

ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ И ТРИОДЫ

М. А. Федоров

В этой статье рассказывается о работе полупроводниковых диодов и триодов (транзисторов), замечательных электронных устройств, изобретение которых стало возможным только после кропотливого изучения свойств определенного типа веществ — полупроводников. Сконструированные несколько десятков лет назад, эти устройства в настоящее время стали незаменимыми элементами огромного количества физических и радиотехнических приборов.

Проводники, полупроводники и диэлектрики

Прежде чем говорить о полупроводниковом диоде и транзисторе, мы расскажем о том, что такое полупроводники, какова их внутренняя структура и какими они обладают свойствами.

Из большого числа известных в природе полупроводников наиболее изучены кремний и германий. Это элементы, стоящие в таблице Менделеева под номерами 14 и 32.

Если говорить строго, то полупроводники — это диэлектрики, только очень слабые диэлектрики. Слабые в том смысле, что их нетрудно превратить в проводники. Для этого существует несколько способов. Основные из них: облучение, бомбардировка полупроводников быстрыми частицами, нагревание и добавление в полупроводник соответствующих примесей.

Как известно, диэлектрики не проводят электрического тока, то есть

являются изоляторами. Если на концах диэлектрического образца создать разность потенциалов и измерить ток, то он будет равен нулю. Это объясняется тем, что внутри образца нет свободных заряженных частиц, например, электронов, которые под действием электрического поля перемещались бы с одного конца образца на другой и таким образом создавали бы ток. Атомы диэлектрика прочно удерживают все свои электроны и уж если отпускают их от себя, то только поменявшись с соседом на другой электрон.

Иное дело в металлах. Основная масса электронов и там ведет себя подобно электронам в диэлектрике. Но некоторые из них, те, которые находятся далеко от положительно заряженных ядер и потому слабо с ними связаны, при образовании кристалла получают такую свободу, что могут бродить по всему кристаллу. Эти электроны называются свободными электронами, или электронами проводимости. Если поместить металл в электрическое поле, то эти

свободные электроны (поскольку они обладают отрицательным зарядом) движутся в сторону, противоположную направлению электрического поля. Электрическое поле, словно ветер, сдувает их на границу образца. Гальванометр, подключенный к такому образцу, регистрирует электрический ток.

Итак, диэлектрик потому не проводит электрический ток, что в нем нет свободных носителей заряда, которые могли бы под действием электрического поля свободно двигаться по образцу.

Но такие заряды мы сами можем создать. Один из способов — подействовать на атом диэлектрика каким-нибудь подходящим «инструментом», например, квантом света — фотоном. Мы облучаем диэлектрик обычным светом, а еще лучше — рентгеновскими лучами, так как они обладают большей энергией, и выбиваем электроны из атомов, образующих кристалл. Таким путем в диэлектрике можно создать свободные электроны и превратить его в проводник. Ясно, что энергия, нужная для освобождения электрона, зависит прежде всего от того, из каких атомов образован кристалл и как эти атомы расположены в кристалле. Другими словами, эта энергия различна для различных кристаллов. Минимальная энергия, необходимая для того, чтобы оторвать электрон от атома и превратить его в свободный, называется энергетической щелью данного диэлектрика. Если на кристалл падает излучение, энергия квантов которого ниже, чем минимальная энергия, необходимая для освобождения электрона, то ни один из электронов не будет выбит и диэлектрик так и останется изолятором.

Различие между диэлектриками и полупроводниками как раз и заключается в величине энергетической щели. Диэлектрики с достаточно малой энергетической щелью называются полупроводниками. Энергетическая щель у полупроводников примерно в 10 раз меньше, чем у ярко выражен-

ных диэлектриков, таких, как, например, алмаз. Ширина энергетической щели для алмаза равна 7 эв, тогда как для германия мы имеем 0,72 эв, а для кремния 1,1 эв *). Таким образом, чтобы полупроводник превратить в проводник, достаточно сообщить его электронам относительно небольшую энергию.

Теперь представьте себе, что мы нагреваем диэлектрик. При повышении температуры средняя энергия электронов увеличится, и многие из них будут иметь энергию, достаточную для преодоления энергетической щели. В результате становится все больше свободных электронов, причем появление свободных электронов происходит тем чаще, чем выше температура, и уж, конечно, чаще всего у тех кристаллов, у которых меньше энергетическая щель, то есть у полупроводников. Потому полупроводники и называют полупроводниками, что при достаточно высоких температурах (например, комнатных $\sim 300^\circ \text{K}$) они являются проводниками, а при понижении температуры превращаются в диэлектрики. Транзисторный приемник перестает работать в трескучий мороз, так как полупроводники, использованные в его транзисторах, превращаются в самые настоящие изоляторы.

Электроны и дырки

Пусть каким-либо образом мы выбили электрон из атома диэлектрика и он улетел от этого атома. В кристалле в этом месте образуется нехватка электрона или, как говорят, дырка. Интересно то, что дырка не стоит на месте, а начинает двигаться и вообще ведет себя подобно положительно заряженной частице. Вот как это происходит.

Когда у какого-то атома не хватает электрона, то электрон может пе-

*) 1 электрон-вольт (1 эв) — энергия, приобретаемая электроном при прохождении разности потенциалов в 1 вольт. $1 \text{ эв} = 1,6 \cdot 10^{-19} \text{ Дж}$.

рескочить с соседнего атома, имеющего ровно столько электронов, сколько нужно (рис. 1, а). Но тогда нехватка электронов обнаруживается на соседнем атоме, то есть дырка перемещается к соседнему атому (рис. 1, б). Конечно, на самом деле это электроны с заполненных атомов перескакивают на соседние незаполненные, но удобнее говорить о движении дырки. Дырка все-таки одна, а отражает движение многих электронов. Движение дырки подобно движению положительно заряженной частицы, потому что, если под действием электрического поля электроны движутся в одну сторону, то их «нехватка» движется в другую.

Будет полезно подчеркнуть разницу между поведением металлов и полупроводников в электрическом поле. Ток в металле создается только свободными электронами. Металл так устроен, что в нем всегда имеется то или иное их количество. В полупроводнике же при достаточно низких температурах свободных электронов нет. Эти свободные электроны

можно создать. Но при этом, отрывая электрон от атома и предоставляя ему свободу, мы создаем и дырку. В дальнейшем эти электроны и дырки движутся независимо друг от друга. Получается так, что ток в полупроводнике складывается из тока электронов и тока дырок.

Может случиться так, что свободный электрон и дырка встретятся, то есть что электрон при своем движении по кристаллу встретит атом, у которого как раз не хватает электрона. Он займет это свободное место, и в этот момент и электрон, и дырка исчезнут. В этом случае говорят, что произошла рекомбинация. Вы вправе спросить, а зачем же тогда создавать электроны и дырки, если они рано или поздно вот так просто друг друга уничтожат? Во-первых, при наложении электрического поля электроны и дырки как частицы, имеющие разные заряды, движутся в противоположные стороны, и мы можем зарегистрировать ток, который они создают еще до того, как рекомбинируют. Во-вторых, мы не из каждого атома выбиваем электроны, а, например, из каждого тысячного. Мы создаем при этом значительный ток (поскольку в каждом кубическом сантиметре кристалла примерно 10^{23} атомов) и оставляем достаточную свободу для движения электронов и дырок.

Полупроводниковые кристаллы выращиваются чрезвычайно чистыми. Это тоже делается для того, чтобы движению электронов и дырок ничего не мешало.

Доноры и акцепторы

Существует и такой способ создания свободных электронов в полупроводнике. Представим себе кристалл германия, в котором один из атомов германия заменен атомом мышьяка. У атомов германия валентность равна 4, а у атомов мышьяка — 5. Когда атом мышьяка оказывается в окружении атомов германия, пятый

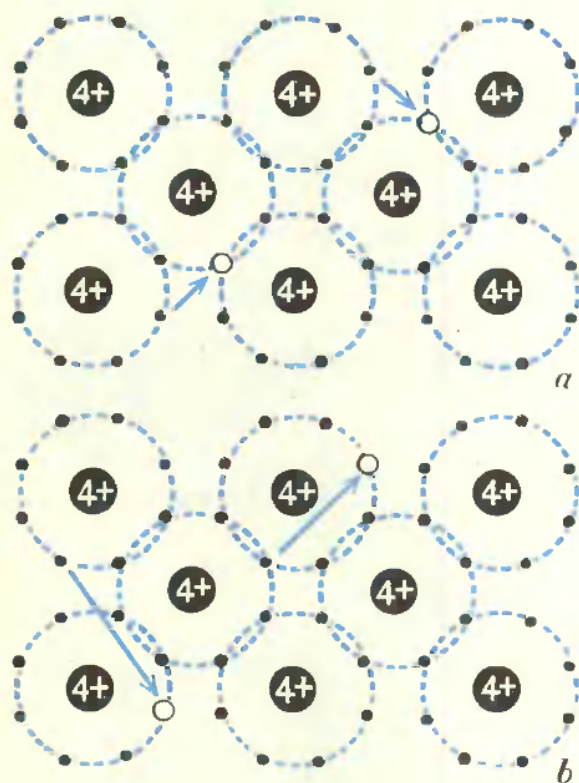


Рис. 1.

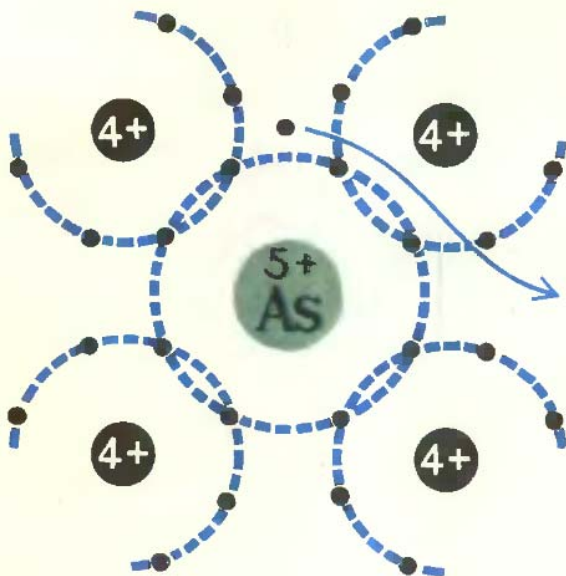


Рис. 2.

электрон мышьяка связан с ним очень слабо (рис. 2). Чтобы его оторвать, достаточно энергии $0,1 \text{ эВ}$, которая не превышает тепловой энергии кристалла. Таким образом, один из электронов атома мышьяка становится свободным и странствует по кристаллу. Вводя в полупроводник примеси, имеющие более высокую валентность, мы всегда можем получить достаточное количество отрицательных носителей тока. В этом случае примесные атомы называются донорами, а сами полупроводники, содержащие примеси донорного типа, называют полупроводниками *n*-типа.

Можно внедрить в кристалл германия трехвалентные атомы, например, алюминия. Атому алюминия, для того, чтобы прочно сидеть в кристалле германия, необходимо иметь четыре внешних электрона, поэтому он захватывает недостающий электрон у соседних атомов германия (рис. 3). В кристалле образуется дырка. Примесные атомы такого типа называются акцепторами *), а полупроводники — полупроводниками *p*-типа.

Полупроводники *n*- и *p*-типа имеют соответственно избыточное количество отрицательных или положительных

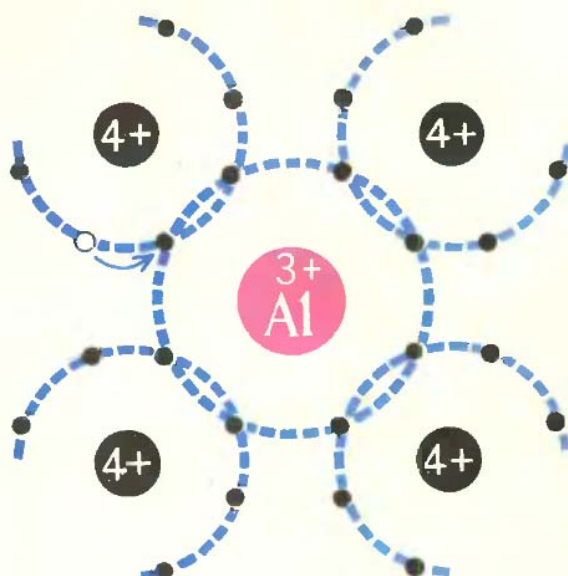


Рис. 3.

ных носителей тока. Избыточное — по сравнению с теми электронами и дырками, которые создаются тепловыми возбуждениями. Однако общий заряд каждого из полупроводников, конечно, равен нулю, поскольку их кристаллы образуются из нейтральных атомов. Отрицательный заряд электронов в полупроводнике *n*-типа компенсируется положительно заряженными донорными центрами, закрепленными в узлах кристаллической решетки, например, атомами мышьяка. В полупроводнике *p*-типа положительный заряд дырок гасится отрицательным зарядом акцепторных центров.

Переходы между полупроводниками

Возьмем два различных образца германия (*p*- и *n*-типа) и приложим их друг к другу *). Нас интересует, что будет происходить на границе раздела (рис. 4).

До тех пор, пока не был создан контакт между образцами, электроны в *n*-области и дырки в *p*-области, подходя к границе образца, отражались от поверхности и возвращались в глубь полупроводника. Как

*) От корня «акцепт» — принимать.

*) В таком случае говорят, что создан *p-n*-переход.

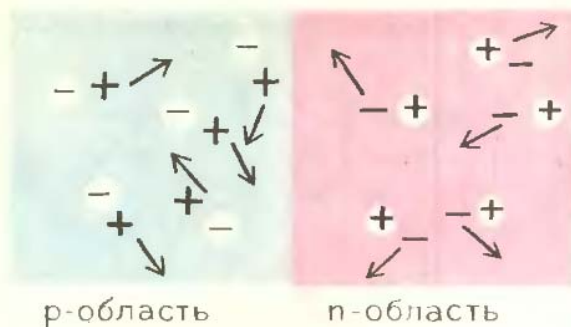


Рис. 4.

только образовался переход, ситуация изменилась. Электроны из n -области, движущиеся по направлению к границе раздела, имеют возможность перескочить в p -область. Аналогично дырки из p -области просачиваются в n -область. Оба эти процесса действуют в одном направлении: они стремятся зарядить германий n -типа положительно, а германий p -типа отрицательно. В самом деле, n -область, полный заряд которой до образования перехода был равен нулю, теперь заряжается положительно, отдавая отрицательно заряженные электроны в p -область. Положительно заряженные дырки, приходя из p -области, только способствуют этому. Германий же p -типа и за счет прихода электронов, и за счет ухода дырок приобретает отрицательный заряд.

Однако весь этот процесс рано или поздно должен остановиться. Когда германий n -типа зарядится положительно, а германий p -типа — отрицательно, то образуется как бы конденсатор, и на границе возникнет электрическое поле, направленное из n -области в p -область (рис. 5). А это значит, что, например, электроны из n -области уже не смогут больше уходить в p -область. Сильное электрическое поле на границе заставит их повернуть обратно. То же самое можно сказать и в отношении дырок. Возникшее электрическое поле понизит их шансы уйти в n -область. Итак, наступит равновесие, при котором германий n -типа будет заряжен положительно, германий p -типа —

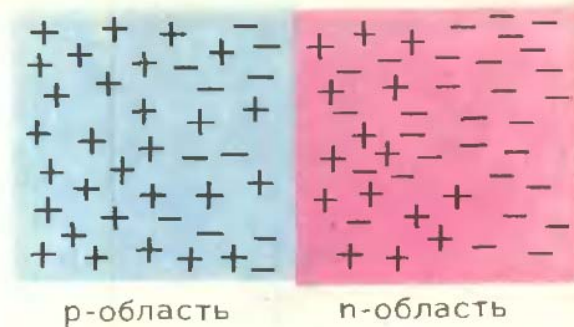


Рис. 5.

отрицательно, а на границе будет сильное электрическое поле, направленное из n -области в p -область.

Если принять потенциал электрического поля на границе за нуль, то потенциал n -области больше нуля, а потенциал p -области меньше нуля (рис. 6), как будто это пластины заряженного конденсатора. Потенциал резко изменяется в той области, где существует электрическое поле, т. е. на границе раздела между германием p - и n -типа. Разность потенциалов в p — n -переходе равна той энергии, которую необходимо иметь электрону для перехода из n -области в p -область, или дырке для перехода из p -области в n -область. Что касается дырок в n -области и электронов в p -области, то если им посчастливится случайно попасть на границу раздела, электрическое поле только подтолкнет их в другую область. Однако если такое случается, то это приводит к уменьшению заряда и n и p -областей, т. е. к уменьшению пере-

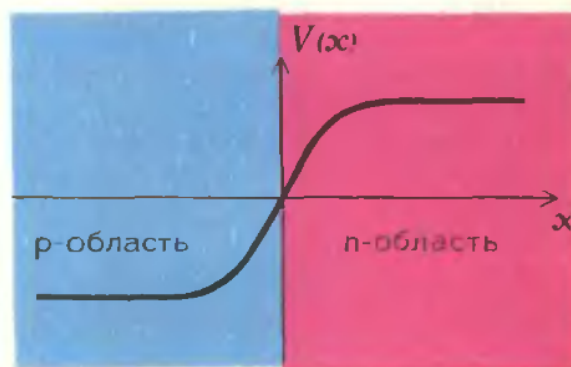


Рис. 6.

пада потенциала на границе и к уменьшению электрического поля. Но тогда найдутся электроны в n -области и дырки в p -области, которые обладают достаточной энергией для преодоления уменьшившегося потенциального барьера*). Они проскочат через барьер, и равновесие будет восстановлено. Мы видим, что равновесие между p - и n -областями динамическое. Между ними все время происходит обмен электронами и дырками, но этот обмен эквивалентен: средний заряд p - и n -областей остается неизменным и, следовательно, остаются неизменными перепад потенциала и электрическое поле на границе. Однако суммарный ток через границу равен нулю. Это равновесие можно нарушить, только создавая на концах p — n -перехода внешнюю разность потенциалов. Рассмотрим теперь этот случай.

Полупроводниковый диод

Образец с p — n -переходом обладает одним важным свойством: он может работать в качестве диода, то есть системы, пропускающей ток только в одном направлении.

Посмотрим, что происходит, если мы включаем p — n -переход в электрическую цепь, то есть создаем на концах образца внешнюю разность потенциалов. Возможны два случая. Либо перепад потенциала на границе p — n -перехода увеличивается, либо уменьшается. Равновесие между p - и n -областями, осуществляемое при помощи обмена электронами и дырками, нарушается, и средний ток через границу раздела отличен от нуля. Величина этого тока будет зависеть от того, увеличивает или уменьшает внешнее напряжение тот перепад потенциала, который существует в переходе сам по себе.

Если внешнее напряжение увеличивает скачок потенциала на границе, то

ясно, что ток через переход чрезвычайно мал. Например, в p -области дырочный ток равен нулю, так как увеличивается величина потенциального барьера, который нужно преодолеть дыркам. Другими словами, увеличивается работа, которую им нужно совершить против сил электрического поля при переходе через границу раздела. Электронный ток в p -области тоже очень мал из-за малого количества электронов. В n -области, где преобладают отрицательные носители, ситуация аналогичная: электрическое поле, отталкивающее электроны от границы, возросло, а дырки не могут обеспечить достаточного тока, поскольку в n -области их ровно столько, сколько электронов в p -области, то есть очень мало.

Если же внешнее напряжение уменьшает величину скачка потенциала, то в p -области найдется большое число дырок, обладающих энергией, достаточной для преодоления ступеньки потенциала, и дырочный ток будет пропорционален числу дырок в p -области, которое очень велико. Точно так же электронный ток будет пропорционален числу электронов в n -области. В результате ток через переход будет очень большим.

Все эти процессы в целом легко понять следующим образом. При смещении равновесия, то есть при увеличении или уменьшении перепада потенциала, обмен электронами и дырками между p - и n -областями нарушается и возникает отличный от нуля суммарный ток через p — n -переход. Этот ток во всех случаях стремится восстановить равновесие, то есть уменьшить перепад потенциала, если он возрос, и наоборот. Однако этого ему не удастся сделать, так как переход включен в электрическую цепь. Средний заряд, переносимый с одной стороны p — n -перехода на другую, не задерживается в образце, а уходит дальше по цепи. Равновесие так и остается нарушенным.

Итак, при включении p — n -перехода под напряжение, обмен заря-

*) См. «Квант» № 5, 1971 г., стр. 8—16.

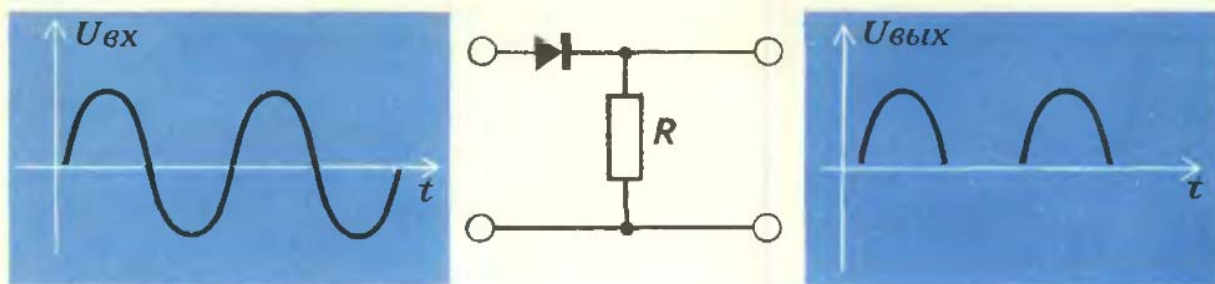


Рис. 7.

дами между p - и n -областями изменяется и через переход протекает ток. Этот ток очень велик при одном знаке внешнего напряжения и очень мал при другом.

Свойство полупроводникового диода пропускать ток только в одном направлении используется, например, для создания выпрямителей — устройств, преобразующих переменный ток в постоянный.

Простейшая схема такого выпрямителя изображена на рисунке 7. Для анализа работы этой схемы нам важен только тот факт, что при одном знаке напряжения сопротивление диода (R_D) практически равно нулю (когда он пропускает огромный ток), а при другом знаке напряжения диод практически не проводит тока и, следовательно, его сопротивление очень велико.

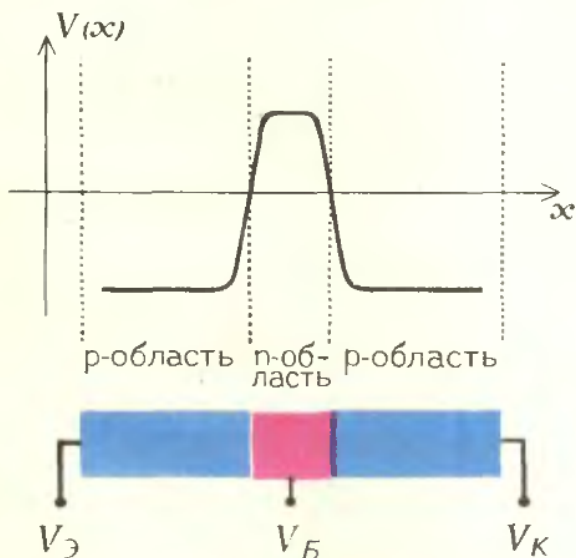


Рис. 8.

Предположим теперь, что диод пропускает ток только слева направо, то есть он открыт, когда напряжение на входе больше нуля, и закрыт, когда оно отрицательно.

При $U_{вх} > 0$ общее сопротивление цепи

$$R_D + R \approx R,$$

и напряжение на выходе будет практически равно входному.

При $U_{вх} < 0$ сопротивление диода бесконечно ($R_D \gg R$), и, следовательно, падение напряжения на сопротивлении R практически равно нулю. Напряжение на выходе равно нулю. Таким образом, наша система преобразовывает знакопеременное напряжение в знакопостоянное.

Транзистор

Теперь мы можем рассмотреть транзистор — систему, которая объединяет в себе два только что рассмотренных перехода. $p-n-p$ -транзистор — это изготовленный, например, из германия небольшой брусок, составленный из трех участков: p -область, n -область и опять p -область (рис. 8). Каждый переход в транзисторе ведет себя так, как описано выше. В частности, в каждом переходе должен наблюдаться перепад потенциала — падение из n -области в каждую из p -областей, так что общее распределение потенциала будет таким, как показано на рисунке 8.

В работающем транзисторе каждая из трех областей подключена к источ-

нику внешнего напряжения. К n -области, которая называется базой, подведен небольшой отрицательный потенциал, а правая p -область (коллектор) подключена к значительно большему отрицательному потенциалу. На рисунке 9 распределение потенциала в работающем p — n — p -транзисторе изображено сплошной линией.

Рассмотрим сначала, что происходит с положительными носителями тока — дырками, потому что именно их поведение управляет работой p — n — p -транзистора. Легко понять, что из левой p -области (эмиттера) на базу пойдет ток положительных носителей, поскольку потенциал базы будет теперь уменьшен на величину V_B и дыркам нужно преодолеть меньший потенциальный горб, чем это было раньше. Таким образом, дырки будут «эмиттироваться» из p -области в n -область. Что будет с ними дальше? Могло бы показаться, что этот ток вытечет из n -области через контакт, подключенный к базе. Но этого не произойдет, поскольку n -область делается достаточно узкой для того, чтобы дырки проскакивали в правую p -область на коллектор; тем более, что при переходе с базы на коллектор потенциал резко падает и на дырки действует сильное электрическое поле, направленное в сторону коллектора. Обратный путь дыркам закрыт. При переходе с коллектора на базу им нужно было бы преодолеть слишком высокий барьер.

Итак, почти весь дырочный ток, вышедший из эмиттера, собирается в области коллектора и только небольшая его часть (доли процента) идет на базу. Естественно, ток через эмиттер равен сумме токов базы и коллектора. Если теперь слегка менять потенциал V_B , то ток через эмиттер будет меняться очень сильно, следовательно, очень сильно будет меняться и ток через коллектор, ведь они примерно равны с точностью до мизерного тока базы. Перед нами усилитель: при помощи небольшого тока I_B , проходящего через электрод

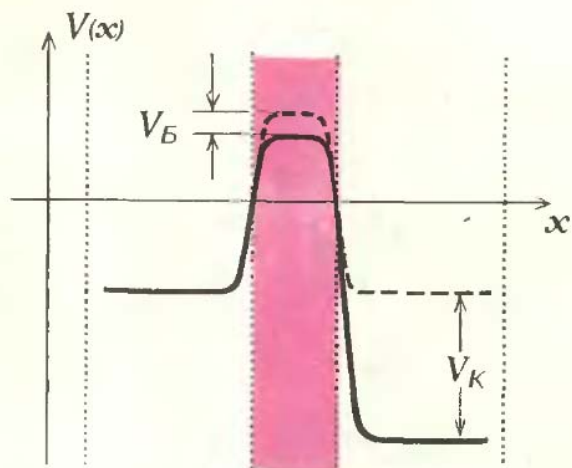


Рис. 9.

базы, мы получим примерно в 100 раз больший ток через коллекторный электрод.

Остается рассмотреть, что происходит с электронами. Электронный ток с базы на коллектор очень мал, поскольку в этом направлении электронам приходится преодолевать очень высокий потенциальный барьер. Но зато при переходе электронов с базы на эмиттер существует подталкивающее электрическое поле, и электронный ток в этом направлении может достигать заметной величины. Это обстоятельство работает против нас, поскольку увеличивает полный ток через базу при заданном коллекторном токе, в результате чего уменьшается усиление. Поэтому транзистор конструируется так, чтобы уменьшить этот ток до минимума.

Все, что рассказано о p — n — p -транзисторе, в равной мере применимо к n — p — n -транзистору, если только переменить знаки потенциалов электродов.

В заключение нужно сказать, что применение полупроводников, конечно, не ограничивается только диодами и транзисторами. Существует очень много других устройств, работа которых основана на тех или иных свойствах полупроводников.