенсивностью поля лигандов и величиной спин-орбитального взаимодействия. Случай слабого поля лигандов реализуется для редкоземельных парамагнитных ионов с внутренней незаполненной f -оболочкой [1-4]. Здесь использование понятия орбитального термина ^{2S+1}L оказывается

неоправданным. Более удобным становится понятие мультиплета, характеризуемого величиной полного вращательного момента парамагнитного иона $J = L + S$. Как и суммарный спиновый момент S , полный момент J характеризует вращательные движения электронов парамагнитного иона. Поэтому в случае редкоземельных парамагнитных ионов для характеристики их магнитных свойств в основном состоянии может быть использована величина J .

Здесь мы рассмотрим подробнее случай сильного поля лигандов.

Этот случай реализуется для парамагнитных ионов с внешней незаполненной d -оболочкой [1-4, 7]. Согласно теории возмущений, широко используемой в квантовой механике, приближенное описание свойств таких ионов в твердых телах базируется на том, что спин-орбитальное взаимодействие в них намного слабее поля лигандов. Поэтому в качестве нулевого приближения используют $(2L+1) \cdot (2S+1)$ состояний полностью вырожденного основного терма рассматриваемого иона (т.е. предполагается, что спин-орбитальное взаимодействие отсутствует). Затем рассматривается влияние поля лигандов на состояния нулевого приближения. Чаще всего поле лигандов приводит к такому расщеплению состояний нулевого приближения, при котором наименьшими энергиями начинают обладать $(2S+1)$ состояний; энергии остальных состояний будет намного выше (по сравнению с интервалами между уровнями основных $(2S+1)$ состояний и даже по сравнению с величиной спин-орбитального взаимодействия). В качестве одной из характеристик для нижней группы состояний используют эффективный спиновый момент $S_{\text{эф}}$, который в описываемом случае совпадает с суммарным спиновым моментом иона S . Другие характеристики можно определить, если учесть спин-орбитальное взаимодействие как возмущение. В первом порядке теории возмущений спин-орбитальное взаимодействие не дает никакого вклада, поскольку среднее значение орбитального момента в данном случае примерно равно нулю. Учет поправок к уровням энергии и к волновым функциям состояний приводит к дополнительным характеристикам, описывающим анизотропию магнитных свойств парамагнитного иона с незаполненной d -оболочкой (d -иона).

В случаях, когда поле лигандов имеет высокую симметрию, нижняя группа достаточно близко расположенных уровней d -иона может содержать несколько орбитальных состояний [16-22]. Несмотря на то, что в этом случае очень эффективной становится взаимодействие парамагнитного иона с колебаниями решетки (эффект Яна-Теллера), последнее не всегда приводит к достаточно большим энергетическим интервалам между уровнями нижней группы состояний. Так как среднее значение орбитального момента в рассматриваемом случае будет сильно отличаться от нуля, на систему нижней группы уровней окажет сильное влияние спин-орбитальное взаимодействие. В результате эти

уровни изменят свое взаимное положение и в самом низу спектра энергий окажутся несколько состояний, которые могут быть охарактеризованы эффективным спиновым моментом $S_{\text{эф}} < S$. Однако и в этом случае при достаточно низких температурах магнитные свойства d -иона будут анизотропными и потребуют для своего описания дополнительных характеристик.

Взаимодействие парамагнитного иона с внешним магнитным полем называется зеемановским взаимодействием. Оно влияет на взаимное расположение магнитных уровней парамагнитного иона, меняя энергетические интервалы между ними. Если одновременно с постоянным внешним магнитным полем парамагнитный ион, расположенный в кристаллической решетке твердого тела, будет подвержен также влиянию переменного электромагнитного поля, то между уровнями будут наблюдаться квантовые переходы. Вероятность таких переходов окажется высокой в случае, если энергетические интервалы между уровнями станут равными кванту электромагнитного поля. Наступает резонанс, носящий название электронного парамагнитного резонанса (ЭПР).

При температурах $T \leq 300\text{K}$ заметно заселенными оказываются только уровни, принадлежащие к нижнему спиновому мультиплету, характеризуемому эффективным спиновым моментом $S_{\text{эф}}$. Условия резонанса для каждой пары уровней этого мультиплета будут зависеть как от величины кванта электромагнитного поля, действующего на кристалл с парамагнитными ионами, так и от величины и направления внешнего постоянного магнитного поля. Факт резонансного поглощения порции энергии электромагнитного поля наблюдается с помощью специальных приборов, называемых спектрометрами ЭПР, а зависимости величины поглощаемой энергии от величины внешнего постоянного магнитного поля называют спектрами ЭПР.

Для того, чтобы выяснить, какая мощность электромагнитной волны поглощается резонирующей парамагнитной системой, необходимо рассмотреть процессы квантовых переходов в этой системе. При этом следует учитывать то, что под воздействием электромагнитной волны возбуждаются квантовые переходы, как с поглощением, так и с излучением энергии на резонансной частоте. Вероятности таких переходов для каждого парамагнитного иона должны быть одинаковыми (обозначим эти вероятности как w_e).

Рассмотрим для примера кристалл с двухуровневыми парамагнитными ионами ($S = 1/2$) [1-3]. Предположим, что концентрация парамагнитных ионов в кристалле равна N . Отношение числа ионов, оказавшихся в состояниях с энергиями E_2 и E_1 ($E_2 > E_1$), в равновесных условиях соответствует распределению Больцмана:

$$\frac{N_2}{N_1} = e^{(-\Delta E / kT_0)}, \quad (1)$$

где $\Delta E = (E_2 - E_1)$ – разность энергий второго и первого уровней, k – постоянная Больцмана, T_0 – температура кристалла, $N_1 + N_2 = N$.

Включим внешнее электромагнитное поле резонансной частоты ($\Delta E = h\nu$). В начальный момент времени величина поглощаемой мощности электромагнитного поля будет определяться равенством:

$$\frac{dW}{dt} = w_e(N_1 - N_2). \quad (2)$$

Из-за преобладания переходов снизу вверх населенности уровней будут стремиться к равным значениям. Однако уменьшению разностей населенностей будут препятствовать нерезонансные переходы с верхнего уровня на нижний уровень (спин-решеточная релаксация). Через какой-то достаточно большой интервал времени с начала включения электромагнитного поля населенности двух спиновых уровней данного иона достигнут значений n_2 и n_1 , причем $n_2 < n_1$. Этим установившимся населенностям ставится в соответствие спиновая температура T_s :

$$\frac{n_2}{n_1} = e^{(-\Delta E / kT_s)}. \quad (3)$$

В новых стационарных условиях величина поглощаемой мощности будет определяться равенством

$$\frac{dW}{dt} = w_e(n_1 - n_2), \quad (4)$$

из которого следует, что величина поглощаемой мощности уменьшилась. Чтобы определить, насколько меньше оказывается поглощаемая мощность с повышением спиновой температуры, найдем скорость изменения разности населенностей рассматриваемых уровней энергии – т.е., найдем решение уравнения

$$\frac{d(n_1 - n_2)}{dt} = -\frac{1}{\tau_1}[(N_1 - N_2) - (n_1 - n_2)], \quad (5)$$

где τ_1 – характеристическое время, называемое временем спин-решеточной релаксации.

Стационарное решение уравнения (5) имеет следующий вид:

$$\frac{n_1 - n_2}{N_1 - N_2} = \frac{1}{1 + 2w_e\tau_1}, \quad (6)$$

где w_e – вероятность перехода одного парамагнитного иона под воздействием электромагнитного поля резонансной частоты. Считая w_e и τ_1 независимыми от времени, приходим к выводу о том, что уменьшение мощности будет зависеть от произведения $w_e \cdot \tau_1$ – чем больше оно, тем сильнее будет насыщаться электронный переход и тем слабее будет сигнал ЭПР.

Кратко рассмотрим вопрос о вероятностях электронных переходов под воздействием сверхвысокочастотного поля [23-25]. Согласно теории возмущений вероятность квантового перехода под воздействием возмущения $\hat{V}(t)$, зависящего от времени, дается формулой :

$$w_{jk} = \frac{1}{\hbar^2} \left| \int_{-\infty}^{+\infty} V_{jk}(t) e^{i\omega_{jk}t} dt \right|^2, \quad (7)$$

где: i и j – индексы, соответствующие начальному $|j\rangle$ и конечному $|k\rangle$ состояниям системы; $V_{jk}(t)$ - матричный элемент оператора возмущения, $V_{jk}(t) \equiv \langle j | \hat{V}(t) | k \rangle$; ω_{jk} - частота, определяемая через разность энергий двух стационарных состояний, между которыми происходит данный квантовый переход, $\omega_{jk} = (E_j - E_k) / \hbar$.

Формула (7) справедлива лишь в первом порядке теории возмущений, т.е. в тех случаях, когда $V_{jk} \gg \hbar\omega_{jk}$. При этом предполагается, что возмущение $\hat{V}(t)$ затухает при $t \rightarrow \pm\infty$. Для определения вероятности перехода на конечный момент времени t верхний предел интеграла полагают равным t , что эквивалентно выключению взаимодействия именно в этот момент времени.

Важным случаем является переход под взаимодействием периодического возмущения:

$$V_{jk}(t) = \tilde{V}_{jk} e^{-i\omega t}. \quad (8)$$

Для простоты включение потенциала можно считать экспоненциальным,

$$V_{jk}(t) = \tilde{V}_{jk} e^{-i\omega t + \lambda t}. \quad (9)$$

В таком случае вероятность квантового перехода определится равенством

$$w_{ji}(t) = \frac{1}{\hbar^2} \left| \int_{-\infty}^t \tilde{V}_{ji} e^{i(\omega_{ji} - \omega)t + \lambda t} dt \right|^2 = \frac{1}{\hbar^2} |\tilde{V}_{ji}|^2 \frac{e^{2\lambda t}}{(\omega_{ji} - \omega)^2 + \lambda^2},$$

откуда в адиабатическом пределе $\lambda \rightarrow 0$ получаем равенство, определяющее вероятность перехода в единицу времени:

$$\frac{d}{dt} [w_{ji}(t)] = \frac{2\pi}{\hbar^2} |\tilde{V}_{ji}|^2 \delta(\omega_{ji} - \omega). \quad (10)$$

В методе ЭПР возмущением является электронное зеемановское взаимодействие электронного магнитного момента парамагнитного центра с магнитной компонентой сверхвысокочастотного поля, действующего на исследуемый образец. В случае парамагнитных центров с магнитной анизотропией величина модуля матричного элемента

$V_{jk}(t)$ оказывается зависящим от направления силовых линий напряженности сверхвысокочастотного магнитного поля относительно исследуемого образца.

Количество наблюдаемых линий ЭПР, их форма, их интенсивность и вид спектра в целом зависят от величины спинового момента парамагнитных центров и от ориентации силовых линий магнитной компоненты сверхвысокочастотного поля относительно линий напряженности постоянного магнитного поля и главных осей симметрии поля лигандов, действующего в позиции парамагнитного иона [26]. Кроме того, в некоторых случаях вид регистрируемого спектра сильно зависит от величины кванта электромагнитного поля. Например, в случае парамагнитного иона со спиновым моментом $S = 1/2$ энергетический интервал между двумя уровнями, $|+1/2\rangle$ и $|-1/2\rangle$, зависит от напряженности внешнего магнитного поля линейно (см. рис. 1.1). Коэффициентом пропорциональности является произведение $g \cdot \beta_e$, где g – фактор спектроскопического расщепления (g -фактор), а β_e – магнетон Бора,

$$\beta_e = \frac{eh}{4\pi m_e}.$$

Таким образом, резонансное значение напряженности постоянного магнитного поля, действующего на кристалл с парамагнитными ионами с $S = 1/2$, будет находиться из равенства

$$H_{\text{рез}} = \frac{h\nu}{g\beta_e}. \quad (11)$$

Из равенства (11) видно, что в случае кубической симметрии парамагнитного центра g -фактор оказывается единственной характеристикой парамагнитного иона. Однако, в случае парамагнитных ионов, локализованных в решетке кристалла в низкосимметричных позициях, вместо одного значения g появляется три значения (g_x, g_y, g_z). Они реализуются при трех взаимно перпендикулярных направлениях вектора напряженности постоянного магнитного поля \vec{H} .

Однако и в данном случае при любом направлении вектора \vec{H} зависимости энергий уровней от H остаются линейными. Поэтому нет необходимости проводить исследования центров на различных частотах сверхвысокочастотного поля.

В случае парамагнитных ионов с большим спиновым моментом число характеристик, определяемых из спектра ЭПР, увеличивается. Например, в случае $S = 3/2$ количество спиновых уровней в составе нижнего спинового мультиплетта парамагнитного иона становится равным четырем. В случае локализации таких ионов, например, в позиции с осевой симметрией поля лигандов диаграммы спиновых уровней энергии будут такими, как это представлено на рис. 2 и рис. 3.

Как видно из обсуждаемых рисунков, вид спектров ЭПР – классический. Такой спектр появляется в тех случаях, когда величина кванта $h\nu$, поглощаемого в момент резонанса, значительно больше величины расщепления спиновых уровней в отсутствие магнитного поля (величины нулевого расщепления ΔE_{zf}). Если же квант сверхвысокочастотного поля окажется одного порядка с начальным расщеплением, резонансное поглощение мощности сверхвысокочастотного поля будет происходить в области R , очерченной пунктирным контуром. В таком случае вид спектра ЭПР будет далеким от классического, а его описание будет представлять проблему, решение которой в общем виде до сих пор не было получено. Можно лишь видеть, что вид спектра будет сильно зависеть от частоты сверхвысокочастотного поля. Можно также догадаться, что путем изменения частоты сверхвысокочастотного поля можно выбрать такое ее значение, которое будет равно ΔE_{zf} . Очевидно, что таким образом легко можно получить одну из характеристик осевого поля лигандов.

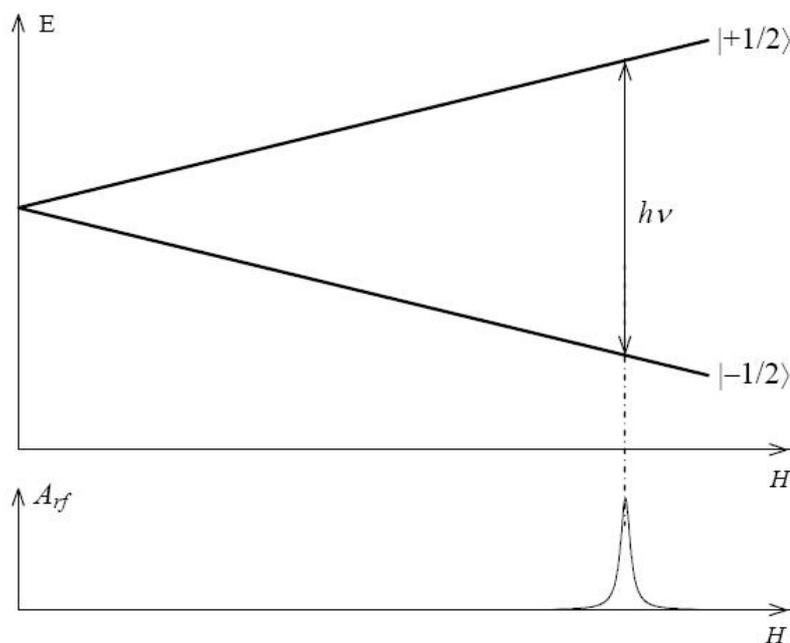


Рис. 1. Зависимости энергий спиновых уровней парамагнитного иона с $S = 1/2$ от напряженности внешнего магнитного поля и вид резонансной линии, наблюдаемой на частоте ν

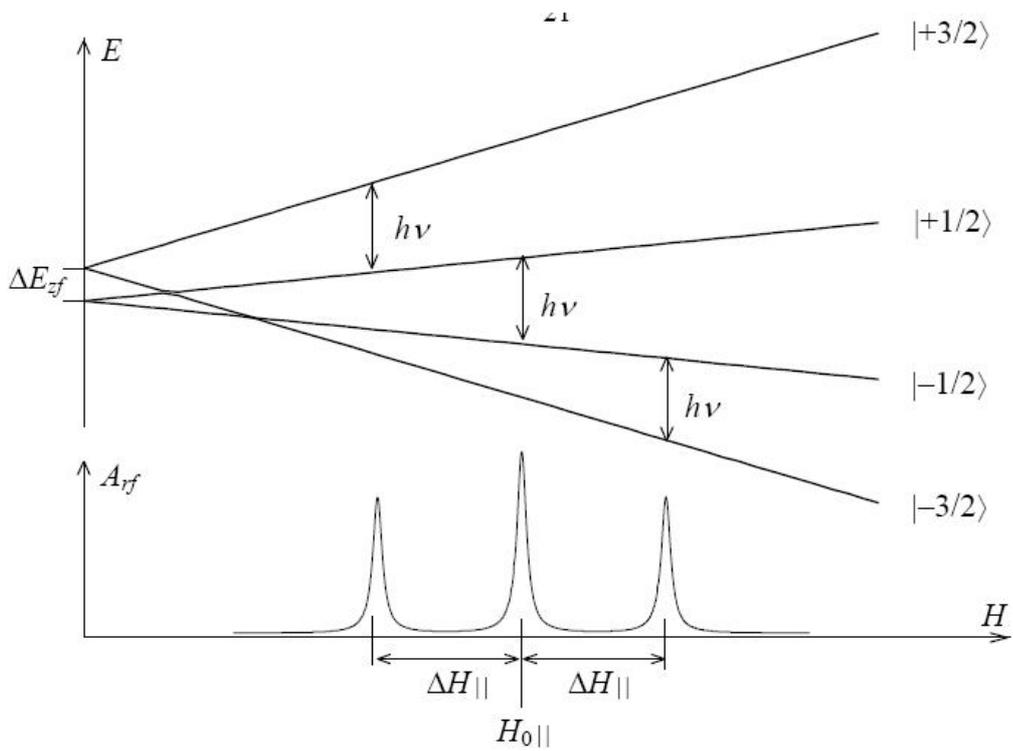


Рис. 2. Уровни энергии парамагнитного иона с $S = 3/2$ и вид линий ЭПР, наблюдаемых на частоте ν в параллельной ориентации

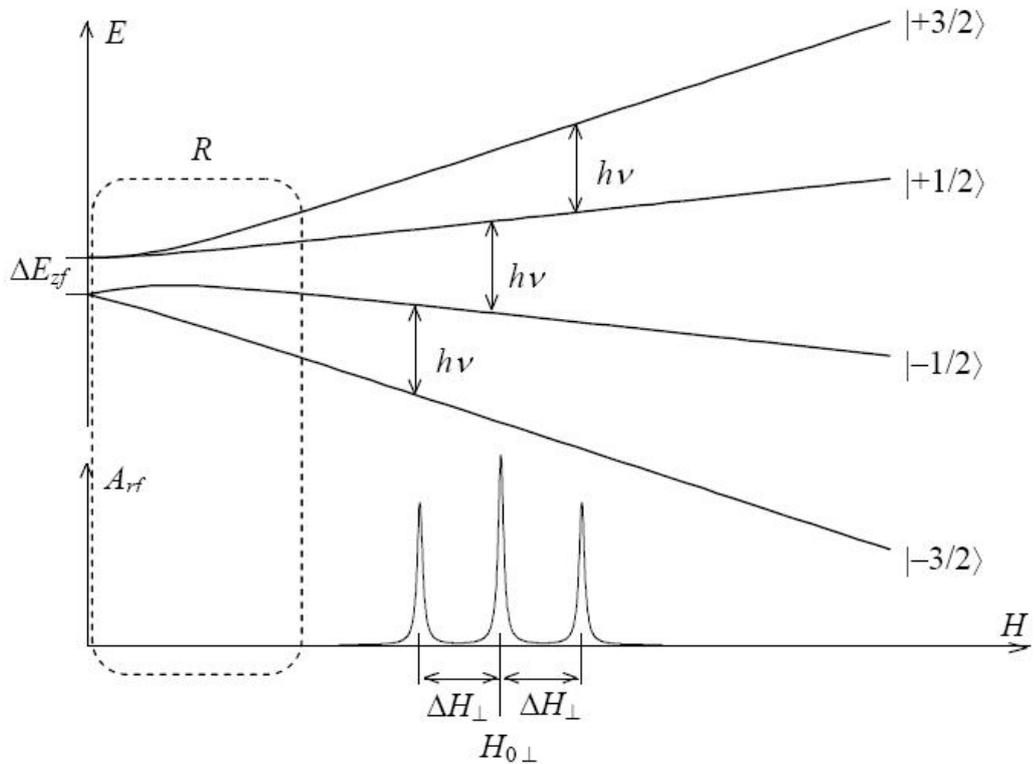


Рис. 3. Уровни энергии парамагнитного иона с $S = 3/2$ и вид линий ЭПР, наблюдаемых на частоте ν в перпендикулярной ориентации

Отметим здесь, что в современных спектрометрах ЭПР регистрируют не сигнал поглощения, а первую производную этого сигнала [1-9]. Поэтому вместо спектра, например, представленного на рис. 1.2, современными спектрометрами ЭПР будет регистрироваться спектр, показанный на рис. 1.4.

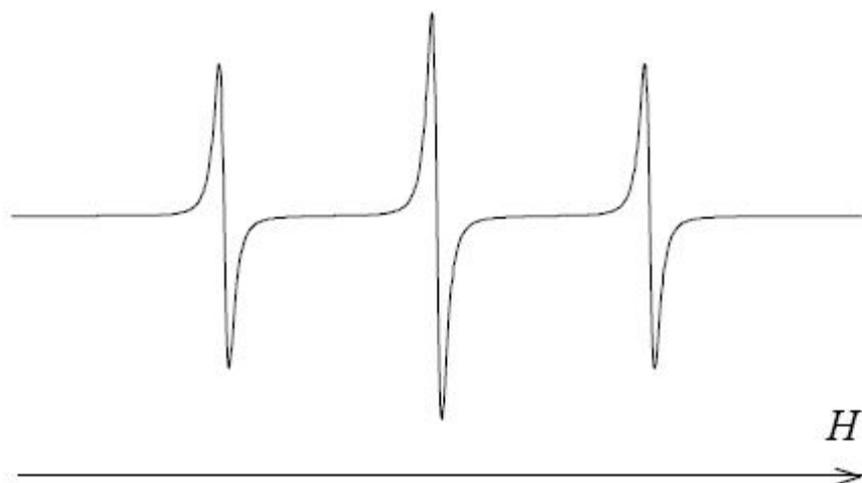


Рис. 4. Вид спектра ЭПР для парамагнитных ионов с $S = 3/2$, полученного на спектрометре с модуляцией постоянного внешнего магнитного поля

Сигнал первой производной получают путем слабой модуляции постоянного магнитного поля на частотах $10 \div 100$ кГц. Для такого способа регистрации спектров ЭПР есть, по крайней мере, две причины. Во-первых, сигнал, снимаемый с кристаллического детектора, бывает очень слабым, поэтому для ее усиления требуются высокочувствительные усилители. Однако сам кристаллический детектор и первые каскады усилителя вносят в сигнал шумы, в спектре которых преобладают низкочастотные составляющие. Исключение таких шумов из состава усиливаемого сигнала облегчается, если он будет представлен в виде высокочастотного несущего сигнала (например, с частотой 100 кГц), модулированного низкочастотным информационным сигналом.

2 Принципы построения современных спектрометров ЭПР и их основные характеристики

Спектрометр ЭПР представляет собой устройство для обнаружения магнитных дипольных переходов. Он позволяет регистрировать поглощение мощности сверхвысокочастотной электромагнитной волны в исследуемом парамагнитном кристалле. Одна из наиболее простых схем спектрометра ЭПР показана на рис. 5 [27].

Эта схема построена по аналогии с оптическими спектрометрами, где монохроматическое электромагнитное излучение проходит через образец и, испытав частичное ослабление вследствие поглощения части мощности на оптических резонансных переходах, попадает на детектор (например, фотоэлемент).

Основное отличие спектрометра ЭПР, представленного на рис. 5, от оптического спектрометра состоит в том, что он требует источника постоянного магнитного поля. Рассматриваемый спектрометр состоит из следующих узлов: клистрон (1), однонаправленный волноводный вентиль (2), ослабитель волны (аттенюатор) (3), направленный ответвитель волны (4), волномер (5), детектор волномера (6), сигнальный детектор (7), проходной резонатор с образцом (8), ярмо электромагнита с токовыми катушками (9), источник модулирующего напряжения (10), усилитель сигнала (11), самописец для регистрации сигнала (12).

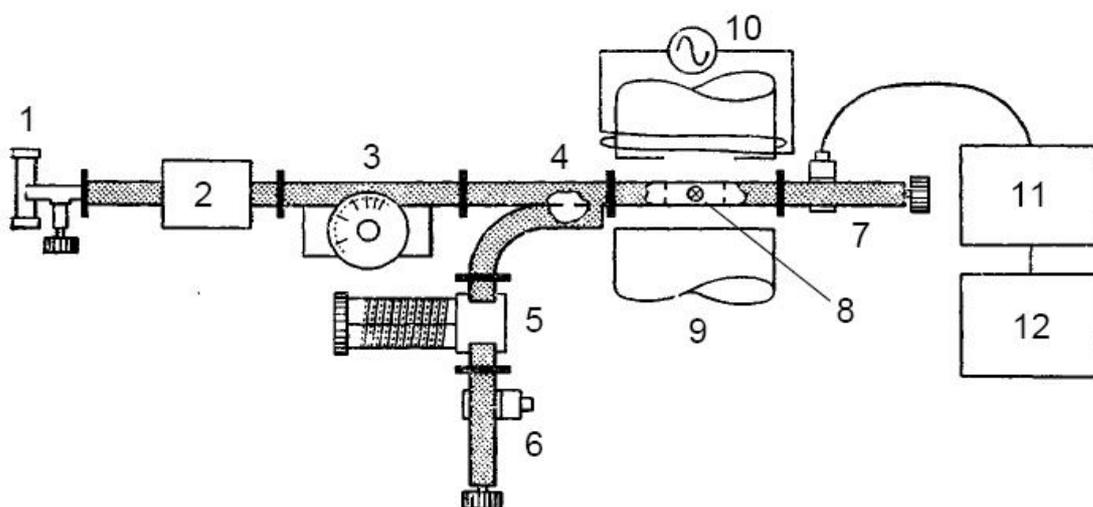


Рис. 5. Схема простейшего спектрометра ЭПР с резонатором

В современных спектрометрах ЭПР используют метод отраженного сигнала; резонатор соединен с сверхвысокочастотным сигналом только одним волноводом. В этих спектрометрах волна подводится к резонатору и, после взаимодействия с образцом, отводится от резонатора по одному и тому же волноводу. После отражения от резонатора волна поступает на

сверхвысокочастотный мост, в качестве которого выступает двойной Т-мост или четырехплечевой циркулятор.

На рис. 6 показана блок-схема спектрометра ЭПР модели JES-SK, выпускавшегося фирмой «Jeol» и снабженного сверхвысокочастотной приставкой для работы в Q-диапазоне.

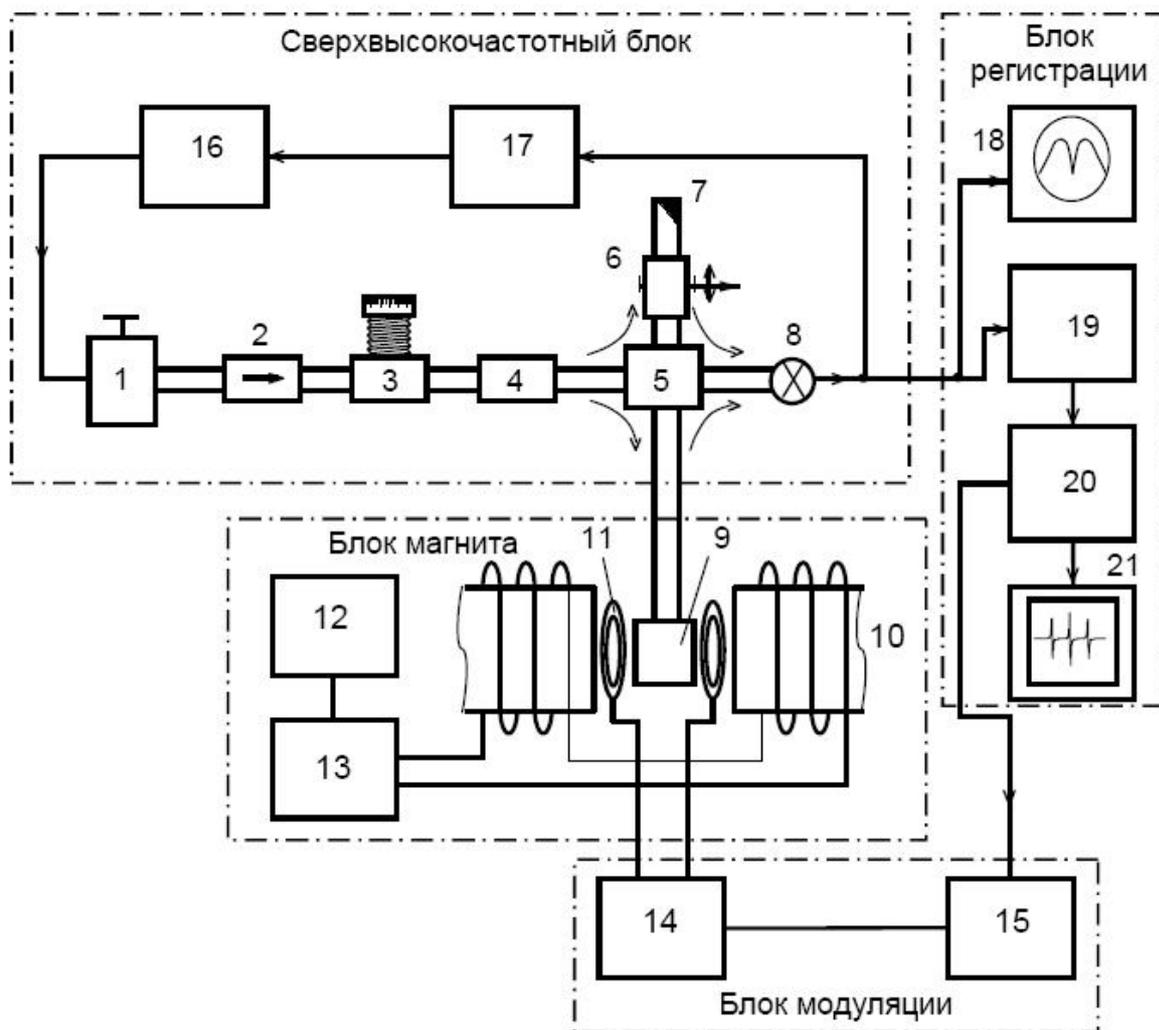


Рис. 6. Блок-схема спектрометра ЭПР модели JES-SK

На рисунке видно, что спектрометр построен по блочному принципу и содержит следующие основные блоки: сверхвысокочастотный блок, блок регистрации сигнала, блок электромагнита, блок модуляции магнитного поля электромагнита. Основными узлами перечисленных блоков являются: 1 – клистронный генератор СВЧ излучения, 2 – волноводный однонаправленный вентиль, 3 – волномер для измерения частоты клистрона, 4 – ослабитель мощности волны (аттенюатор), 5 – двойной Т-мост, 6 – перестраиваемый отражатель волн, 7 – согласованная нагрузка, 8 – диодный детектор, 9 – объемный резонатор с исследуемым образцом, 10 – стальное ярмо и обмотки электромагнита, 11 – катушки модуляции поля электромагнита, 12 – блок сканирования магнитного поля, 13 – источник питания электромагнита, 14 – усилитель мощности сигнала модуляции магнитного поля, 15 –

генератор сигнала модуляции, 16 – блок питания клистронного генератора, 17 – блок автоматической подстройки частоты клистрона, 18 – осциллограф для настройки блока сверхвысокой частоты, 19 – усилитель сигнала, 20 – синхронный детектор, 21 – самописец или иной регистрирующий прибор.

Запись спектров ЭПР на современных спектрометрах ЭПР, работающих в режиме непрерывного облучения образца (режим “Continuous Wave”), осуществляется путем прохождения через резонанс в процессе медленного изменения напряжённости магнитного поля, создаваемого электромагнитом. Сила тока в электромагните регулируется электронным блоком сканирования магнитного поля (12), управляющим источником питания электромагнита (13). Для улучшения основного параметра детектируемого сигнала, отношения «сигнал/шум», используется высокочастотная модуляция магнитного поля. Для чего в межполюсном зазоре электромагнита помещены модулирующие катушки (11), питаемые от источника модуляционного напряжения. Этот источник содержит два блока – усилитель мощности сигнала модуляции магнитного поля (14) и генератор сигнала модуляции (15), работающие на частотах 10 – 100 кГц).

Если бы не было модуляции магнитного поля, ток детекторы был бы пропорционален магнитной восприимчивости образца в резонаторе спектрометра, $\chi(H)$. Поскольку, при высокой индуктивности обмотки электромагнита, запись спектра ЭПР производится в процессе медленного линейного изменения тока электромагнита, то изменения тока детектора при такой процедуре записи также происходят очень медленно. Поэтому усиление сигнала на выходе детектора пришлось бы осуществлять усилителями постоянного тока. Но это привело бы к примешиванию к усиливаемому сигналу низкочастотных шумов, создаваемых собственными элементами усилителя и вибрациями частей сверхвысокочастотного тракта.

Модуляция магнитного поля позволяет использовать узкополосные усилители переменного напряжения вместо широкополосных усилителей, что приводит к значительному улучшению отношения «сигнал/шум» на выходе усилительного каскада.

При высокочастотной модуляции напряжённость внешнего магнитного поля имеет вид:

$$H(t) = H'(t) + H_M \cos(\omega_M t), \quad (12)$$

где $H'(t)$ – постоянная составляющая поля магнита (медленно меняющаяся в процессе развертки), ω_M – частота модуляции, а H_M – амплитуда модуляции магнитного поля ($H_M \ll |H'(t)|$).

При развертке магнитного поля по закону (12) ток детектора (зависящий от магнитной восприимчивости образца в резонансе $\chi(H')$), определяется выражением:

$$I_D \sim \chi(H') + \frac{d\chi(H')}{dH} \cdot H_M \cos(\omega_M t) + \frac{1}{2} \frac{d^2\chi(H')}{dH^2} \cdot H_M^2 \cos^2(\omega_M t). \quad (13)$$

В результате вместо резонансной линии в форме сигнала поглощения, на информационном поле регистрирующего прибора спектрометра ЭПР будет наблюдаться кривая, соответствующая первой производной сигнала поглощения (см. рис. 7).

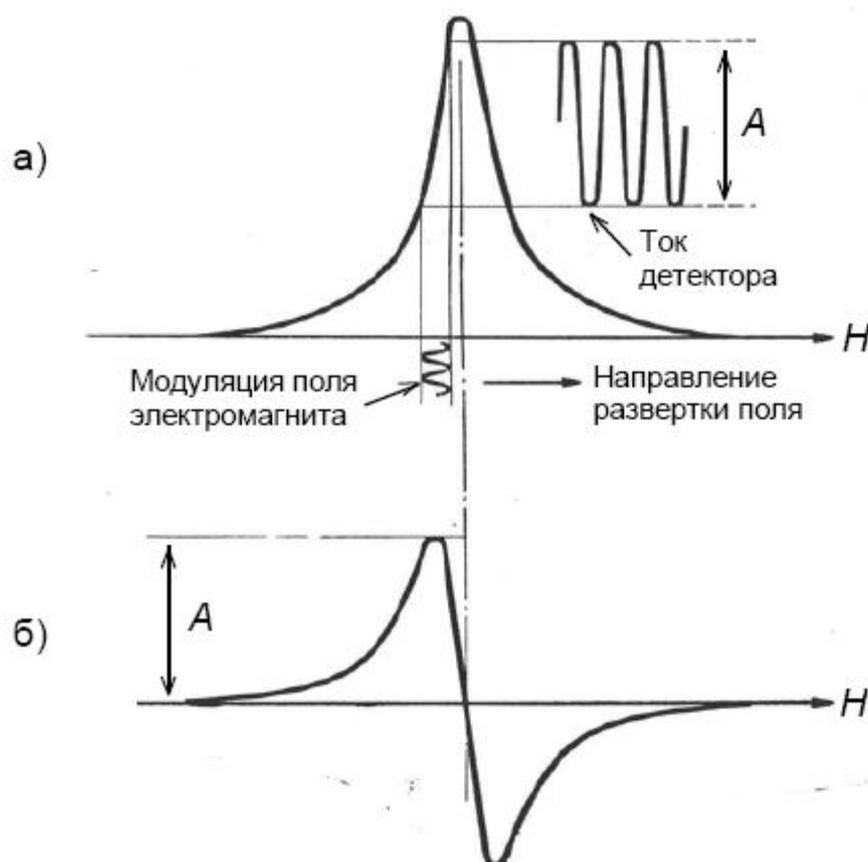


Рис. 7. Форма резонансной линии ЭПР, зарегистрированной в режиме без модуляции (а) и с модуляцией (б) поля электромагнита

Таким образом, на вход узкополосного усилителя поступает сигнал с частотой ω_M , модулированный по амплитуде. При этом напряжение на выходе детектора соответствует первой производной сигнала поглощения ЭПР-излучения образцом. Так как производная сигнала поглощения в точках экстремума спектра меняет знак, для регистрации сигнала используется фазочувствительный (синхронный) детектор.

Частотные шумы спектрометра могут существенным образом влиять на чувствительность [28-34]. Поэтому в современных спектрометрах ЭПР используется система

автоматической подстройки частоты (АПЧ) сверхвысокочастотного генератора к частоте рабочего резонатора. Во многих случаях в качестве генератора используют клистрон. Для подстройки частоты генерации на отражатель клистрона одновременно с постоянным напряжением подаётся небольшое (10 мВ) переменное высокочастотное напряжение (с частотой АПЧ), вызывающее частотную модуляцию генерируемых клистроном колебаний. В результате этого сверхвысокочастотное напряжение, отраженное от отражателя, оказывается модулированным по амплитуде. Следовательно, регистрируемый детектором сигнал (сигнал ошибки) также будет модулирован по амплитуде. Сигнал ошибки и опорный сигнал, поданный с генератора системы АПЧ, поступают на фазочувствительный детектор (ФД). На выходе ФД появляется корректирующее напряжение, которое подаётся на отражатель клистрона. Изменение напряжения на отражателе приводит к изменению генерируемой частоты.

Для передачи СВЧ-поля от клистрона к образцу в ЭПР-спектрометре используется прямоугольный волновод [35-40]. Структура магнитного и электрического полей в прямоугольном волноводе изображена на рис. 8. Геометрические размеры (ширина и высота) волновода сравнимы с длиной волны передаваемого СВЧ-излучения. Для записи типа электромагнитных волн используются следующие обозначения: H_{mn} – поперечные электрические волны, где числа m и n обозначают число полуволн, укладываемых вдоль ширины a и высоты b волновода. Для концентрации СВЧ-поля в месте расположения образца и для увеличения эффективности взаимодействия парамагнитных центров с электромагнитным полем, образец помещают в резонатор. В спектрометрах ЭПР часто используется прямоугольный резонатор. Тип электромагнитных колебаний в резонаторе H_{102} (рис. 9).

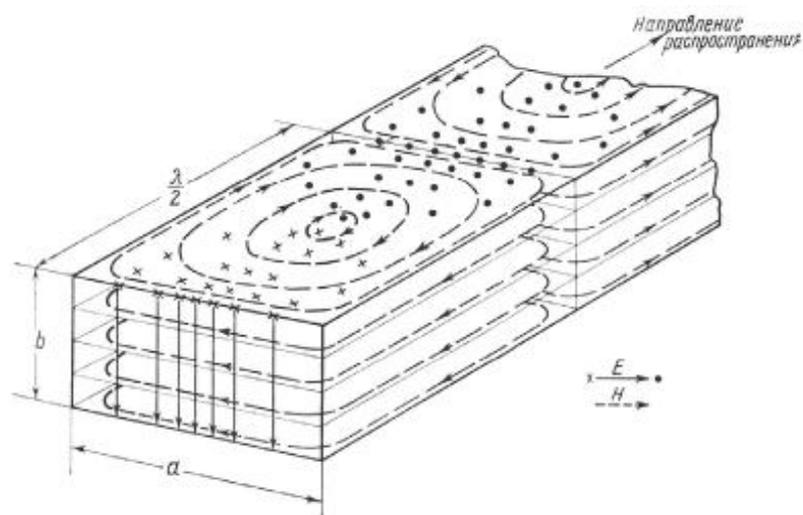


Рис. 8. Структура волны типа H_{10} в прямоугольном волноводе

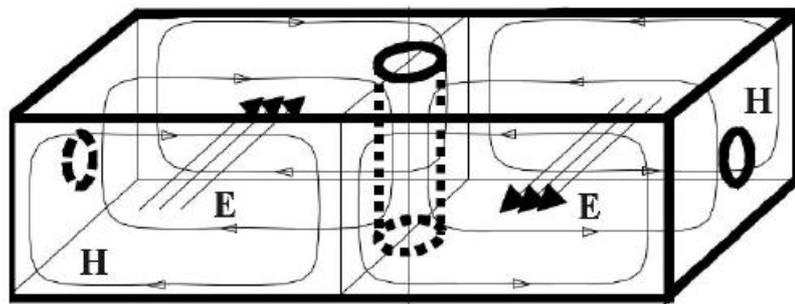


Рис. 9. Структура электромагнитного поля в прямоугольном резонаторе с модой H_{102}

В современных спектрометрах используется балансная схема подключения полупроводникового детектора [27]. Это объясняется тем, что зависимость параметра «сигнал/шум» в выходном сигнале детектора от тока детектора имеет максимум при достаточно малой величине протекающего через детектор тока, что соответствует малым мощностям падающего на детектор СВЧ-излучения. Добиться максимума в этой зависимости позволяет отдельная регулировки мощностей отраженного СВЧ-излучения, попадающего на детектор от образца, и излучения поступающего от клистрона. Одной из простейших схем является двойной Т-мост (рис. 10).



Рис. 10. Балансная схема с волноводным двойным Т-мостом

При точной балансировке моста мощность, поступающая от клистрона в плечо 1, распределяется поровну между плечами 2 и 3, не попадая в плечо 4. Путём небольшой разбалансировки моста (например, изменением волнового сопротивления плеча 2) можно

добиться нужной величины поступающей на детектор мощности. Прохождение через резонанс в образце приведёт к изменению волнового сопротивления плеча 3, что вызовет дополнительную разбалансировку моста и изменение величины падающей на детектор мощности СВЧ-излучения и, следовательно, изменению тока на выходе детектора.

Теперь коротко охарактеризуем продукцию наиболее известных в настоящее время производителей спектрометров ЭПР [41-44].

В настоящее время существует ряд предприятий и компаний, занимающихся серийным выпуском спектрометров ЭПР. Одним из ведущих производителей в настоящее время является германская компания «Bruker Corporation». Известным производителем спектрометров ЭПР остается также японская фирма «Jeol». Среди российских производителей получили известность УП «Конструкторское бюро специальной техники» Белорусского государственного университета, фирма «Адани», ООО Конструкторское бюро «Физэлектронприбор» в г. Самара.

Компания «Bruker Corporation» в настоящее время изготавливает три группы ЭПР-спектрометров: ELEXSYS, EMX Plus/Micro и e-scan. Модели серии EMX выпускаются с 1995 года. В рамках линейки EMX выпускаются две модели: EMX Plus и EMX Micro.

Модель EMX Plus имеет следующие технические характеристики:

- работает в X-диапазоне (9,3 ГГц);
- возможна установка значения центрального поля с разрешением 24 бита (что соответствует точности <math><1 \text{ мГц}</math>);
- используется цифровое кодирование сигнала с разрешением 24 бита
- протяжка магнитного поля может осуществляться в диапазонах любой требуемой ширины (начиная со 100 мГц и до максимально допустимой для данной модели магнита);
- сильные и слабые сигналы могут регистрироваться при одинаковых значениях усиления, вне зависимости от времени конверсии (conversion time)
- спектрометр имеет двойной сигнальный канал для одновременного детектирования 1-й и 2-й производных сигнала;
- чувствительность спектрометра достигает до 2000:1 (согласно международному стандарту Weak-Pitch);
- спектрометр может быть использован для CW-ENDOR спектроскопии

Модель EMX Micro имеет практически те же характеристики, но имеет более компактную конструкцию.

В линейку спектрометров ELEXSYS-II входят следующие модели: ELEXSYS-II E500 (X-диапазон, $\approx 9,3$ ГГц), ELEXSYS-II E540 (X-диапазон), ELEXSYS-II E560 (X-диапазон), ELEXSYS-II E580 (базовая система X-диапазона, но с возможностью работы в диапазонах

W (≈ 95 ГГц), Q (≈ 36 ГГц), S (≈ 3 ГГц) и L (≈ 1 ГГц)), ELEXSYS-II E600 (W) и ELEXSYS-II E680 (W), ELEXSYS-II E780 (263 ГГц).

Эта группа спектрометров ЭПР ориентирована на решение реальных спектроскопических задач, требующих современного программного обеспечения и сложнейшего оборудования.

Спектрометры X-диапазона серии ELEXSYS-II E500 оснащаются подсистемой SuperX и новым модулем обработки сигналов (Signal Processing Unit, SPU).

Технология SuperX включает в себя высокомогущный сдвоенный источник СВЧ-излучения типа диод Ганна со сверхнизким уровнем шума, а также резонатор с высокой добротностью. Наличие данных элементов в значительной степени увеличивает чувствительность СВ-спектрометров в X-диапазоне. Под мерой чувствительности здесь подразумевается отношение сигнала откалиброванного образца Weak Pitch к шуму. У спектрометра E500 это величина составляет 3000:1.

Модуль обработки сигналов спектрометра ELEXSYS-II E500 представляет собой интегральное устройство, способное реализовать любую современную методику анализа спектров СВ-ЭПР. Модуль имеет 8 входных сигнальных каналов и несколько различных режимов синхронизации, что позволяет легко осуществлять даже совершенно нестандартные СВ-ЭПР эксперименты.

Спектрометр ELEXSYS-II E500 позволяет осуществлять цифровую синхронизацию с детектированием до пяти производных сигнала одновременно и работать на частотах модуляции до 1 МГц. В его конструкции предусмотрены сдвоенные входные каналы для одновременной регистрации сигналов поглощения и рассеяния. Он имеет генератор быстрой развертки и систему для модулированного и прямого детектирования, допускает внутреннюю и внешнюю синхронизацию. На нем возможно проведение стационарных и времяразрешенных экспериментов. Для анализа спектров ЭПР в спектрометре имеется программный пакет Херг. Рабочая станция под управлением ОС семейства LINUX является базой для программного пакета Херг и для программного сервера сбора данных, управляющего всеми аппаратными событиями в реальном времени.

Спектрометр ELEXSYS-II E540 представляет собой прибор, предназначенный для молекулярной томографии.

Спектрометры ELEXSYS-II E560 сконструированы как спектрометры двойного-электронно-ядерного резонанса (ENDOR). Поэтому в их наименовании присутствует аббревиатура «ENDOR»: ELEXSYS-II E560 DICE-II ENDOR/TRIPLE. Приборы типа ELEXSYS-II DICE-II широко используются в современной ENDOR-спектроскопии. В их основе лежит технология DICE (Digitally Computed Excitation). В рамках концепции DICE

используется модуляция радиоизлучения типа AM, FM или FSK (Frequency Shift Keying), либо модуляция магнитного поля. В импульсной версии системы E560D-P применяется технология прямого цифрового синтеза частоты (DDS). Применяемая в системах DICE-II модуляция типа FSK позволяет работать даже с довольно большими амплитудами модуляции частоты радиоимпульса, что особенно удобно для широких ENDOR-сигналов. Метод EIE (ENDOR-induced EPR) и опциональный метод TRIPLE (обыкновенный или специальный вариант) также весьма полезны при распутывании сложных комбинаций спектров.

Приборы ELEXSYS-II E580 построены на базовой FT/CW-системе, ориентированной для импульсного многочастотного и мультрезонансного ЭПР. Наряду с выпуском в 1997 году очередной импульсной модели ELEXSYS E580 (в 1987 году был налажен выпуск спектрометра ESP380), компания Bruker Corporation начала производство высокоскоростных цифровых устройств PatternJet и SpecJet. Тогда впервые появилась возможность усреднять спад свободной индукции (FID) и спиновое эхо со скоростью, задаваемой временами электронной спиновой релаксации, а не аппаратными ограничениями.

Устройства следующего поколения SpecJet-II и PatternJet-II уже выпущены на рынок и обладают увеличенной скоростью и точностью сбора импульсных данных. Данные разработки, в сочетании с чувствительностью, универсальностью и надежностью микроволновых FT/CW-мостов и резонаторов серии Flexline, делают ELEXSYS-II E580 одним из передовых исследовательских инструментов для изучения парамагнитных частиц во всех областях науки.

В FT/CW-спектрометрах X-диапазона ELEXSYS E580 также имеется возможность мультчастотных расширений. Приставки на дополнительные частоты подсоединяются к базовой системе X-диапазона. На сегодняшний день возможны расширения на W-диапазон (E680), Q-диапазон (SuperQ-FT), S-диапазон, (SuperS-FT) и L-диапазон (SuperL-FT).

Особенностью моделей ELEXSYS-II E600 и E680 является возможность их использования в качестве ЭПР- и ENDOR-спектрометров в высоких полях на частоте 94 ГГц. Модель E600 оптимизирована для проведения стационарных ЭПР-экспериментов на частоте 94 ГГц, а модель E680 способна работать как в стационарном, так и в импульсном режимах. Их электронная схема обеспечивает высокую точность рабочей частоты, что необходимо при любых измерениях g-фактора. Достаточно высокая гибкость системы является следствием модульной архитектуры, применимой для создания, как спектрометров CW-режима, так и спектрометров комбинированного CW/FT-режима. В этих приборах реализована возможность быстрого переключения между режимами CW и FT, а также между X- и W-диапазонами. Микроволновый резонатор W-диапазона серии TeraFlex способен

работать при температурах от 4 К до 300 К. Исследуемые образцы можно менять при любой температуре. Высокоточное согласование и настройка резонатора, а также механика позиционирования образца, обеспечивают простоту управления. Для согласования моста и резонатора на экране рабочей станции отображается резонансный профиль шириной 400 МГц. Все эти аппаратные возможности в полной мере поддерживаются программным пакетом Херг, предлагающим простой графический пользовательский интерфейс. Поскольку для работы в W-диапазоне необходимо сильное магнитное поле, спектрометры ELEXSYS-II E680 работают со сверхпроводящими магнитами. В них используется разработанный специально для нужд ЭПР магнит EPR 6T SC.

Рассмотрим характеристики спектрометров ЭПР, выпускаемые УП «Конструкторское бюро специальной техники» БГУ. Предприятие УП "КБСТ" БГУ (г. Минск) выпускает спектрометры электронного парамагнитного резонанса "ESR70-03 XD/1" (9,2-9.5 ГГц) и "ESR70-03 XD/2" (9,1-9.6 ГГц). Данные устройства представляют собой полностью автоматизированные приборы, предназначенные для физико-химических исследований в научных и производственно-технологических целях, и применяется в медицине, экологии, геологии, химии, металлургии и других отраслях. Приборы могут дополнительно комплектоваться системой стабилизации температуры образца в резонаторе. Они снабжены малогабаритными электромагнитами с воздушным (величина поля до 0,4 Т) или водяным (величина поля до 0.8 Т) охлаждением. Их СВЧ-тракт выполнен по гомодинной схеме с использованием балансного детектора, р-і-n аттенюатора, генератора на диоде Ганна. Он оборудован системой автоматической настройки частоты фазы и режима согласования.

Следующей моделью, которую следует упомянуть, являются спектрометры электронного парамагнитного резонанса «CMS 8400» выпускаемые фирмой ADANI. Данный прибор предназначен для ЭПР-анализа в автономном и неавтономном режиме в лабораторных или заводских условиях. Он обеспечивает достаточно высокую чувствительность и разрешающую способность, сравнимые с аналогичными параметрами стационарных ЭПР-спектрометров, превышающих CMS 8400 по цене и габаритным размерам. Ключевыми особенностями спектрометра являются компактный электромагнит и компактный СВЧ-тракт. Компактное исполнение спектрометра и малая потребляемая мощность позволяют использовать его в различных условиях, вплоть до полевых (в передвижном исполнении). Автоматизация управления и обработки результатов позволяют потребителю легко адаптировать спектрометр к решению различных задач.

Специально разработанный малошумящий генератор СВЧ-мощности на основе диода Ганна является ключевым элементом микроволнового моста спектрометра ЭПР. Центральная частота генератора - 9.35 ГГц, диапазон перестройки -300 МГц. Применение

стандартных, но специально отбираемых СВЧ-элементов микроволнового моста и наличие канала смещения детекторов СВЧ позволяют обеспечивать высокую чувствительность спектрометра во всем диапазоне микроволновой мощности СВЧ. СВЧ - тракт обеспечен при смене измеряемого образца автоматической системой настройки параметров (частота СВЧ, связь СВЧ - тракта с измерительным резонатором и фаза в канале смещения детекторов СВЧ).

В спектрометре «CMS 8400» используется оригинальная конструкция прямоугольного резонатора H_{102} с рабочей частотой 9.4 ГГц и ненагруженной добротностью 5000. Устройство связи позволяет автоматически поддерживать критическую связь резонатора с СВЧ - трактом при установке различных ампул, капилляров и плоских кювет с образцами, а также сосудов Дьюара с максимальным диаметром 9.5 мм. Поляризуемое магнитное поле в резонаторе модулируется с частотой 100 кГц и имеет максимальную амплитуду модуляции 1 мТ. Система регистрации использует 100 кГц модуляцию магнитного поля с синхронным детектированием сигнала на частоте модуляции и последующим 12 - битным преобразованием. Коэффициент усиления, постоянная времени сглаживания и амплитуда модуляции магнитного поля программно управляются. Устройство задания магнитного поля устанавливает и поддерживает с высокой стабильностью напряженность поляризуемого магнитного поля на измеряемом образце. Центральное значение, амплитуда и время развертки магнитного поля программно управляются. Для стабилизации и управления магнитным полем используется датчик Холла, размещенный в межполюсном зазоре электромагнита.

Конструкция компактного электромагнита с полюсными наконечниками диаметром 60 мм позволяет получать магнитные поля в 15 мм зазоре до 0,7 Т. Специальная конфигурация полюсных наконечников и магнитопровода обеспечивают высокую однородность поляризуемого магнитного поля в объеме исследуемого образца. Достижимая однородность такая же, как и у электромагнитов с диаметром полюсных наконечников 100 мм и больше. Источник питания с импульсным преобразованием обеспечивает ток электромагнита до 4 А. Источник снабжен системой защиты от перегрева электромагнита. Управление источником осуществляется устройством задания магнитного поля.

Особо отметим здесь широкодиапазонный субмиллиметровый спектрометр ЭПР оригинальной конструкции, разработанный сотрудниками Казанского физико-технического института В.Ф. Тарасовым и Г.С. Шакуровым [47,48]. Блок схема этого спектрометра представлена на рис. 11. Он создан по оптическому принципу передачи сверхвысокочастотного излучения на исследуемый образец, поэтому способен работать в очень широком диапазоне частот (79 ÷ 535 ГГц). СВЧ излучение генерируется сменной

лампой обратной волны (1), помещенной между полюсами электромагнита (2). Через открытый конец волноводного вывода лампы обратной волны электромагнитная волна излучается в пространство. Часть излучения попадает на фторопластовую линзу (3), формирующую плоскую волну. Далее волна проходит через модулятор (4), поляризатор (5) или (6) и попадает на фокусирующую фторопластовую линзу (7). Металлическое зеркало (8) отклоняет электромагнитный луч на 90° и направляет его на систему фторопластовых линз (9) и, далее, на образец (10) и болометр (11). Образец и болометр погружены в жидкий гелий, залитый в криостатную систему (12). Магнитное поле, воздействующее на исследуемый образец, создается электромагнитом (13). Модуляция магнитного поля осуществляется подачей на модуляционные катушки (14) переменного напряжения частотой 15 кГц.

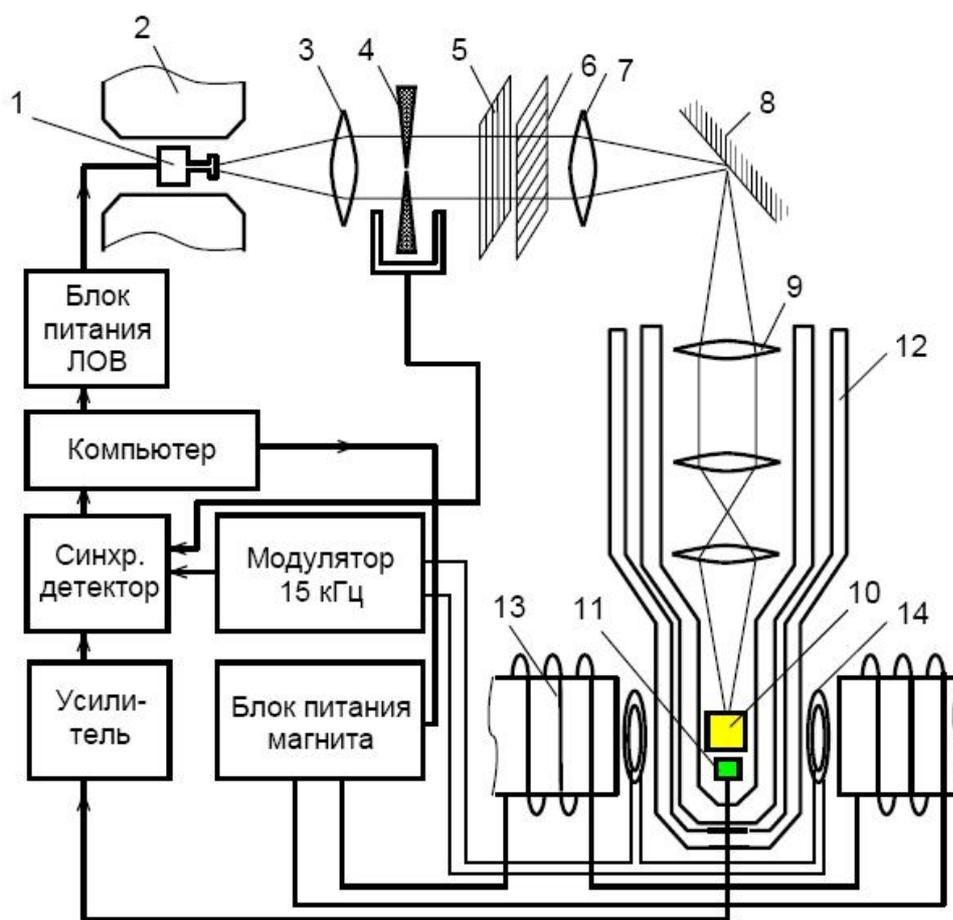


Рис. 1.11. Блок-схема перестраиваемого спектрометра ЭПР с квазиоптической системой передачи электромагнитной волны на исследуемый образец

Имея широкие возможности в перестройке частоты, описанный спектрометр обладает и некоторыми недостатками. Один из них – это отсутствие объемного резонатора. Сверхвысокочастотная волна передается на исследуемый образец по открытому

пространству с помощью системы фторопластовых линз и проходит через образец лишь один раз. Из-за отсутствия резонатора чувствительность данного спектрометра оказывается невысокой, что не позволяет изучать образцы небольшого объема. Другой недостаток – это слабый электромагнит, создающий постоянное магнитное поле в образце. Из-за этого данный спектрометр не может быть использован для изучения образцов с малыми расщеплениями спиновых уровней в нулевом магнитном поле.

2. Теоретическое описание спектров ЭПР в рамках формализма спинового гамильтониана.

Описание магнитных свойств парамагнитных ионов в твердых телах обычно осуществляется в рамках формализма спинового гамильтониана [1-16, 50-55]. При этом парамагнитный ион и ближайшие к нему лиганды рассматривают как единую систему, называемую парамагнитным центром.

Спиновый гамильтониан, представляет собой определенную совокупность спиновых операторов различной степени. Каждый из таких операторов входит в спиновый гамильтониан со своим коэффициентом, называемым параметром гамильтониана. Число и вид таких операторов зависят, прежде всего, от величины эффективного спинового момента исследуемых парамагнитных центров (ПЦ) и от их симметрии.

Операторы спинового гамильтониана можно сгруппировать по степени их влияния на характеристики спектров ЭПР. При такой группировке следует ограничить величины напряженности внешнего магнитного поля в пределах того диапазона, который обычно реализуется электромагнитами существующих спектрометров ЭПР. Оказывается, что в указанном диапазоне магнитных полей максимальное влияние на характеристики спектров ЭПР оказывают операторы спинового гамильтониана, зависящие от величины эффективного спинового момента $S_{\text{эф}}$. Эта группа операторов спинового гамильтониана может быть разбита на три подгруппы.

К первой можно отнести спиновые операторы, зависящие только от проекций электронного спинового момента, $\hat{S}_x, \hat{S}_y, \hat{S}_z$. Наличие таких операторов приводит к частичному расщеплению уровней нижнего электронного спинового мультиплетта даже в отсутствие внешнего магнитного поля. По мере понижения симметрии ПЦ число операторов этой подгруппы увеличивается. Данная подгруппа операторов приводит к появлению в спектрах ЭПР так называемой тонкой структуры, а расщепления в отсутствие внешнего постоянного магнитного поля часто называют начальными расщеплениями или расщеплениями в нулевом магнитном поле.

Ко второй подгруппе следует отнести операторы, зависящие как от электронного спинового момента, так и от величины и направления внешнего магнитного поля. Эта совокупность операторов представляет собой гамильтониан зеемановского взаимодействия и определяет основную часть взаимодействия электронного магнитного момента ПЦ с внешним магнитным полем.

К третьей подгруппе относятся операторы, одновременно зависящие от электронных и ядерных спиновых моментов. Эти операторы описывают сверхтонкую структуру (СТС) и лигандную сверхтонкую структуру (ЛСТС) спектров ЭПР.

Для практических целей нередко используется феноменологический подход к построению спинового гамильтониана. Из симметричных соображений находится общий вид зависимости гамильтониана от компонент электронного спина, $\hat{S}_x, \hat{S}_y, \hat{S}_z$, и компонент напряженности внешнего постоянного магнитного поля, H_x, H_y, H_z . При этом коэффициенты полученного полинома оказываются неопределенными параметрами, которые в дальнейшем должны быть определены из эксперимента. Для этого строят графики экспериментальных угловых зависимостей резонансных значений напряженности внешнего магнитного поля для всех линий наблюдаемого спектра ЭПР. Зная точное значение частоты электромагнитного поля в резонаторе спектрометра, методом наименьших квадратов (или, если это возможно, из аналитических уравнений) находят значения параметров спинового гамильтониана, которые можно рассматривать как характеристики изучаемой парамагнитной системы.

Ниже приведем общий вид спиновых гамильтонианов для различных групп симметрии кристаллического поля (имеются в виду редкоземельные ионы с $S \geq 3$; члены, соответствующие слабым взаимодействиям – сверхтонкое, квадрупольное и ядерное зеемановское взаимодействия, не включены, поскольку они не часто проявляются в реальных спектрах ЭПР):

1) кубическая –

$$H_S = B_4^0(O_4^0 - 5O_4^0) + B_6^0(O_6^0 - 21O_6^4) + g\beta_e \vec{H} \cdot S; \quad (14)$$

2) гексагональная –

$$H_S = B_2^0 O_2^0 + B_4^0 O_4^0 + B_6^0 O_6^0 + B_6^6 O_6^6 + g_{\parallel} \beta_e H_z S_z + g_{\perp} \beta_e (H_x S_x + H_y S_y); \quad (15)$$

3) тригональная –

$$H_S = B_2^0 O_2^0 + B_4^0 O_4^0 + B_4^3 O_4^3 + B_6^0 O_6^0 + B_6^3 O_6^3 + B_6^6 O_6^6 + g_{\parallel} \beta_e H_z S_z + q_{\perp} \beta_e (H_x S_x + H_y S_y); \quad (16)$$

4) тетрагональная –

$$H_S = B_2^0 O_2^0 + B_4^0 O_4^0 + B_4^4 O_4^4 + B_6^0 O_6^0 + B_6^4 O_6^4 + \\ + g_{\parallel} \beta_e H_z S_z + q_{\perp} \beta_e (H_x S_x + H_y S_y); \quad (17)$$

5) орторомбическая –

$$H_S = B_2^0 O_2^0 + B_2^2 O_2^2 + B_4^0 O_4^0 + B_4^2 O_4^2 + B_4^4 O_4^4 + B_6^0 O_6^0 + B_6^4 O_6^4 + \\ + \beta_e (g_x H_x S_x + g_y H_y S_y + g_z H_z S_z). \quad (18)$$

В случаях еще более низкой симметрии в спиновом гамильтониане появляются термы типа $B_n^m O_n^m$ с нечетными значениями m и термы $C_n^m \Omega_n^m$ ($n = 2, 4, 6; m = 1, 2, 3, 4, 5, 6$). Но, как следует из теории поля лигандов, в реальных кристаллах коэффициенты C_n^m по величине оказываются намного меньше соответствующих коэффициентов B_n^m , поэтому приближенное описание спектров ЭПР низкосимметричных парамагнитных центров оказывается возможным на основе спинового гамильтониана (18). При этом учтено то, что от термов $B_n^m O_n^m$ с нечетными значениям m можно избавиться путем подходящего выбора системы координат представления спинового гамильтониана.

Если у редкоземельных ионов $S = 2$ или $S = 5/2$, то в спиновых гамильтонианах оставляются термы $B_n^m O_n^m$ с $n = 2, 4$ и $m \leq n$, а в случаях ионов с $S = 1$ и $3/2$ – оставляются термы $B_n^m O_n^m$ с $n = 2$ и $m \leq n$.

В случае ионов переходной группы (d -ионов) спиновые гамильтонианы соответствующих групп симметрии будут отличаться от приведенных выше отсутствием термов $B_n^m O_n^m$ с $n = 6$. При этом число термов в спиновых гамильтонианах, как и в случае редкоземельных ионов, зависит от величины их спинового момента. Если у переходного иона $S = 2$ или $S = 5/2$, то в спиновых гамильтонианах будут присутствовать все термы $B_n^m O_n^m$ с $n = 2$ и 4 и $m \leq n$, а в случаях ионов с $S = 1$ и $3/2$ – оставляются термы $B_n^m O_n^m$ с $n = 2$ и $m \leq n$.

Здесь следует отметить, что в выражения (14) – (18), кроме членов, описывающих СТВ и ЛСТВ спектров ЭПР, не включены также члены, представляющие собой произведения различных степеней операторов \hat{S}_x, \hat{S}_y и \hat{S}_z и компонент вектора напряженности внешнего постоянного магнитного поля, H_x, H_y, H_z . Это объясняется тем, что в подавляющем большинстве случаев влияние этих термов практически не проявляется в наблюдаемых спектрах ЭПР. Но, тем не менее, в редких случаях термы такого рода необходимы для описания спектров высокоспиновых ПЦ, локализованных в кубическом поле лигандов [56]. В таких случаях используют обобщенный спиновый гамильтониан (см., например, [3]), построенный в соответствии с наблюдаемой симметрией магнитных свойств.

Здесь мы не рассматриваем обобщенный спиновый гамильтониан, поскольку здесь основной интерес представляют низкосимметричные центры, в которых влияние термов $C_{ij}^{nm} \hat{S}_i^n H_j^m$ обычно полностью маскируется влиянием термов $B_n^m O_n^m$.

Следует подчеркнуть, что спиновые гамильтонианы (14)–(18) представлены в каноническом виде (т.е., в своих главных осях симметрии). Эти оси далеко не всегда совпадают с главными осями симметрии кристаллической решетки, в которую парамагнитные ионы были внедрены как примеси [50-55].

Рассмотрим простейший пример практического использования формализма спинового гамильтониана – расчет зависимости положения резонансной линии от направления постоянного внешнего магнитного поля для парамагнитного иона с $S=1/2$.

В случае центров с ромбической симметрией магнитных свойств и с $S=1/2$ спиновый гамильтониан содержит только операторы электронного зеемановского взаимодействия:

$$\hat{H}_S = \beta_e (g_x H_x \hat{S}_x + g_y H_y \hat{S}_y + g_z H_z \hat{S}_z). \quad (19)$$

Матрица гамильтониана (19) имеет размерность 2×2 и легко может быть приведена к диагональному виду посредством преобразования координат.

В результате такого преобразования можно получить спиновый гамильтониан, содержащий только z -компоненту оператора спинового момента S :

$$\hat{H}'_S = \beta_e g_{\text{эф}} |\vec{H}| S_z,$$

где

$$g_{\text{эф}} = \sqrt{(g_x^2 \cos^2 \varphi \sin^2 \theta + g_y^2 \sin^2 \varphi \sin^2 \theta + g_z^2 \cos^2 \theta)} \quad (20)$$

– эффективный g -фактор,

φ и θ – два угла Эйлера, определяющие направление вектора напряженности внешнего магнитного поля \vec{H} относительно главных осей симметрии парамагнитного центра. В выражении (20) произведение $(\cos \varphi \sin \theta)$ соответствует косинусу угла между вектором \vec{H} и осью x , $(\sin \varphi \sin \theta)$ соответствует косинусу угла между \vec{H} и осью y , а $\cos \theta$ – косинусу между \vec{H} и осью z . Для краткости эти направляющие косинусы обозначим как

$$\cos \varphi \sin \theta = l_x, \quad \sin \varphi \sin \theta = l_y, \quad \cos \theta = l_z. \quad (21)$$

Эффективный g -фактор (20) в новых обозначениях приобретает следующий вид:

$$g_{\text{эф}} = \sqrt{(g_x^2 l_x^2 + g_y^2 l_y^2 + g_z^2 l_z^2)}. \quad (22)$$

Энергии спиновых состояний $|\pm 1/2\rangle$ определяются равенством

$$\varepsilon^{\pm} = \pm \frac{1}{2} \beta_e g_{\text{эф}} |H|. \quad (23)$$

Из последнего равенства видно, что уровни энергии оказываются линейно зависящими от напряженности внешнего магнитного поля. В условиях резонанса

$$h\nu = \Delta\varepsilon = \varepsilon^+ - \varepsilon^- = \beta_e g_{\text{эф}} H_{\text{рез}} \quad (24)$$

для главных ориентаций вектора напряженности внешнего магнитного поля, $\vec{H} \parallel x$, $\vec{H} \parallel y$ и $\vec{H} \parallel z$, мы получаем равенства:

$$h\nu = \beta_e g_x H_{\text{рез1}}, \quad h\nu = \beta_e g_y H_{\text{рез2}}, \quad h\nu = \beta_e g_z H_{\text{рез3}}, \quad (25)$$

где $H_{\text{рез1}}$, $H_{\text{рез2}}$, $H_{\text{рез3}}$ – резонансные значения внешнего магнитного поля в ориентациях $\vec{H} \parallel x$, $\vec{H} \parallel y$ и $\vec{H} \parallel z$, соответственно. Откуда легко определить компоненты ромбического центра, g_x , g_y , g_z .

При развертке спектра по магнитному полю, угловая зависимость положения резонансной линии от направления вектора \vec{H} определяется равенством

$$H_{\text{рез}}(i) = \frac{h\nu}{\beta_e g_{\text{эф}}(i)}, \quad (26)$$

где i соответствует номеру ансамбля центров, а направляющие косинусы вектора \vec{H} определяют величину $g_{\text{эф}}$ в выбранном направлении.

Из равенства (26) видно, что при анизотропном g -тензоре ($g_x \neq g_y \neq g_z$) при определенных направлениях вектора \vec{H} резонансная линия будет находиться в экстремальных положениях. Например, при $g_x > g_y > g_z$ положение линии в направлении \vec{H} , определяемом углом $\theta = 0^\circ$ (при произвольном φ), будет соответствовать максимальному резонансному значению напряженности магнитного поля. Такую точку на угловой зависимости $H_{\text{рез}}(\theta, \varphi)$ спектроскописты называют двойным сильнополевым экстремумом (подразумеваются два независимых вращения вектора \vec{H} вокруг двух взаимно перпендикулярных направления, характеризуемых углами θ и φ). В этом же случае ($g_x > g_y > g_z$), в направлении \vec{H} , соответствующем углам $\theta = 90^\circ$ и $\varphi = 0^\circ$, в угловой зависимости $H_{\text{рез}}(\theta, \varphi)$ получится экстремальная точка, называемая двойным минимумом. В направлении, соответствующем углам $\theta = 90^\circ$ и $\varphi = 90^\circ$, оказывается точка, называемая экстремумом типа «минимум-максимум».

Очевидно, что даже в таком простом случае число линий может стать большое одной (до 24), если центр окажется с низкой симметрией магнитных свойств, а кристалл, содержащий этот вид центров, будет иметь симметричную кристаллическую решетку. В

таком случае при одном и том же направлении постоянного магнитного поля в угловых зависимостях эффективного g-фактора направляющие углы в равенствах типа (1.21) должны отсчитываться от осей локальных систем координат, в которых представлен спиновый гамильтониан каждого из парамагнитных центров.

Лекция 6. Оптическая спектроскопия как метод изучения зонной структуры материалов.

1 Основы зонной теории полупроводниковых материалов

Для идеальной кристаллической решетки, в узлах которой находятся ядра атомов, волновая функция Φ , которая описывает состояние этой решетки, зависит от координат всех электронов \vec{r}_i и от координат всех ядер \vec{R}_j , т. е.

$$\Phi = \Phi(\vec{r}_1 \vec{r}_2 \vec{r}_3 \dots \vec{r}_i; \vec{R}_1 \vec{R}_2 \vec{R}_3 \dots \vec{R}_j). \quad (1)$$

Принимая во внимание (1), уравнение Шрёдингера для кристаллической решетки можно записать в следующем виде:

$$\sum_i \frac{\hbar_i^2}{2m} \nabla_i^2 \Phi + \sum_j \frac{\hbar^2}{2M} \nabla_j^2 \Phi + (E - U) \Phi = 0, \quad (2)$$

где m – масса электрона, M – масса ядра атома, ∇ – оператор Лапласа, E – полная энергия системы, U – потенциальная энергия взаимодействия всех электронов и ядер, $\hbar = h/2\pi$, h – постоянная Планка.

Решить уравнение (2) в общем виде невозможно из-за огромного числа переменных \vec{r}_i и \vec{R}_j . Поэтому в зонной теории твердого тела используется ряд допущений, позволяющих упростить уравнение (2), это адиабатическое приближение, которое позволяет разделить переменные \vec{r}_i и \vec{R}_j : одно из которых относится только к электронам, другое – к ядрам.

В результате упрощений уравнение (2) для каждого электрона твердого тела запишется:

$$\left[-\frac{\hbar^2}{2m} \nabla_i^2 + U(\vec{r}_i) \right] \Psi(\vec{r}_i) = E_i \Psi(\vec{r}_i), \quad (3)$$

где $\Psi(\vec{r}_i)$ – волновая функция электрона,

$E_i(\vec{r}_i)$ – полная энергия электрона,

$U(\vec{r}_i)$ – потенциальная энергия электрона.

Для свободного электрона уравнение (3) запишется

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{2m}{\hbar^2} E \psi, \quad (4)$$

Общим решением уравнения (4) является волновая функция следующего вида

$$\psi(x, y, z) = \exp i(k_x x + k_y y + k_z z), \quad (5)$$

Возьмем первую и вторую производные выражения (5):

$$\frac{\partial \Psi}{\partial x} = ik_x \exp i(k_x x + k_y y + k_z z), \quad (6)$$

$$\frac{\partial^2 \Psi}{\partial x^2} = -k^2 \Psi, \quad (7)$$

где

$$k^2 = (k_x^2 + k_y^2 + k_z^2).$$

Подставляя (7) в (4), получим:

$$-k^2 \Psi = -\frac{2m}{\hbar^2} E \Psi$$

или

$$k^2 = \frac{2m}{\hbar^2} E \Psi. \quad (8)$$

Из (8) найдем энергию:

$$E = \frac{\hbar^2 k^2}{2m}. \quad (9)$$

Для раскрытия физического смысла величины k выразим энергию электрона через импульс:

$$E = \frac{mv^2}{2} = \frac{(mv)^2}{2m} = \frac{p^2}{2m}. \quad (10)$$

Сравнивая (9) и (10), получим

$$p^2 = \hbar^2 k^2, \quad p = \hbar k. \quad (11)$$

Длина волны де Бройля λ определяется из выражения

$$\lambda = \frac{h}{mv} = \frac{h}{p}, \quad (12)$$

$$p = \frac{h}{\lambda}, \quad (13)$$

Из (11) и (13) будем иметь

$$\frac{h}{\lambda} = \hbar k = \frac{h}{2\pi} k, \quad (14)$$

Тогда

$$k = \frac{2\pi}{\lambda}. \quad (15)$$

Из (15) следует, что k есть число длин волн, укладывающихся на длине 2π , которое называется волновым вектором.

Из (9) следует, что энергия свободного электрона пропорциональна квадрату волнового вектора и есть уравнение равнобочной параболы. Отсюда следует, что энергия электрона в отсутствие потенциального поля параболически зависит от волнового вектора (рис. 1).

В периодическом силовом поле потенциальная энергия $U(x, y, z)$ является периодической функцией координат, поэтому волновая функция электрона должна удовлетворять следующему уравнению:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{2m}{\hbar^2} [E - U(x, y, z)] \psi = 0.$$

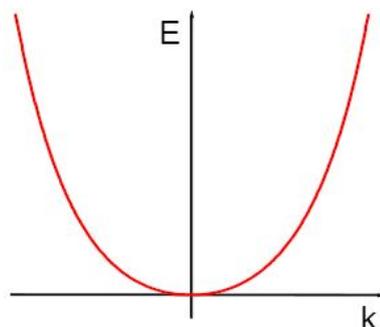


Рис. 1. Зависимость энергии E свободной частицы от волнового вектора k .

Для решения уравнения (16) теория вводит три приближения.

1. Кристаллическая решетка твердого тела идеальна: силовое поле является строго периодическим.
2. Электроны не взаимодействуют друг с другом, а взаимодействуют только с кристаллической решеткой (одноэлектронное приближение).
3. По сравнению с электронами атомные остовы совершают колебательные движения очень медленно и их движением можно пренебречь (адиабатическое приближение).

Решением (16) при указанных допущениях является волновая функция следующего вида:

$$\psi(x, y, z) = \exp i(k_x x + k_y y + k_z z) \varphi_k(x, y, z). \quad (17)$$

В выражении (17) первый множитель такой же, как и для свободного электрона в (5), второй же множитель (функция Блоха) зависит не только от координат, но и от волнового вектора k . По физическому смыслу (17) есть плоская волна длиной $\lambda = 2\pi/k$, которая распространяется в направлении волнового вектора и модулируется периодичностью кристаллической решетки.

Из выражения (17) следует, что в твердом теле имеются только вполне определенные значения энергии E , для которых волновой вектор k принимает вещественные значения. Для других же значений энергии k есть мнимая величина. Значит, в пространстве волнового вектора энергии электронов не могут быть непрерывной функцией k , существуют запрещенные энергии.

Схематически выражение (17) представлено на рис. 2.

Следовательно, в твердом теле существуют зоны разрешенных энергий, отделенные друг от друга запрещенными полосами. Это значит, что электроны не могут обладать энергией, отвечающей уровню внутри полосы между разрешенными зонами.

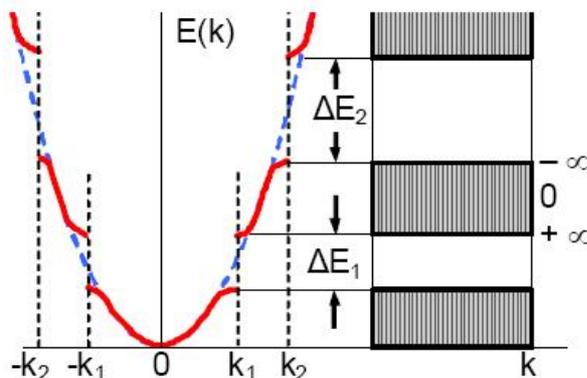


Рис. 2. Значения энергий электрона, находящегося в поле кристалла.

Основные оптические характеристики полупроводников.

Для описания оптических явлений в полупроводниках достаточно использовать две характеристики: показатель преломления n и показатель поглощения k .

Показатель преломления n определяется как отношение скорости распространения световой волны в вакууме c к ее скорости в среде (кристалле) v :

$$n = \frac{c}{v}. \quad (18)$$

Показатель поглощения или коэффициент экстинкции k определяется через линейный коэффициент поглощения α и длину волны падающего света λ :

$$k = \frac{1}{4\pi} \alpha \lambda. \quad (19)$$

Величина n характеризует преломление световой волны в кристалле с учетом ее частичного поглощения. Поэтому n является комплексной величиной:

$$n = n^0 - ik, \quad (20)$$

где n^0 – показатель преломления в отсутствие поглощения.

Справедливость выражения (20) можно доказать, используя выражение, которое описывает изменение волнового вектора световой волны, которая распространяется по кристаллу:

$$\vec{E}(x, t) = \vec{E}_0 \exp \left[i\omega \left(\frac{t-x}{v} \right) \right], \quad (21)$$

где \vec{E}_0 – амплитуда волны, прошедшей в образец, ω – ее круговая (циклическая) частота.

Из (18) и (20) найдем $1/\nu$ и, подставив в (21), получим

$$\vec{E} = \vec{E}_0 \exp\left(-\frac{\omega k x}{c}\right) \exp i\omega\left(t - \frac{n^0 x}{c}\right). \quad (22)$$

При $k = 0$, т. е. в отсутствие поглощения волна, описываемая формулой (2.22), переходит в плоскую волну, которая распространяется с постоянной амплитудой и фазовой скоростью c/n , что имеет место для непоглощающих сред.

Обозначим (рис. 3) через I_0 интенсивность света падающего на поверхность кристалла полупроводника, через I_T интенсивность света прошедшего через кристалл. Измеряемая на опыте величина, называемая коэффициентом пропускания, определяется как отношение этих интенсивностей:

$$T = \frac{I_T}{I_0}. \quad (23)$$

Коэффициент пропускания T связан с поглощением света, которое в полупроводниках характеризуется законом Бугера – Ламберта – Бера:

$$I_T = I_0 e^{-\alpha d}, \quad (24)$$

где d – толщина полупроводника.

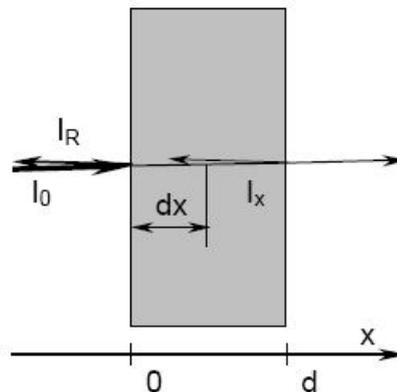


Рис. 3. К выводу закона Бугера – Ламберта – Бера

Выражение (23) описывает пропускание света без учета отражения от поверхности кристалла. Если же имеет место частичное отражение светового потока, то вместо I_0 в (2.23) будет входить величина $(1-R)I_0$, где R – коэффициент однократного отражения света, равный отношению интенсивности отраженной световой волны I_R к интенсивности падающей I_0 волны:

$$R = \frac{I_R}{I_0}. \quad (25)$$

Таким образом, $(1 - R)$ характеризует долю интенсивности световой волны, вошедшую через освещаемую поверхность внутрь кристалла. В реальных кристаллах

конечной толщины свет отражается не только от освещаемой поверхности, но и от противоположной ей, другой поверхности плоского образца. В результате в толще образца свет испытывает внутреннее многократное отражение.

На рис. 4 показаны световые потоки в образце при многократном отражении. После каждого акта отражения интенсивность прошедшего света уменьшается и в конечном счете значение I_T определяем как сумму убывающей геометрической прогрессии:

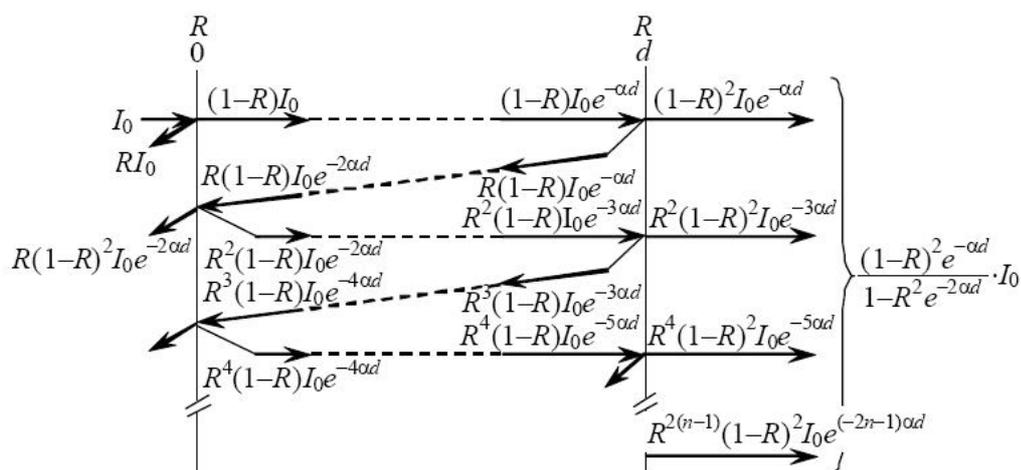


Рис. 4. Световые потоки в образце при многократном отражении.

$$I_T = I_0 (1 - R)^2 e^{-\alpha d} \sum_{k=1}^{\infty} R^{2(k-1)} e^{2(k-1)\alpha d} e^{-2(k-1)\alpha d}. \quad (26)$$

После суммирования получим

$$I_T = I_0 \frac{(1 - R)^2 e^{-\alpha d}}{1 - R^2 e^{-2\alpha d}}. \quad (27)$$

Виды поглощения света в кристаллах.

Механизмы взаимодействия света с полупроводником в зависимости от энергии падающего излучения E и соответственно от длины волны λ можно разделить на несколько групп (рис. 5). При воздействии света на кристалл в нем происходит взаимодействие квантов света с двумя подсистемами – атомной и электронной, из которых состоит кристалл.

В области фотонов высоких энергий (наиболее короткие длины волн) особенности спектров поглощения обусловлены возбуждением электронов из валентной и более глубоко лежащих зон в зону проводимости. При несколько меньшей энергии света поглощение сопровождается переходами электронов из валентной зоны в зону проводимости. При дальнейшем увеличении λ энергия электромагнитного излучения E уменьшается, и поглощение света в кристалле происходит за счет нескольких механизмов, в том числе с

участием примесных уровней и путем взаимодействия квантов света со свободными носителями заряда. Таким образом, взаимодействие квантов света с электронной подсистемой весьма многообразно: собственное, экситонное, примесное и поглощение свободными носителями заряда.

В области наиболее длинных волн происходит взаимодействие света с атомами кристаллической решетки – свет поглощается на оптических колебаниях кристаллической решетки, что приводит к превращению квантов света в энергию фононов; т. е. при поглощении фотона в кристалле появляются один (однофононное) или несколько фононов (многофононное поглощение). Такой тип взаимодействия получил название решеточного поглощения. Рассмотрим далее более подробно некоторые виды поглощения электромагнитного излучения кристаллическими телами.

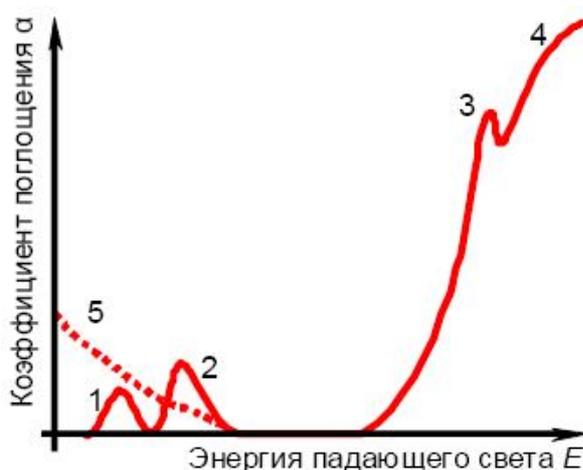


Рис. 5. Спектральная зависимость коэффициента поглощения: 1 – поглощение на оптических колебаниях решетки; 2 – ионизация примесных атомов; 3 – экситонное поглощение; 4 – собственное поглощение; 5 – поглощение свободными носителями

Собственное поглощение.

Если при поглощении полупроводником кванта света электроны валентной зоны, приобретая дополнительную энергию, превышающую или равную ширине запрещенной зоны (E_g), т. е. $\hbar\omega \geq E_g$, переходят в зону проводимости, то такое поглощение называется *собственным*, или *фундаментальным*.

При взаимодействии электрона с электромагнитным полем излучения должны выполняться два закона – закон сохранения энергии и закон сохранения квазиимпульса или волнового вектора:

$$E_n = E_p + \hbar\omega, \quad (28)$$

$$\vec{p}_n = \vec{p}_p + \vec{p}_{\text{фот}}, \quad (29)$$

где $\vec{p}_n = \hbar k_n$ – импульс электрона, переброшенного в зону проводимости, $\vec{p}_p = \hbar k_p$ – импульс дырки, образовавшейся в валентной зоне, $\vec{p}_{\text{фот}}$ – импульс фотона.

Импульс фотона примерно на три порядка меньше импульса электрона, поэтому в выражении (29) им можно пренебречь и тогда (29) можно записать в следующем виде:

$$\hbar k_n \approx \hbar k_p, \quad k_n \approx k_p. \quad (30)$$

Соотношения (30) называют правилами отбора для электронных переходов; они показывают, что в процессе взаимодействия электрона с полем излучения возможны только такие переходы, при которых волновой вектор электрона сохраняется. Эти переходы получили название вертикальных или прямых межзонных переходов (переход 1 на рис. 6).

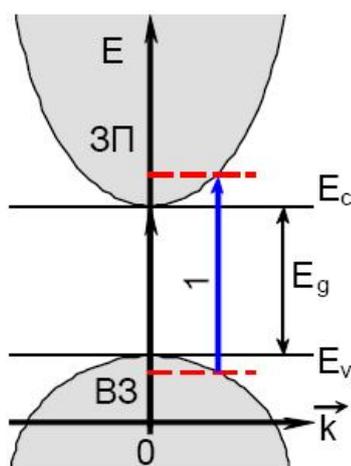


Рис. 6. Оптические переходы при собственном поглощении в прямозонных полупроводниках

Они характерны для прямозонных полупроводников, у которых минимум зоны проводимости и максимум валентной зоны находятся при одном и том же значении волнового вектора. Такими полупроводниками, в частности, являются ZnSe, CdS, GaAs, AgIn₅S₈.

У непрямозонных полупроводников (Si, Ge, GaP, CdP₂, In₂S₃) экстремумы зоны проводимости и валентной зоны находятся при различных значениях волнового вектора k , а такие переходы называют непрямыми межзонными переходами (переход 2 на рис. 7).

Такие переходы происходят с участием третьей квазичастицы – фонона, представляющего собой квант тепловых колебаний кристаллической решетки. В этом случае законы сохранения энергии и импульса будут иметь вид

$$E_n = E_p + \hbar\omega \pm E_{\text{фон}}, \quad (31)$$

$$p_n = p_p + p_{\text{фот}} \pm p_{\text{фон}}. \quad (32)$$

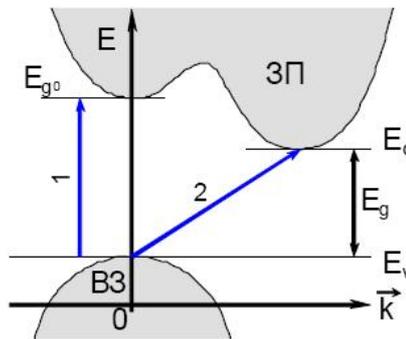


Рис. 7. Оптические переходы при собственном поглощении в непрямозонных полупроводниках

Знак «плюс» относится к процессам, протекающим с поглощением фонона, «минус» – с испусканием. В то же время в случае достижения энергией фотона значения $\hbar\omega \geq E_{g0}$ в таких полупроводниках возможны и прямые межзонные переходы (переход 1 на рис.7).

Экситонное поглощение.

При поглощении света полупроводником возможно и такое возбуждение электрона валентной зоны, при котором он не переходит в зону проводимости, а образует с дыркой связанную систему. Такая система получила название экситон. Если размеры экситона велики по сравнению с постоянной кристаллической решетки, то взаимодействие электрона и дырки можно представить как кулоновское взаимодействие двух точечных зарядов, ослабленное в $\epsilon\epsilon_0$ раз, где ϵ – относительная диэлектрическая проницаемость кристалла, ϵ_0 – электрическая постоянная. Экситон сходен с возбужденным атомом водорода. В обоих случаях около единичного положительного заряда движется электрон, и энергетический спектр является дискретным (рис.8).

Экситонные уровни располагаются в запрещенной зоне у дна зоны проводимости.

Экситон может перемещаться по кристаллу. Однако это не движение данной пары электрон-дырка, а перемещение возбужденного состояния атома кристалла, которое передается от атома к атому посредством квантово-механического резонанса: когда электрон, находящийся в возбужденном состоянии, при взаимодействии с фотоном или фононом возвращается в нормальное состояние, а освободившаяся при этом энергия передается электрону соседнего атома, который переходит на тот же уровень. Так как экситон образован парой электрон-дырка, то движение экситонов не создает электрического тока.

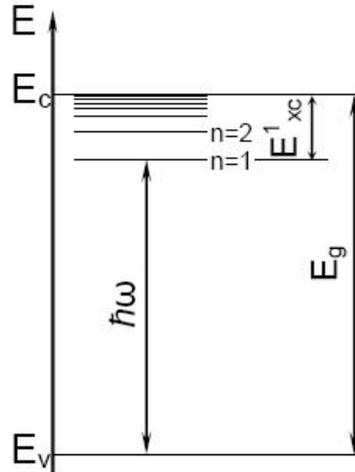


Рис. 8. Энергетический спектр экситонных состояний.

Поскольку, как уже было сказано, экситоны являются электрически нейтральными системами, постольку их возникновение в полупроводниках не приводит к появлению дополнительных носителей заряда и поглощение света не сопровождается увеличением проводимости полупроводников. При столкновении же с фононами, примесными атомами и другими дефектами кристаллической решетки полупроводника экситоны или рекомбинируют, или «диссоциируют». В первом случае возбужденный электрон переходит в валентную зону, а энергия возбуждения передается кристаллической решетке или излучается в виде квантов света. Во втором случае образуется пара носителей заряда – электрон и дырка, которые обуславливают повышение электропроводности полупроводников.

Если оптический переход прямой и осуществляется при $k = 0$, экситон будет характеризоваться водородоподобной серией линий поглощения, которые удовлетворяют соотношению

$$\hbar\omega = \hbar\omega_{\infty} - \frac{E_{xc}}{n^2}, \quad (33)$$

где n – главное квантовое число, $\hbar\omega_{\infty}$ совпадает с E_g при $k = 0$.

Поэтому (33) можно переписать в виде

$$\hbar\omega = E_g - \frac{E_{xc}}{n^2}, \quad (34)$$

где E_{xc} – энергия связи экситона, которая в свою очередь определяется по формуле:

$$E_{xc} = \frac{e^4 m_r}{32\pi^2 \hbar^2 \epsilon^2 \epsilon_0^2}, \quad (35)$$

где приведенная эффективная масса экситона определяется равенством:

$$m_r = \frac{m_n m_p}{m_n + m_p}.$$

В прямозонных полупроводниках экситонному поглощению соответствует узкий пик на краю собственного поглощения (рис. 9,а), который уширяется с увеличением температуры.

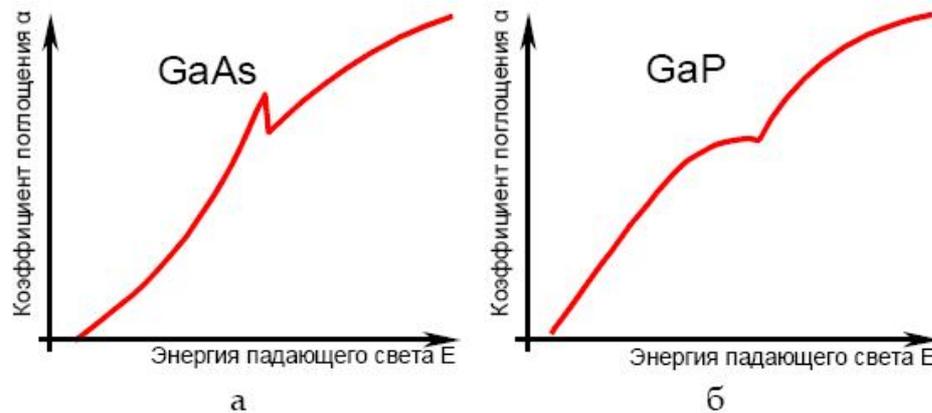


Рис. 9. Экситонное поглощение в прямо- (а) и непрямоzonных (б) полупроводниках.

В случае не прямых переходов экситоны образуются с поглощением или с испусканием фононов. Соответственно в непрямоzonных полупроводниках образование экситона проявляется в возникновении на краю полосы собственного поглощения ступеньки (рис. 2.9, б).

Образование экситонов и, следовательно, экситонное поглощение наблюдается при низких температурах; при более высоких температурах связь между электроном и дыркой разрывается и образуется пара свободных носителей.

Диссоциация экситонов с образованием свободных электронов и дырок может также происходить при их взаимодействии с фотонами, обладающими достаточной энергией.

Поглощение света свободными носителями заряда.

При освещении полупроводника светом определенной длины волны электроны зоны проводимости и электроны не полностью заполненной валентной зоны могут осуществлять переходы внутри зоны с одного уровня на другой (рис. 10,а).

Такие внутризонные переходы происходят с изменением волнового вектора и нарушением правил отбора. В силу закона сохранения импульса они осуществляются только тогда, когда наряду с поглощением фотона происходит поглощение или испускание фонона (т. е. происходит взаимодействия с решеткой или с ионами примесей).

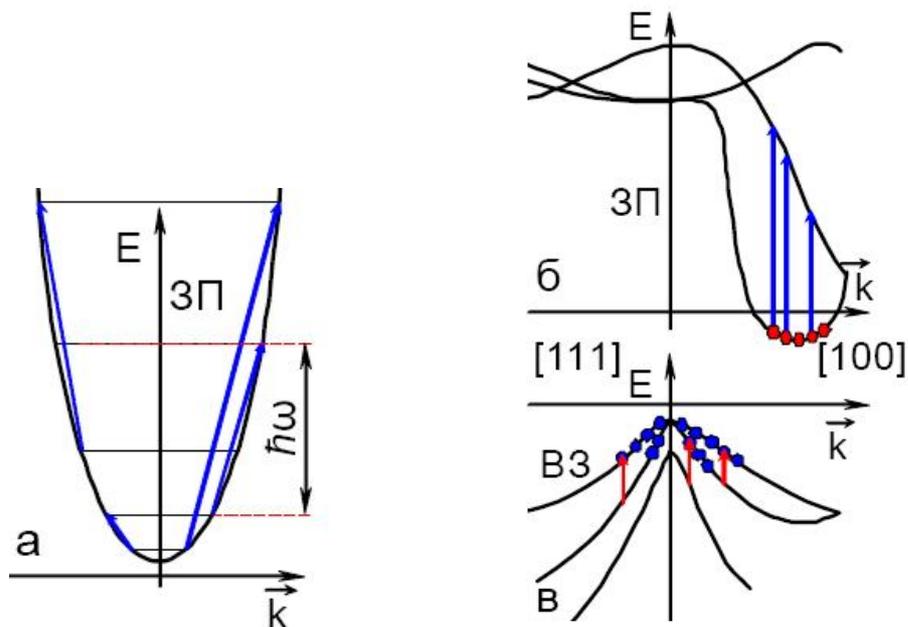


Рис. 10. Оптические переходы при поглощении света свободными носителями заряда

Поглощение свободными носителями заряда происходит следующим образом. Носители заряда ускоряются в электрическом поле световой волны и, рассеиваясь на дефектах кристаллической решетки, передают им свою энергию, т. е. энергия световой волны переходит в теплоту вследствие эффекта Джоуля – Ленца. Это поглощение пропорционально концентрации свободных носителей заряда n , квадрату длины волны падающего света λ и обратно пропорционально дрейфовой подвижности μ :

$$\alpha = \frac{n\lambda^2}{\mu}. \quad (36)$$

Если энергетические зоны у полупроводника сложны (Ge, Si, GaP, AlAs), то на поглощение свободными носителями заряда накладывается поглощение, обусловленное переходами между отдельными подзонами зоны проводимости (рис. 2.10, б) и валентной зоны (рис. 2.10, в).

Примесное поглощение.

В примесных полупроводниках под действием света может происходить переброс электронов с примесных уровней в зону проводимости (рис. 11) и из валентной зоны на примесные уровни, расположенные в запрещенной зоне (рис. 11).

Могут также наблюдаться не прямые переходы примесный уровень – зона с поглощением или испусканием фонона. Поглощение, обусловленное ионизацией или возбуждением примесных центров в кристалле, называется примесным. Граница этого

перехода сдвинута в длинноволновую область тем сильнее, чем меньше энергия соответствующего перехода.

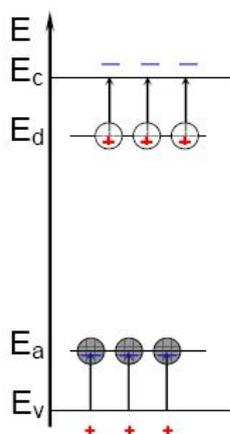


Рис. 11. Переходы при примесном поглощении

Следует отметить, что если примесные атомы уже ионизированы, то указанное поглощение наблюдаться не будет. Так как температура ионизации примеси падает с уменьшением энергии ее ионизации, то для наблюдения длинноволнового поглощения необходимо охлаждение полупроводника до низких температур. Например, спектр примесного поглощения германия легированного золотом (энергия ионизации примеси $E_{Au} = 0,08$ эВ, граница поглощения $\lambda = 9$ мкм) наблюдается при 77 К, а при легировании сурьмой ($E_{Sb} = 0,01$ эВ, $\lambda = 135$ мкм) примесное поглощение наблюдается при 4 К.

Коэффициент примесного поглощения $\alpha_{пр}$ при $\lambda < \lambda_{max}$ зависит от λ и пропорционален концентрации примеси $N_{пр}$. Поэтому примесное поглощение удобно характеризовать сечением поглощения $\sigma_{пр}$:

$$\sigma_{пр} = \frac{\alpha_{пр}(\lambda)}{N_{пр}}. \quad (37)$$

Максимального значения $\sigma_{пр}$ достигает вблизи края примесного поглощения. В максимуме $\sigma_{пр} \approx 10^{-16} - 10^{-15}$ см² и $\alpha_{пр} \approx 1 - 10$ см⁻¹ при $N_{пр} = 10^{16}$ см⁻³. Получить большее значение $\alpha_{пр}$ не удается из-за низких растворимостей примесей в полупроводниках. Например, предел растворимости золота в германии составляет $\sim 10^{16}$ см⁻³.

2. Фотопроводимость в полупроводниках

Фотопроводимость – это эффект увеличения электропроводности образца под действием падающего излучения. Величина электропроводности определяется концентрацией носителей заряда n и их подвижностью μ :

$$\sigma = q \cdot n \cdot \mu,$$

где q – заряд электрона.

В процессе поглощения света происходит рост концентрации свободных носителей заряда и проводимость полупроводника возрастает.

Из всех рассмотренных механизмов поглощения только три сопровождаются увеличением числа носителей – собственное поглощение, примесное и экситонное (в случае диссоциации экситона). Поглощение света свободными носителями так же, как и поглощение на оптических колебаниях кристаллической решетки, сопровождаются увеличением тепловых колебаний кристаллической решетки. В результате процессы рассеяния носителей заряда усиливаются, их подвижность падает, и проводимость полупроводника уменьшается.

Эффект Франца – Келдыша.

Эффект заключается в изменении местоположения края полосы собственного поглощения при действии на образец внешнего поля. Во внешнем поле энергетические уровни полупроводника приобретают наклон, пропорциональный его напряженности E (рис. 12). При этом становится возможным туннельное просачивание электронов из точки x_2 в x_3 и из точки x_3 в x_1 . Поэтому в точке x_3 электрон может перейти из состояния E_p в состояние E_n под действием энергии кванта света с меньшей величиной, чем E_g .

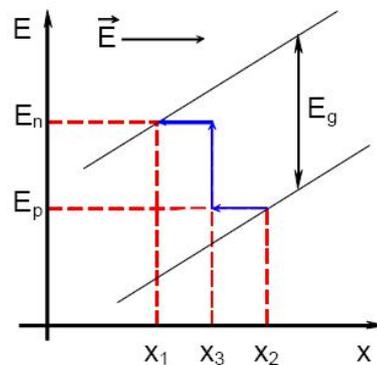


Рис. 12. К объяснению эффекта Франца – Келдыша

В заключение приведем примерную последовательность действий, направленных на экспериментальное изучение материалов методом оптической спектроскопии. При изучении оптического поглощения полупроводников они обычно сводятся к следующему:

1. Изучаются основные механизмы поглощения света материалом, изложенные выше.
2. Из справочной литературы находят данные (для материалов, подобных исследуемому) по спектральной зависимости пропускания (T , %) от длины волны (λ , нм).
3. С помощью измерений находят экспериментальную зависимость пропускания (T , %) от длины волны (λ , нм).

4. По выражению (38) с использованием какой либо компьютерной программы производится расчет коэффициента поглощения α :

$$\alpha = \frac{1}{d} \ln \left[\frac{(1-R)^2}{2T} + \sqrt{\left(\frac{(1-R)^2}{2T} \right)^2 + R^2} \right], \quad (38)$$

где d – толщина образца, T – пропускание (относительные единицы), R – коэффициент отражения.

5. Полученные данные заносятся в таблицу, подобную нижеприведенной.

λ , нм	T , %	λ , мкм	T , отн. ед.	α , см ⁻¹	$\hbar\omega$, эВ	$(\alpha \cdot \hbar\omega)^{1/2}$, см ⁻²	$(\alpha \cdot \hbar\omega)^2$, см ⁻¹

6. Определяется энергия падающего фотона, используя расчетную формулу:

$$\hbar\omega = \frac{1,24}{\lambda},$$

где λ – длина волны падающего света в мкм.

7. Проводится расчет значений $(\alpha \cdot \hbar\omega)^{1/2}$ и $(\alpha \cdot \hbar\omega)^2$ и строятся графические зависимости $(\alpha \cdot \hbar\omega)^{1/2}$ и $(\alpha \cdot \hbar\omega)^2$ от энергии фотона ($\hbar\omega$). Определяется ширина запрещенной зоны (E_g) для непрямых и прямых межзонных переходов, используя метод экстраполяции прямолинейных участков этих зависимостей до пересечения с осью энергий.

8. Производится расчет коэффициента преломления материала, для чего используется график зависимости $T(\lambda)$. На данном графике строим две касательные таким образом, как показано на рис. 13. Далее из каждого максимума (T_{max}^1) и минимума (T_{min}^2) проводим перпендикуляры к касательным, как показано на рис. 13.

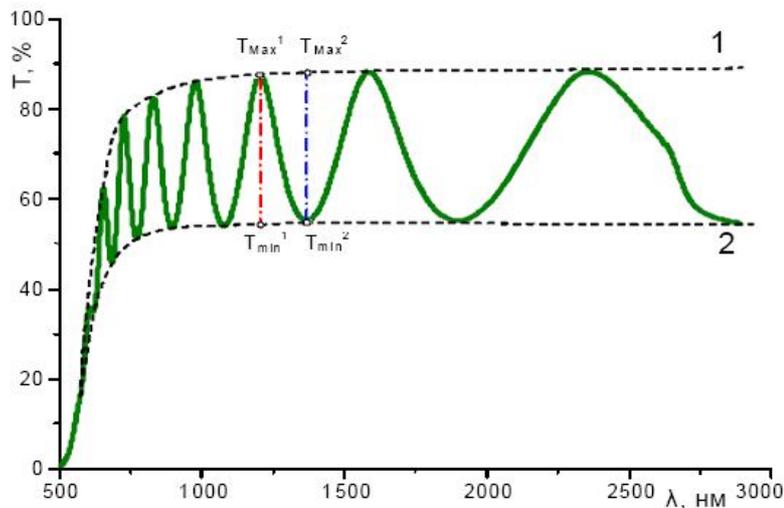


Рис. 13. К расчету коэффициента преломления

Полученные результаты вносятся в таблицу, подобную ниже следующей.

№ п/п	λ , нм	T_{max}^n , %	T_{min}^n , %	n
1				
2				
...				
n				

Далее по формулам (40) и (41) рассчитываем коэффициенты преломления n для полученного материала:

$$n = \left[N + (N^2 - s^2)^{1/2} \right]^{1/2}, \quad (40)$$

$$N = 2s \frac{T_{max} - T_{min}}{T_{max} T_{min}} + \frac{s^2 + 1}{2}, \quad (41)$$

где s – коэффициент преломления подложки (в качестве подложки использовали стекло, поэтому коэффициент преломления $s = 1,5$), T_{max} – значение максимума, T_{min} – значение минимума. В этих зависимостях каждому максимуму соответствует свой минимум (для T_{max}^1 соответствует T_{min}^1) и каждому минимуму соответствует свой максимум.

9. По полученным данным построить график зависимости коэффициента преломления n от длины волны λ , нм.

10. Используя полученные данные, пересчитать коэффициент преломления R по формуле

$$R = \frac{(n_{cp} - 1)^2}{(n_{cp} + 1)^2}, \quad (42)$$

где n_{cp} – среднее значение коэффициента преломления.

Лекция 7. Кристаллическая структура материалов по данным рентгеноструктурного анализа.

Открытие в 1912 г. Лауэ и его сотрудниками дифракции рентгеновских лучей привело к разработке одного из самых совершенных прямых методов изучения твердых материалов – рентгеноструктурного анализа. Метод основан на получении и анализе дифракционной картины, возникающей в результате интерференции рентгеновских лучей, рассеянных электронами атомов облучаемого объекта.

Основной задачей рентгенофазового анализа является идентификация различных фаз в их смеси на основе анализа дифракционной картины, даваемой исследуемым образцом. Широкое распространение из-за простоты и универсальности получил метод порошка. Постепенное усовершенствование методики рентгенофазового анализа, улучшение конструкций рентгеновских камер и рентгеновских установок, широкое распространение легкозаменяемых электронных трубок привело к повсеместному внедрению рентгенофазового анализа в практику работы химических и технических лабораторий.

Рентгеновские камеры и дифрактометры различаются способом регистрации дифракционной картины. В камерах эта картина фиксируется на фотографическую пленку, в дифрактометрах – с помощью различного типа счетчиков (сцинтилляционных, пропорциональных, полупроводниковых).

Широкое применение способов регистрации рентгеновских лучей (счетчики) и усовершенствование фотографического метода (фокусирующие камеры – монохроматоры) расширило границы чувствительности фазового анализа и увеличило надежность идентификации фаз в самых сложных случаях.

К достоинствам рентгеновского фазового анализа должны быть отнесены: высокая достоверность и экспрессность метода; то, что метод – прямой, так как в принципе основан не на косвенном сравнении с какими-либо эталонами или изменениями свойств, а непосредственно указывает на кристаллическую структуру вещества, наилучшую характеристику каждой фазы, наконец, этот метод не требует большого количества вещества; анализ можно проводить без нарушения образца или детали; кроме того, метод допускает оценку количества фаз в смеси.

Здесь мы рассмотрим методику получения порошковых рентгенограмм, с последующим качественным и количественным анализом. В приложении приводятся справочные данные для расшифровки дифрактограмм некоторых соединений, часто применяемых в технологиях оксидных и силикатных материалов.

Дифракция рентгеновских лучей.

К рентгеновским лучам относят излучение, занимающее участок электромагнитного спектра от нескольких сотен до десятых долей ангстрема (10^{-10} м) или от нескольких десятков до сотых долей нанометра (10^{-9} м). Это жесткое коротковолновое излучение. Расстояние между атомами в кристаллической решетке твердых тел колеблется от единиц до полутора десятков ангстрем. Прохождение рентгеновских лучей через вещество сопровождается разными видами взаимодействия, одним из которых является рассеяние

рентгеновских лучей: вещество, которое подвергается действию рентгеновского излучения, испускает вторичное излучение, длина волны которого равна длине волны падающих лучей (когерентное рассеяние). Каждый изолированный атом рассеивает излучение равномерно во все стороны в виде концентрических сфер. Если падающая волна рентгеновского излучения перпендикулярна атомному ряду, то все атомы ряда одновременно излучают электромагнитные колебания. Поскольку расстояние между атомами соизмеримо с длиной волны вторичного когерентного излучения, то кристалл может служить для него дифракционной решеткой. Энергия этого излучения рассеивается в разных направлениях с различной интенсивностью: по одним направлениям усиливается, по другим ослабляется и даже полностью гасится. Усиление колебаний происходит по тем направлениям, где разность хода рентгеновских лучей равна целому числу волн или четному числу полуволн. Это правило (условие интерференции) справедливо для любого излучения. В результате образуется серия плоских волн, которые распространяются в особых направлениях.

Реальные кристаллы не имеют гипотетического линейного порядка. Геометрические условия дифракции от трехмерно расположенных атомов более жестки, чем в случае одномерной цепочки, от которой может дифрагировать любая волна, в случае если ее $\lambda < d$. Английские ученые Уильямы Брегги (отец и сын) и русский ученый Г.В.Вульф показали, что дифрагированный луч можно рассматривать как результат отражения от одной из плоскостей атомной решетки. Любая трехмерная решетка рассматривается как совокупность бесконечного числа параллельных атомных плоскостей, расположенных на равном расстоянии друг от друга. Отражение лучей будет происходить не только от внешней поверхности, а от всех атомных плоскостей, так как рентгеновский луч, в отличие от оптического излучения, проникает вглубь кристалла.

Рассмотрим одну из таких систем плоскостей, характеризующихся межплоскостным расстоянием d (рис.1). Каждая плоскость отражает луч под одним и тем же углом θ , так как рентгеновские лучи распространяются в веществе практически без преломления.

При отражении лучи могут интерферировать в том случае, если их разность хода ($\Delta(OB)$) будет равна целому числу волн $\Delta(OB) = n\lambda$. Из прямоугольных треугольников AOO' и BOO' видно, что $AOB = 2d \cdot \sin\theta$.

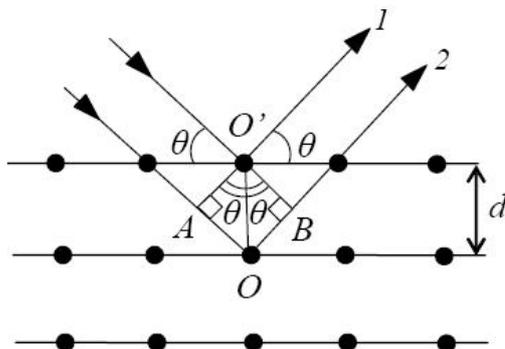


Рис. 1. Отражение рентгеновских лучей в кристалле по Брэггу

Следовательно, условие дифракции рентгеновских лучей (Вульфа -Брэггов) будет иметь вид:

$$2d \cdot \sin\theta = n\lambda. \quad (1)$$

n - называют порядком отражения ($n = 1, 2, 3$).

Это уравнение лежит в основе всех рентгеновских методов исследования кристаллических тел. Его используют для расчета межплоскостных расстояний d , которые собственно являются характеристикой конкретной кристаллической решетки. Величины θ и λ - данные эксперимента. Выполнение условия Вульфа - Брэггов при изменении угла падения рентгеновского излучения регистрируется на рентгенограмме возникновением максимума или почернением на фотопленке. Интенсивность этого максимума или почернения соответствует интенсивности отраженного луча, которое в свою очередь зависит от количества атомов, составляющих данное семейство плоскостей, то есть от "заселенности" атомной плоскости. Поэтому интенсивность отраженного луча также является характеристикой изучаемого объекта.

Методы съемки рентгенограмм и аппаратура.

Методы съемки рентгенограмм делят на фотографические (с регистрацией рентгеновского излучения на фотопленку) и дифрактометрические. В первом случае дифракционная картина одновременно регистрируется на фотопленке и представляет собой ряд concentric пар дужек (черных полос).

Дифрактометрические методы съемки рентгенограмм отличаются от фотографических тем, что дифракционная картина регистрируется последовательно во времени. В данном случае используется счетчик отраженных рентгеновских лучей, который

перемещается по окружности таким образом, что угол дифракции θ при этом непрерывно изменяется.

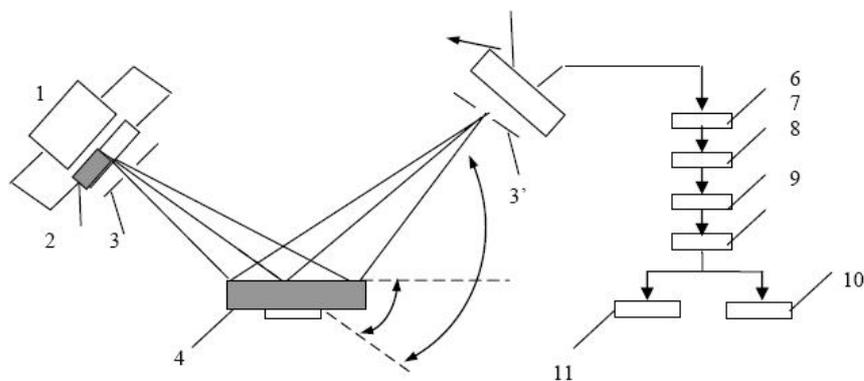


Рис. 2. Принципиальная схема дифрактометра.

1– источник высокого напряжения, 2– рентгеновская трубка, 3,3'– диафрагмы, 4– образец, 5– счетчик квантов, 6– фотоэлектронный умножитель, 7– усилитель, 8– дискриминатор, 9– пересчетная схема, 10– самопишущее устройство, 11– интенсиметр.

Для получения интенсивных рефлексов на рентгенодифрактограмме необходимо использовать фокусирующие методы съемки, при которых в достаточно узкую щель счетчика попадает рентгеновское излучение, отраженное от образца с относительно большой поверхностью. При фокусировке по схеме Брэгга-Брентано, которая используется, например, в дифрактометре ДРОН - 3 (рис. 2) образец плоский расположен в центре окружности I постоянного радиуса, по которой движется счетчик 5 и на которой находится рентгеновская трубка 2. При этом образец 4 вращается одновременно со счетчиком таким образом, чтобы поверхность образца все время была касательной к окружности фокусировки II, на которой в данный момент находятся фокус рентгеновской трубки, центр образца и входная щель счетчика. Это условие выполняется, если угловая скорость вращения счетчика в два раза превышает угловую скорость вращения образца. Следовательно, если образец поворачивается на угол θ , то угол поворота счетчика будет 2θ . Измерение углов поворота осуществляется с помощью гониометра.

Разность потенциалов между анодом и катодом рентгеновской трубки 2 создается генераторным устройством I. Рентгеновский луч, пройдя диафрагмы 3 и 3', отразившись от образца 4, попадает в счетчик 5, где преобразуется в излучение видимой области. Световой сигнал в фотоэлектронном умножителе 6 преобразуется в электрический, усиливается в 7, дискриминатором 8 отделяются импульсы, не отвечающие характеристическому излучению. После дискриминатора импульсы регистрируются пересчетной схемой 9 и записываются

самописцем 10 на диаграммную ленту. С помощью интенсиметра 11 фиксируется значение интенсивности отраженного рентгеновского излучения, пропорциональное количеству его квантов.

Типичный вид рентгенограммы, полученной на дифрактометре, представлен на рис.3. Значения угла θ для расчета d отвечающего любому рефлексу, определяются положением центра тяжести пика. Чаще всего, если пик (рефлекс) достаточно узкий и симметричный, пользуются положением максимума. На основе анализа рентгенограммы образца часто в повседневной практике, так же как и в научной литературе приводят не сами рентгенограммы, а штрих-диаграммы вещества или химического соединения. Штрих-диаграмма представляет собой линейчатый спектр интенсивностей рефлексов при определенных значениях межплоскостных расстояний d , определенных из первичной рентгенограммы. Штрих-диаграмма строится в определенном масштабе по отношению к исходной рентгенограмме и имеет вид, приведенный на рис.4.

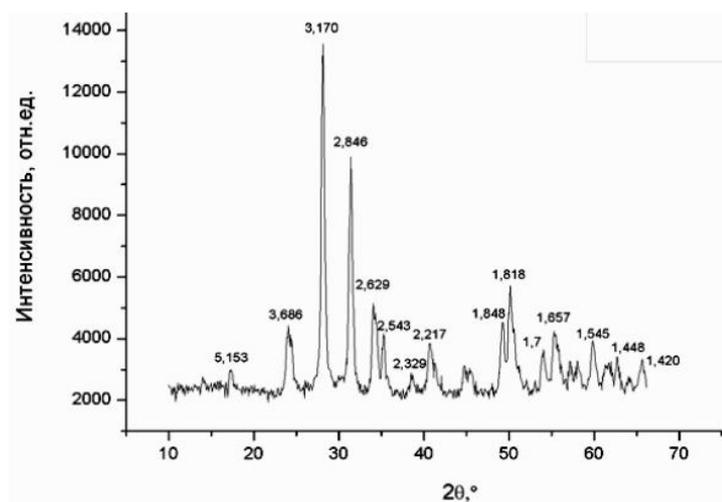


Рис. 3. Дифрактограмма кристаллической фазы керамики(ZrO_2)

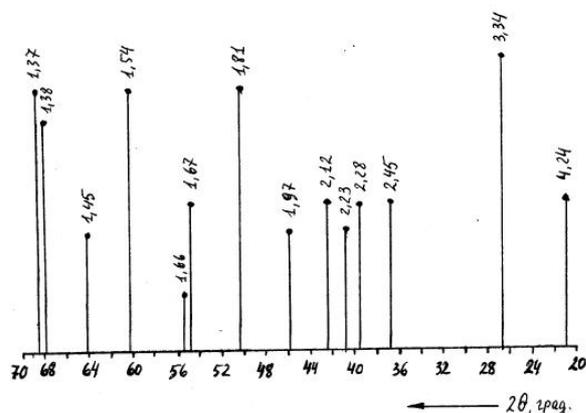


Рис. 4. Штрихрентгенограмма (в \AA) β - кварца

Подготовка образцов.

Для получения достаточно хороших рентгенограмм необходимо тщательно готовить образцы для съемки. Самым распространенным методом съемки является метод порошка, по которому образец является поликристаллическим телом, полученным из тонко измельченного порошка. Для съемки на фотопленке порошок наносится на нить или набивается в капилляр диаметром 0,7 мм.

Для съемки дифрактограмм (рентгенограмм на самописец) порошок насыпается и фиксируется в углублении специальной кюветы из кварцевого стекла, если для фиксации не достаточно простого придавливания, применяют органические клеи (БФ, цапонлак и т.д., а также обычные технические масла). Кроме того, порошок можно спрессовать в виде таблетки диаметром до 25 мм или применить для съемки плоский образец в твердом виде.

Важным фактором, определяющим чувствительность метода, является размер кристалликов исследуемого вещества. Поэтому следует обратить внимание на тщательность растирания порошка, так как порошок, состоящий из крупных кристаллов, дает нечеткие, малоинтенсивные рентгенограммы.

Растирать порошок следует в агатовой ступке агатовым пестиком (для исключения загрязнения пробы) до прохождения через сито 10000 отв./см² (0063; 63 мкм). Оптимальный размер кристаллов около 5 - 10 мкм. Вместе с тем в некоторых случаях нельзя сильно перетирать пробу, так как сильное воздействие (особенно с давлением) приводит к нарушению кристаллической структуры препарата, появлению напряжений в кристаллах (а значит, к ухудшению качества рентгенограммы). В некоторых случаях при растирании пробы происходит полиморфное превращение. Это явление можно использовать при исследовании стабильности материалов. Но, следует помнить

также, что в случае, если размер кристалликов ниже определенного (менее 0,1 мкм), то интерференционные линии могут быть размыты; и при малом количестве фазы её линии могут сливаться с фоном.

Керамика, ситалл или изделия на основе вяжущих материалов чаще всего представляют собой поликристаллический материал. Поэтому при фазовом анализе можно использовать непосредственно поверхность образца или воспользоваться методом порошка.

Методика расшифровки рентгенограмм.

Основное назначение расшифровки рентгенограмм - определение значений d/n (d) и относительных интенсивностей рефлексов.

При съемке рентгенограмм на дифрактометрах, например, типа ДРОН -3 на диаграммной ленте самописца автоматически проставляются реперные линии через 0,5 или 1 градус (2θ). Для определения угла 2θ , отвечающего данному рефлексу, измеряют расстояние от максимума пика (рефлекса) до ближайшей реперной линии с меньшим углом и прибавляют к значению угла для данной реперной линии соответствующее этому расстоянию изменение угла 2θ до максимума рефлекса. По найденному значению угла 2θ определяют θ и по этому углу по справочным таблицам для данного анода (антикатада) находят межплоскостное расстояние d/n , соответствующее анализируемому пику, например, для $\theta = 10$, межплоскостное расстояние $d/n = 4,4394 \text{ \AA}$ (0,4439 нм). Значение d можно так же получить расчетом из уравнения Вульфа-Брэгга.

$$d = \frac{n\lambda}{2 \sin \Theta} \quad (2)$$

Относительную интенсивность рефлексов определяют следующим образом. Находят площадь между контуром каждого пика и уровнем фона (базисной линией), интенсивность пика, для которого она максимальна, принимают за 100. Относительные интенсивности остальных пиков находят как

$$I_{\text{отн}} = (S/S_{\text{max}}) \cdot 100 \quad (3)$$

Площадь пика приближенно может быть определена по формуле:

$$S = h \frac{l}{2}; \quad (4)$$

где: h - высота пика, мм или см; $l/2$ – его ширина на половине высоты ($h/2$).

Если оценка интенсивности рефлексов не требует высокой точности, то вместо значений S можно воспользоваться значениями h (что часто применяется на практике). Такая замена приводит к тем меньшему искажению значений интенсивности, чем меньше значение 2θ .

Возможности метода рентгенографии.

В зависимости от поставленной задачи с помощью рентгенографического анализа можно решать различные вопросы для большого круга тугоплавких неметаллических и силикатных материалов. Чаще всего в технологии силикатов анализу подвергаются поликристаллические вещества. Следует отметить, что полная расшифровка структуры такого материала возможна только в исключительных случаях. Однако исследование поликристаллического вещества методами рентгенографии позволяет решать следующие задачи:

1. Определять присутствующие кристаллические фазы (качественный рентгенофазовый анализ);

2. Определять количественное содержание кристаллических фаз в многофазном материале (количественный рентгенофазовый анализ);
3. Определять размеры элементарной ячейки;
4. Определять микронапряжения кристаллической решетки, размеры зерен поликристаллов и блоков мозаичной структуры;
5. Исследовать текстуру вещества;
6. Изучать степень дальнего порядка и распределение катионов в кристаллическом веществе;
7. Определять тип твердого раствора, степень его упорядоченности, границы растворимости.

Рентгеновский фазовый анализ может быть с успехом применен для изучения исходных материалов (природного и техногенного происхождения) всех производств силикатных и оксидных материалов; его можно использовать также для анализа синтезированных продуктов в процессе их изготовления и их изменения в процессе эксплуатации

Далее будут рассмотрены варианты решения наиболее часто встречающихся задач, возникающих при исследовании силикатных и других тугоплавких неметаллических материалов.

Определение качественного фазового состава вещества.

Фазовым анализом называется установление числа фаз в данной системе и их идентификация. Рентгеновский метод фазового анализа основан на том, что каждое кристаллическое вещество дает специфическую интерференционную картину с определенным количеством, расположением и интенсивностью интерференционных линий, которые определяются природой и расположением атомов в данном веществе. Так как почти нет двух кристаллических веществ, которые обладали бы одинаковой во всех отношениях кристаллической структурой, то рентгенограммы почти однозначно характеризуют данное вещество и никакое другое. В смеси нескольких веществ каждое из них дает свою картину рентгеновской дифракции независимо от других. Полученная рентгенограмма смеси представляет собой сумму ряда рентгенограмм, которые получились бы, если бы на одну и ту же пленку поочередно снимали каждое вещество в отдельности.

Часто при проведении рентгенофазового анализа качественный химический состав исследуемого вещества известен или можно сделать предположение о нем. Исходя из химического состава образца, можно составить список кристаллических фаз, которые могут присутствовать в исследуемом веществе, или список систем, к которым такие фазы могут относиться. Например, материал, полученный в результате твердофазового взаимодействия при термической обработке смеси CaO и β -кварца (SiO_2), может содержать CaO, CaCO_3 ,

$\text{Ca}(\text{OH})_2$, полиморфные разновидности SiO_2 , а также различные силикаты кальция. Эти соединения могут быть рассмотрены в системе $\text{CaO} - \text{SiO}_2$.

Другим примером является то, что в продуктах кристаллизации магнезиально - кальциевых силикатов (стекол) могут присутствовать модификации SiO_2 , силикаты кальция и магния, а присутствие свободных CaO и MgO практически нереально. Данные процессы могут рассматриваться в системе $\text{CaO} - \text{MgO} - \text{SiO}_2$. Следует учитывать возможность взаимодействия исследуемого вещества при его хранении или в ходе предыдущей тепловой обработки с SiO_2 , H_2O и другими реагентами, содержащимися в окружающей атмосфере. Например, калиево - силикатные стекла и продукты их кристаллизации могут в результате контакта с атмосферой воздуха подвергаться частичной гидратации.

Допустим на основании имеющихся данных о химическом составе исследуемого вещества (или исходя из предварительного петрографического анализа) предполагается, что в нем возможно присутствие кристаллических фаз А, В, С, Д, Е. Значения d/n и I для этих фаз, приведенные в справочниках [23] являются основой для анализа (табл.1).

Таблица 1. Значения d/n и I_0 трех наиболее интенсивных рефлексов фаз А, В, С, Д, Е по справочным данным

Кристаллическая фаза	d/n , нм, (I_0)
А	0,202 (10,0); 0,327 (5,0); 0,288 (5,0)
В	0,322 (10,0); 0,580 (8,0); 0,417 (4,0)
С	0,362 (10,0); 0,202 (10,0); 0,264 (8,0)
Д	0,1868 (10,0); 0,404 (7,0); 0,202 (5,0)
Е	0,321 (10,0); 0,265 (10,0); 0,404 (9,0)

Экспериментально полученная на дифрактометре рентгенограмма исследуемого вещества пусть имеет вид (рис. 5), экспериментальные значения d/n , определенные по значениям θ , и I - в табл.1. Номера рефлексов на рис.5 и в табл.1 соответствуют друг другу. Выбрав по табл. 2 рефлекс с наибольшей интенсивностью ($d/n = 0,321$ нм), устанавливаем (см.табл.1), что он может принадлежать фазе В или Е. Сравнивая значения d/n и I для принадлежащих

этим фазам рефлексов с экспериментальными (табл.1), видим, что нельзя исключить наличие ни одной из этих фаз. Для уточнения обращаемся к полным наборам эталонных значений d/n и I для фаз В и Е (табл.3). Всем рефлексам из этого набора для фазы В, кроме малоинтенсивного рефлекса с $d/n = 0,242$ нм, отвечают пики на экспериментальной рентгенограмме, что указывает на ее присутствие в образце. В тоже время один из интенсивных рефлексов фазы Е ($d/n = 0,291$ нм) отсутствует на ней.

Таблица 1. Экспериментальные значения d/n и I_0 исследуемого вещества и результаты его рентгенофазового анализа

№ рефлекса	d/n , нм	I_0	Принадлежность рефлекса кристаллическим фазам (по итогам анализа)
1	0,581	9,0	В
2	0,521	1,2	В
3	0,472	3,6	С, Д
4	0,418	6,7	В
5	0,404	3,8	Д
6	0,361	7,5	С
7	0,321	10,0	В
8	0,264	6,0	С
9	0,256	4,2	В
10	0,251	4,0	С
11	0,229	1,3	С
12	0,213	1,0	С
13	0,202	8,5	С, Д
14	0,1872	6,5	В
15	0,1643	1,0	С
16	0,1514	0,7	В, Д
17	0,1350	2,7	С
18	0,1333	2,8	С
19	0,1247	0,5	Д
20	0,1173	2,0	С, Д
21	0,1060	3,0	С
22	0,1044	1,8	В
23	0,0988	1,0	С
24	0,0982	1,0	Д
25	0,0941	0,5	С

Поэтому, несмотря на наличие остальных пиков, отвечающих фазе В, делаем вывод о ее отсутствии в исследуемом веществе. Далее выбираем наиболее интенсивный из пиков на экспериментальной рентгенограмме, не принадлежащих фазе В. Этот пик с $d/n = 0,202$ нм, $I = 8,5$. Он может принадлежать фазе А или С. Два остальных рефлекса фазы А ($d/n = 0,327$ и $0,288$ нм; см.табл.2) на экспериментальной рентгенограмме отсутствуют, что позволяет сделать заключение об отсутствии фазы А. Сравнение полного набора рефлексов фазы С (см.табл.3) с экспериментальными данными (см.табл.2) подтверждает ее присутствие. Сравнив рефлексы фазы Д с пиками экспериментальной рентгенограммы, убеждаемся в

присутствии и этой фазы. Часть ее пиков совпадает с пиками, принадлежащими другим фазам; некоторые рефлексы отсутствуют на экспериментальной рентгенограмме ввиду их малой интенсивности. Установленная по итогам анализа принадлежность всех рефлексов экспериментальной рентгенограммы кристаллическим фазам В, С и Д показана в табл. 2.

Таблица 3 - Справочные наборы значений d/n и I_0 фаз В, С, Д, Е

d/n , нм	I_0						
Фаза В		Фаза С		Фаза Д		Фаза Е	
0,580	8,0	0,454	4,0	0,471	3,0	0,404	9,0
0,519	2,0	0,362	10,0	0,404	7,0	0,321	10,0
0,417	4,0	0,264	8,0	0,344	0,5	0,291	8,0
0,322	10,0	0,251	4,0	0,273	1,0	0,265	10,0
0,256	5,0	0,229	2,0	0,202	5,0	0,255	6,0
0,242	0,3	0,213	1,5	0,1868	10,0	0,230	3,0
0,1874	3,0	0,202	10,0	0,1510	2,0	0,1872	6,0
0,1512	1,0	0,1645	1,2	0,1248	3,5	0,1349	3,0
0,1045	1,5	0,1350	4,0	0,1172	1,5	0,1060	2,0
		0,1333	4,0	0,1010	1,0		
		0,1176	1,0				
		0,1060	4,0				
		0,0988	1,0				

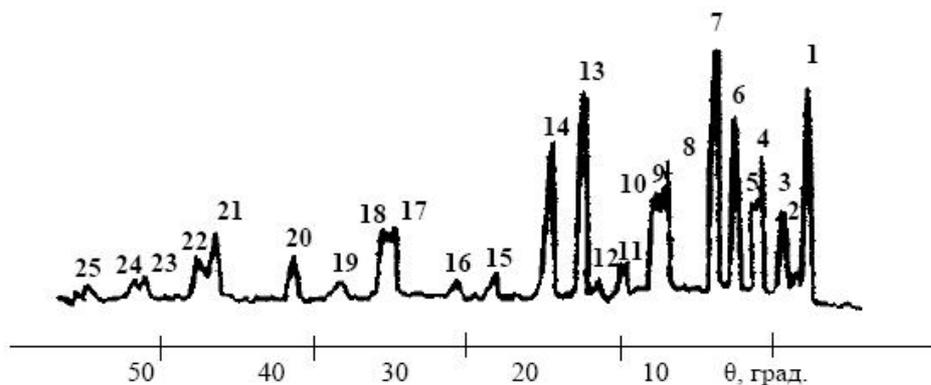


Рис. 5. Рентгенограмма многофазного образца

Лекция 8. Метод термо-ЭДС: физические принципы и применения для исследования проводящих материалов.

Термоэлектрические эффекты (в неоднородном температурном поле и под влиянием тока)

Работа термоэлектрических холодильных машин и генераторов базируется на термоэлектрических явлениях. К их числу относятся эффекты Зеебека, Пельтье и Томсона. Эти эффекты связаны, как с превращением тепловой энергии в электрическую, так и с превращением энергии электрического тока в холод.

Эффект Зеебека состоит в том, что если в разомкнутой электрической цепи, состоящей из нескольких разнородных проводников, на одном из контактов поддерживать температуру T_r (горячий спай), а на другом температуру T_x (холодный спай), то при условии $T_r > T_x$ на концах цепи возникает термоэлектродвижущая сила E . При замыкании контактов в цепи появляется электрический ток.

Эффект Пельтье заключается в том, что при пропускании постоянного тока через термоэлемент, состоящий из двух проводников или полупроводников, в месте контакта выделяется или поглощается некоторое количество теплоты (в зависимости от направления тока). Когда электроны переходят из материала p -типа в материал n -типа через электрический контакт, им приходится преодолевать энергетический барьер и забирать для этого энергию у кристаллической решетки (холодный спай). Наоборот, при переходе из материала n -типа в материал p -типа электроны отдают энергию решетке (горячий спай).

Эффект Томсона состоит в том, что при протекании электрического тока через проводник или полупроводник, в котором создан градиент температуры, в дополнение к теплоте Джоуля выделяется или поглощается (в зависимости от направления тока) некоторое количество теплоты. Физическая причина данного эффекта связана с тем, что энергия свободных электронов зависит от температуры. Поэтому на горячем спае электроны приобретают более высокую энергию, чем на холодном. Плотность свободных электронов также растет при повышении температуры, вследствие чего возникает поток электронов от горячего конца к холодному. На горячем конце накапливается положительный заряд, на холодном – отрицательный. Перераспределение зарядов препятствует потоку электронов и при определенной разности потенциалов совсем его останавливает. Аналогично протекают вышеописанные явления и в веществах с дырочной проводимостью, с той лишь разницей, что на горячем конце накапливается отрицательный заряд, а на холодном – положительно заряженные дырки. Поэтому в веществах со смешанной проводимостью эффект Томсона оказывается несущественным.

Взаимосвязь между рассматриваемыми термоэлектрическими эффектами представлена на рис.1.

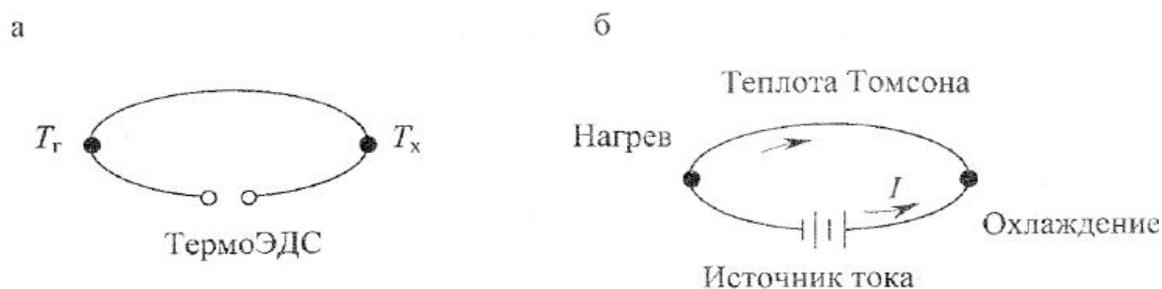


Рис. 1. Термоэлектрические эффекты:
а – эффект Зеебека; б – эффект Пельтье

Является общепризнанным, что термоэлектричество берет свое начало от открытия Зеебеком в 1821 году термоэлектродвижущих сил. Поэтому в настоящее время этот эффект носит его имя. Также стало общепринятым считать, что с именем Зеебека связано и начало открытия термоэлектричества. Однако в действительности, немецкий ученый, уроженец г. Ревеля (ныне Таллин), Томас Йохан Зеебек (1770-1831) занимался изучением магнитного поля Земли.

Опыты Зеебека сводились к следующему. Он использовал два разнородных материала в виде проволоки или стержней из висмута или сурьмы и меди, приводил их в контакт и нагревал с одной стороны (рис.1). В результате Зеебек наблюдал возникновение магнитного поля, которое фиксировалось по отклонению магнитной стрелки. Из этого опыта Зеебек сделал вывод, что «разность температур в местах соприкосновения металлической цепи является источником освобождающегося магнетизма, причиной магнитных действий». При этом величина и направление действия на магнитную стрелку зависели от степени нагрева и от материала, из которого был сделан исследуемый контур. Зеебек назвал этот эффект термомагнетизмом. Он считал, что это явление как нельзя лучше объясняет возникновение земного магнетизма: магнитные свойства земли он связывал с действием разницы температур между полюсами и экватором в разных участках земной коры. Отчет об исследованиях Зеебека был опубликован только в 1825 году в трудах Берлинской академии наук.

Но первый термоэлемент создал Эрстед и именно он изобрел первую термоэлектрическую батарею (генератор) на основе пары металлов свинец-висмут. У него в эксперименте были изготовлены 6 термопар, «горячие» спаи которых нагревались, а «холодные» спаи охлаждались в стаканах водой (рис.2).

Таким образом, открытие термоэлектричества на основе опытов Зеебека скорее можно присвоить не самому Зеебеку, а Эрстеду, давшему правильное объяснение этим

опытам. Интересно отметить, что Зеебек до конца жизни был противником термоэлектричества, хотя Эрстед никогда не оспаривал первенство открытия эффекта, признавая его за Зеебеком.

В основу метода исследования физических свойств проводящих материалов, метода термо-ЭДС, положен эффект, названный именем Зеебека и связанный с возникновением термоэлектродвижущей силы (термо-э.д.с. Зеебека) в цепи, имеющем два различных проводника А и В, спаи которых имеют разные температуры T_1 и T_2 . Эффект Зеебека применяется для прямого преобразования тепловой энергии в электрическую и в термометрии. На рис. 2 изображена схема, имеющая два проводника А и В (материалы этих проводников различны), где температуры спаев равны T_0 и T .

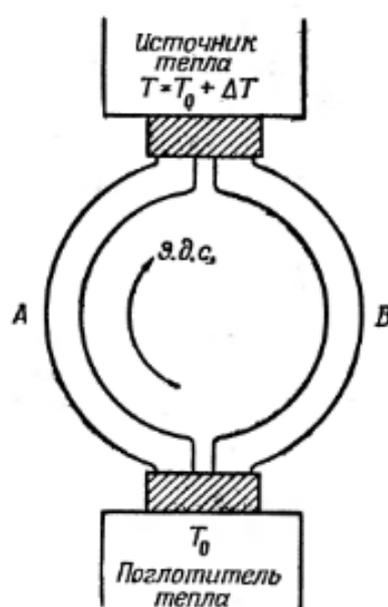


Рис. 2. Схема, демонстрирующая эффект Зеебека

При наличие между спаями перепада температур, термо-э.д.с. пропорциональна разности температур. Она зависит от природы материалов, из которых составлена цепь. Уравнение для термо-э.д.с имеет вид:

$$E_{AB} = S_{AB} \Delta T, \quad (1)$$

где S_{AB} — относительный коэффициент Зеебека для материалов А и В. Величины E_{AB} и S_{AB} являются положительными, тогда ток у горячего спая протекает от проводника А к проводнику В. Для любой пары однородных проводников, подчиняющихся закону Ома, величина термо-э.д.с. зависит от природы проводников и от температуры спаев и не зависит от распределения температуры между этими спаями. Имеем,

$$E_{T_0 T} = E_{T_0 T_1} + E_{T_1 T}, \quad (2)$$

где T_1 — температура произвольной точки на проводниках А и В, T_0 и T — температуры спаев. Данное соотношение называется «законом промежуточных температур».

Если для одной пары проводников А и В справедливо выражение $E_{AB} = S_{AB} + \Delta T$, а для другой пары В и С справедливо соотношение $E_{BC} = S_{BC} + \Delta T$, то э.д.с. в цепи, составленной из проводников А и С, будет иметь вид

$$E_{AC} = E_{AB} + E_{BC} = (S_{AB} + S_{BC})\Delta T. \quad (3)$$

Происходит алгебраическое суммирование относительных коэффициентов Зеебека. Это свойство называется «законом промежуточных проводников» и приводит к выводу, что относительный коэффициент Зеебека для любой пары проводников равен разности двух коэффициентов. Они называются абсолютные коэффициенты Зеебека, т. е.

$$S_{AB} = S_{A_1} - S_{B_1}. \quad (4)$$

Когда цепь замкнута, в ней течет электрический ток (который называется термоток I_T). Изменение знака у разности температур спаев сопровождается изменением направления термотока (рис. 3).

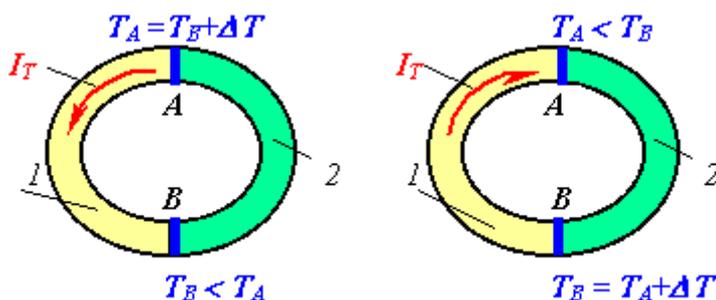


Рис. 3. Возникновение термоиндуцированного тока в двух проводниках, спаянных на концах при различных температурах контактов.

Направление термотока определяется следующим действием: в нагретом спае ток протекает от материала с меньшим значением коэффициент термо-ЭДС к материалу с большим коэффициентом термо-ЭДС. В качестве примера рассмотрим термопары железо (1) - константан (2), абсолютные коэффициенты термо-ЭДС которых равны: $E_1 = +15$ мкВ/К (для железа) и $E_2 = -38$ мкВ/К (для константана). Ток в горячем спае течёт от константана к железу (от материала 2 к материалу 1). Этот случай изображён на рис. 2 для электрической цепи (когда $E_2 < E_1$).

В проводнике градиент температуры создаст градиент концентраций «холодных» и «горячих» носителей заряда. В результате чего возникают два диффузионных потока носителей, направленные вдоль и против градиента температуры. Скорости диффузии и

концентрации «горячих» и «холодных» носителей заряда различны. На одном конце проводника создаётся избыточный положительный заряд, а на другом — отрицательный. Поле этих зарядов устанавливает стационарное состояние. Число носителей, проходящих через поперечное сечение образца в обоих направлениях, одинаково. Возникает диффузионная термо-э.д.с., которая определяется температурной зависимостью концентрации носителей заряда и их подвижностью μ , обусловленной характером их взаимодействия с фононами и примесями. В полупроводниках термо-э.д.с. обусловлена зависимостью подвижности от T и концентрации электронов и дырок от T . Вклад в термо-э.д.с., связанный с температурной зависимостью концентрации носителей, превышает вклад, обусловленный различием в $\mu(T)$. Поэтому термо-э.д.с. в полупроводниках значительно выше, чем в металлах.

Термо-э.д.с. обусловлена тремя причинами:

- 1) температурной зависимостью уровня Ферми. Тем самым появляется контактная составляющая термо-э.д.с.;
- 2) диффузией носителей заряда от горячего конца к холодному, которая определяет объёмную часть термо-э.д.с.;
- 3) процессом увлечения электронов фононами, благодаря которому появляется ещё одна составляющая - фононная.

Контактная часть термо-э.д.с.

Когда оба спая термоэлемента находятся при одной и той же температуре, контактные разности потенциалов равны и направлены в противоположные стороны. Они компенсируют друг друга. Если же температура спаев различна, то будут неодинаковы и внутренние контактные разности потенциалов. Это приведёт к нарушению электрического равновесия и появлению контактной термо-э.д.с. (E_{12}^k):

$$E_{12}^k = \int_{T_B}^{T_A} S_{12}^k dT = \int_{T_B}^{T_A} (S_1^k - S_2^k) dT \quad (5)$$

$$S^k = -\frac{1}{e} \frac{\partial E_F}{\partial T} \quad (6)$$

где E_F - энергия Ферми; k - постоянная Больцмана; e - заряд электрона.

Объёмная часть термо-э.д.с.:

Объёмная часть термо-э.д.с. связана с неоднородным распределением температуры в проводнике. Если градиент температуры поддерживается постоянным, то через проводник будет идти постоянный поток тепла. Возникнет диффузионный поток электронов,

направленный против градиента температуры. В результате, концентрация электронов на горячем конце будет уменьшаться, а на холодном увеличиваться. Внутри проводника возникает электрическое поле E_T , направленное против градиента температуры, препятствующее дальнейшему разделению зарядов (рис. 4).

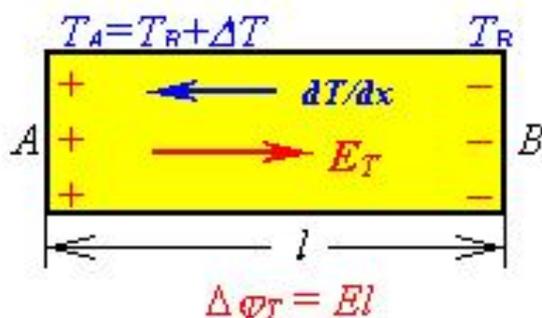


Рис. 4. Возникновение термо-э.д.с. в однородном материале вследствие пространственной неоднородности температуры

Фононная часть термо-э.д.с.

При наличии градиента температуры вдоль проводника возникает дрейф фононов, направленный от горячего конца к холодному. Сталкиваясь с электронами, фононы передают им направленное движение, увлекая их за собой. В результате, вблизи холодного конца образца будет накапливаться отрицательный заряд, а на горячем – положительный, до тех пор, пока возникшая разность потенциалов не уравновесит эффект увлечения. Эта разность потенциалов и представляет собой дополнительную составляющую термо-э.д.с., вклад которой при низких температурах становится определяющим. Фононные и диффузные слагаемые термо-э.д.с. имеют один и тот же знак, а контактная термо-э.д.с. противоположна им по знаку.

Коэффициент Зеебека.

Коэффициент Зеебека выражает термо-э.д.с, возникающую при разности температур в один градус. Оптимальное значение коэффициента Зеебека, соответствующее максимальной термоэлектрической добротности, равно приблизительно 200 мкВ/°К. У полупроводников значения коэффициента Зеебека находятся в пределах 50 — 1000 мкВ/°К, а у материалов, предназначенных специально для термоэлектрических преобразователей, равны 100 — 400 мкВ/°К.

Обратить внимание:

- 1) в полупроводнике **p-типа** $S > 0$ (положительный)
- 2) в полупроводнике **n-типа** $S < 0$ (отрицательный)

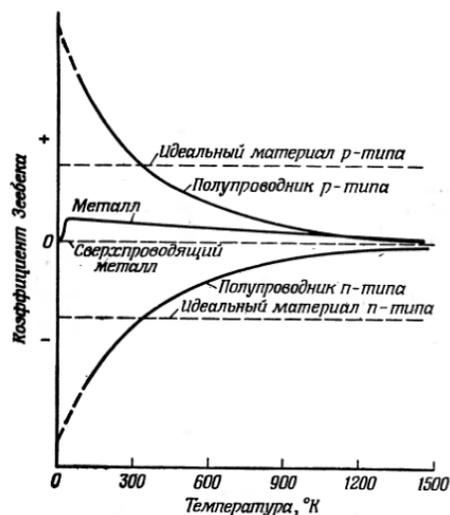


Рис.5. Зависимость коэффициента Зеебека от температуры.

Кривые, изображённые на рис. 5, дают качественную картину для материалов различных типов. При высоких температурах значения коэффициентов Зеебека для всех материалов по абсолютной величине уменьшаются, приближаясь к нулю. При очень низких температурах некоторые металлы становятся сверхпроводниками.

На рис. 6 изображено устройство для измерения коэффициента Зеебека. В пластине из исследуемого материала создается градиент температуры. На пластине укрепляются две одинаковые термопары, находящиеся с ней в электрическом и тепловом контакте.

Температурный градиент устанавливается таким, чтобы разность температур $T_2 - T_x$ составила несколько градусов. После определения температур T_2 и T_x две ветви из одинакового материала присоединяются к потенциометру. Частное от деления полученной э. д. с. на разность температур представляет собой коэффициент Зеебека материала пластины относительно материала ветвей термопары. В этих измерениях обычно используются термопары, одна из ветвей которых изготовлена из меди или платины.

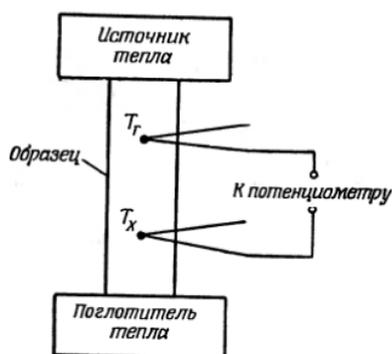


Рис. 6. Схема прибора для определения коэффициента Зеебека.

Для сравнения термоэлектрических материалов и определения характеристик термоэлектрического устройства пользуются понятием термоэлектрической добротности Z отдельно взятого материала

$$Z = \frac{S^2}{\rho k_T}, \quad (7)$$

где S — коэффициент Зеебека, ρ — удельное электрическое сопротивление и k_T — коэффициент теплопроводности.

В выражении для термоэлектрической добротности имеется квадрат коэффициента Зеебека, поэтому особое внимание уделяют на уменьшение погрешностей при его измерении. Наибольшие погрешности возникают при измерении разности температур. Неоднородность проводников термопары также может привести к неправильным результатам. Также источником ошибок является окисление в области контакта между термопарой и исследуемым образцом.

Нередко в технической литературе и в практической термометрии бытует **ошибочное мнение** о том, что термоЭДС возникает **в области спая двух проводников**. Однако появление термоЭДС обусловлено физическими процессами в однородном проводнике при наличии градиента температуры. Упрощенно, возникновение термоЭДС в металлах и сплавах можно объяснить на основе теории электронного газа П.Друде (1900 г.), где распределение электронов по скоростям внутри проводника описывается статистикой Максвелла-Больцмана. Металл рассматривается [1, с.317-328] как решетка атомов, на внешних орбиталях которых находятся валентные электроны, слабо связанные электромагнитными силами с ядром атома. Валентные электроны можно считать свободными частицами, т.к. они легко могут переходить от одного атома к другому. В металлах эти электроны называют еще электронами проводимости. Система свободных электронов подчиняется определенным статистическим закономерностям и находится в равновесии при отсутствии внешних воздействий. Средняя скорость электронов равновесной системы равна нулю, хотя каждый из электронов обладает конечной энергией и скоростью, пропорциональной локальной температуре металла.

При изменении температуры вдоль проводника система электронов отклоняется от равновесного состояния. Электроны в области горячего конца проводника приобретают более высокие кинетические энергии и скорости, а их средняя скорость становится отличной от нуля и направлена в сторону области с более низкой температурой. Поскольку электроны являются носителями заряда, наличие такой скорости приведет к возникновению электрического тока. Но электрическая цепь разомкнута, и поэтому электрический ток существует лишь до тех пор, пока в более холодной области не накопится заряд (т.е. на

концах проводника будут разные количества электронов), достаточный для создания замедляющего электрического поля. Это поле противодействует дальнейшему накоплению заряда и в точности компенсирует влияние градиента температуры на среднюю скорость электронов. Когда достигается новое равновесное состояние, электрический ток прекращается, средняя скорость электронов снова равна нулю.

Таким образом, при наличии градиента температуры в длинном и тонком проводнике возникает электрическое поле, направленное навстречу градиенту температуры. А в замкнутой цепи, составленной из двух разнородных проводников, будет течь электрический ток. Различие в числе электронов на концах проводника и ток замкнутой цепи существуют до тех пор, пока есть градиент температуры. Соответственно, градиент потенциала электрического поля, собственно и являющийся термоэлектродвижущей силой, не может возникнуть без температурного градиента. Обычно называют этот механизм основной причиной возникновения термо-ЭДС и называют его **диффузионной составляющей термо-ЭДС**. Она **доминирует в чистых металлах** при температурах выше температуры Дебая (характеристическая температура, при которой для данного вещества становятся существенными квантовые эффекты) и является основным механизмом возникновения термо-ЭДС во всем температурном диапазоне применения для сплавов с содержанием легирующих элементов в несколько процентов.

Такой подход согласуется с современными представлениями физики твердого тела, термо- и электродинамики сплошных сред, и лежит в основе многих работ по термоэлектрической термометрии.

Для чистых металлов при низких (<273К) температурах существенна и **вторая составляющая возникновения термоЭДС - фононная**. Вблизи конца проводника с большей температурой увеличивается энергия тепловых колебаний атомов, что приводит к распространению колебаний в сторону холодного конца. Фононы - кванты энергии колебаний кристаллической решетки, сталкиваясь с электронами, передают им часть своей энергии и как бы увлекают их за собой, создавая направленное движение носителей тока от более нагретого конца проводника к более холодному. Величина фононной составляющей термо-ЭДС сначала растет при увеличении температуры вследствие роста равновесной концентрации фононов, а затем уменьшается из-за увеличения числа фонон-фононовых столкновений. Максимальный вклад фононной составляющей в величину результирующей термо-ЭДС приходится у металлов на область температур от $0,1T_d$ до $0,2 T_d$, где T_d – температура Дебая. Необходимо отметить, что уже при температурах близких к 0°C фононная составляющая термо-ЭДС в металлах близка к нулю.

Некоторые авторы выделяют еще одну составляющую термо-ЭДС термопары, обусловленную **контактной разностью потенциалов**. Контактная разность потенциалов возникает при соприкосновении двух разнородных металлов вследствие перехода электронов из одного металла в другой. Условием равновесия между соприкасающимися металлами является равенство полных энергий, соответствующих уровням энергии Ферми в металлах. Величина энергии Ферми – это минимальная энергия электронов, при которой все разрешенные энергетические уровни в атоме конкретного металла заполнены электронами. Энергия Ферми, теоретически, зависит от температуры. Поэтому скачок потенциала при переходе из одного металла в другой для спаев, находящихся при разных температурах, неодинаков, и сумма скачков потенциала отлична от нуля, чем и объясняют возникновение термо-ЭДС. Одни авторы считают эту составляющую значимой, однако не приводят количественных значений, т.к. расчет уровней Ферми очень сложен. Другие авторы считают ее основной причиной возникновения термо-ЭДС. При этом сторонники такого подхода сами отмечают слабую зависимость энергии Ферми от температуры, которая заметно проявляется на уровне 10^4 К. Следует также отметить, что в физике твердого тела контактную разность потенциалов относят к поверхностным эффектам в металлах и не связывают ее с механизмом возникновения термо-ЭДС.

Рассмотрим пример исследования температурной зависимости энергии Ферми методом термо-ЭДС.

Целью рассматриваемого исследования является изучение физической сущности явления термо-ЭДС. Покажем, что по данным измерений можно определить температурную зависимость коэффициента термо-ЭДС, рассчитать энергию Ферми и эффективную массу плотности состояний, исследовать температурную зависимость энергии Ферми. В простейшем виде схема опыта выглядит так, как это изображено на рис. 7.

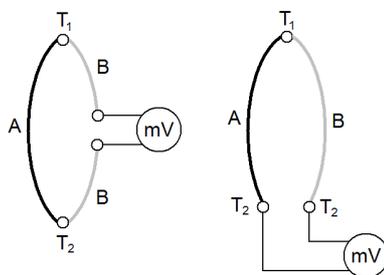


Рис. 7. Появление тока в цепи разнородных проводников

В замкнутой цепи при этом возникает ток, называемый термоэлектрическим, а концах разомкнутой цепи появляется разность потенциалов, называемая термоэлектродвижущей

силой. По имени первооткрывателя это явление получило название эффекта Зеебека. В относительно небольшом температурном интервале термоЭДС пропорциональна разности температур контактов (спаев).

$$U_{AB} = \alpha_{AB}(T_2 - T_1),$$

где α_{AB} – коэффициент пропорциональности дифференциальной или удельной термоЭДС. Значение α_{AB} зависит от природы соприкасающихся проводников и температуры.

Используя эффект Зеебека для измерения температур или проводя исследования методом термо-ЭДС, следует учитывать и эффект Пельтье, открытый в 1834 году. Он представляет собой эффект, обратный эффекту Зеебека. Сущность его состоит в том, что при прохождении тока в цепи, состоящей из различных проводников, в местах контакта в дополнение к теплоте Джоуля выделяется или поглощается (в зависимости от направления тока) некоторое количество тепла, пропорциональное прошедшему через контакт количеству электричества и некоторому коэффициенту Π . Этот коэффициент зависит от природы находящихся в контакте материалов. Он назван коэффициентом Пельтье.

Причина возникновения явления Пельтье заключается в том, что средняя энергия электронов, участвующих в переносе тока в двух материалах, находящихся в контакте, различна, несмотря на то, что уровни их электрохимического потенциала совпадают. Рассмотрим контакт электронного полупроводника и металла. Предположим, что направление поля соответствует направлению перехода электронов из полупроводника в металл (рис.8).

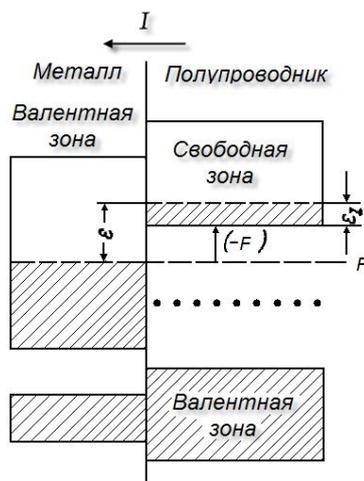


Рис.8. Схема расположения энергетических уровней на контакте полупроводника с металлом(ϵ - энергия , выделяющаяся в виде тепла при переходе одного электрона из полупроводника в металл.

Если бы электроны, находящиеся на примесных уровнях, могли бы точно также перемещаться под действием электрического поля, как электроны в зоне проводимости, то в

среднем энергия электронов, участвующих в электрическом токе, в полупроводнике равнялась бы энергии электронов на уровне Ферми в металле. При таких условиях переход электронов в металл не нарушил бы теплового равновесия в последнем; но электроны на примесных уровнях неподвижны, а энергия свободных электронов в зоне проводимости значительно выше энергии на уровне химического потенциала. Перейдя в металл, эти электроны опускаются до уровня Ферми, отдавая при столкновениях свою избыточную энергию атомам металла. Выделяющаяся при этом теплота и есть теплота Пельтье. Так как электроны приходят в тепловое равновесие в результате нескольких десятков соударений в непосредственной близости от контакта, то и вся теплота Пельтье выделяется почти на самом контакте. При противоположном направлении тока весь процесс идет в обратном направлении.

Если вдоль проводника, по которому протекает электрический ток, существует перепад температур, причем направление тока соответствует движению электронов от горячего конца к холодному, то, переходя из более горячего участка в более холодный, электроны передают избыточную энергию окружающим атомам, чем вызывают нагрев проводника- выделение тепла; при обратном направлении тока электроны, проходя из более холодного участка в более горячий, пополняют свою энергию за счет окружающих атомов. Этим объясняется явление Томсона.

Термо-ЭДС в контуре складывается из трёх составляющих. Первая из них обусловлена температурной зависимостью внутренней контактной разности потенциалов U_i , которая связана с температурным изменением положения уровня Ферми. В металлах с увеличением температуры уровень Ферми, хотя и слабо, но смещается вниз по энергетической шкале. Поэтому на холодном конце однородного проводника он должен располагаться выше, чем на более нагретом. Следствием смещения уровня Ферми и является возникновение контактной составляющей термоэдс.

Дифференциальная термо-ЭДС, соответствующая этой составляющей:

$$\alpha_k = -\frac{1}{e} \frac{\partial F}{\partial T}, \quad F - \text{уровень Ферми}$$

Вторая составляющая термо-ЭДС, называемая объёмной, обусловлена диффузией носителей заряда от горячих спаев к холодным. Средняя кинетическая энергия электронов в металле также несколько изменяется с температурой. Электроны, сосредоточенные на более горячем конце, обладают несколько большей кинетической энергией и большей скоростью теплового движения по сравнению с носителями заряда холодного конца. Поэтому они в большем количестве диффундируют в направлении температурного градиента.

Диффузионный поток электронов в этом случае, перенося заряд из горячего конца в холодный, создаёт между ними разность потенциалов.

Третья составляющая эдс возникает в контуре вследствие увлечения электронов квантами тепловой энергии (фононами). Этот эффект открыт Л.Э.Гуревичем в 1945 г. и состоит в следующем. При наличии градиента температуры в проводнике возникает дрейф фононов от горячего конца к холодному, совершающийся с некоторой средней скоростью $V_{др}$

Существование такого дрейфа приводит к тому, что электроны, рассеиваемые на фононах, сами начинают совершать направленное движение от горячего конца к холодному примерно с той же скоростью $V_{др}$. Накопление электронов на холодном конце проводника и обеднение электронами горячего конца вызывает появление термоэдс.

При низких температурах эта составляющая термоэдс может в десятки и сотни раз превосходить объёмную и контактную составляющие.

И в однородном проводнике, т.е. изготовленном из одного металла, при наличии перепада температур на концах его также возникает разность потенциалов в силу рассматриваемых причин. Её значение, отнесённое к единичной разности температур, называется абсолютной удельной термоэдс.

В термопарном контуре относительная удельная термоэдс представляет собой разность абсолютных удельных термоэдс, составляющих проводников:

$$\alpha_{AB} = \alpha_A - \alpha_B,$$

где α_A и α_B – абсолютные удельные термоЭДС контактирующих металлов А и В.

Квантовая теория твёрдого тела даёт следующее выражение удельной термоЭДС:

$$\alpha_{AB} = \frac{\pi^2}{3} \frac{\kappa}{e} \frac{\kappa T}{F}.$$

При комнатной температуре отношение $\kappa T / F$ имеет значение порядка 10^{-3} . α_A составляет несколько мкВ/К.

Для определения абсолютных термоэдс в качестве эталона используют свинец, у которого термоэлектрические свойства выражены очень слабо, или платину, которая в чистом виде отличается высокой стабильностью свойств. Для нормального проводника горячий спай заряжается положительно (это характерно для многих простых металлов). Ток в горячем спае проводника и свинца направлен от проводника к свинцу. Термоэдс проводника в этом случае считается отрицательной.

Рассмотрим причины возникновения термо-ЭДС в полупроводниках.

В полупроводнике при наличии градиента температуры возникают электронный и дырочный токи, которые обусловлены действием электрохимического потенциала.

Если полупроводник нагрет неравномерно, то средняя энергия носителей заряда и их концентрация в нём будет больше там, где выше температура. Градиент температуры в однородном полупроводнике приводит к градиенту средней энергии носителей заряда и градиенту их концентрации, вследствие чего возникает диффузионный поток носителей заряда, т.е. возникает электрический ток.

В разомкнутой цепи в стационарном состоянии плотность тока во всех точках образца равна 0. При измерении термо-ЭДС на границах полупроводника и измерительных металлических электродов существуют контактные разности потенциалов, которые не равны друг другу, вследствие существующего температурного градиента. Измерительный прибор отметит ЭДС, равную термо-ЭДС полупроводника и разности контактных потенциалов измерительных электродов.

Чтобы исключить последнее слагаемое, следует определить термоэдс, как градиент электрохимического потенциала:

$$\nabla(F - e\varphi),$$

где F – уровень Ферми, φ – электрохимический потенциал.

Величина $\nabla(F/e)$ равна разности контактных потенциалов в граничных точках, если считать, что полупроводник и металл на контакте находятся в термодинамическом равновесии, тогда дифференциальная термоэдс α определяется как

$$\alpha = \frac{|\nabla(\varphi - \frac{F}{e})|}{\nabla T}.$$

Коэффициентом термоэдс обычно называется термоэлектрическая разность потенциалов, приходящаяся на один градус:

$$\alpha = \frac{\Delta U}{\Delta T} \left[\frac{B}{K} \right].$$

Полная плотность тока в полупроводнике из решения кинетического уравнения Больцмана для полупроводника со сферической симметрией зон и предположения, что рассеяние осуществляется на акустических фононах, может быть представлена выражением:

$$j = n\mu_n \left\{ \nabla(F - e\varphi) + \left(2k_0 - \frac{F}{T} \right) \nabla T \right\} + p\mu_p \left\{ \nabla(F - e\varphi) + \left(2k_0 - \frac{F+E_g}{T} \right) \nabla T \right\},$$

где n и p – концентрации электронов и дырок, μ_n и μ_p – соответствующие подвижности.

Полагая ток равным нулю, можно получить выражение для дифференциальной термоэдс полупроводника:

$$\alpha = -\frac{k_0}{e} \frac{1}{n\mu_n + p\mu_p} \left\{ n\mu_n \left(2 - \frac{F}{k_0 T} \right) - p\mu_p \left(2 - \frac{F+E_g}{k_0 T} \right) \right\}.$$

Выражения для концентрации электронов и дырок невырожденного полупроводника:

$$n = \frac{2(2\pi m_n k_0 T)^{3/2}}{h^3} e^{\frac{F}{k_0 T}},$$

$$p = \frac{2(2\pi m_p k_0 T)^{3/2}}{h^3} e^{\frac{E_g - F}{k_0 T}},$$

При абсолютном нуле температуры в собственном полупроводнике уровень Ферми (химического потенциала) проходит строго посередине между дном зоны проводимости и потолком валентной зоны. С ростом температуры уровень Ферми изменяется по закону:

$$F = -\frac{E_g}{2} + \frac{1}{2} kT \left(\frac{N_v}{N_c} \right)$$

На рис. 9 представлена температурная зависимость уровня Ферми в электронном полупроводнике.

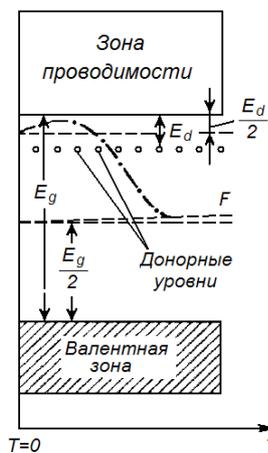


Рис. 9. Температурная зависимость уровня Ферми в полупроводнике п-типа

При температуре абсолютного нуля уровень Ферми находится строго посередине между дном зоны проводимости и донорным уровнем. С ростом температуры происходит ионизация примеси, электроны с донорного уровня переходят в зону проводимости, уровень Ферми при этом сначала поднимается, а затем, по мере роста температуры, начинает опускаться. При дальнейшем подъеме температуры основную роль начинает играть собственная проводимость, и дальнейший ход уровня Ферми совпадает с законом изменения его для случая собственной проводимости. Совершенно аналогичные выводы можно получить для примесного дырочного полупроводника (рис. 10). В этом случае при нуле температур уровень Ферми проходит посередине между потолком валентной зоны и акцепторным уровнем. При повышении температуры уровень Ферми поднимается выше акцепторных уровней, они оказываются заполненными электронами, перешедшими из валентной зоны. Уровень Ферми начинает приближаться к середине запрещенной зоны. Концентрация собственных электронов и дырок с ростом температуры увеличивается и основную роль начинает играть собственная проводимость.

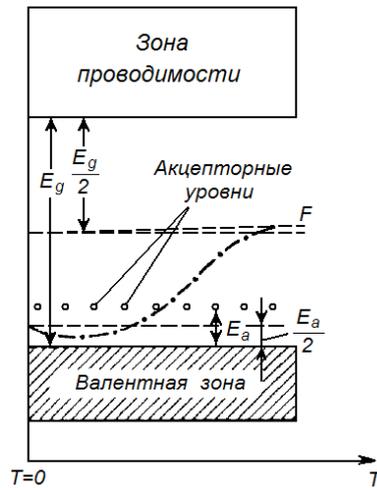


Рис. 10. Температурная зависимость уровня Ферми в полупроводнике р-типа

ТермоЭДС полупроводника определяется двумя слагаемыми, каждое из которых соответствует вкладу, вносимому электронами и дырками, причём эти слагаемые имеют противоположные знаки.

В случае электронного полупроводника на горячем конце возникает положительный объёмный заряд, поскольку электроны диффундируют от горячего конца к холодному. В случае дырочного полупроводника знак коэффициента термоЭДС обратный (рис.11).

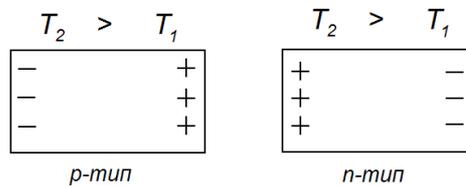


Рис. 11. Распределение зарядов в полупроводниках п- и р- типов при наличии градиента температуры.

Для собственного полупроводника $n = p = n_i$ и $F = E_g / 2$, поэтому величина термо-ЭДС полупроводника с собственной проводимостью определяется лишь шириной запрещённой зоны и соотношением подвижностей электронов и дырок:

$$\alpha = -\frac{k_0}{e} \left(\frac{b-1}{b+1} \right) \left\{ 2 + \frac{E_g}{2k_0T} \right\}, \quad b = \frac{\mu_n}{\mu_p}.$$

Формулы для определения коэффициента термоэдс для полупроводников “n”- и “p”-типа могут быть представлены в следующем виде:

$$\alpha_n = -\frac{k_0}{e} \left[\left(\frac{5}{2} - S \right) - \frac{F}{k_0T} \right],$$

$$\alpha_p = -\frac{k_0}{e} \left[\left(\frac{5}{2} - S \right) - \frac{E_g - F}{k_0T} \right],$$

При рассеянии носителей заряда на акустических фононах $S = 1/2$. При рассеянии на ионизированных примесях – $S = -3/2$.

Для невырожденного электронного полупроводника с простой параболической зоной коэффициент термоэдс может быть определен из следующего выражения:

$$\alpha_n = -\frac{k}{e} \left(r + 2 + \ln \frac{2(2\pi m_n kT)^{3/2}}{h^3 n} \right)$$

Эта формула была впервые выведена в 1940 году советским физиком Писаренко и носит его имя. Совершенно аналогичный, с точностью до знака, вид имеет выражение для термоэдс дырочного полупроводника:

$$\alpha_p = \frac{k}{e} \left(r + 2 + \ln \frac{2(2\pi m_n kT)^{3/2}}{h^3 n} \right)$$

В случае смешанной проводимости, когда электрический ток переносится электронами и дырками, термоэдс, как правило, значительно ниже. Если концентрация или подвижность носителей одного знака, например, электронов, больше, чем другого, то они диффундируют на холодный конец в большом количестве до тех пор, пока возникшее вследствие этого поле (тормозящее электроны и ускоряющее дырки) не уравнивает оба потока. Коэффициент термоэдс в этом случае определяется выражением:

$$\alpha = \frac{\alpha_p \mu_p p - \alpha_n \mu_n n}{\mu_p p + \mu_n n}$$

Приведенное значение энергии Ферми при рассеянии на акустических фононах для полупроводника “n”-типа определяется выражением

$$F^* = \frac{F}{kT} = \frac{\alpha_n e}{k_0} + 2$$

Для дырочного полупроводника используется та же формула с учетом знака коэффициента термо-ЭДС.

В случае собственной проводимости коэффициент термоэдс имеет отрицательный знак, поскольку подвижность электронов больше, чем подвижность дырок и вклад электронной составляющей соответственно больше.

С ростом температуры значение коэффициента термоэдс при высоких температурах уменьшается, а в полупроводниках р-типа при температуре, близкой к температуре собственной проводимости, меняет знак.

На рис. 12 представлены температурные зависимости коэффициентов термо-ЭДС в полупроводниках р- и n-типа.

Из эксперимента по определению коэффициента термоэдс можно рассчитать эффективную массу плотности состояний.

$$m^* = \frac{\left[\frac{nh^3}{2} \exp\left(-\frac{\alpha e}{k_0} - 2\right) \right]^{2/3}}{2\pi k_0 T}.$$

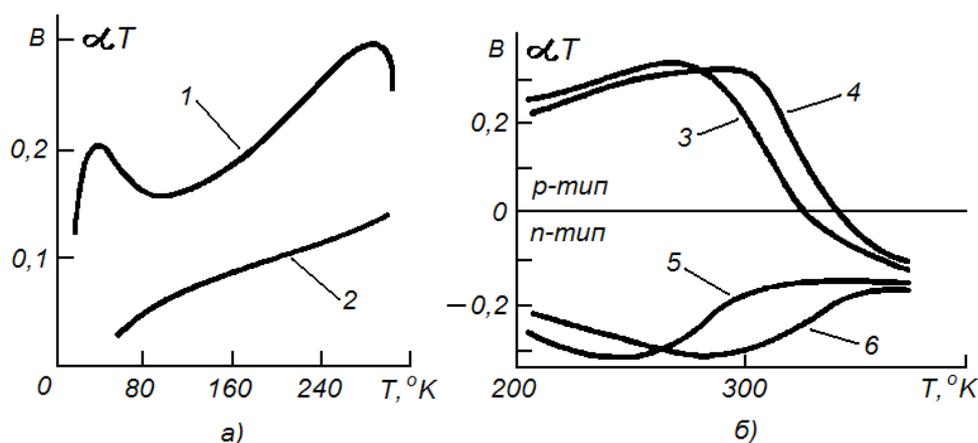


Рис. 12. Экспериментальные зависимости термоэдс от температуры образца германия при низких (а) и высоких температурах (б): кривая 1 – $N_a = 1,9 \cdot 10^{14} \text{см}^{-3}$, кривая 2 – $N_a = 6 \cdot 10^{17} \text{см}^{-3}$, кривая 3 – $N_a = 8 \cdot 10^{13} \text{см}^{-3}$, кривая 4 – $N_a = 1,6 \cdot 10^{14} \text{см}^{-3}$, кривая 5 – $N_d = 8 \cdot 10^{12} \text{см}^{-3}$, кривая 6 – $N_d = 8,9 \cdot 10^{13} \text{см}^{-3}$.

Температурная зависимость коэффициента термо-ЭДС может быть исследована на программно-аппаратном комплексе, структурная схема которого показана на рис. 13.

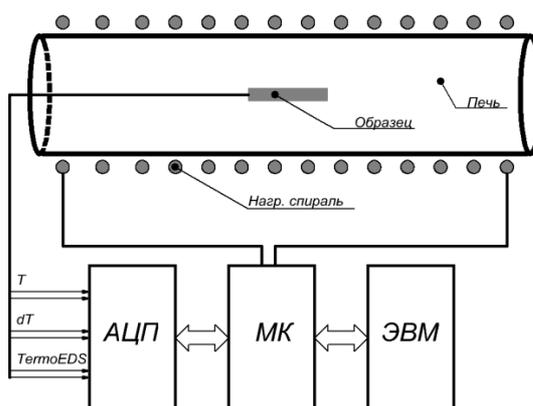


Рис. 13. Блок-схема лабораторной установки «Исследование температурной зависимости коэффициента термо-ЭДС»

Для получения зависимости термоэдс от температуры образец германия (рис. 14) с размерами $1.4 \times 4.5 \times 15 \text{мм}^3$ помещается в трубчатую печь.

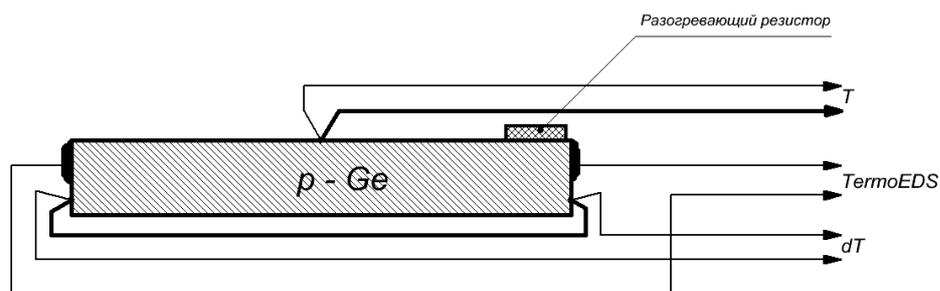


Рис. 14. Схема расположения измерительных элементов на образце.

Один из концов образца подогревается с помощью приклеенного резистора для поверхностного монтажа SMD-1210 100 Ом. Цифровой регулятор температуры, реализованный на микроконтроллере (МК), поддерживает заданный градиент температуры вдоль образца на уровне 1-3°C. На торцах образца нанесены индиевые контакты, с помощью которых производится измерение термоэдс. Для изучения температурной зависимости термоэдс, печь в течение 30-40 минут линейно нагревается от комнатной температуры до 410 К. В процессе нагрева образца через заданный температурный интервал осуществляется измерение температуры образца, разности температур и термоэдс. Температура образца T измеряется термопарой медь-константан, второй конец которой находится при комнатной температуре. Разность температур dT измеряется дифференциальной термопарой медь-константан присоединенной к торцам образца. Все сигналы, поступающие с образца (T , dT , термоэдс), с помощью аналого-цифрового преобразователя (АЦП) преобразуются в цифровую форму и поступают далее в МК и ЭВМ.

Измерения проводятся обычно в следующей последовательности:

1. Включаются компьютер и блок измерения. Для обеспечения стабильной работы электронных блоков и получения воспроизводимых результатов измерения необходимо осуществить прогрев приборов не менее 10 минут.
2. В блоке «Задание начальных условий» устанавливается разность температур холодного и горячего концов образца и заданный шаг измерения.
3. Производится калибровка измерительной системы. Калибровка необходима для исключения аддитивной и мультипликативной погрешностей измерительной системы. Необходимо нажать кнопку «Calib» блока «Управление измерением». Через несколько секунд загорится индикатор, что означает завершение процесса калибровки.
4. Включается определенная разность температур. При этом нельзя переходить к следующему пункту пока разность температур не достигнет заданного значения и колебания температуры относительного заданного значения превышают ± 0.02 К. Контроль за

температурой осуществляется по индикатору «Разность температур» блока «Текущее состояние образца».

5. Включается нагрев образца. С этого момента начинается запись температурной зависимости разности температур холодного и горячего концов образца (dT) и величины термоэдс с заданным температурным интервалом в таблицу и отображение этой зависимости на графике.

6. Если процесс измерения температурной зависимости не завершить вручную кнопкой «Нагрев образца», то он завершится автоматически при достижении температуры 410 К.

7. Сохранение результатов измерения (кнопка «Save»).

8. Питание блока измерения можно выключить только тогда, когда температура образца будет менее 310 К.

Список измерительных процедур следующий:

1. Измеряется термо-ЭДС полупроводника без включения общего подогрева.

2. Провести измерения температурной зависимости коэффициента термо-ЭДС,

$$\alpha = \frac{U_{T \text{ ЭДС}}}{dT}$$

и строится график.

3. Рассчитывается значение приведенной энергии Ферми (F^*) и по результатам строится график температурной зависимости приведенного уровня Ферми.

4. Определяется ход температурной зависимости уровня Ферми в рамках зонной диаграммы полупроводника,

$$F = F^* \cdot kT$$

5. Рассчитывается эффективная масса плотности состояний при комнатной температуре.