

Московский физико-технический институт  
Кафедра общей физики

Лекция 11

КОНТАКТНЫЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ.  
ПОСТРОЕНИЕ ЭНЕРГЕТИЧЕСКИХ ДИАГРАММ КОНТАКТОВ  
ПОЛУПРОВОДНИКОВ.

заметки к лекциям по общей физике

В.Н.Глазков

Москва

2021

В данном пособии представлены материалы к лекции по теме «Контактные явления в полупроводниках» из курса «Квантовая макрофизика», преподаваемого на кафедре общей физики МФТИ.

Пособие не претендует на полноту изложения материала и в основном является авторскими заметками к лекциям, оно содержит основные сведения по этой теме курса. Основная цель этой лекции проиллюстрировать метод построения энергетических диаграмм контактов полупроводников (изгиба зон).

Материал по этой теме содержится в главе 11 книги Киттеля [1] и в методичке А.И.Морозова «Физика твёрдого тела» [2]. Вопросы построения энергетических диаграмм гетеропереходов рассмотрены в учебнике Демиховского «Физика квантовых низкоразмерных структур» [3], пособии [4] и в Нобелевской лекции Г.Крёмера [5]. Обсуждение физики p-n перехода имеется в лабораторном практикуме [6]. В качестве подробного классического учебника по физике полупроводников можно рекомендовать учебник Бонч-Бруевича [7]

### **Основные понятия этой лекции:**

- построение схемы изгиба зон на границе контакта полупроводников
- энергетическая диаграмма p-n перехода, контакта металл-полупроводник, гетероструктуры
- возможность формирования потенциальных ям для квазичастиц на границе гетеропереходе

## Оглавление

Условия равновесия носителей заряда в неоднородной системе (напоминание).....	4
Энергетическая диаграмма полупроводника и положение уровня химического потенциала в полупроводнике (напоминание).....	4
Статические свойства р-п перехода.....	6
Перераспределение заряда в р-п переходе. Качественное рассмотрение.....	7
Условия равновесия и изгиб зон на р-п переходе.....	8
Постоянство электрохимического потенциала.....	8
«Изгиб зон» на р-п переходе.....	9
Анализ перераспределения зарядов на р-п переходе при $T=0$ .....	11
О протекании тока через р-п переход.....	12
Туннельный диод: отрицательное дифференциальное сопротивление.....	15
Контакт полупроводник-металл.....	17
Гетеропереход.....	19
Правило Андерсона для контакта двух полупроводников.....	19
Классификация гетеропереходов.....	20
Построение энергетической диаграммы гетероперехода AlAs-GaAs. Случай номинально чистых полупроводников. ....	21
Построение энергетической диаграммы гетероперехода AlAs-GaAs. Случай сильно легированного n-AlAs. ....	23
Полевой транзистор.....	25

## Список литературы

- 1: Ч.Киттель, Введение в физику твёрдого тела., ,
- 2: А.И.Морозов, Физика твёрдого тела, МИРЭА (есть на сайте кафедры), 2008
- 3: В.Я.Демиховский, Г.А.Вугальтер, Физика квантовых низкоразмерных структур, Москва, Логос, 2000
- 4: Шиляев П.А., Павлов Д.А., Полупроводниковые структуры: гетеропереход. Учебно-методическое пособие., Нижегородский государственный университет им. Н.И.Лобачевского, 2009
- 5: Herbert Kroemer, Quasi-Electric Fields and Band Offsets: Teaching Electrons New Tricks, Nobel Prize Lecture, (2000)
- 6: Ф.Ф.Игошин, Ю.А.Самарский, Ю.М.Ципенюк, Лабораторный практикум по физике: Квантовая физика., Физматкнига, 2012
- 7: В.Л.Бонч-Бруевич и С.Г.Калашников, Физика полупроводников, М.: Наука, 1990
- 8: H.J.Round, A note on carborundum, Electrical World, 19, 309(1907)
- 9: Leo Esaki, Long Journey into Tunnelling, Nobel Prize Lecture, (1973)
- 10: F.Braun, Ueber die Stromleitung durch Schwefelmetalle, Annalen der Physik und Chemie, 153, 556(1874)
- 11: Semiconductors on NSM , <http://www.ioffe.ru/SVA/NSM/Semicond/>
- 12: Основы физики полупроводниковых структур пониженной размерности , [http://ftn-mipt.itp.ac.ru/attachments/123\\_Dolgoplov.pdf](http://ftn-mipt.itp.ac.ru/attachments/123_Dolgoplov.pdf)
- 13: Э.В.Девятков, Основы физики низкоразмерных систем и режима квантового эффекта Холла., МФТИ, 2015

*Whenever I teach my semiconductor device course, one of the central messages I try to get across early is the importance of energy band diagrams. I often put this in the form of «Kroemer's Lemma of Proven Ignorance»:*

- If, in discussing a semiconductor problem you cannot draw an **Energy Band Diagram**, this shows that you don't know what you are talking about with the corollary.*
- If you can draw one, but don't, then your audience won't know what you are talking about.*

H.Kroemer, Nobel Prize Lecture, 2000 [5]

## **Условия равновесия носителей заряда в неоднородной системе (напоминание)**

Основная цель этой лекции — рассмотреть явления, возникающие на контакте полупроводника с другим материалом (металлом или другим полупроводником). Естественно, при этом возникает вопрос о перераспределении зарядов в такой неоднородной системе и об условиях равновесия.

Так как в целом образец является электронейтральным — заряд делокализованных электронов полностью компенсируется зарядом ионного остова кристалла, то при перераспределении зарядов между разными частями образца в образце возникнет электрическое поле. «Цена» добавления носителя заряда в некоторой точке образца равна значению химического потенциала в этой точке. Кроме этого, если в образце имеется электрическое поле, то необходимо учесть электростатическую потенциальную энергию заряда в этом поле.

В результате для электронов условием равновесия оказывается постоянство электрохимического потенциала по образцу:  $\mu(\vec{r}) - e\phi(\vec{r}) = \text{const}$ , знак «минус» здесь явно выражает отрицательность заряда электрона. Физический смысл этого равенства в том, что в равновесии при переносе частицы возможный выигрыш за счёт изменения химического потенциала  $\mu(\vec{r})$  в точности равен проигрышу в электростатической потенциальной энергии.

## **Энергетическая диаграмма полупроводника и положение уровня химического потенциала в полупроводнике (напоминание).**

При рассмотрении большинства вопросов этой лекции мы не будем рассматривать тонкие вопросы о структуре спектра полупроводника, количестве электронных и дырочных экстремумов в зонной структуре, о прямозонном или не прямозонном характере спектра. Для упрощения эффективную массу электронов и дырок будем считать изотропной. При рассмотрении примесей мы также будем полагать, что примесной уровень неглубок

(находится близко от потолка или дна соответствующей зоны) и единственен.

Для этих целей оказывается удобно и наглядно характеризовать полупроводник упрощённой энергетической диаграммой (рисунок 1), на которой указывается положение потолка валентной зоны, дна зоны проводимости, положение примесных уровней и положение уровня химического потенциала.

Нам будет также важно положение уровня минимальной энергии электрона в вакууме. Его расположение есть функция материала. Положение уровня минимальной энергии электрона в вакууме относительно электронных состояний полупроводника характеризуют величиной *сродства к электрону*  $\chi$ , по определению равной расстоянию от дна зоны проводимости до уровня минимальной энергии в вакууме. Эта величина обычно порядка электронвольта, как и работа выхода в металлах. По смыслу сродство электрона равно минимальной работе, необходимой на то, чтобы выбить электрон из зоны проводимости в вакуум — то есть именно эта величина будет определять красную границу внешнего фотоэффекта на полупроводнике при  $T \neq 0$ . Однако, необходимо отличать сродство электрона от *термодинамической работы выхода*, равной расстоянию от уровня химпотенциала до уровня минимальной энергии электрона в вакууме.

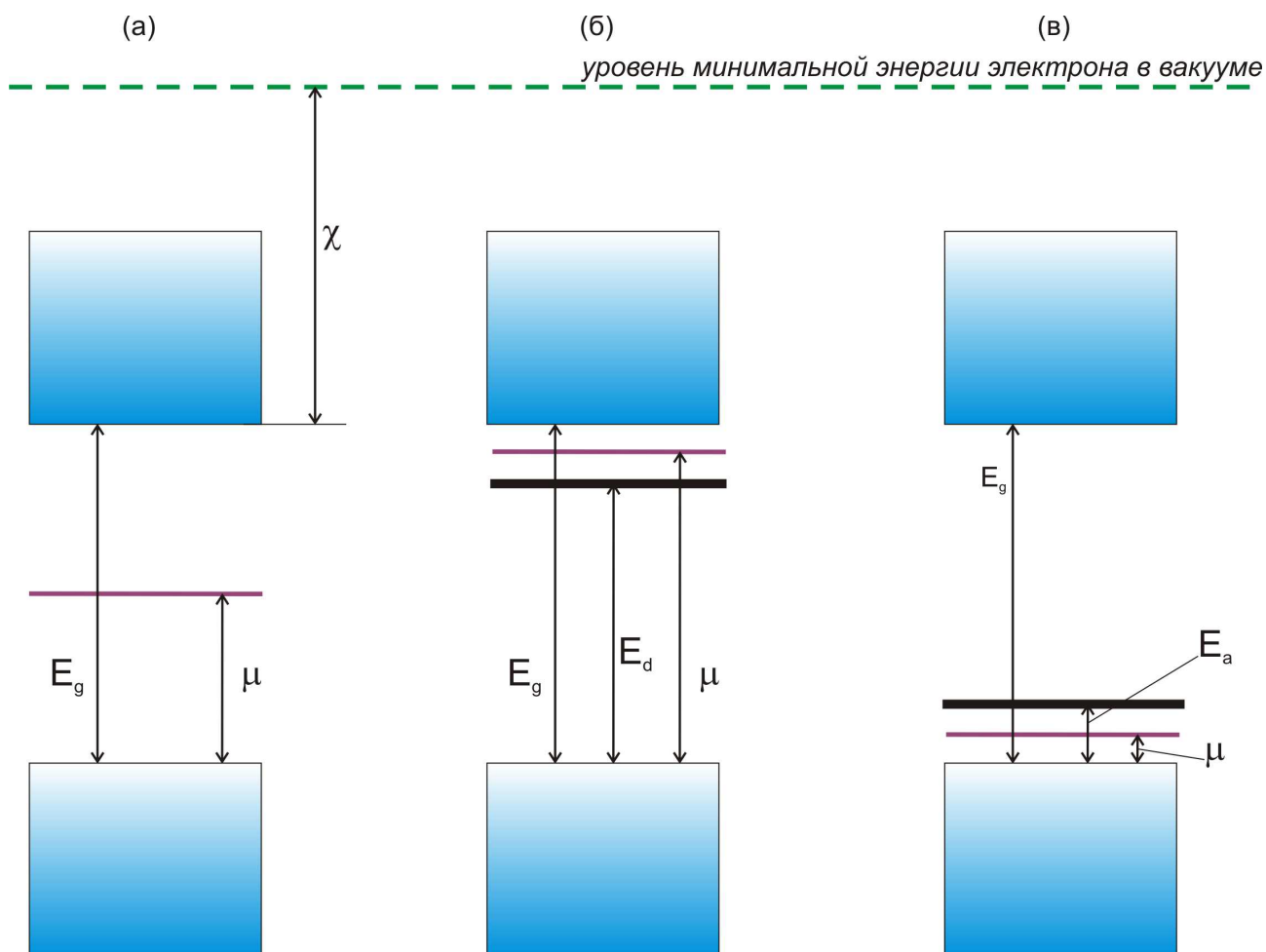


Рисунок 1: Упрощённое изображение зонной схемы полупроводника: (а) чистый полупроводник, (б) полупроводник с примесью донорного типа, (в) полупроводник с примесью акцепторного типа.  $E_g$  — ширина запрещённой зоны,  $\mu$  — уровень химпотенциала,  $E_{d,a}$  — уровни донорной и акцепторной примеси. Положение химпотенциала показано для случая  $T=0$ . Положение минимального уровня энергии электрона в вакууме показано зелёным пунктиром,  $\chi$  — сродство к электрону.

Напомним также, что при  $T=0$  уровень химпотенциала в чистом полупроводнике находится посередине запрещённой зоны, а в легированных полупроводниках с единственным примесным уровнем — посередине между примесным уровнем и ближайшей зоной. При изменении температуры уровень химпотенциала может смещаться, направление смещения зависит от параметров полупроводника (соотношения статфакторов зон, концентрации примесей).

## Статические свойства р-п перехода.

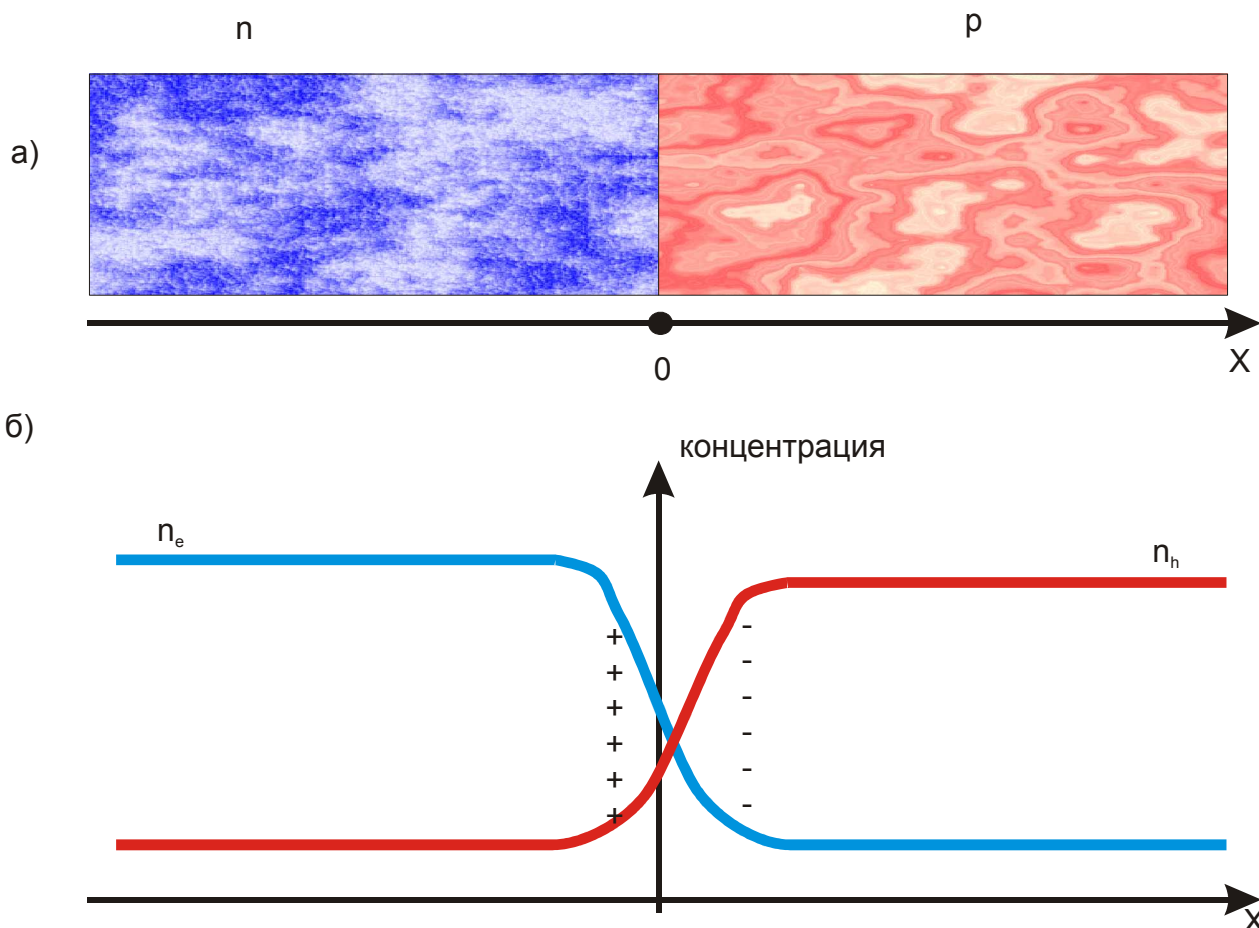


Рисунок 2: (а) контакт полупроводников р-типа и п-типа. (б) изменение концентрации электронов и дырок на р-п переходе, знаками схематически показаны области перераспределённого заряда.

Рассмотрим контакт легированных полупроводников р и п типа, созданных из одного «родительского» материала. Будем считать, что ширина запрещённой зоны при легировании не изменилась (случай контакта полупроводников с разной шириной запрещённой зоны будет рассмотрен далее). Для простоты рассуждений примем геометрию контакта плоской — пусть р и п области представляют собой брусочки одинакового сечения, соединённые торцами. Пусть также контакт будет атомно-гладким, то есть нет никакой переходной области между полупроводниками.<sup>1</sup>

Тогда задача становится одномерной и все переменные (концентрации электронов и дырок,

<sup>1</sup> В реальности р-п переходы получают, например, ионной имплантацией примесей нужного типа во время литографического процесса создания полупроводникового прибора. Поэтому реальная геометрия границы контакта полупроводников, вообще говоря, устроена не столь идеально. При этом, однако, качественные выводы не изменятся.

электростатический потенциал и уровень химического потенциала) зависят только от координаты  $X$  (рисунок 2). Принципиально задача об установлении равновесия в такой системе похожа на задачу о контактной разности потенциалов для металлов, некоторые отличия связаны с наличием запрещённой зоны и двух типов носителей заряда.

Рассмотрим основные физические явления, возникающие на p-n переходе.

### **Перераспределение заряда в p-n переходе. Качественное рассмотрение.**

В полупроводнике n-типа при конечной температуре число подвижных электронов больше числа подвижных дырок, а в полупроводнике p-типа число подвижных дырок больше числа подвижных электронов.<sup>2</sup> При приведении их в контакт начнётся диффузия зарядов из богатой ими области в бедную: электроны будут проникать в p-область, дырки — в n-область. В результате вблизи от перехода нарушится условие электронейтральности: в n-области стало (по сравнению с изолированным полупроводником n-типа в тех же условиях) больше дырок и меньше электронов, в p-области стало (по сравнению с изолированным полупроводником p-типа в тех же условиях) больше электронов и меньше дырок. Одновременно с этой диффузией квазичастиц из «богатой» в «бедную» области возникает задерживающая электроны разность потенциалов: n-область приобретает положительный заряд, p-область — отрицательный заряд. В результате электроны притягиваются обратно в n-область и установится некоторое равновесие.

Эта картина становится совсем простой при  $T \rightarrow 0$ , когда концентрация подвижных носителей стремится к нулю: при переходе носителя заряда в область другого типа он оставляет за собой «родительский» примесной ион. Соответственно, в n-области возникают полностью ионизованные (электроны ушли из n-полупроводника) положительно заряженные примесные ионы, а в p-области — полностью ионизованные (захватившие пришедшие электроны) отрицательно заряженные примесные ионы.

Таким образом, в приповерхностном слое p-области появляется избыточный отрицательный заряд, а в приповерхностном слое n-области избыточный положительный заряд (конечно же эти заряды равны по модулю, нарушения электронейтральности «в целом» не происходит). Этот двойной слой называют также *запорным слоем*. Электрическое поле этих зарядов препятствует дальнейшему перераспределению заряда и устанавливается равновесие (рисунок 2-б). При  $T=0$  эти избыточные заряды создаются электронами и дырками, локализованными на ионизованных примесях, при  $T \neq 0$  происходит размытие области ионизованных примесей.

В грубом приближении приповерхностные заряды в p-n переходе можно рассматривать как плоский конденсатор с распределённым зарядом (его ёмкость является одной из технических характеристик p-n перехода, она обычно составляет 1...100 пФ). Напряжение на этом «конденсаторе» (потенциал p-области относительно потенциала n-области отрицательный<sup>3</sup>) есть контактная разность потенциалов между полупроводниками. Характерный поперечный размер области перераспределения заряда составляет от 0.1-10 мкм в обычных переходах до 10 нм в сильнолегированных.

Напомним, что уже это качественное рассмотрение объясняет одностороннюю проводимость

2 Так как нас интересуют только делокализованные носители заряда, то, конечно, речь идёт об электронах в зонах проводимости и о дырках в валентной зоне. Полная электронейтральность образца обеспечивается тем, что «недостающие» дырки или электроны локализованы на соответствующей примеси.

3 Удерживающее электрическое поле направлено из n-области в p-область, так как  $\vec{E} = -\vec{\nabla} \varphi$  получаем, что электрический потенциал убывает от n-области к p-области.

p-n перехода. Если приложить положительный потенциал к p-области, то электрическое поле вызовет дрейф носителей заряда от p-n перехода, запирающее действие перераспределённых зарядов ослабнет и новые электроны и дырки перейдут через переход. Если приложенный потенциал превысит контактную разность потенциалов (напряжение «открытия» p-n перехода), то станет энергетически выгоден переход электронов через контакт. То есть в цепи потечёт ток. При обратной полярности подключения, наоборот, приложенное электрическое поле только усиливает запирающее действие перераспределённых зарядов и тока не будет (точнее, ток будет гораздо меньше из-за малости концентрации неосновных носителей заряда).

В условиях, когда p-n переход открыт и по нему течёт ток, ток вдали от контакта в p-области переносится в основном дырками, движущимися (по рисунку 2) справа налево, а в n-области — в основном электронами, движущимися слева направо. При этом в области перехода происходит рекомбинация электронов и дырок. Как мы уже обсуждали при обсуждении экситонов, рекомбинация может сопровождаться излучением света — область p-n перехода может стать источником фотонов. Это практически приводит нас к созданию светодиода — необходимо только создать условия, чтобы интенсивность излучения превзошла интенсивность поглощения, то есть создать инверсную заселенность возбужденных уровней. Цвет испускаемого излучения определяется шириной запрещённой зоны.<sup>4</sup>

## **Условия равновесия и изгиб зон на p-n переходе.**

### **Постоянство электрохимического потенциала.**

Как уже обсуждалось качественно, при соединении полупроводников p-типа и n-типа произойдёт перераспределение зарядов до достижения нового положения равновесия. Условием термодинамического равновесия в присутствии электрического поля является постоянство по системе электрохимического потенциала<sup>5</sup>, который для электронов равен  $\mu - e\phi$  (знак минус связан с отрицательностью заряда электрона). Различие в положении уровня химпотенциала в соединяемых полупроводниках будет компенсироваться электростатическим потенциалом, возникающим при перераспределении зарядов.

Будем считать, что электростатический потенциал меняется достаточно медленно (область перераспределённых зарядов гораздо шире межатомного расстояния). Тогда химпотенциал является локальной функцией, описывает локальное равновесие между электронами и дырками и плавно меняется. Это условие с запасом выполняется в полупроводниках с невысоким уровнем легирования.

Невырожденность электронного газа в полупроводнике позволяет переписать условие постоянства электрохимического потенциала в более удобной для этой задачи форме. Так как концентрация электронов на дне зоны проводимости  $n_e = Q_e e^{-(E_g - \mu)/T}$ , то  $\mu = T \ln n_e + const$  (здесь константа вообще говоря зависит от температуры). Поэтому в полупроводнике, находящемся при одной температуре  $T \ln n_e - e\phi = const$ . Аналогично

4 Историческим курьёзом является то, что возникновение свечения на диодном контакте было впервые замечено ещё до появления диодов — в 1907 году британский исследователь Г.Раунд, работавший в лаборатории Г.Маркони, заметил, что в контакте металла и карборунда (SiC) возникает свечение [8]. Как будет обсуждаться далее, контакт металл-полупроводник обладает диодными свойствами и это активно использовалось в ранних детекторах радиоволн (англ. *cat's whisker detector*).

5 Работа по добавлению к системе частицы (что есть термодинамическое определение химпотенциала) изменяется на величину электростатической потенциальной энергии. В равновесии эта работа должна быть постоянна по всей системе. Однако локальные равновесные концентрации частиц в различных состояниях по-прежнему определяются локальным химпотенциалом.



можно показать, что для дырок  $T \ln n_h + e\phi = \text{const}$ . Из этих равенств в частности следует, что отношение равновесных концентраций носителей вдали от контакта двух полупроводников  $T \ln \frac{n_e^{(1)}}{n_e^{(2)}} = T \ln \frac{n_h^{(2)}}{n_h^{(1)}} = e \Delta \phi_{12} = e(\phi_1 - \phi_2)$ , где  $\Delta \phi_{12}$  установившаяся контактная разность потенциалов.

### «Изгиб зон» на p-n переходе.

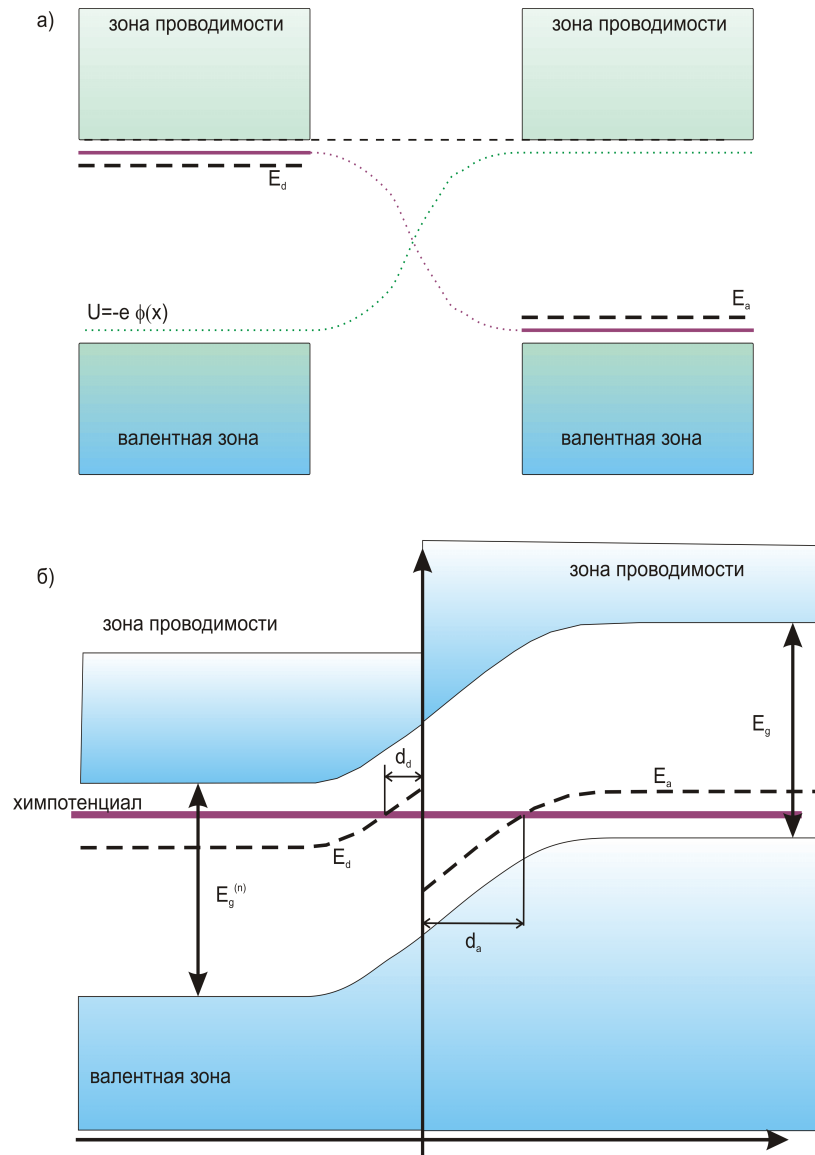


Рисунок 3: Энергетическая диаграмма p-n перехода для случая контакта полупроводника одного типа с различным легированием. Слева n-область, справа p-область, как и на рисунке 2. (а) Образцы находятся достаточно далеко друг от друга, зоны выравнены по общему уровню энергии электрона в вакууме, уровни химпотенциала соединены схематически, также схематически показана зависимость электростатического потенциала от координаты. (б) Изгиб зон в представлении энергетической диаграммы p-n перехода для полной энергии.

На большом удалении от перехода концентрация зарядов равновесная и электростатический

потенциал перераспределённых зарядов не зависит от координаты: в n-области уровень химического потенциала находится вблизи дна зоны проводимости, в p-области — вблизи потолка валентной зоны. При соединении полупроводников электрохимический потенциал  $\mu - e\phi = \text{const}$  должен быть постоянен. Это означает, так как  $\mu_n > \mu_p$ , что на большом удалении от контакта электростатический потенциал  $\phi$  n-области положителен относительно потенциала p-области. Такой знак напряжения на переходе соответствует нашим качественным рассуждениям выше. Так как электростатический потенциал непрерывен, то из условия  $\mu - e\phi = \text{const}$  также следует непрерывность изменения химического потенциала  $\mu$ : в окрестности перехода уровень химического потенциала должен плавно измениться от положения вблизи дна зоны проводимости в n-области к положению вблизи потолка валентной зоны в p-области (рисунок 3-а).

Удобно оказывается перестроить эту схему. Заметим, что в энергетической диаграмме учтена только *кинетическая* энергия электронов (с учётом каких-то перенормировок за счёт взаимодействия с решёткой, но без учёта электрического поля зарядов на p-n переходе). При перераспределении зарядов к этой энергии добавляется *потенциальная* энергия электрона в электрическом поле  $U = -e\phi$ . Тогда имеет смысл построить энергетическую диаграмму для полной энергии, «добавив» потенциальную энергию к уже отражённой в нашем рисунке зонной структуре. В силу условия  $\mu - e\phi = \text{const}$  это построение приобретёт простой графический смысл: уровень химического потенциала станет постоянен по рисунку (теперь это электрохимический потенциал), а так как ширина запрещённой зоны и расстояние от уровня химического потенциала до потолка или дна зоны или до примесного уровня при этом перестроении не меняются<sup>6</sup>, то возникает так называемый изгиб зон (рисунок 3-б). По нашему рисунку изгиб зон — это подъём потолка валентной зоны и отстоящих от него на постоянную (в пределах своей области) величину дна зоны проводимости и положения примесного уровня при движении слева направо. «Изгиб зон» возникает на масштабах длины, много больших межатомного расстояния.

---

6 Если считать контакт между материалами идеально резким и атомно-гладким, то ширина запрещённой зоны и положение примесного уровня это константы материала, от перераспределения заряда не зависящие, и скачком изменяются на границе.

## Анализ перераспределения зарядов на p-n переходе при $T=0$ .

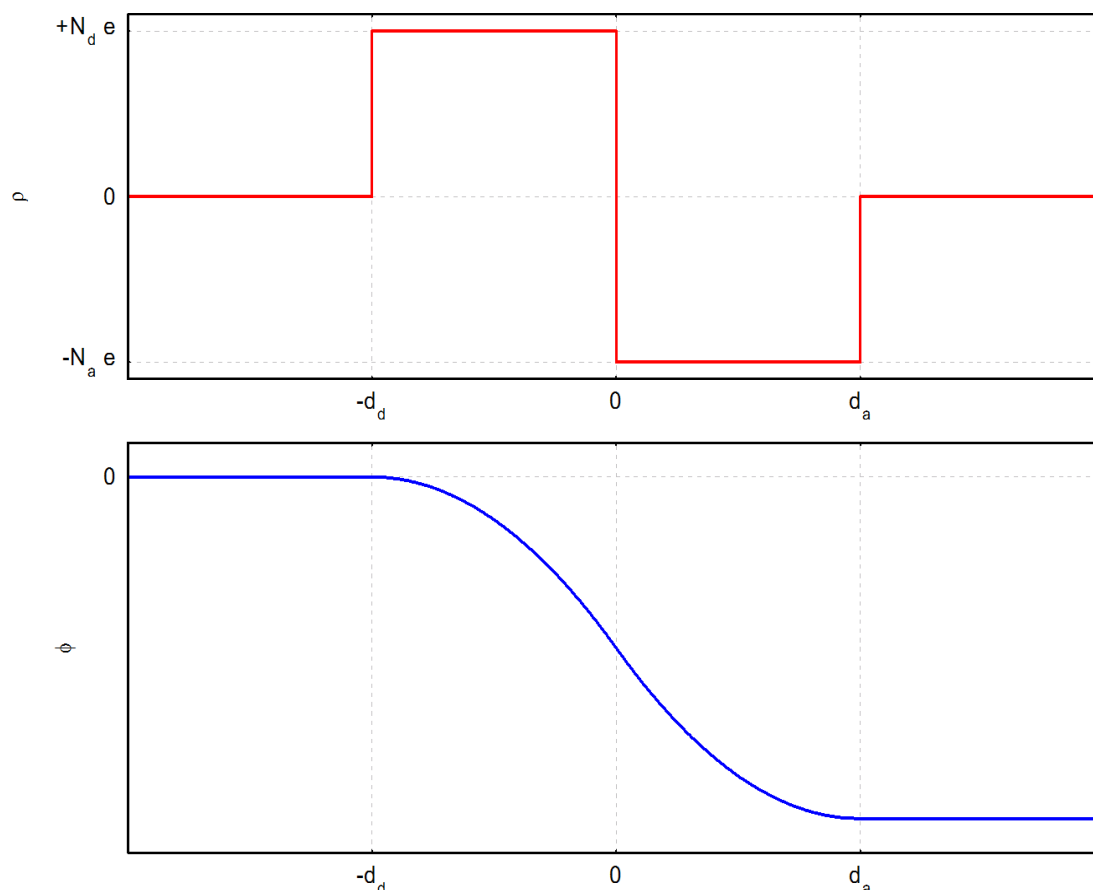


Рисунок 4: Сверху: распределение плотности зарядов ионизованных примесей на p-n переходе при  $T=0$ . Снизу: зависимость электростатического потенциала от координаты.

Обратите внимание, что уровень химического потенциала на рисунке 3 пересекает примесные уровни донорной и акцепторной примеси. По определению химического потенциала, при  $T=0$  все электронные состояния ниже уровня химического потенциала заняты, а все электронные состояния выше уровня химического потенциала свободны. Таким образом, в этой энергетической диаграмме уже содержится информация о перераспределении зарядов в окрестности p-n перехода.

Рассмотрим подробнее, как устроено перераспределение зарядов на p-n переходе. Для простоты рассмотрим случай  $T=0$ .<sup>7</sup> Свободных носителей заряда при этом нет, однако из-за изгиба зон на некоторой глубине  $d_d$  в n-области донорные уровни оказались выше уровня химического потенциала, а на некоторой глубине  $d_a$  в p-области акцепторные уровни оказались ниже уровня химического потенциала (рисунок 3-б). В этих областях ионы окажутся полностью ионизованы (рисунок 4). Если  $N_d$  и  $N_a$  концентрации донорной и акцепторной примеси (считаем что эти примеси дают или забирают один электрон), то для электростатического потенциала имеем в n-области (положительные ионы)

<sup>7</sup> Точнее, считая температуру очень низкой. Мы пренебрегаем в нашем рассуждении появлением делокализованных носителей, но в то же время небольшая конечная температура нам нужна, чтобы обеспечить механизм отрыва носителей от родительской примеси для перехода в более энергетически выгодное состояние по другую сторону p-n перехода.

$$\epsilon \frac{d^2 \varphi}{dx^2} = -4\pi e N_d \quad \text{и в } p\text{-области (отрицательные ионы)} \quad \epsilon \frac{d^2 \varphi}{dx^2} = 4\pi e N_a, \quad \text{где } \epsilon -$$

диэлектрическая проницаемость среды (типично для полупроводников  $\epsilon \simeq 10$ ). Решение этих уравнений имеет внутри слоя ионизованных примесей вид переходящих друг в друга ветвей парабол (рисунок 4):

$$\varphi(x) = \begin{cases} 0, & x < -d_d \\ -2\pi e N_d (x + d_d)^2 / \epsilon, & -d_d < x < 0 \\ \left( 2\pi e N_a (x - d_a)^2 - 2\pi e N_a d_a^2 - 2\pi e N_d d_d^2 \right) / \epsilon, & 0 < x < d_a \\ \left( -2\pi e N_a d_a^2 - 2\pi e N_d d_d^2 \right) / \epsilon, & x > d_a \end{cases}.$$

Для нахождения ширины ионизованных слоёв из условия электронейтральности имеем  $d_a N_a = d_d N_d$ , а из условия  $\mu - e\varphi = \text{const}$  полная разность потенциалов (контактная разность потенциалов)  $\Delta\varphi = \varphi_d - \varphi_a = \frac{1}{e}(\mu_d - \mu_a) = \frac{E_g + E_d - E_a}{2e} \approx \frac{E_g}{e}$  задана свойствами материала.

Отсюда:

$$\begin{aligned} N_a d_a^2 + N_d d_d^2 &= \epsilon \frac{\Delta\varphi}{2\pi e} \\ d_a^2 &= \epsilon \frac{\Delta\varphi}{2\pi e} \frac{N_d}{N_a(N_a + N_d)} \\ d_d^2 &= \epsilon \frac{\Delta\varphi}{2\pi e} \frac{N_a}{N_d(N_a + N_d)} \end{aligned}$$

Потенциал границы раздела:

$$\varphi(0) = -\frac{N_a}{N_a + N_d} \Delta\varphi,$$

а ёмкость на единицу площади перехода:

$$C = \frac{N_a d_a}{\Delta\varphi} = \sqrt{\frac{\epsilon}{2\pi e \Delta\varphi}} \sqrt{\frac{N_a N_d}{N_a + N_d}} \approx \sqrt{\frac{\epsilon}{2\pi E_g}} \sqrt{\frac{N_a N_d}{N_a + N_d}}.$$

Обратите внимание, что при равной концентрации примесей  $N_a = N_d$  и  $\varphi(0) = -\frac{\Delta\varphi}{2}$ , а при разном уровне допирования слой ионизованных примесей шире в области с меньшей концентрацией примеси и большая часть контактной разности потенциалов «набегает» в области с меньшей концентрацией примеси. Ширина запирающего слоя оказывается обратно пропорциональна квадратному корню из концентрации примеси.

## О протекании тока через р-п переход.

В рамках лекционного курса мы не будем подробно анализировать вопросы проводимости р-п перехода. Подробное рассмотрение этого вопроса есть в «Лабораторном практикуме по общей физике» [6]. Ограничимся здесь лишь указанием на то, что в энергетической диаграмме для полной энергии (с постоянным уровнем химпотенциала) также наглядно содержится информация о выпрямляющих свойствах р-п перехода (рисунок 5).

Для того, чтобы перестроить энергетическую диаграмму в присутствие источника

напряжения, заметим сразу, что эта ситуация неравновесная — источник всё время совершает работу для поддержания заданной разности потенциалов и по переходу течёт ток. Поэтому применение условия равновесия  $\mu - e\phi = \text{const}$  не имеет смысла. Приложение напряжения к переходу приводит к двум эффектам: возникает ток носителей, поддерживаемый источником, и изменяется картина перераспределения зарядов на границе. Эти явления можно разделить при рассмотрении: перераспределение зарядов создаёт некоторый рельеф потенциала внутри образца, а ток носителей происходит по этому рельефу, под действием вынуждающей силы внешнего источника.

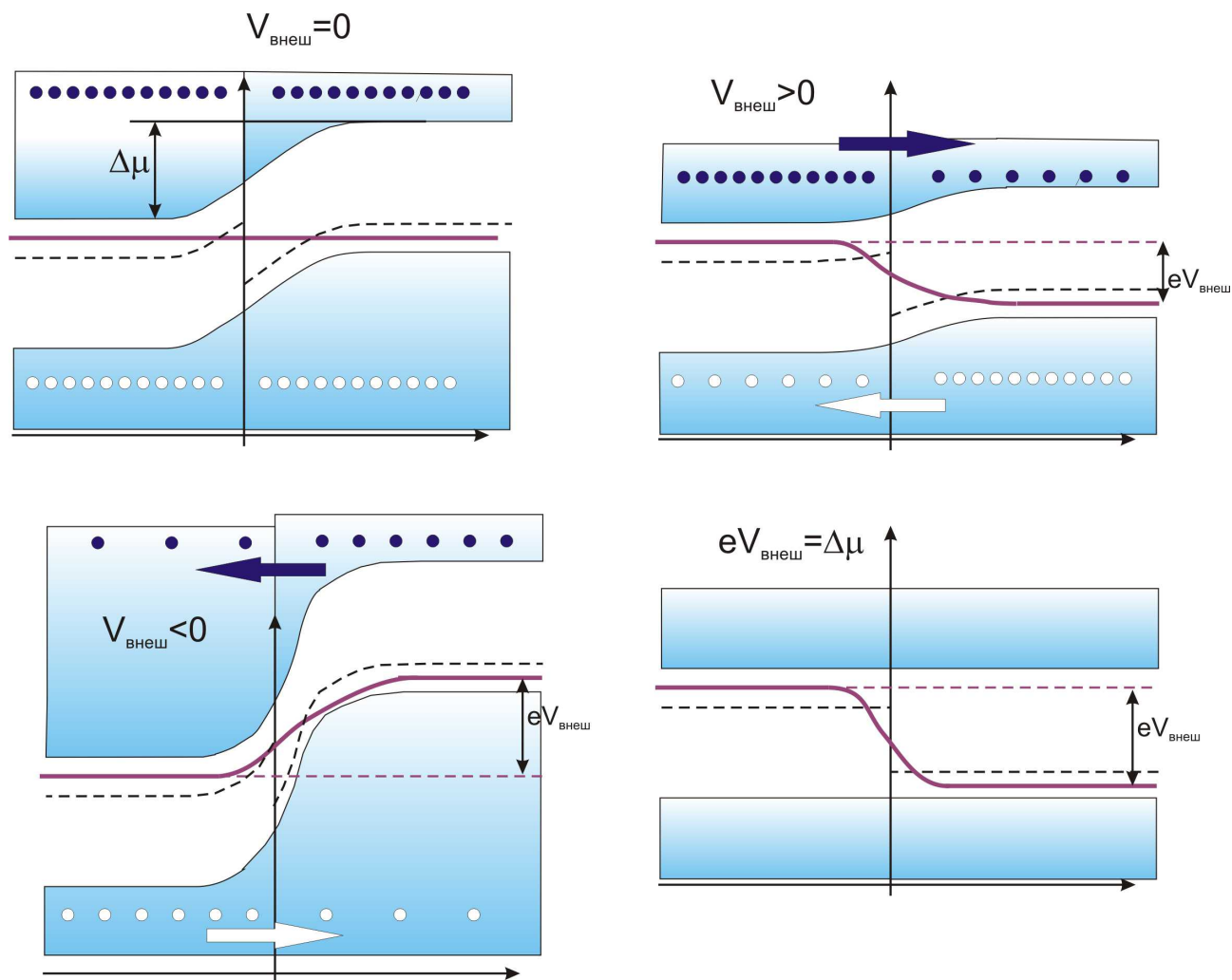


Рисунок 5 Построение схемы энергетической диаграммы для p-n перехода с приложенным внешним напряжением. Открытые и закрытые кружки условно показывают концентрации дырок и электронов на одном уровне полной энергии в разных случаях.

Приложение к переходу напряжения  $V_{\text{внеш}}$  (для определённости считаем, что потенциал p-области поддерживается равным нулю и приложенное напряжение определяет потенциал n-области) от внешнего источника приводит к тому, что теперь при переносе электронов между приводимыми в контакт полупроводниками мы должны учитывать, что дно зоны проводимости в n-области дополнительно сместилось относительно уровня минимальной энергии электрона в вакууме на  $-eV_{\text{внеш}}$ . На столько же сместится и положение химического потенциала, примесного уровня и потолка валентной зоны, являющихся функцией материала.

Это означает, что на языке энергетической диаграммы для полной энергии (с постоянным

уровнем химпотенциала в отсутствие приложенного напряжения) мы должны для установления нового статического рельефа потенциала сместить на  $-eV_{\text{внеш}}$  уровень химпотенциала в р-области вдали от перехода и достроить по непрерывности его зависимость от координаты в переходной области. Точный расчёт выходит за рамки курса. Понятно, что непрерывно (и с одинаковой зависимостью от координаты) будут меняться и положения границ зон, и положения примесных уровней.

В отсутствие внешнего напряжения концентрация электронов на одном уровне нашей энергетической диаграммы одинакова в обоих полупроводниках: она определяется только расстоянием до этого уровня от спрямлённого в этом представлении уровня химпотенциала (схематически показано количеством кружков на рисунке 5). То же самое верно для дырок. Поэтому, естественно, ток не течёт (нет перепада концентраций, токи слева-направо и справа-налево точно компенсируют друг друга). При приложении напряжения уровень химического потенциала сдвигается и (для  $V_{\text{внеш}} > 0$ , для определённости) на том же уровне энергии в полупроводнике р-типа меньше электронов, чем в полупроводнике n-типа (и, соответственно, больше дырок). Это различие концентраций вызывает электронный ток слева направо по нашему рисунку и дырочный ток справа налево. При отрицательной полярности, очевидно, сменится направление тока.

Рассуждение выше было сделано в предположении, что р-п контакт создан полупроводниками с одинаковыми статфакторами зон. Может показаться, что для «несимметричного» перехода вывод мог бы оказаться другим. Однако, полученный вывод более общий: более строгое рассмотрение похоже на рассмотрение туннельного тока в различных структурах, проводившееся на предыдущей лекции, когда мы получали, что туннельный ток пропорционален произведению плотностей состояний. На контакте двух образцов мы всегда должны сравнивать токи электронов «слева-направо» и «справа-налево». Такой диффузный ток электронов с энергией  $\epsilon$  в одном направлении пропорционален числу электронов с данной энергией, способных перейти через границу, но, в то же время, и числу незанятых состояний с этой энергией на втором контакте:

$$I_{1 \rightarrow 2}(\epsilon) \propto N_1(\epsilon) n(\mu_1, \epsilon) \times N_2(\epsilon) (1 - n(\mu_2, \epsilon)),$$

здесь  $N_{1,2}(\epsilon)$  - плотности состояний в материале 1 и 2, соответственно, а  $n(\mu, \epsilon) = \frac{1}{e^{(\epsilon - \mu)/T} + 1}$  - фермиевская функция распределения.

Полный ток равен разности диффузных токов:

$$I \propto N_1(\epsilon) N_2(\epsilon) (n(\mu_1, \epsilon) (1 - n(\mu_2, \epsilon)) - n(\mu_2, \epsilon) (1 - n(\mu_1, \epsilon))) = N_1(\epsilon) N_2(\epsilon) (n(\mu_1, \epsilon) - n(\mu_2, \epsilon)).$$

В отсутствие внешнего напряжения на обоих контактах совпадает расстояние от уровня химпотенциала до уровня энергии электронов, переходящих через контакт. Поэтому тока, естественно, нет. При приложении внешнего напряжения смещение уровней химпотенциала в контактах нарушит это равенство и возникнет ток через контакт.

При небольшом напряжении между полупроводниками при  $V_{\text{внеш}} > 0$  могут переходить только электроны, энергия которых выше уровня дна зоны проводимости в полупроводнике р-типа, и дырки, энергия которых на диаграмме ниже уровня потолка валентной зоны в полупроводнике n-типа. Так как в основном основные носители заряда сосредоточены вблизи экстремума своих зон, таких носителей при небольшой температуре мало (напомним, что характерная контактная разность потенциалов порядка 1В, а 1эВ эквивалентен 11600К). Поэтому при малых приложенных напряжениях ток через р-п переход мал.

Картина радикально меняется, если  $eV_{\text{внеш}} \geq \Delta\mu$ . В этом случае все основные носители

могут переходить через переход. Переход открывается и ток резко возрастает.

При обратной полярности подключения переход всегда заперт для основных носителей заряда: только малая часть высокоэнергетичных электронов из зоны проводимости полупроводника n-типа может попасть в зону проводимости полупроводника p-типа. Однако при такой полярности возникает интересная возможность туннельного перехода «настоящего» электрона из валентной зоны полупроводника p-типа в зону проводимости полупроводника n-типа. Такой туннельный переход на языке квазичастиц соответствует рождению пары электрон-дырка на p-n переходе, являясь аналогом «пробоя вакуума». Для возникновения такого эффекта необходимо добиться узости промежутка, в котором происходит изгиб зон, что требует применения сильно легированных полупроводников.

Наконец, заметим, что выпрямляющее действие p-n перехода оказывается однозначным следствием изгиба зон. Изгиб зон возникает вследствие различия положения химпотенциала в соединяемых полупроводниках. Следовательно, при повышении температуры, когда становится главным вклад собственных носителей заряда и химпотенциал смещается к центру запрещённой зоны, выпрямляющее действие p-n перехода пропадёт. Грубой оценкой температуры, при которой это произойдёт, является совпадение концентрации примеси с концентрацией термоактивированных собственных электронов в зоне проводимости (или собственных дырок в валентной зоне).

## Туннельный диод: отрицательное дифференциальное сопротивление.

Вкратце остановимся на особенностях вольт-амперной характеристики туннельного диода. За экспериментальное изучение туннелирования носителей заряда в полупроводниковом контакте Л.Есаки была присуждена Нобелевская премия по физике 1973 года [9]. Отличием туннельного диода от рассмотренного выше p-n перехода является сильное (в работах Л.Есаки использовались полупроводники с концентрацией примеси на уровне  $10^{18} \dots 10^{19} \text{ 1/см}^3$ ) легирование обоих контактов. В результате ширина запертого слоя уменьшается до 10 нм, а уровень химпотенциала смещается в зону проводимости в полупроводнике n-типа и в валентную зону в полупроводнике p-типа. Фактически, допирование оказывается настолько сильным, что полупроводник n-типа становится «почти» металлом с электронной проводимостью, а полупроводник p-типа - «почти» металлом с дырочной проводимостью.

Энергетическая диаграмма такого диода строится аналогично рассмотренному случаю (рисунок 6). При приложении прямого напряжения в отличие от обычного диода возникает возможность протекания тока при значениях напряжения меньших напряжения открытия диода: может течь туннельный ток электронов из зоны проводимости сильно допированного n-полупроводника в свободные состояния валентной зоны сильно допированного p-полупроводника (рисунок 6-b). При увеличении напряжения так, что заполненные состояния в зоне проводимости полупроводника n-типа уже не пересекаются со свободными состояниями в валентной зоне полупроводника p-типа, но и ещё не пересекаются со свободной зоной проводимости полупроводника p-типа, ток должен пропасть (рисунок 6-с). При дальнейшем увеличении напряжения диод откроется и будет вести себя аналогично обычному диоду. Если для простоты положить, что ширина полосы занятых состояний в зоне проводимости полупроводника n-типа совпадает с шириной полосы свободных состояний в валентной зоне полупроводника p-типа  $E_{fn} = E_{fp}$ , то туннельный ток будет максимален при определённом значении приложенного напряжения, когда эти полосы точно совпадают. При дальнейшем увеличении напряжения перекрытие начнёт уменьшаться и туннельный ток

начнёт спадать. Таким образом, в области малых напряжений (меньших напряжения открытия «нормального» диода) возникнет участок на котором  $\frac{dI}{dU} < 0$  (рисунок 6). При обратной полярности включения при малых напряжениях ток через диод в основном также туннельный.

Отрицательное дифференциальное сопротивление приводит к неустойчивости электрической цепи, в которой такой элемент находится, и может использоваться для создания генераторов.

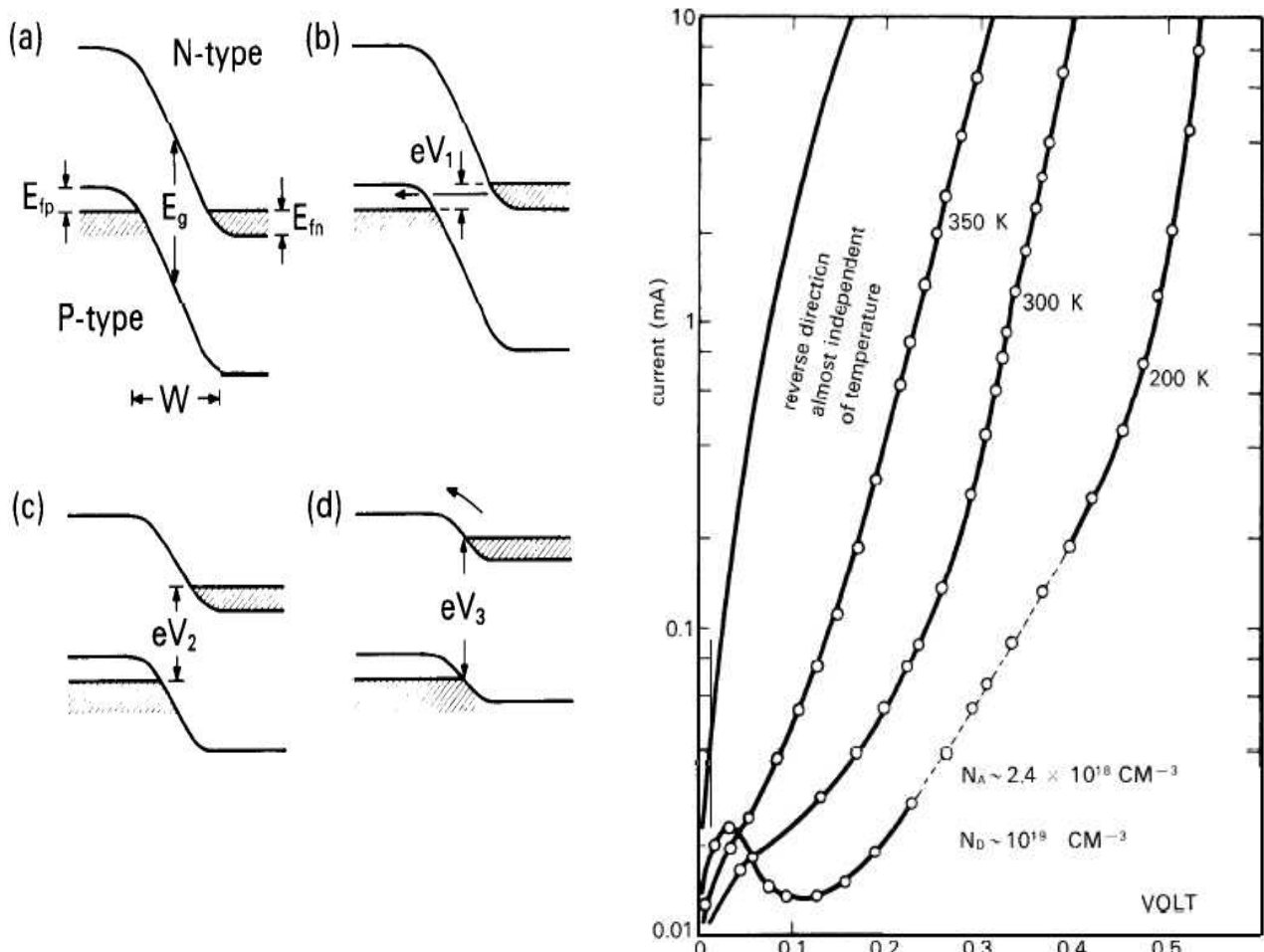


Рисунок 6: (слева) Энергетические диаграммы туннельного диода при различных значениях прямого напряжения. (справа) Пример вольт-амперной характеристики туннельного диода, снятой при различных температурах. Из нобелевской лекции Л.Есаки [9].



## Контакт полупроводник-металл.

Для того, чтобы проиллюстрировать построение энергетических диаграмм, рассмотрим также контакт полупроводник-металл. Для определённости рассмотрим полупроводник n-типа.

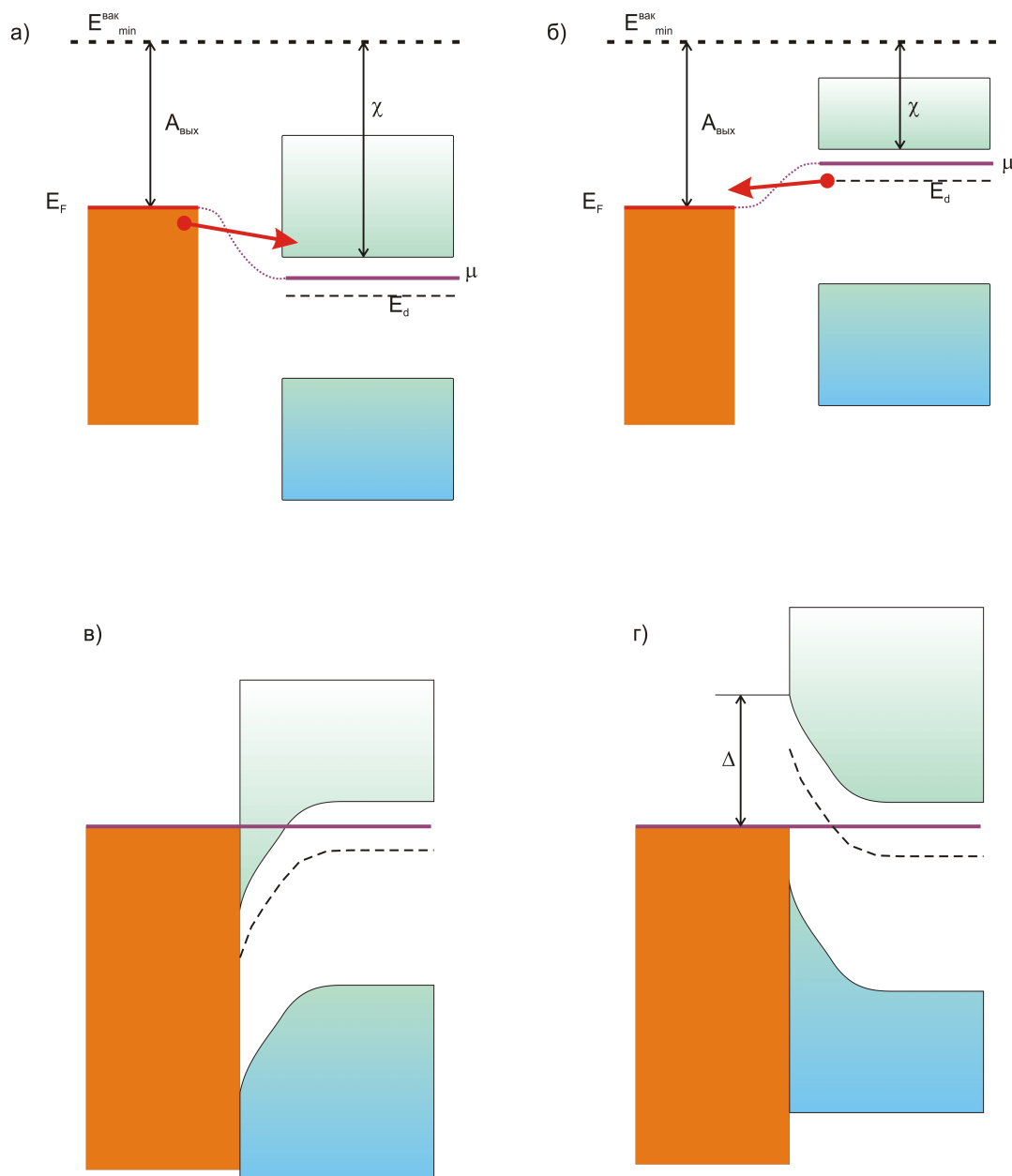


Рисунок 7: Схема энергетической диаграммы контакта полупроводник-металл. Верхний ряд (а, б): построение с постоянным уровнем энергии электрона в вакууме. Нижний ряд (в, г): построение с постоянным уровнем химпотенциала. Слева (а, в): случай, когда энергия Ферми металла попадает в зону проводимости полупроводника n-типа. Справа (б, г): случай, когда энергия Ферми металла попадает в запрещённую зону полупроводника n-типа.

При контакте разнородных материалов возникает вопрос о выборе общего начала отсчёта энергии. Правило Шоттки-Мота (аналог рассматриваемого далее правила Андерсона для полупроводников с той же аргументацией и применимое только на качественном уровне, так

как не учитывает деталей поведения электронов вблизи поверхности контакта<sup>8)</sup> подсказывает нам, что нужно принять для соединяемых материалов общий уровень энергии электрона в вакууме и отложить от него вниз работу выхода для металла (получим уровень поверхности Ферми) и сродство к электрону для полупроводника (получим дно зоны проводимости).

В этом упрощённом подходе возможны два принципиально различных случая (рисунок 7): энергия Ферми металла попадает в запрещённую зону полупроводника или не попадает (попадает в зону проводимости или валентную зону). В первом случае (рисунок 7-б) в металл перейдёт часть электронов с донорного уровня (мы рассматриваем полупроводник n-типа) и появится приповерхностный заряд положительно ионизованных примесей в полупроводнике и отрицательных избыточных электронов в металле. Во втором случае (рисунок 7-а) произойдёт или перетекание электронов из металла в зону проводимости или из валентной зоны в металл и также возникнет приповерхностный заряд.

Однако между этими двумя случаями есть принципиальное отличие, становящееся очевидным в представлении энергетической диаграммы с постоянным уровнем химпотенциала (рисунок 7-в, г). В обоих случаях возникает изгиб зон вблизи контакта. Но для случая, когда энергия Ферми металла попадает в запрещённую зону изгиб зон таков, что на границе возникает барьер для электронов (рисунок 7-г), высота этого барьера  $\Delta$  может измеряться долями электронвольта — то есть быть много больше тепловой энергии, ширина барьера определяется аналогично p-n переходу толщиной слоя ионизованных дефектов. Если пренебречь туннелированием под барьером, то ток через такой контакт будет течь только при приложении к полупроводнику напряжения определённой полярности (положительной по нашему рисунку) и величиной большей  $U \simeq \Delta / e$ . Конечная температура, конечная вероятность туннелирования сделают вольт-амперную характеристику более плавной, но она всё равно останется нелинейной и будет обладать выпрямляющими свойствами<sup>9</sup> p-n перехода. Возникающий барьер называют барьером Шоттки, а работающий на этом принципе диод — диодом Шоттки. В случае же попадания энергии Ферми в зону проводимости (или валентную зону) контакт получается безбарьерный, заметной нелинейностью вольт-амперной характеристики не обладает и называется омическим контактом. Получение омических контактов необходимо при подключении полупроводниковых приборов к внешним цепям, но, как видно из наших рассуждений, получение такого контакта требует специального согласования зонных структур металла и полупроводника и является часто отдельной технологической задачей.

---

8 Во многих случаях на границе с металлом формируются дополнительные состояния внутри примесной зоны «подтягивающие» уровень химпотенциала в полупроводнике вблизи контакта на центр запрещённой зоны. Поэтому, в частности, нетривиальной технологической проблемой является формирование чисто омического (безбарьерного) контакта полупроводник-металл.

9 Выпрямляющие свойства контакта металл-полупроводник известны с конца 19 века. Одним из изобретателей этого типа диодов был Фердинанд Браун [10], нобелевский лауреат 1909 года (совместно с Маркони, за вклад в развитие беспроводной телеграфии). Привычные нам полупроводниковые диоды с p-n переходом начали появляться только в 30-х годах 20 века. См. также сноску на стр.8.

## Гетеропереход.

Пока мы рассматривали контакт двух полупроводников с одинаковой зонной структурой, но с разным легированием. Что произойдёт, если соединить разнородные полупроводники?

Такие системы, называемые гетероструктурами<sup>10</sup> активно применяются в технике (например, полупроводниковые лазеры строятся на базе гетероструктур). Кроме того, в физике низкоразмерных электронных систем гетероструктуры и созданные на их основе устройства являются одним из способов создания низкоразмерных электронных систем.

### ***Правило Андерсона для контакта двух полупроводников.***

При соединении полупроводников разного типа (с различным химическим составом и зонной структурой) возникает вопрос о выборе одного начала отсчёта энергии в нашей задаче — до этого мы рассматривали полупроводники, отличающиеся только легированием, в которых зонные структуры «собственных» электронов совпадали и можно было выбрать начало отсчёта энергии произвольно (лишь бы одинаково в обоих полупроводниках). Точное решение этой задачи требует полного расчёта или знания зонной структуры и её заполнения с учётом перераспределения заряда на границе. То есть, требуется совместное решение квантового уравнения Шредингера и электростатического уравнения Пуассона. Эта задача слишком сложна для даже качественного анализа в нашем курсе.

Практически удобным оказывается правило Андерсона<sup>11</sup>: нужно отсчитать в сторону уменьшения от нулевой энергии свободного электрона в вакууме интервалы энергии, равные средству соединяемых материалов к электрону, эти уровни дадут положения дна зоны проводимости. Правило Андерсона (и родственное ему правило Шоттки-Мота) рассматривает два близко находящихся, но не соединённых полупроводника (полупроводник и металл в правиле Шоттки-Мота). Его смысл в том, что переход электронов с одного полупроводника на другой будет продолжаться, пока можно выиграть энергию, затратив работу выхода на выемку электрона из одного образца в вакуум, но выиграв большую работу выхода, помещая этот электрон в другой образец. Это подразумевает ненулевую температуру, так как работа выхода определяется здесь как выемка электрона из зоны проводимости. Наличие общего относительно обоих образцов уровня энергии в вакууме позволяет сформулировать это правило. В результате, на соединяемых образцах появится некоторый заряд, не локализованный вблизи перехода, и возникнет разность потенциалов между образцами (в образцах появляются свободные заряды и потенциал на образце становится постоянным). При образовании р-п перехода (при соединении образцов) предположения лежащие в основе правила Андерсона нарушаются и оно становится верно лишь приближённо. Кроме того, применение правила Андерсона к находящимся в контакте образцам предсказывает нефизический скачок потенциала на границе. Тем не менее, оно оказывается простым и удобным при качественном построении зонной структуры контакта полупроводников.

---

<sup>10</sup> За создание полупроводниковых гетероструктур была присуждена Нобелевская премия 2000 года Ж.Алферову, Х.Кромеру и Д.Килби.

<sup>11</sup> R.L.Anderson, 1960

## Классификация гетеропереходов.

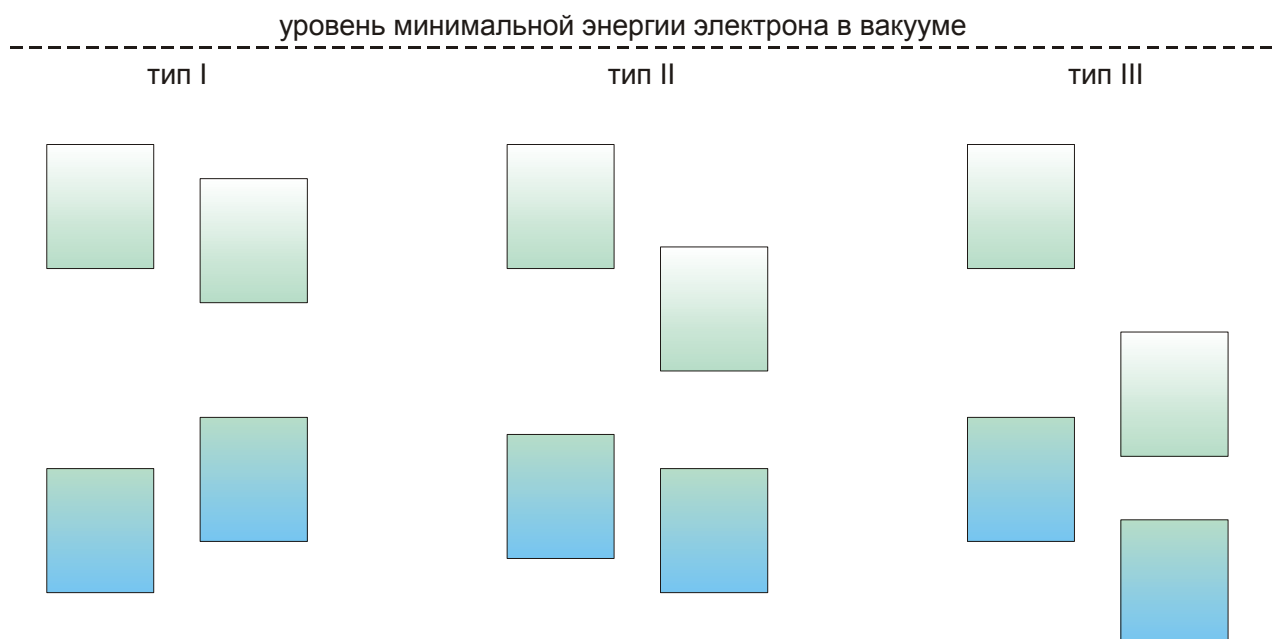


Рисунок 8: Схематическое изображение расположения зон соединяемых полупроводников на гетеропереходах разного типа.

Теперь мы потребуем несколько специальные свойства от используемых полупроводников. Во-первых, соединяемые полупроводники должны быть различными (с разными значениями ширины запрещённой зоны или с разными значениями сродства к электрону). Во-вторых, нам нужна идеальная (атомно-гладкая) граница между этими полупроводниками. Для её создания необходимо, чтобы периоды кристаллической решётки обоих полупроводников были близки, в противном случае получатся неконтролируемые механические напряжения и дефекты на границе. Кроме того, нам необходимо, чтобы эти полупроводниковые материалы были технологичными, была возможность получения таких материалов высокой чистоты (с большими подвижностями носителей) и была возможность их легирования. Для определения взаимного расположения зон воспользуемся приближением правил Андерсона.

В зависимости от соотношения сродства к электрону и ширины запрещённых зон в соединяемых полупроводниках возможны несколько различных вариантов: запрещённая зона полупроводника с меньшей шириной запрещённой зоны (узкозонного полупроводника) может лежать внутри запрещённой зоны полупроводника с большой шириной запрещённой зоны (широкозонного полупроводника), запрещённая зона узкозонного полупроводника может перекрываться с валентной зоной или зоной проводимости широкозонного полупроводника и, наконец, запрещённая зона узкозонного полупроводника полностью попадает в валентную зону или зону проводимости широкозонного полупроводника (рисунок 8). Эти переходы называют, в перечисленном порядке, гетеропереходами I, II и III типа<sup>12</sup>. Мы ограничимся рассмотрением наиболее распространённого гетероперехода I типа, реализуемого в структурах на основе  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ . Переход II типа реализуется, например, в гетеропереходе  $\text{InP-In}_{0.52}\text{Al}_{0.48}\text{As}$ , а переход III типа в гетеропереходе  $\text{InAs-GaSb}$  [4].

<sup>12</sup> Также называют: охватывающим, ступенчатым и разрывным гетеропереходом для гетеропереходов I, II и III типа, соответственно.

## **Построение энергетической диаграммы гетероперехода AlAs-GaAs. Случай номинально чистых полупроводников.**

Требуемым свойствам удовлетворяют полупроводниковые соединения из семейства  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ . В этом семействе оказывается возможным вырастить кристаллы<sup>13</sup> высокой чистоты во всём диапазоне концентраций алюминия от арсенида галлия до арсенида алюминия. Период решётки изменяется от 5.6533 Å в чисто галлиевом соединении до 5.6611 Å в чисто алюминиевом арсениде, ширина запрещённой зоны<sup>14</sup> изменяется от 1.42 эВ в арсениде галлия до 2.16 эВ в арсениде алюминия, сродство к электрону равно 4.1 эВ для GaAs и 3.5 эВ для AlAs [11].

Алюминий и галлий оба являются элементами основной подгруппы третьей группы таблицы Менделеева, так что в наивной картине донорного или акцепторного допирования всё семейство  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  должно вести себя как чистый полупроводник. Однако оказывается, что даже в максимально очищенном виде GaAs является полупроводником p-типа, а AlAs – полупроводником n-типа. Это связано с небольшой остаточной концентрацией каких-то примесей или кристаллических дефектов. Таким образом, при  $T=0$  уровень химпотенциала в чистом GaAs оказывается вблизи потолка валентной зоны, а в чистом AlAs — вблизи дна зоны проводимости. Это может оказываться существенно при рассмотрении свойств гетероструктур, используемых в низкотемпературных экспериментах.

Теперь, ровно также как мы рисовали зонную структуру p-n перехода мы можем изобразить схематическую структуру зон на контакте номинально чистых AlAs и GaAs (рисунок 9). Дно зоны проводимости для GaAs должно быть по правилу Андерсона на 0.6 эВ ( $\chi_{\text{GaAs}} - \chi_{\text{AlAs}} = 4.1\text{эВ} - 3.5\text{эВ} = 0.6\text{эВ}$ ) ниже, чем в AlAs, на практике это отличие оказывается чуть меньше и равно примерно 0.4 эВ. Соответственно, потолок валентной зоны в GaAs окажется примерно на 0.3 эВ выше, чем в AlAs. Эти скачки сохранятся в нашем построении, они описывают скачок потенциала между соединяемыми полупроводниками, связанный с выравниванием положения зон относительно уровня энергии электрона в вакууме.<sup>15</sup>

---

13 Кубическая решётка со структурой алмаза, но в отличие от настоящего алмаза из 8 позиций атомов в кубической элементарной ячейке 4 позиции занято мышьяком, а 4 галлием или алюминием.

14 Интересно, хотя и не имеет отношения к нашим дальнейшим рассуждениям, что при  $x < 0.45$  (включая чистый GaAs) полупроводник является прямозонным, а при  $x > 0.45$  - непрямоzonным [11].

15 Скачок потенциала конечно же нефизичен и исчезнет при точном решении задачи о распределении заряда на гетеропереходе. Его возникновение — это расплата за использование упрощённого правила Андерсона. Однако качественно получаемый результат оказывается правильным.

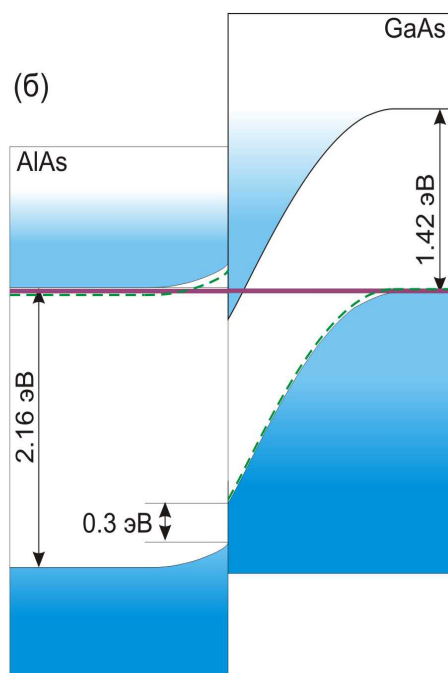
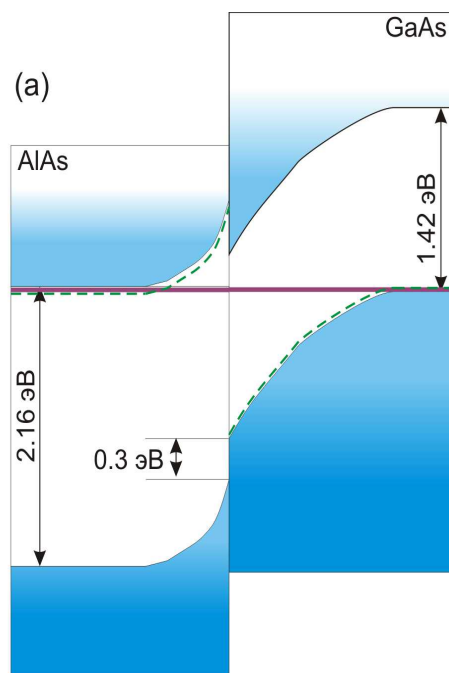
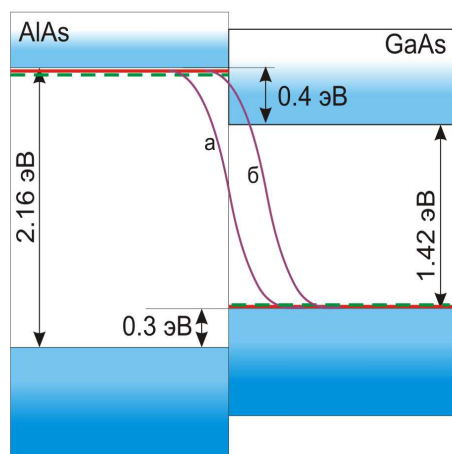


Рисунок 9: Построение энергетической диаграммы контакта номинально чистых AlAs и GaAs. В случае (а) химический потенциал не пересекает зону проводимости узкозонного GaAs, в случае (б) – пересекает. Ширина запрещённой зоны и разрывы зон показаны в масштабе.

При приведении полупроводников в контакт возникнет перераспределение зарядов, аналогичное случаю р-п перехода: часть примесей в AlAs ионизуется и электроны перейдут в GaAs, заполняя вакансии. Под действием электростатического потенциала перераспределённых зарядов возникает изгиб зон. В зависимости от концентрации примесных центров равновесный уровень химического потенциала может пересечь или не пересечь зону проводимости узкозонного полупроводника GaAs. Если такое пересечение произошло (случай (б) на рисунке 9), то в принципиальном отличии от случая р-п перехода

между однотипными полупроводниками при  $T=0$  часть электронов от донорных примесей AlAs, переходя в GaAs, оказывается не захвачена акцепторными центрами арсенида галлия, а попадает в зону проводимости — т.е. становится делокализована. Однако доступная делокализованным носителям область в зоне проводимости ограничена вглубь AsGa тем расстоянием, на котором уровень химического потенциала остаётся больше энергии дна зоны проводимости.

Обратите внимание, что из-за того, что в GaAs вдали от контакта уровень химпотенциала близок к потолку валентной зоны, то, в случае пересечения уровнем химпотенциала с зоной проводимости, дно зоны проводимости на границе оказывается ниже потолка валентной зоны. Это означает, что возможен и переход электронов из валентной зоны GaAs вдали от контакта в зону проводимости на границе контакта. Механизм накопления носителей на границе за счёт ионизации примесей в соседнем полупроводнике называют *механизмом аккумуляции*, механизм перераспределения электронов из валентной зоны в зону проводимости – *инверсией*. Соответственно, если основной вклад даёт один из механизмов, то формирующийся слой делокализованных электронов называют слоем аккумуляции или слоем инверсии.<sup>16</sup> Более широкий слой, в котором примесной уровень в GaAs оказывается ниже уровня химпотенциала и все примеси ионизируются называют *обеднённым слоем* (концентрация вакансий в GaAs в этом слое уменьшается).

### **Построение энергетической диаграммы гетероперехода AlAs-GaAs. Случай сильно легированного n-AlAs.**

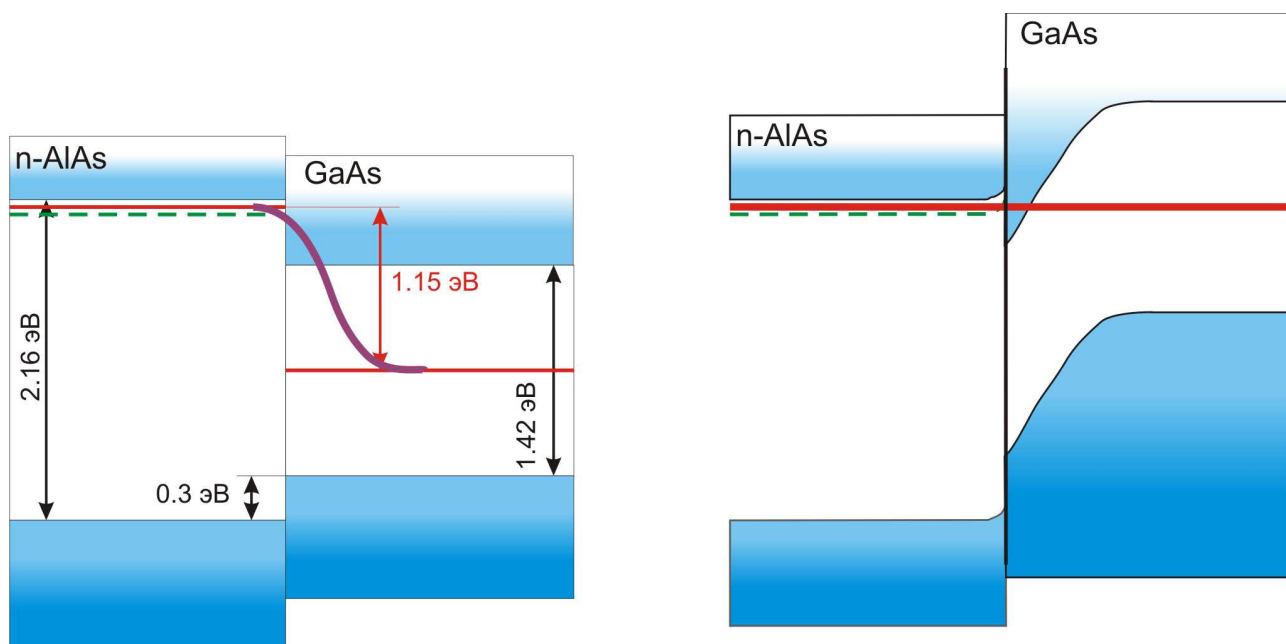


Рисунок 10: Построение энергетической диаграммы для гетероперехода между сильно легированным n-AlAs и чистым GaAs. Ширина запрещённой зоны и разрывы зон показаны в масштабе.

Напомним из рассмотрения р-п перехода, что ширина области ионизованных примесей тем меньше, чем больше концентрация примесей. Это позволяет, искусственно создав дополнительные донорные примеси в слое AlAs, добиться того, что слой ионизованных примесей в AlAs тонок и зоны AlAs почти не изгибаются: практически весь изгиб зон происходит в узкозонном GaAs. Это с неизбежностью означает, что на границе гетероперехода изогнутое дно зоны проводимости в GaAs окажется примерно на 0.4эВ ниже

<sup>16</sup> Иногда оба случая называют инверсным слоем.

уровня химического потенциала. То есть, мы ожидаем появления слоя делокализованных электронов на границе контакта.

Кроме того, из-за большой разности концентраций примесей в целенаправленно сильно легированном n-AlAs и номинально чистом GaAs оказывается правильнее считать GaAs идеальным полупроводником с химпотенциалом, расположенным посередине запрещённой зоны. При формировании слоя делокализованных электронов подавляющее большинство электронов с примесей в AlAs перейдут в слой в зоне проводимости GaAs, а не займут места на акцепторных центрах. Простым объяснением этого является невозможность перехода электрона по, в общем-то, диэлектрической матрице полупроводника при  $T=0$  (нулевая температура предположена, так как обсуждаемый механизм является одним из способов формирования двумерного электронного газа, изучаемого обычно в низкотемпературных экспериментах) далеко от границы раздела для заполнения вакансий на акцепторных центрах.<sup>17</sup>

Энергетическая диаграмма строится аналогично уже рассмотренным случаям (рисунок 10). Теперь дно зоны проводимости в GaAs всегда выше потолка валентной зоны и накопление носителей в слое делокализованных электронов происходит только за счёт ионизации примесей в n-AlAs.

В направлении перпендикулярном границе раздела для электронов в зоне проводимости возникает потенциальная яма. Такую яму можно аппроксимировать для упрощения треугольной потенциальной ямой с потенциалом  $U(z) = \begin{cases} +\infty, & z < 0 \\ kz, & z > 0 \end{cases}$ . Эта ограниченность приведёт к размерному квантованию движения в направлении, перпендикулярном границе раздела. Если уровень химического потенциала окажется выше некоторых квантовых уровней, то соответствующие зоны размерного квантования окажутся заполнены при  $T=0$ .

Контролируя на этапе создания гетероперехода количество примесных центров в AlAs можно регулировать свойства этой ямы. При достаточно низких температурах окажутся заселены только нижние уровни размерного квантования и формирующийся вблизи границы в GaAs электронный газ будет проявлять двумерные свойства. Отметим также принципиальную особенность этой технологии: легируется слой арсенида алюминия, а двумерный электронный газ формируется в слое чистого арсенида галлия. Это позволяет достичь высокой подвижности носителей заряда в таких структурах.

На гетеропереходе между чистым узкозонным полупроводником и широкозонным сильно легированным полупроводником р-типа можно показать аналогичными построениями, что уровень химпотенциала пересечёт валентную зону чистого полупроводника. Соответственно, на границе контакта часть электронов из валентной зоны уйдёт на акцепторные центры р-полупроводника. Другими словами, на границе контакта появится квантовая яма для дырок, для которых также можно добиться условий размерного квантования.

---

<sup>17</sup> Конечно это не строгое объяснение. Можно относиться к этому факту просто как к экспериментальному наблюдению.



## Полевой транзистор.

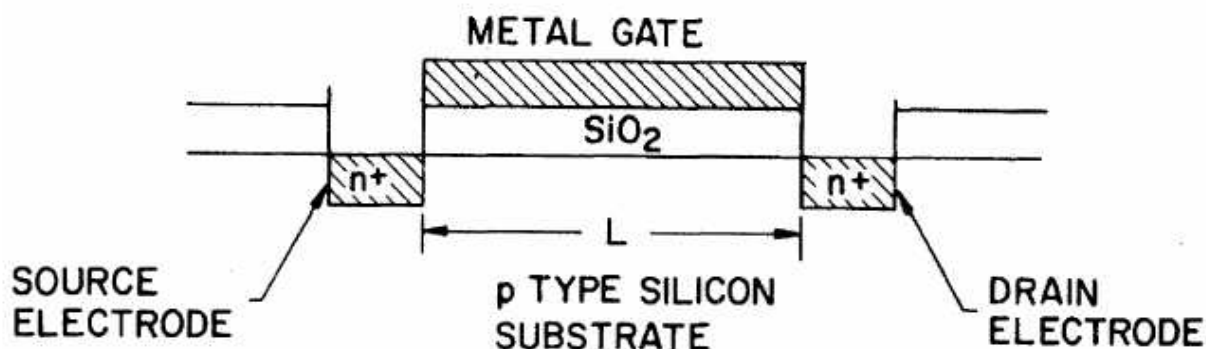


Рисунок 11: МОП-транзистор, используемый для получения двумерного электронного газа. Из [12].

Исторически первой структурой, в которой был получен двумерный электронный газ является кремниевый полевой транзистор. Для его получения на слой полупроводника (кремния n или p-типа) наносится слой диэлектрика (оксида кремния  $\text{SiO}_2$ ), поверх которого напыляется металлический электрод (называемый затвором, «gate» в англоязычной литературе). Такую структуру (рисунок 11) также называют МОП-транзистором (металл-оксид-полупроводник), а в англоязычной литературе иногда обозначают аббревиатурой MOSFET (по первым буквам определения metall-oxide-semiconductor field effect transistor).

Изменяя потенциал электрода можно создать электрическое поле внутри полупроводника, которое будет либо притягивать, либо отталкивать электроны от границы. Возникающее в полупроводнике перераспределение зарядов будет экранировать поле электрода, в результате возникнет изгиб зон вблизи границы, который может привести к образованию потенциальной ямы для электронов.

Для простоты будем считать, что при нулевом потенциале затвора уровень Ферми металла и уровень химпотенциала в полупроводнике совпадают.<sup>18</sup> Ширину запрещенной зоны в прослойке оксида считаем очень большой. Диэлектрические проницаемости оксида и полупроводника могут различаться, например для оксида кремния  $\text{SiO}_2$   $\epsilon \approx 3.8$ , а для кремния  $\epsilon \approx 12$ . Это отличие необходимо учитывать при расчёте электрического поля в МОП-структуре.

Отметим, что, в отличие от p-n перехода под напряжением, постоянного тока через МОП-структуру не течёт. То есть, рассматривая распределение электронов мы имеем дело со стационарной задачей и в равновесии для электронов внутри полупроводника  $\mu - e\phi = \text{const}$ . Будем считать, что потенциал полупроводника вдали от контакта поддерживается равным нулю.

<sup>18</sup> Наша цель в этом разделе показать, что подстройкой потенциала затвора можно создать квантовую яму для носителей заряда. В общем случае опять надо стартовать с правила Шоттки-Мотта для выбора общего начала отсчёта энергии при построении зонной структуры. Несовпадение уровней химпотенциала полупроводника и энергии Ферми металла приведёт к возникновению некоторого изгиба зон и в отсутствие внешнего напряжения. Этот «собственный» изгиб зон можно скомпенсировать приложением подходящего компенсирующего напряжения  $V_{\text{комп}}$  к затвору, тогда все дальнейшие рассуждения будут верны, только вместо приложенного напряжения  $V_{\text{внеш}}$  необходимо подставить отличие приложенного напряжения и напряжения компенсации  $(V_{\text{внеш}} - V_{\text{комп}})$ .

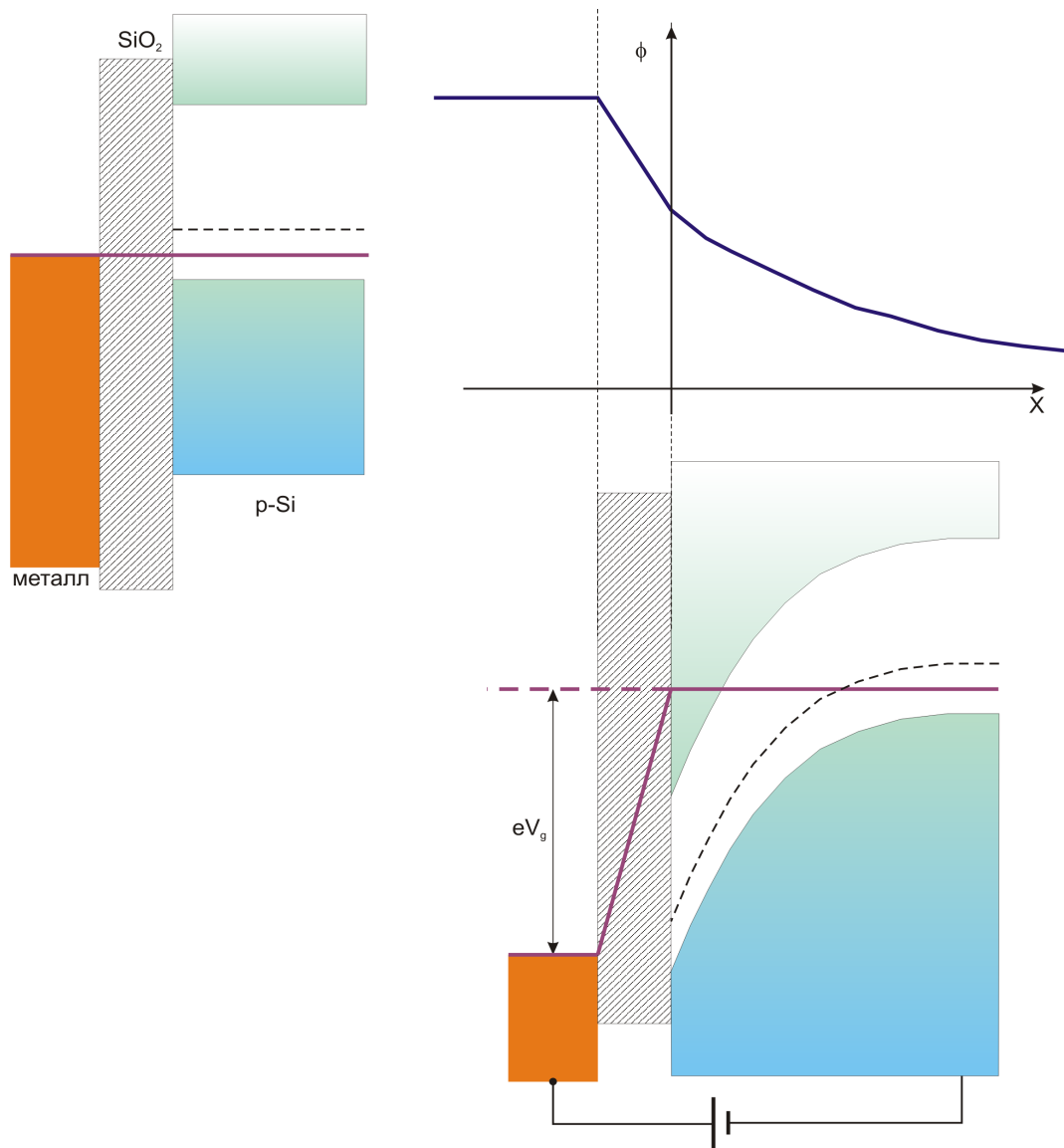


Рисунок 12: Энергетическая диаграмма полевого транзистора. Слева сверху: при равных потенциалах полупроводника и металла. Справа сверху: зависимость электрического потенциала от расстояния от границы при приложении к МОП-структуре внешнего напряжения (положительный потенциал металла). Справа внизу: представление энергетической диаграммы для полной энергии.

Будем для определённости рассматривать МОП-структуру с полупроводником р-типа и ограничимся случаем положительного (то есть притягивающего электроны) напряжения на затворе. Металлический затвор является эквипотенциальным. Если бы полупроводник был металлом с большим количеством носителей заряда, то и он бы был эквипотенциален и всё изменение потенциала происходило бы в слое оксида. Из-за малого количества носителей заряда в полупроводнике изменение потенциала в полупроводнике происходит плавно. При низкой температуре условиями на потенциал являются:

$$\epsilon_{ox} \left( \frac{d\phi}{dx} \right)_{ox} = \epsilon_{semi} \left( \frac{d\phi}{dx} \right)_0,$$

$$\epsilon_{semi} \frac{d^2\phi}{dx^2} = 4\pi e n(x),$$

здесь  $\epsilon_{ox,semi}$  - диэлектрические проницаемости оксида и полупроводника, граница считается плоской, ось  $X$  направлена вглубь полупроводника и ноль отсчёта координаты считается на границе полупроводник-оксид. Уравнение Пуассона написано для  $p$ -полупроводника (это определяет выбор знака в правой части, сравните с анализом  $p$ - $n$  перехода на стр. 11),  $n(x)$  — это и концентрация электронов локализованных на акцепторных центрах, и концентрация электронов в приповерхностном инверсном слое, если он образуется. Эти электроны будут инжектированы из внешней цепи при подключении источника напряжения.

Таким образом, потенциал  $\phi(x)$  имеет излом на границе полупроводник-оксид, для контакта  $\text{SiO}_2\text{-Si}$  наклон графика  $\phi(x)$  в оксиде втрое круче. Это позволяет в принципе найти зависимость потенциала от координаты в МОП-структуре, в случае появления приповерхностного слоя это потребует решения самосогласованной задачи (концентрация электронов в приповерхностном слое определяется формой потенциала). Мы ограничимся качественным построением энергетической диаграммы, для которого примем как данность, что большая часть изменения потенциала происходит внутри полупроводника (это вопрос подбора параметров МОП-структуры, из опыта известно, что такие структуры существуют).

Тогда далее по той же логике, что и для стационарного распределения зарядов в  $p$ - $n$  переходе, перестроим схему зонной структуры от «кинетической» к полной энергии, что соответствует вычитанию зависимости  $e\phi(x)$  (рисунок 12). В этом представлении в полупроводнике уровень химического потенциала (теперь это электрохимический потенциал) станет постоянным, границы зон и примесной уровень искривятся вниз, электрохимический потенциал металла опустится на  $eV_{внеш}$ .<sup>19</sup> Внутри диэлектрика в силу линейности  $\phi(x)$  зависимость  $\mu(x)$  станет также линейна. Из-за изменения электростатического потенциала внутри металла относительное расположение уровня Ферми в металле и потолка валентной зоны изменится по сравнению с тем, как они были расположены при нулевом напряжении.

По полученной диаграмме видно, что инжектируемые в полупроводник электроны сначала занимают места на акцепторных примесях, формируя обеднённый слой, а при достаточно большом напряжении и занимая места в зоне проводимости вблизи границы. Электроны в приграничном слое частично инжектируются из внешней цепи, частично могут переходить из валентной зоны вдали от поверхности, соотношение этих вкладов зависит от особенностей технологического решения конкретного устройства.

Аналогичным образом управляющие электроды-затворы могут быть добавлены и к гетеропереходу, что даст возможность регулировать свойства формирующейся на гетеропереходе ямы.

Недостатком кремниевого полевого транзистора является меньшая чем в гетероструктурах на основе арсенида галлия подвижность носителей заряда. Преимуществом является возможность управления параметрами квантовой ямы, технологичность изготовления слоя диэлектрика — это просто контролируемое окисление поверхности кремния и высокое качество диэлектрика: к слою  $\text{SiO}_2$  толщиной 100 нм можно приложить напряжение до 100 В без заметных утечек, а для  $\text{AlAs}$  эта величина на порядок меньше и составляет около 10В на 100 нм [13].

Во избежание недоразумений поясним почему мы говорим как о транзисторе о структуре, при описании которой пока упоминается лишь один электрод. В действительности нужны и

---

<sup>19</sup> Отличие уровня электрохимического потенциала металла от уровня электрохимического потенциала в полупроводнике не противоречит условию  $\mu - e\phi = \text{const}$  так как электроны могут перейти из полупроводника в металл только через источник напряжения, который совершит над ними работу.

дополнительные электроды — в рассмотренном примере нам нужно как-то «ввести» в акцепторный полупроводник дополнительные электроны для получения двумерного электронного газа, для изучения транспортных свойств этого двумерного газа нужно несколько электродов. Схематический пример такого прибора из [12] показан на рисунке 11. В простейшем случае создаются два электрода по разные стороны от накрытого затвором пространства, эти электроды традиционно называют исток и сток (в англоязычной литературе «source» и «drain»). Эти электроды получают селективным внедрением примесей n-типа, создающих избыток электронов в областях электродов. При понижении под затвором дна зоны проводимости ниже уровня химпотенциала эти электроны занимают появившиеся состояния вблизи границы полупроводник-диэлектрик.

При высоких температурах в отсутствие размерного квантования можно просто заметить, что глубина разрешённого для электронов слоя под затвором определяется потенциалом затвора. Соответственно, изменяя затворное напряжение можно изменять сечение канала, соединяющего исток и сток и управлять током в цепи сток-исток. Это и есть основное свойство транзистора в радиотехнике. С развитием технологий получения полупроводниковых структур именно полевые транзисторы являются сейчас основой большинства электронных схем, обеспечивая лучшие показатели по размерам и энергоэффективности, чем «классические» транзисторы на p-n переходах.