§8. Контакт двух полупроводников различного типа проводимости. Кристаллический диод

Широкое использование полупроводников обусловлено, в основном, теми физическими процессами, которые происходят на контакте примесных полупроводников с различным типом проводимости. Такой контакт называют электронно-дырочным переходом, или *p-n* переходом.

Под контактом двух полупроводников понимают не механическое соприкосновение полупроводников *n*-, и *p*- типов, а создание внутри кристалла резкой границы этих областей. Существует несколько методов формирования *p-n* – переходов. Рассмотрим их суть с физической точки зрения.

Метод сплавления. На примесный кристалл германия n- типа накладывается кусочек кристалла индия, затем производится прогрев до температуры 500-600^0 С. Расплавившийся индий растворяет в себе германий, в результате чего образуется полупроводник, который после отвердевания будет обладать проводимостью *p*- типа. На границе двух кристаллов с различной проводимостью образуется *p-n* – переход (конечно, весь технологический процесс происходит более сложно, здесь намечена лишь его основная схема).

Метод диффузии. В этом методе используется процесс диффузии одного вещества в другое через границу их соприкосновения. При нагреве полупроводника *p*- типа через одну из его поверхностей происходит диффузия примеси (можно взять кристалл с *n*-типом проводимости и диффундировать в него акцепторную примесь). В зависимости от температуры процесса и его длительности, на определенной глубине кристалла возникает *p-n*-переход.

Используется также **эпитаксиальный метод**, суть которого заключается в осаждении на кристалл с помощью химических реакций из окружающей газообразной или жидкой среды тонкого слоя пленки того же полупроводника, но с другой, примесной проводимостью.

Наконец, широкое применение нашел **метод ионного легирования**, когда необходимая примесь внедряется с помощью ионного пучка. Практически все технологические процессы весьма сложны и доступны не всем промышленным странам. От качества изготовления *p-n*-переходов во многом зависит и качество радиоэлектронной полупроводниковой аппаратуры. В нашей стране все эти методы разработаны в промышленных масштабах. Получив одним из указанных способов *p-n*-переход,

рассмотрим его равновесное (не рабочее) состояние (рис. 50).

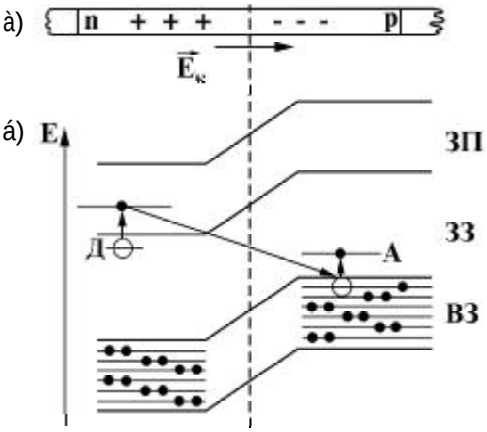

Рис. 50.

Так как в полупроводнике n-типа имеется избыток электронов (по отношению к полупроводнику *p*-типа), а в полупроводнике *p*-типа – «дырок», поэтому электроны из полупроводника n-типа сразу же начинают диффундировать в полупроводник *p*-типа. А из этого полупроводника возникает диффузия «дырок» в полупроводник *n*-типа. В приграничной области со стороны полупроводника *p*-типа появляются электроны, а по другую сторону контакта, в полупроводнике *n*-типа – «дырки». В результате граничная область со стороны полупроводника p-типа получает избыточный отрицательный заряд, а со стороны полупроводника n-типа - положительный (рис. 50а). На рисунке (50б) изображено для упрощения картины перемещение лишь одного электрона, который, попав в «дырку», аннигилирует (исчезает как свободный заряд). Но так как обе части кристалла в исходном состоянии были нейтральны, то уход электрона из полупроводника *n*-типа нарушает эту нейтральность, и полупроводник *n*-типа вблизи контакта приобретает локализованный положительный объемный заряд. Аналогичные рассуждения можно провести относительно появления объемного отрицательного заряда в приконтактной области полупроводника *p*-типа. Локализация объемных зарядов обусловлена тем, что эти заряды связаны с локализованными в кристаллической решётке донорами и акцепторами.

В контактном поле возникает двойной электрический слой, возникает контактное электрическое поле $\vec{E}_k$, на рис. 50а оно направлено слева направо, из полупроводника *n*–типа в полупроводник *p*-типа. Но это контактное поле, само порожденное диффузионным потоком основных носителей заряда, будет тормозить их движение, и даже может вызвать появление дрейфового тока неосновных носителей заряда из полупроводника *p*-типа – электронов- в полупроводник n-типа

(соответственно из полупроводника *n*-типа – «дырок» – в полупроводник *p*-типа). Когда эти два встречных потока уравновесят друг друга, говорят о наступлении динамического равновесия на *p-n*-переходе. При этом в приконтактной области по обе стороны контакта образуются слои, обедненные основными носителями заряда (ведь часть их продиффундировала в соседнюю область). Эта переходная область получила название запорного слоя, её ширина Δx порядка 10^{-5} см. А если учесть, что контактная разность потенциалов Δj достигает величины 0,1-0,5 В, то напряженность контактного поля оказывается очень большой величиной

$$E_k = \frac{\Delta j}{\Delta x} \approx 10^5 B / см.$$

Как видно из рис.50 возникновение p-n-перехода сопровождается поднятием уровней (зон) в полупроводнике p-типа по отношению к уровням (зонам) полупроводника *n*-типа. Разберемся в этом явлении, так как такие сдвиги энергетических состояний постоянно будут встречаться нам в дальнейшем.

Дело в следующем. Диффузия каждого последующего электрона из полупроводника *n*-типа в полупроводник p-типа требует преодоления отталкивающего электрического поля ранее продиффундировавших электронов. Поэтому электроны, перешедшие в полупроводник *p*-типа по отношению к своему первоначальному состоянию будут обладать избыточной энергией, что на зонной схеме и изображается относительным поднятием всех уровней в полупроводнике *p*-типа по отношению к уровням в полупроводнике *n*-типа.

Рассмотрим теперь, что произойдет на *p-n*-переходе, если он окажется во внешнем электрическом поле. Можно выделить два крайних случая: 1) внешнее поле направлено в ту же сторону, что и контактное поле ; 2) внешнее поле направлено навстречу контактному полю.

Случай 1. Запорное состояние *p-n*-перехода.

Если внешнее поле , направлено в ту же сторону, что и контактное поле, то под его воздействием «дырки» из граничного слоя полупроводника *n*-типа будут возвращены в полупроводник *p*-типа и отодвинуты далее внутрь этого полупроводника. Аналогично и электроны, оказавшиеся в результате диффузии из полупроводника *n*-типа в полупроводнике *p*-типа, будут возвращены в полупроводник n-типа и отодвинуты от границы раздела полупроводников (напомним: элект-

роны движутся в направлении против внешнего поля!). Переходная область обедняется носителями заряда, это равноценно увеличению омического сопротивления *p-n*-перехода, электрический ток практически не будет проходить через *p-n*-переход. Такое состояние p-n-перехода называют запорным (рис. 51).

На рисунке 51 изображена вольтамперная характеристика p-n- перехода, обсуждаемый участок обозначен буквами Оа. Наличие хотя и малого, но все же отличного от нуля тока, обусловлено движением не основных носителей, т.е. диффузией через p-n-переход «горячих» электронов и «дырок», которые в результате тепловой флуктуации могут приобрести достаточную энергию, чтобы преодолеть внешнее поле. Но таких «горячих» зарядов очень мало, поэтому и говорят, что p-n-переход в рассматриваемом случае заперт. Но, как ясно из предыдущих рассуждений, нагрев p-n-перехода разрушит его запорное состояние. Так для германиевого p-n-перехода достаточно нагреть кристалл до температуры 150^0 С, чтобы разрушить запорное состояние *p-n*-перехода. Отсюда следует очевид-

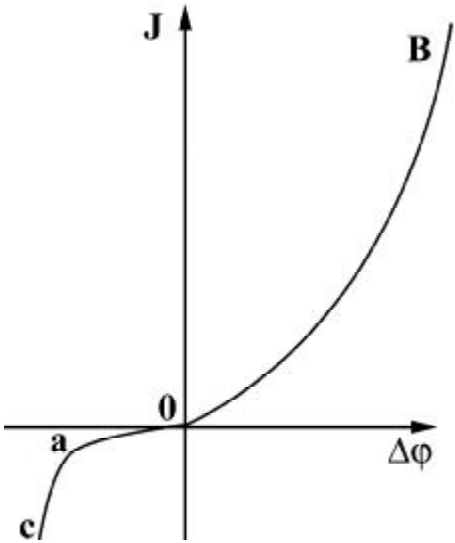

Рис. 51.

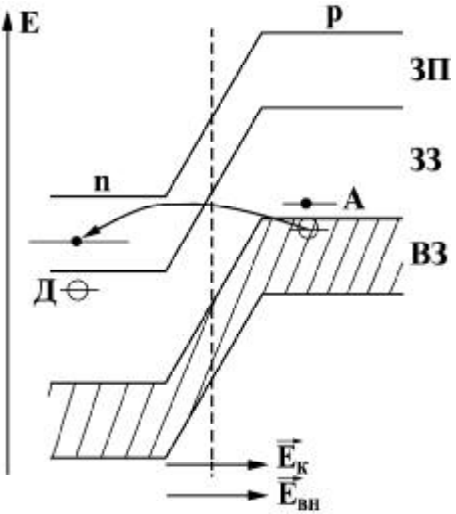

Рис. 52.

ный практический совет: полупроводниковые устройства – транзисторы, телевизоры и т.д., нельзя располагать вблизи нагревательных устройств, на прямом пути солнечных лучей.

На рис. 52 представлена зонная схема *p-n*-перехода, находящегося в запорном состоянии. При сравнении с равновесным состоянием p-n-перехода (рис.50) обнаруживается изменение положения энергетических зон в полупроводниках разной проводимости. Это есть результат воздействия внешнего электрического поля на энергетические состояния: электронам энергетически «выгодно» перейти из полупроводника *p*-типа в полупроводник *n*-типа, такой переход приводит к уменьшению их энергии, так как переход совершается в более низкие энергетические состояния.

Случай 2. Пропускное состояние *p-n*-перехода.

Если внешнее электрическое поле приложено к *p-n*-переходу в направлении, противоположном контактному полю, то состояние *p-n*-перехода существенно изменяется по сравнению с его запорным состоянием (рис. 53).

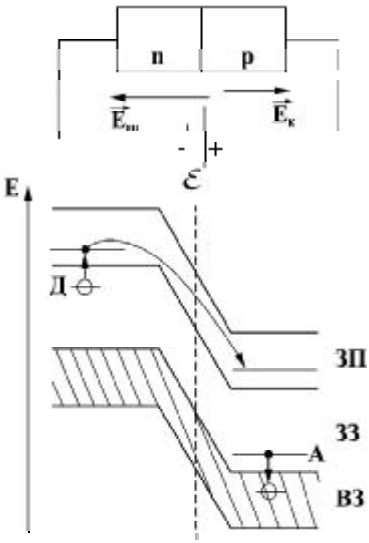

Рис. 53.

Уровни и зоны в полупроводнике *p*-типа оказываются существенно ниже уровней и зон в полупроводнике *n*-типа (ср. с рис. 50), что символизирует энергетическую выгодность движения электронов слева направо. Под действием внешнего электрического поля такого направления основные носители заряда в каждом полупроводнике приобретают направленное движение к границе соприкосновения полупроводников. Область перехода обогащается носителями заряда, что эквивалентно уменьшению омического сопротивления. В результате этого ток через *p-n*-переход, - пропускной ток – будет возрастать с увеличением внешнего электрического поля по сложному закону, отличному от закона Ома (на рис. 51 вольт-амперная характеристика пропускного тока изображена участком ОВ). Объективной характеристикой *p-n*-перехода является коэффициент усиления, который определяется как отношение про-

пускного тока к запорному. Обычно он имеет величину порядка 10^9, т.е. *p-n* - переход практически обладает односторонней проводимостью. И в этом его основное применение как выпрямителя переменного тока (кристаллический диод). Очевидны и физические, и экономические преимущества кристаллического диода перед вакуумным: значительные меньшие габариты, относительная дешевизна, долговечность, механическая прочность (но – боязнь перегрева!!!).

О влиянии повышения температуры на *p*-n-переход говорилось выше. Возможен и пробой *p-n*-перехода при увеличении обратного поля. Хотя это и сужает область использования *p-n*-перехода в качестве выпрямителя, вместе с тем, его же можно использовать в качестве физической основы работы полупроводникового прибора – стабилитрона, служащего для стабилизации (ограничения верхнего значения) напряжения. На рис. 51 участок ас соответствует пробою *p-n*-перехода.

§ 9. Туннельный диод

В туннельных диодах (ТД) концентрация как акцепторных, так и донорных примесей в полупроводниках около *p-n*- перехода в тысячи раз больше, чем в обычных диодах $\sim 10^{19}\,см^{-3}$. В результате этого энергетические уровни в таких участках оказываются в одной части близкими к энергетическому уровню акцепторной примеси, а в другой – к донорной. Если сильно легированный участок имеет ширину, близкую к ширине *p-n*- перехода в обычных диодах (порядка 0.02 мм), то никаких особых эффектов не возникает. Однако при ширине этого участка порядка 150 $\overset{0}{A}$ и менее возникает так называемый туннельный переход, когда электрону не требуется дополнительной энергии. Поэтому (ТД) могут работать при более низких температурах, чем обычные диоды, в которых возбуждение электронов с донорных уровней в ЗП и с уровней ВЗ на акцепторные происходит за счёт энергии теплового движения ($\sim kT$).

Туннельные диоды работают лучше обычных в области высоких и сверхвысоких частот.

На рис. 54 изображена вольтамперная характеристика туннельного диода. На участке АВ при увеличении U ток уменьшается, т.е. на этом участке диод имеет отрицательное сопротивление. Благодаря это-

му ТД используется в качестве переключателей, усилителей или генераторов колебаний.

Последнее объясняется тем, что у обычных диодов быстродействие ограничивается временем дрейфа носителей. В ТД переход электронов происходит почти мгновенно и быстродействие их ограничивается паразитными ёмкостью и индуктивностью вводов. У ТД обратный ток достигает большой величины уже при малом обратном напряжении из-за увеличения числа электронов, способных совершать туннельный переход. Туннельный эффект в полупроводниках был открыт в 1958 г. японским физиком Л.Эсаки.

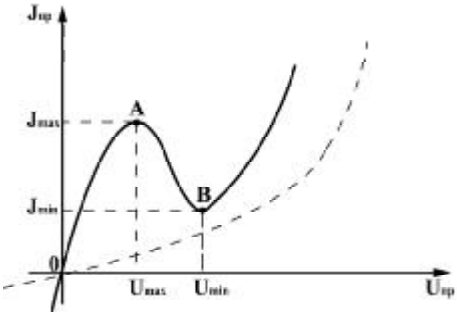

Рис. 54.

§ 10. Диоды Ганна

В 1963 г. американский физик Дж. Ганн обнаружил явление генерации электромагнитных сверхвысокочастотных колебаний в кристалле арсенида галлия *GaAs* под действием сильного постоянного электрического поля.

В кристалле *GaAs* имеются две зоны проводимости *n*- типа, сдвинутые относительно друг друга на величину энергии 0.36 эВ.

Рассмотрим физику работы диода Ганна, анализируя вольтамперную характеристику, изображённую на рис. 55.

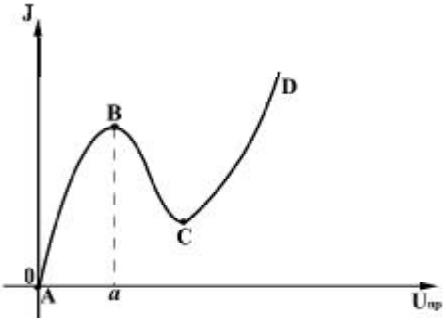

Рис. 55.

При отсутствии внешнего напряжения электроны находятся в нижней ЗП, где они обладают повышенной подвижностью. При увеличении напряжения на участке Оа подвижность электронов возрастает, увеличивается и ток – это соответствует участку ОВ вольтамперной характеристики.

При дальнейшем увеличении внешнего постоянного напряжения происходит возбуждение электронов и переход их в верхнюю ЗП. В этой зоне подвижность электронов уменьшается, уменьшается и ток, протекающий через кристалл (участок ВС).

Участок с отрицательным динамическим сопротивлением (участок ВС) появляется лишь при динамических режимах в СВЧ диапазоне. Период колебаний тока *T* зависит от дрейфовой скорости *V* электрона и длины образца *L*:

$$T = L/V = \frac{1}{n}.$$

Так в *GaAs* длиной 50-30 мкм частота колебаний оказывается в интервале 0.3-2 ГГц.

Диоды Ганна используются для генерации сверхвысокочастотных колебаний при прохождении постоянного тока.

§ 11. Кристаллический триод - транзистор

Интенсивное развитие полупроводниковой электроники началось с 1948 г., когда был создан полупроводниковый усилительный прибор – транзистор (его, создателями были ученые Дж. Бардин, В.Браттейн и В.Шокли, удостоенные за это изобретение Нобелевской премией), аналог вакуумного триода, но обладающего всеми преимуществами полупроводниковых приборов: миниатюрность, экономичность, долговечность.

Транзистор состоит из кристалла, в котором имеются три области: по краям кристалла находятся премесные полупроводники одного типа проводимости, между ними – чрезвычайно тонкий слой полупроводника другого типа проводимости. Существуют транзисторы как (*n-p-n*)-, так и (*p-n-p*)- типов. Формально кристаллический триод можно рассматривать как парный диод.

На рис. 56 представлена зонная схема транзистора, не включенного в электрическую цепь. Зонное объяснение работы транзистора основано на объяснении работы двух диодов, находящихся в разных состояниях: один диод – левый на схеме – находится в пропускном состоянии, другой – правый на схеме в запорном (рис. 57).

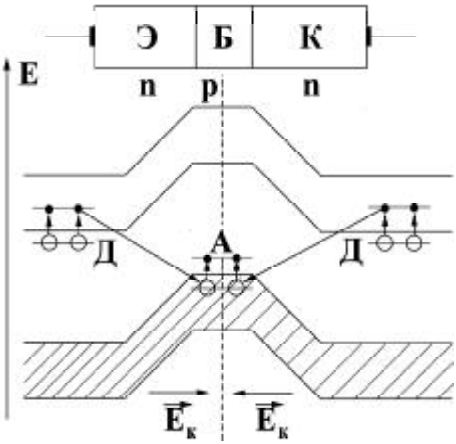

Рис. 56.

Малая толщина полупроводника *p*-типа (рассматривается работа транзистора *n-p-n* – типа) позволяет носителям заряда, прошедших переход ЭБ

(левый *n-p*-переход), по инерции пройти полупроводник Б и попасть через переход БК (правый *p-n*-переход) в правый полупроводник К и резко изменить там концентрацию носителей заряда синхронно с изменением тока в переходе ЭБ, который, в свою очередь, синхронно изменяется с изменением напряжения U_1 на входе схемы. Принято называть левый полупроводник эмиттером, правый – коллектором, средний – базой, что отчасти указывает на функциональную роль этих частей. На сопротивлении R_2 будет происходить падение напряжения U_2. Подбором батарей E_1 и E_2 и сопротивлений R_1 и R_2 можно добиться того, что напряжение U_2 будет больше U_1, (это следует из соотношения $U_2=I_2 \cdot R_2$). Данный транзистор работает как усилитель напряжения (и мощности), коэффициент усилителя может достигать нескольких сотен. Аналогично можно рассматривать работу транзистора *p-n-p*-типа.

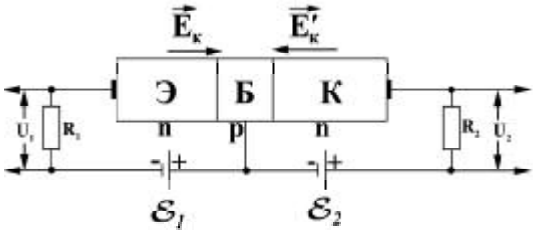

Рис. 57.

§ 12. Классификация транзисторов

Полупроводниковая электроника является основной частью современных устройств от карманного приёмника до современных компьютеров. Поэтому целесообразно рассмотреть различные варианты транзисторов.

Слово транзистор составлено из двух английских слов: *transfer* – переносить, преобразовать и *resistor* – сопротивление.

В качестве исходного кристалла используется германий или кремний. Диапазон рабочих частот – от низкой до высокой, по мощности – от малой до большой.

Рассмотренный выше (§11) транзистор *n-p-n*- типа , как и транзистор *p-n-p*- типа относятся к биполярным транзисторам. В них, как правило, имеется два взаимодействующих *n-p* перехода, усилительные свойства этих транзисторов обусловлены явлением инжекции неосновных носителей заряда. Постоянные напряжения к электродам транзистора подводятся от внешних источников. Возможны три схемы подключения транзистора к источникам питания (рис.58) в зависимости от того, какой из электродов является общим для входной и выходной цепей: с общей базой ОБ (58а), общим эмиттером ОЭ (58б) и общим коллектором ОК (58в).

На схемах (рис.58) применены транзисторы *p-n-p*- типа, во всех случаях коллектор имеет отрицательный, а эмиттер – положительный потенциал по отношению к базе. При включении транзистора по схеме (58а) происходит усиление сигнала по напряжению и мощности; по схеме (58б) – достигается усиление по току. Все усиления происходят за счёт энергии подключённых источников.

Помимо биполярных транзисторов, существует класс полевых транзисторов. Если в первых физические процессы связаны с движением носителей заряда обоих знаков, то в полевых транзисторах принцип их работы основан на управлении движением носителей заряда только одного знака (основных носителей).

Элементы полевого транзистора (см. рис. 59) получили иное название, нежели элементы биполярного транзистора.

Канал – это область полупроводникового кристалла, в которой поток носителей заряда регулируется изменением её поперечного сечения. Истоком называется электрод полевого транзистора, через который в канал втекают носители заряда, коллектируемые при выходе из канала другим электродом – стоком. Электрод, к которому прикладывается управляющее напряжение, называется затвором.

Различают два типа полевых транзисторов: полевые транзисторы с управляющими *p-n*- переходами и полевые транзисторы с изолированным затвором (МДП- транзисторы). Аббревиатура МДП означает структуру металл - диэлектрик – полупроводник. Очень часто в качестве диэлектрика используется окисел (в частности, двуокись кремния SiO_2), поэтому в литературе нередко встречается термин МОП транзистор (металл – окисел – полупроводник). МДП транзисторы в свою очередь подразделяются на транзисторы со встроенным (собственным) каналом и транзисторы с индуцированным каналом.

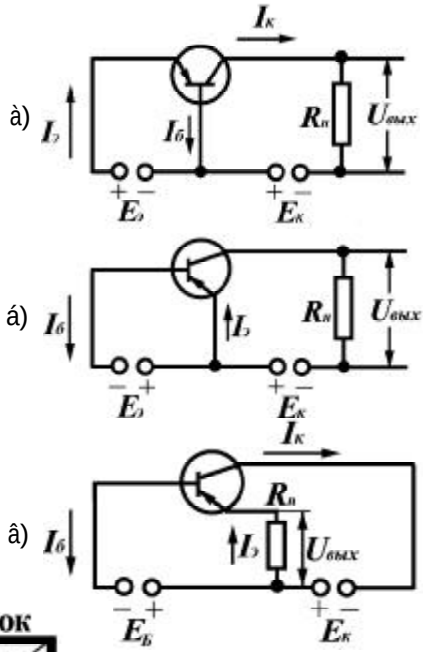

Рис. 58.

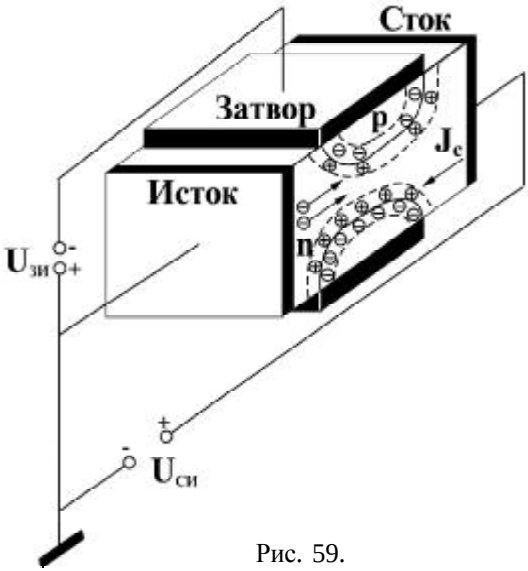

Рис. 59.

Все полевые транзисторы различают также по виду проводимости канала: транзисторы с каналом *p*- или *n*- типа.

Устройство полевых транзисторов трёх типов показано на рис. 60. В транзисторе с

управляющим *p-n*- переходом канал образован частью кристалла *n*- полупроводника, с меньшим поперечным сечением. В *n*- полупроводнике созданы *p*- области, на границе которых с *n*- кристаллом образуется *p-n*- переход.

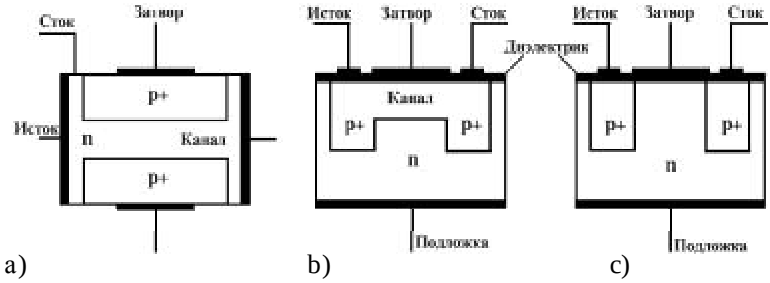

Рис. 60.

В МДП транзисторе со встроенным каналом (рис.60b) *p*- канал и области стока и истока (*p*+ - области) выполнены в процессе изготовления прибора. В МДП транзисторе с индуцированным каналом (рис. 60c) сам канал не создаётся технологическим путём. Под действием электрического поля, возникающего в результате приложения напряжения к затвору, в части *n*- полупроводника, вблизи поверхности, между стоком и истоком образуется тонкий инверсный слой с дырочной проводимостью – *p*- канал.

Транзистор с управляющим *p-n*- переходом может быть выполнен также на основе *p*- полупроводника (с каналом *p*- типа), а транзисторы с изолированным затвором – с областями стока и истока *n*- типа, образованными в кристалле с дырочной проводимостью.

Как видно из рис.60 b) и c), затвор в МДП транзисторах отделён от кристалла полупроводника слоем диэлектрика. Поэтому такие приборы и называют транзисторами с изолированным затвором.

МДП – транзисторы – четырёхэлектродные приборы: четвёртым электродом – подложкой – служит кристалл полупроводника, на основе которого выполнен транзистор.

Как уже отмечалось выше, ток в полевых транзисторах в отличие от биполярных обусловлен лишь движением основных носителей заряда, причём движение это имеет дрейфовый характер. В связи с этим частотные свойства полевых приборов и особенности их работы в импульсных режимах зависят от иных физических параметров и процессов, чем в биполярных транзисторах.

Вторая особенность, отличающая полевые транзисторы от биполярных, - это принцип управления током в приборе с помощью электрического поля. Это поле создаётся обратным напряжением на управляющем *p-n*- переходе или напряжением на затворе в МДП транзисторах. И в том и в другом случае токи в управляющей цепи (цепи затвора) весьма малы, и входное дифференциальное сопротивление прибора велико: $10^8 - 10^{10}$ *Ом* в транзисторах с управляющим *p-n*- переходом и $10^{10} - 10^{12}$ *Ом* в МДП транзисторах. В этом отношении полевые транзисторы близки к электронным лампам. Поэтому усилительные свойства полевых транзисторов принято оценивать не коэффициентом передачи тока, как в биполярных транзисторах, а, как в электронных лампах, крутизной характеристики, определяющей зависимость тока стока (тока в выходной цепи) от напряжения, приложенного ко входной цепи (цепи затвора).

§ 13. Фотоэлектрические приборы

Кристаллические диоды и транзисторы являются основными элементами современной электронной техники в силу их дешевизны, миниатюрности, прочности и долговечности. Помимо прямого назначения (диоды для выпрямления переменных импульсов, транзисторы - для усиления электрических сигналов), диоды и транзисторы используются и для многих других целей.

Рассмотрим использование кристаллического диода для превращения энергии электромагнитных волн в электрическую энергию. Среди разных путей использования этого эффекта – фотогенераторы, преобразующие энергию видимого света в электрическую (фото-ЭДС) (рис.61). Фотоэлектрический эффект (вырывание электронов из металлического образца под действием светового потока) был открыт Г. Герцем в 1887 г. Законы этого явления детально изучил наш соотечественник А. Г. Столетов.

На основе классических представлений о волновой природе света не удалось обосновать все законы фотоэффекта. Лишь в 1905 году великий физик XX века А. Эйнштейн, исходя из представлений о существовании квантов (корпускул, частиц) электромагнитного поля, обосновал эти законы, установив формулу $h\nu = A + \frac{mn^2}{2}$, где $h\nu$ - энергия кванта, падающего на тело, A - работа выхода, работа, которую необходимо совершить по удалению электрона с поверхности облучаемого металла (мы рассматриваем так называемый внешний фотоэффект), $\frac{mn^2}{2}$ - кинетическая энергия, которой будет обладать электрон на выходе из тела.

Наряду с внешним фотоэффектом в полупроводниках обнаружен внутренний фотоэффект, при котором электроны, получившие энергию фотона. не покидают тело, но отрываются от своей структурной частицы. Очевидно, что в этом случае энергия квантов может быть меньше работы выхода A. Но, вместе с тем, достаточной, чтобы возбудить электроны из валентной зоны полупроводника в зону проводимости. На рис. 61а изображена принципиальная схема полупроводникового фотоэлемента. При отсутствии освещения *p-n*-переход нахо-

дится в равновесном состоянии, возникает контактная разность потенциалов, направленная от полупроводника *n*-типа к полупроводнику *p*-типа. Металлическая подложка *M* берется очень тонкой, чтобы световые лучи могли проникнуть до полупроводника *p*-типа. При освещении энергия кванта передается электронам валентной зоны полупроводника, которые, поглащая энергию, возбуждаются, переходя в зону проводимости (рис. 61б). Эти электроны из зоны проводимости пролупроводника *p*-типа могут перейти в металлическую пластинку. В результате динамическое равновесие, возникшее на *p*-*n*-переходе, нарушается. Электроны из полупроводника *n*-типа снова получают возможность переходить в полупроводник *p*-типа, так как там в валентной зоне возникли свободные места («дырки»). Разность потенциалов на *p*-*n*-переходе увеличивается благодаря световому воздействию, и если цепь фотоэлемента замкнута, то в ней возникает фототок, соответствующая ЭДС – называется фотоэлектродвижущей силой (фото-ЭДС). Лучшие фотоэлементы (кремниевые) могут иметь высокий КПД преобразования лучистой энергии в электрическую – 22 – 23%. Помимо кремниевых, используются германиевые,-меднозакисные, селеновые и др. полупроводниковые фо-

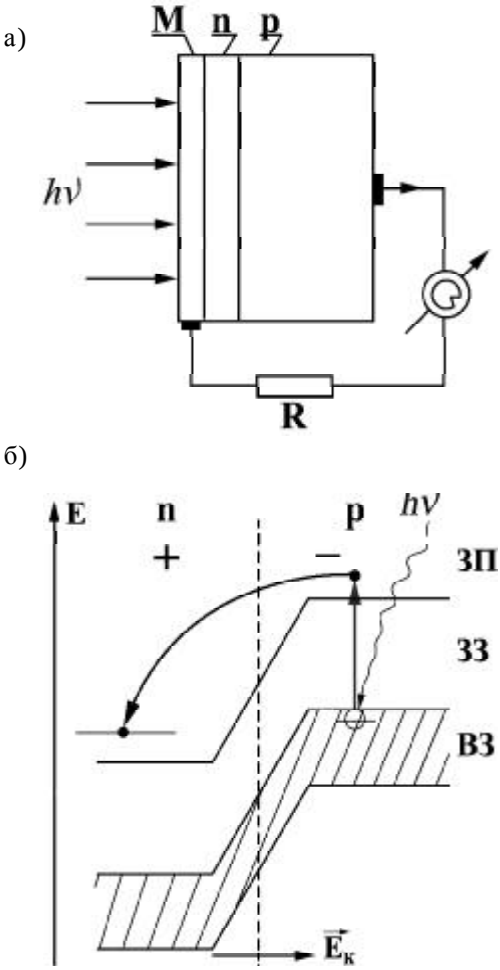

Рис. 61.

тоэлементы. Они более чувствительны, чем фотоэлементы с внешним фотоэффектом (вакуумные фотоэлементы). Основным недостатком полупроводниковых фотоэлементов является их большая инертность, т.е. запаздывание появления фототока от начала облучения *p-n*-перехода, что связано с временным процессом диффузии электронов и «дырок», возникших на *p-n*-переходе при облучении его светом. *P-n*-переход является основной частью фотодиодов (рис.62).

Как видно из принципиальной схемы включения фотодиода, он включается в цепь в запорном направлении последовательно с внешним источником тока (рис.62а).

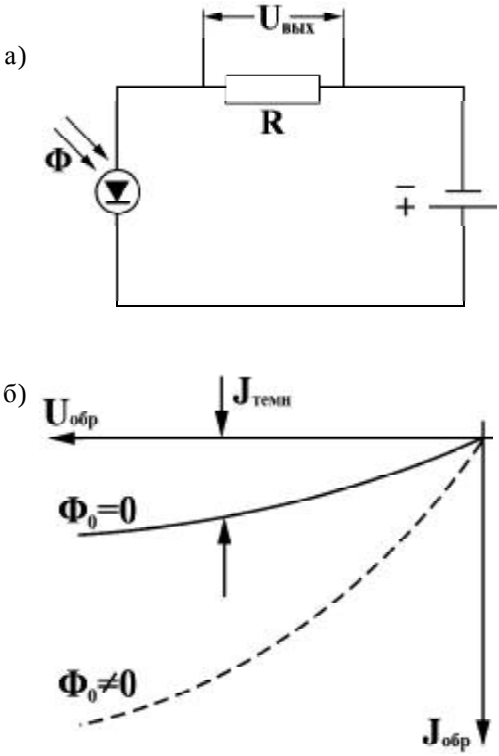

Рис. 62.

При отсутствии светового потока через фотодиод протекает незначительный ток обратного, запорного направления – темновой ток. При освещении *p-n*-перехода возникают дополнительные носители заряда, обратный ток увеличивается в зависимости от интенсивности освещения. На нагрузочном сопротивлении увеличивается падение напряжения, что регистрируется дополнительными устройствами. У полупроводниковых фотодиодов перед вакуумными фотодиодами те же преимущества, что и у других полупроводниковых приборов: относительная дешевизна в изготовлении, миниатюрность, высокая чувствительность и возможность использования источников тока с малым рабочим напряжением.

Если вместо одиночного *p-n*-перехода использовать двойной p-n-переход (транзистор), то можно значительно усилить фототок. Этот эффект используется в управляющих системах.

На рис. 63 изображен фототранзистор *p-n-p*-типа. При изображенном способе включения вывод базы фототранзистора остается свободным. При освещении базы (полупроводник n-типа) в ней появляются свободные электроны и «дырки». Последние, продиффундировав к правому *p-n*-переходу, подхватываются внешним электрическим полем (создаваемым батареей), в результате чего возрастает коллекторный ток, возрастает и выходное напряжение , которое подается в следующие элементы схемы (они не изображены на рис.63).

Если вместо освещения *p-n*-перехода видимым светом использовать гамма-, рентгено- или радиоактивное излучение, то можно получить датчики соответствующих излучений-дозиметры.

При прохождении электрического тока через *p-n*-переход происходит инжектирование в полупроводник p-типа не основных носителей заряда - электронов (аналогично, в полупроводник *n*-типа инжектируются не основные в нем носители заряда – «дырки»), которые встречаясь с основными носителями в данном полупроводнике, рекомбинируют с ними. В ряде полупроводников выделяющаяся энергия не передается кристаллической решётке, а выделяется в виде квантов излучения – фотонов. В таких полупроводниках, как *SiC, GaAs, InAs, InSb*, пропускание тока в прямом направлении через *p-n*-переход, сопровождается свечением. Такие диоды получили название светодиодов. Они нашли применение в световых табло, в счётно-решающих устройствах для ввода и вывода информации, там, где требуются долговечные и потребляющие малую мощность световые индикаторы.

Особенно перспективно использование полупроводника с *p-n*-переходом в качестве рабочего тела квантового генератора. Впервые такое устройство было изготовлено в 1951 г. двумя российскими учёными А.М. Прохоровым и Н.Г. Басовым и американским учёным Ч. Таунсом. За это изобретение все трое были удостоены Нобелевской премией. Но ещё в 1916 г. А. Эйнштейн теоретически предсказал возможность работы такого устройства, в котором каким-то образом большая часть атомов должна быть переведена в возбужденное состояние (инверсное состояние). Если в объеме такого вещества окажется квант излучения с энергией, равной разности энергий возбужденного и основного состояний структурных частиц рабочего тела, то воз-

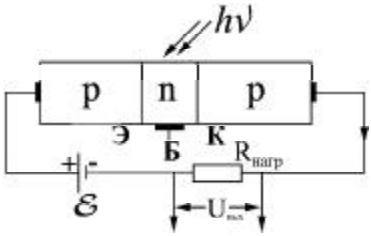

Рис. 63.

никнет процесс индуцированного (вынужденного) перехода возбужденных частиц в основное состояние с испусканием огромного числа фотонов с одной и той же частотой – возникает мощный поток монохроматического излучения. Благодаря особому устройству квантового генератора (эти детали выходят за рамки данного пособия), излучение выходит остро направленным.

В полупроводниковых квантовых генераторах (лазерах – испускающих видимый свет, мазерах – испускающих радиоволны) используется свойство *p-n*-перехода пропускать электрический ток в одном направлении. Инверсное состояние электронов создается при пропускании тока в прямом направлении через *p-n*-переход. Энергия электрического поля, создаваемого внешним источником тока, приводит к возбуждению электронов из валентной зоны в зону проводимости (рис.64).

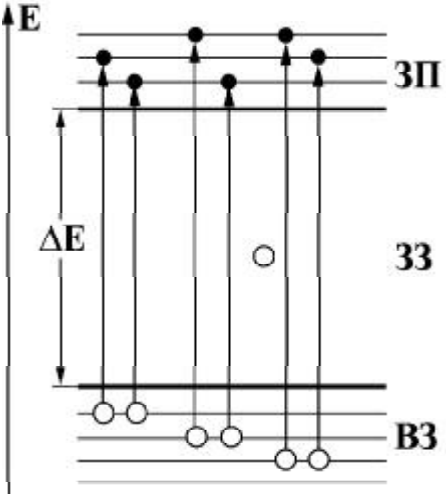

Рис. 64.

При рекомбинации хотя бы одной пары электрон + «дырка» рождается квант с энергией , равной ширине запретной зоны. Повторное поглощение этого кванта практически невозможно, так как этой энергии было бы достаточно для возбуждения электрона из верха валентной зоны на дно зоны проводимости. Но эти состояния или уже свободны от электронов (в ВЗ), или уже заняты (в ЗП). Поэтому родившийся квант станет затравкой для индуцированной рекомбинации электронов ЗП с «дырками» ВЗ. Рождается мощное монохроматическое излучение. Полупроводниковые генераторы монохроматического излучения (лазеры и мазеры) обладают практически КПД=100%. Это обусловлено тем, что в полупроводниковом генераторе происходит прямое преобразование электрической энергии в энергию когерентного (монохроматического) излучения. Они могут работать и при очень низких температурах, так как инверсия в них создаётся не тепловыми процессами.

§14. «Зонная схема» физического вакуума

В 1928 г. английский физик П. Дирак объеденил квантовую механику и специальную теорию относительности, построив новую физическую теорию – квантовую электродинамику. В своей теории П. Дирак получил необычный результат: полная энергия электрона может быть как положительной, так и отрицательной. Последний результат никак не укладывался в представления физиков. Согласно специальной теории относительности, наименьшей положительной энергией, которой может обладать элементарная частица – это энергия покоя этой частицы $E_0 = mc^2$. Теория П.Дирака утверждала, что могут быть состояния с энергией $E<-mc^2$. Таким образом, интервал значений энергии-

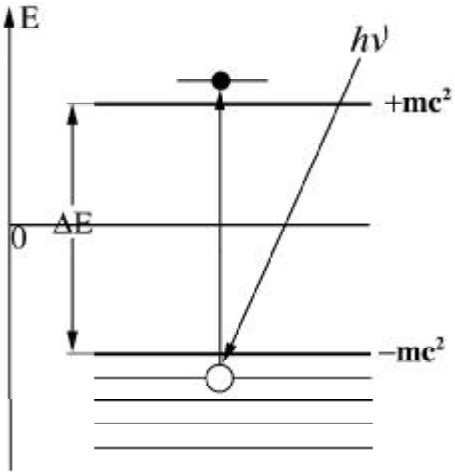

Рис. 65.

$$-mc^2 < \Delta E < +mc^2$$

запрещён. Если изобразить это утверждение графически (рис.65), то мы получим нечто, что напоминает нам зонную схему диэлектрика.

Согласно Дираку, все состояния с отрицательной энергией заполнены электронами, причём, этот фон не проявляет себя непосредственно (этот фон получил название «море Дирака»). Но если на «море Дирака» воздействовать жёсткими гамма-лучами, то электроны фона могут возбудиться и перейти в состояние с положительной энергией. Очевидно, что энергия гамма-квантов должна быть не меньше $2mc^2$ (см. рис65). Освободившееся место, в «море Дирака» было названо «дыркой», которая, совпадая с электроном по всем своим свойствам, должна отличаться от него знаком заряда. В 1932 г. в космических лучах (частицах, приходящих на Землю из космоса) были обнаружены частицы, во всём подобные электронам, но отличающиеся от них знаком заряда. Эти частицы, полу-

чившие название «позитроны», стали сопоставлять «дыркам» в «море Дирака». Впоследствии было установлено, что у всех частиц имеются античастицы. И каждой частице можно сопоставить своё «море Дирака». Вся эта сложная система из «морей Дирака» получила в современной физике название «физический вакуум». Эта специальная физическая система проявляет себя в ряде экспериментов, что говорит о её реальности, но со свойствами, существенно отличающимися от свойств известных видов материи (вещества и различных полей). Но более подробный разговор о физическом вакууме выходит за пределы возможностей данной книги. Для первоначального ознакомления с этим физическим объектом читателя можно отослать к интересной книге Р. Подольного «Нечто по имени Ничто».

§ 15. Эффект Холла (классический)

Добавление к названию параграфа «классический» оказалось чрезвычайно важным спустя 100 лет после открытия обсуждаемого эффекта в 1879 г. физиком Е. Холлом. Никто не подозревал, что в этом явлении могут проявиться квантовые законы.

Но сначала рассмотрим классический эффект Холла, который находит большое практическое (и, как оказалось) научное применение. Например, мы неоднократно говорили о том, что «дырки» обладают положительным эффективным зарядом. Эффект Холла позволяет определить это экспериментально.

Рассмотрим схему опыта по обнаружению эффекта Холла. Имеется проводник прямоугольного сечения, по нему идёт ток (рис. 66). Перпендикулярно току (и боковым граням проводника) приложено магнитное поле – это простейший вариант расположения векторов плотности тока $\vec{j}$ и магнитного поля индукции $\vec{B}$. Дадим сначала качественное описание явления.

Пусть для определённости носители заряда имеют положительный знак (технический ток). Тогда, учитывая направление тока – слева направо – и магнитного поля (перпендикулярно чертежу), по правилу «левой руки» устанавливаем, что на заряды будет действовать дополнительная сила $\vec{F}$ - магнитная составляющая силы Лоренца, которая отклонит положительные заряды к верхней грани проводника. Между верхней и нижней гранями возникает разность потенциалов – ЭДС Холла. В данном случае эта ЭДС направлена сверху вниз (рис. 66а).

Пусть теперь носители заряда заряжены отрицательно (рис. 66 б). В том же внешнем электрическом поле, которое определяет ток в цепи, отрицательно заряженные частицы будут двигаться против поля. Так как изменились знаки и у заряженных частиц, и у скорости, то направление $\vec{F}$, действующей на движущиеся частицы, имеющие отрицательный заряд, по-прежнему будет направлена снизу вверх. На верхней грани теперь накопятся отрицательно заряженные частицы (рис. 66 б). Теперь ЭДС Холла (разность потенциалов, возникающая между верхней и нижней гранями) будет направлена снизу вверх. Таким образом, по направлению ЭДС Холла можно установить, какой знак имеют носители заряда.

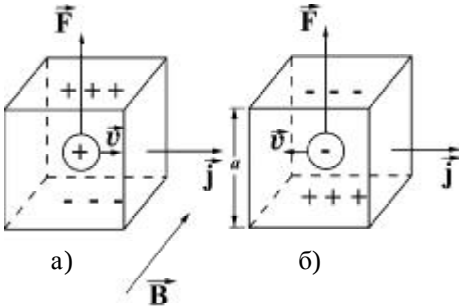

Рис. 66.

Дадим теперь математическое описание явления Холла. Будем исходить из экспериментально полученной формулы для ЭДС Холла:

$$\Delta j = R_H Bja \quad , \qquad (3.15.1)$$

где B – индукция поперечного магнитного поля, j – плотность тока (технического), a – высота образца, R_H - постоянная Холла – коэффициент пропорциональности.

На положительно заряженную частицу, движущуюся слева направо со скоростью v, действует сила

$$F_л = enB.$$

Под действием этой силы заряженные частицы будут отклоняться к верхней грани, на противоположной грани таких зарядов будет меньше, между гранями возникает разность потенциалов, соответствующее ей электрическое поле будет направлено сверху вниз, его напряжённость равна

$$E = \frac{\Delta j}{a} \ ,$$

где a – высота образца.

Это поле будет действовать на положительно заряженные части-

цы с силой F, направленной сверху вниз:

$$F = eE = \frac{e\Delta j}{a},$$

она будет противодействовать накоплению положительно заряженных частиц у верхней грани проводника. В состоянии динамического равновесия сила F уравновесит силу $F_л$:

$$F = F_л, \text{ или } enB = \frac{e\Delta j}{a},$$

откуда

$$\Delta j = nBa \cdot$$

Учитывая, что

$$j = enn \text{ и } n = \frac{j}{en},$$

для Dj можно написать:

$$\Delta j = \frac{1}{en} jBa \cdot \qquad (3.15.2)$$

Сравнивая формулы (3.15.1) и (3.15.2), получаем для постоянной Холла (если движутся положительно заряженные частицы)

$$R_H = \frac{1}{en} \cdot \qquad (3.15.3)$$

Зная R_H, можно определить как концентрацию носителей заряда, так и знак их заряда (при отрицательно заряженных носителях заряда R_H <0).

Если применить полученный результат к полупроводникам, то в полупроводниках *n*- типа постоянная Холла отрицательна (исторически сложилось так, что этот эффект Холла назвали нормальным), у дырочных полупроводников $R_H > 0$.

Так как удельная проводимость $s = enU$, где U – подвижность носителей заряда, то, используя формулу (3.15.3), получаем возможность определить подвижность носителей заряда:

$$U = R_H \cdot s \quad .$$

В предыдущих рассуждениях предполагалось, что все носители заряда имеют одну и ту же скорость v. Но это справедливо лишь для металлов и вырожденных полупроводников. В невырожденных полупроводниках скорости носителей заряда распределены по Максвеллу. Поэтому для невырожденных полупроводников более строгие расчёты дают следующую формулу для постоянной Холла:

$$R_H = \frac{A}{en},$$

где A порядка единицы, например, для германия, кремния и серого олова A=1.17 при рассеянии на фононах (рассеяние на фононах определяет величину s).

В полупроводниках со смешанной проводимостью постоянная Холла меньше, чем в полупроводниках с одним типом носителей заряда. В этом случае

$$R_H = \frac{A}{e} \cdot \frac{U_p^2 p - U_n^2 n}{\left(U_p p + U_n n\right)^2} \quad , \qquad (3.15.4)$$

где n и p – концентрация носителей заряда, U_n и U_p - их подвижности.

В зависимости от величины членов в числителе (3.15.4) постоянная Холла будет больше или меньше нуля. Для собственных полупроводников, для которых n=p , имеем:

$$R_H = \frac{A}{en} \cdot \frac{U_p - U_n}{U_p + U_n}.$$

У полупроводников постоянная Холла больше, чем у металлов, что связано с меньшей концентрацией заряда в полупроводниках, а подвижность их, наоборот, в полупроводниках значительно выше, чем в металлах.

У большинства металлов R_H колеблется в пределах от 10^{-11} до 10^{-10} $м^3/Кл$. Аномально высокое значение R_H имеют элементы пятой группы в таблице Менделеева: $Bi, Sb, As.$ У висмута, напри-

мер, R_H почти в 10^4 раз больше, чем у меди. У ряда металлов (Na, Cs, Cu, Ag) постоянная Холла имеет отрицательное значение (нормальный эффект Холла), у других (Be, Cd, Zn и др.) – положительное.

Нормальный эффект Холла (то есть R_H <0) будет у тех металлов, у которых зона проводимости меньше чем наполовину заполнена электронами, которые ведут себя как отрицательно заряженные частицы с положительной эффективной массой. Если же ЗП заполнена почти полностью, то на оставшихся в ней незаполненных уровнях носители заряда ведут себя как частицы, обладающие положительной эффективной массой и положительным зарядом. Такие металлы имеют дырочную проводимость, поэтому R_H >0. Такой эффект Холла называют аномальным. Он обнаружен у цинка, кадмия, бериллия, висмута и др.

Эффект Холла получил практическое применение для измерения магнитной индукции постоянных и переменных полей, знака, подвижности, концентрации носителей заряда.

§16. Термомагнитные и термоэлектрические явления

1) Эффект Эттингаузена (1884 г.).

При выводе условия равенства силы Лоренца и силы Холла (см.§15):

$$enB = eE_H ,$$

предполагалось, что все электроны имеют одну и ту же скорость. В действительности это соотношение справедливо для некоторой средней скорости $\overline{n}$. Более быстрые электроны отклоняются к одной грани и нагревают этот край образца, медленные электроны отклоняются к другой и забирают энергию у решётки , охлаждая эту грань. В результате в проводнике возникает градиент температуры. Этот эффект получил название эффекта Эттингаузена. Он используется для понижения температуры одной грани проводника (более чем на $100^0 C$).

2) Эффект Нернста (1884 г.).

Так как в магнитном поле движущиеся электроны закручиваются (сила $F_Л = enB$ выполняет роль центростремительной силы) с радиусом окружности:

$$R = \frac{mn}{eB},$$

то чем больше скорость электронов, тем больше и радиус закручивания. В результате одна грань (передняя) обогащается быстрыми электронами, она нагревается, соответственно задняя грань охлаждается (туда «добираются» медленные электроны. Возникает градиент температуры. Этот эффект получил название эффекта Нернста.

3)Эффект Зеебека (1823 г.).

Электроны играют роль не только в электропроводности, но и в переносе энергии. Поэтому естественна связь между электрическими явлениями и термической неоднородностью металла или полупроводника.

Пусть электрическая цепь состоит из двух неоднородных проводников, контакты которых поддерживаются при разных температурах. При малой разности температур термоЭДС, возникающая в этой цепи, будет равна

$$\Delta j = a\left(T_2 - T_1\right),$$

где $a = \dfrac{d(\Delta j)}{dT}$ называется дифференциальной, или удельной термоэлектродвижущей силой.

Естественно, что эффект Зеебека зависит от свойств соприкасающихся проводников и температуры.

Существуют три причины, обуславливающие эффект Зеебека:

1) наличие градиента температуры, вызывающего движение носителей заряда (объёмная составляющая Δj);

2) изменение положения уровня Ферми при нагревании, то есть изменение максимальной энергии электронов при нагревании (контактная составляющая Δj , обусловленная переходом электронов из металла с большей энергией Ферми в металл с меньшей энергией Ферми);

3) увлечение электронов фононами (эффект, предсказанный советским физиком Л. Гуровичем в 1945 г.).

В полупроводниках термоЭДС может на несколько порядков быть выше,чем у металлов. Это видно из следующей формулы, связывающей дифференциальный коэффициент термоЭДС с теплоёмкостью электронного газа

$$a \sim \frac{C_e}{ne} .$$

Действительно в металлах теплоёмкость электронов очень мала (вплоть до самых низких температур, см.Гл.1 §14) и на несколько порядков выше концентрация носителей заряда, чем в полупроводниках. При низких температурах существенен вклад эффекта Гуревича.

4.Эффект Пельтье (1834 г.)

В результате различия средних энергий электронов в разных проводниках, составляющих электрическую цепь, при прохождении тока в месте контакта разнородных проводников помимо «джоулева тепла» выделяется (поглощается) дополнительное количество теплоты. Из опыта следует, что это количество энергии Пельтье равно:

$$Q_{\Pi} = \Pi Jt ,$$

где Π - коэффициент Пельтье.

Между эффектами Пельтье и Зеебека существует непосредственная связь: в цепи из разнородных проводников разность температур вызывает появление электрического тока, а проходящий термоток создаёт дополнительную разность потенциалов. Дж. Томсон установил связь между этими коэффициентами

$$a = \frac{\Pi}{T} .$$

5. Эффект Томсона (1854 г.).

Этот эффект был предсказан теоретически. Если в однородном проводнике существует градиент температуры, то при пропускании тока помимо «джоулева тепла» происходит поглощение или выделение дополнительного «количества энергии». Этот эффект линейный в зависимости от тока J, поэтому его знак зависит от направления тока по отношению к направлению градиента температуры

$$Q_T = tJ(T_2 - T_1)t ,$$

где t носит название коэффициента Томсона.

Явление объясняется тем, что носители заряда переносят не только электрический заряд, но и энергию.