

ОГЛАВЛЕНИЕ	
ПРЕДИСЛОВИЕ.....	6
Список общих обозначений и сокращений.....	8
Глава 1. ЭЛЕКТРИЧЕСКИЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ..	10
1.1. Энергетические зоны.....	10
Электронные состояния атомов и твёрдых тел.....	10
Модель ковалентной связи.....	13
Модель энергетических зон.....	16
Модель примесного полупроводника.....	20
Выводы.....	23
1.2. Свободные носители заряда.....	24
Функция распределения Ферми–Дирака.....	24
Максимальная плотность разрешённых состояний.....	28
Концентрация свободных носителей заряда.....	30
Равновесный полупроводник.....	30.....
Собственный полупроводник.....	33
Квазиуровни Ферми, произведение концентраций.....	34
Примесный полупроводник.....	35
Температурная зависимость концентрации носителей.....	37
Положение уровня Ферми и тип проводимости.....	40
Изгиб энергетических зон и электрическое поле.....	41
Заряд поверхности.....	42
Распределение концентрации носителей заряда по энергиям.....	44
1.3. Рекомбинация носителей заряда.....	46
1.4. Электрические токи в полупроводниках.....	51
Диффузионный ток.....	52
Дрейфовый ток.....	53
Уравнения токов.....	56
Температурная зависимость тока, температурный коэф.....	57
Сопrotивление полупроводника.....	59
Соотношение Эйнштейна.....	61
1.5. Уравнения непрерывности.....	63
Вывод уравнений.....	64
1.6. Задачи об инжекции и экстракции неосновных носителей.....	66
Задача 1.1. Полубесконечный образец.....	67
Образец конечных размеров.....	70
Задача 1.2. Экстракция неосновных носителей.....	72
Задача 1.3. Двусторонняя инжекция/экстракция носителей....	73
Контрольные вопросы.....	76
Глава 2. ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ	77
НА ЭЛЕКТРОННО - ДЫРОЧНЫХ ПЕРЕХОДАХ	
2.1. Физические основы работы идеального $p-n$ -перехода.....	
Формирование. Диффузионно-дрейфовое равновесие.....	

## Глава 1

### ЭЛЕКТРИЧЕСКИЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

#### 1.1. Энергетические зоны

##### Электронные состояния атомов и твёрдых тел

Зарождение электроники естественно связывать с открытием электрона английским физиком Дж. Дж. Томсоном в 1897 г. Более века тому назад, изучая катодные лучи в трубках Крукса, он показал, что эти лучи представляют собой поток отрицательных зарядов. Позднее эти заряды были названы электронами<sup>1</sup>. Электрон стал первой в истории физики экспериментально обнаруженной элементарной частицей. Само слово “электрон” ввел в обиход ирландский физик и математик Дж. Стоней, который впервые высказал идею о дискретности электричества<sup>2</sup>, дал количественную оценку минимального электрического заряда и в 1891 г. элементарный электрический заряд назвал электроном. Правда, первоначально это название относилось к заряду отрицательного одновалентного иона.

Современные представления об электронных состояниях атомов основаны на квантовой модели атома, предложенной датским физиком Нильсом Бором, и понимании места, которое занимает соответствующий химический элемент в Периодической таблице химических элементов Д. И. Менделеева. В основе квантовой модели атома лежат постулаты Н. Бора о дискретности энергетических состояний электронов в атоме и предложенные им представления об оболочечной структуре атомов, соответствующие периодической системе химических элементов. Квантовая модель атома явилась естественным развитием, основанным на идеях Макса Планка о дискретности энергии, планетарной модели атома Резерфорда.

Периодический закон химических элементов был открыт Д. И. Менделеевым в 1869 г. почти за 30 лет до открытия электрона. Периодическая таблица химических элементов включала в то время чуть более 60 элементов и немало свободных мест<sup>3</sup>. Однако, как считал Д. И. Менделеев, “периодическому закону – будущее не грозит разрушением, а только надстройки и развитие обещает ...”.

Действительно, исследования рентгеновских спектров, выполненные английским физиком Г. Мозли, помогли установить физический смысл атомного номера и понять, что свойства элементов состоят в периодической зависимости от заряда ядра. Представления Н. Бора об оболочечной структуре электронной системы атома прояснили смысл номера периода Таблицы. Номер периода оказался равным числу электронных оболочек в атоме. Принцип исключения Паули, который допускал пребывание на электронной орбите не более 2-х электронов с противоположными спина-

ми, позволил установить алгоритм заполнения электронных оболочек многоэлектронных атомов.

Таким образом, если рассматривать Периодическую таблицу химических элементов с точки зрения электронной структуры атомов, то суммарное число электронов в атоме равно порядковому номеру элемента. Число электронных оболочек в атоме равно номеру периода (строки) элемента. Число электронных уровней в каждой оболочке равно номеру оболочки. Номер группы (столбца) равен числу валентных электронов во внешней, наиболее удаленной от ядра, электронной оболочке.

Для обозначения электронных уровней в каждой оболочке принято пользоваться буквенными символами, заимствованными из спектроскопии. Это  $s, p, d, f$  - уровни.

С учётом принципа Паули на дискретных электронных уровнях атома, может находиться строго определенное число электронов (Таблица 1.1). Всего в оболочке  $2n^2$  разрешённых состояний, где  $n$  – номер электронной оболочки,  $n = 1, 2, \dots, N$ ,  $N$  – номер периода.

Т а б л и ц а 1.1

Электронный уровень	$s$	$p$	$d$	$f$	...
Максимальное число разрешённых состояний	2	6	10	14	...

Энергетическая диаграмма изолированного многоэлектронного атома показана на рис. 1.1. Электронная конфигурация изолированного атома формируется в соответствии с принципом Паули за счёт последовательного заполнения электронами разрешённых дискретных энергетических

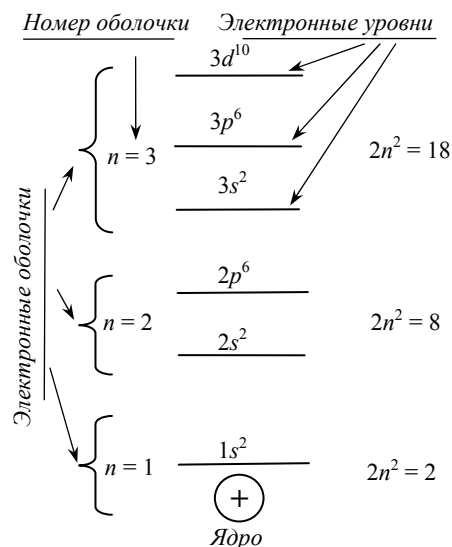


Рис.1.1 Структура энергетических уровней многоэлектронного гипотетического атома из третьего периода  $N = 3$ . На электронных уровнях показано максимально возможное число электронов. Конкретное число электронов и электронных уровней на внешней, третьей, оболочке зависит от номера группы, в которой находится элемент.

уровней. Причём, как указывалось, номер периода равен числу электронных оболочек. Номер группы равен числу валентных электронов. Номер элемента равен общему числу электронов в атоме.

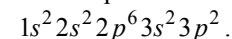
Цифра перед буквенным символом электронного уровня обозначает номер оболочки, показатель степени – максимально возможное число электронов (разрешенных состояний уровня). Гипотетический атом из третьего периода содержит три электронные оболочки, расположенные на возрастающем удалении от ядра атома. Орбиты электронов разных электронных уровней одной оболочки различаются формой (круговая, эллиптическая). Орбиты электронов одного энергетического уровня различаются ориентацией. Электроны на одной орбите имеют противоположные направления спинов. Электроны наиболее удаленных от ядра уровней внешней оболочки называются валентными. Именно валентные электроны взаимодействуют с другими атомами при формировании более сложных атомных структур.

Интересующий нас фрагмент периодической таблицы химических элементов, содержащий элементы III-V групп, показан в таблице 1.2. В твердотельной электронике чаще всего используются монокристаллы кремния (Si) и германия (Ge).

Т а б л и ц а 1.2

Период \ Группа	III	IV	V
2	5 B $2s^2 2p^1$	6 C $2s^2 2p^2$	
3	13 Al $2s^2 2p^1$	14 Si $3s^2 3p^2$	15 P $3s^2 3p^3$
4	31 Ga $4s^2 4p^1$	32 Ge $4s^2 4p^2$	33 As $4s^2 4p^3$
5	49 In $5s^2 5p^1$	50 Sn $5s^2 4p^2$	51 Sb $5s^2 5p^3$
6		52 Pb $6s^2 6p^2$	

Кремний находится в четвертой группе третьего периода. Значит, он имеет три электронных оболочки и 4 валентных электрона в третьей оболочке. Всего 14 электронов, что соответствует атомному номеру кремния. Электронная конфигурация атома кремния выглядит следующим образом:



Обратим внимание, что все элементы IV группы таблицы Менделеева имеют одинаковую конфигурацию внешней электронной оболочки:

$Ns^2Np^2$ , содержащую 4 электрона из 8 возможных. Аналогично, элементы V группы имеют конфигурацию  $Ns^2Np^3$  и содержат 5 внешних электронов, на один больше. Элементы III группы имеют 3 валентных электрона, на один меньше, чем элементы IV группы.

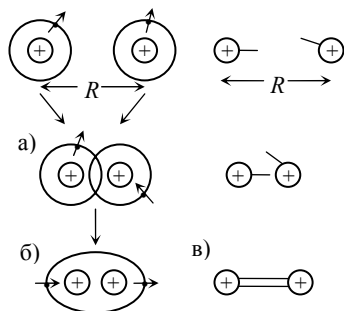
### Модель ковалентной связи

В твёрдотельной электронике электроны выполняют функцию не только свободных носителей заряда, формирующих токи проводимости, но и являются теми элементами, которые играют основную роль в процессе объединения индивидуальных атомов в твёрдое тело. Электроны выполняют роль привязных ремней, реализующих межатомные связи, благодаря которым из совокупности индивидуальных атомов создаются пространственно упорядоченные периодические структуры. Причём образование свободных носителей заряда (СНЗ) происходит в органическом единстве с формированием этих связей.

Различают ионную, металлическую и ковалентную связи. При ионной связи часть электронов одного атома перемещается к другому атому. В результате образуется противоположно заряженные ионы, взаимодействие которых консолидирует атомы в твёрдое тело. При металлической связи кристаллическая решётка положительно заряженных ядер окружена отрицательным электронным газом. В случае ковалентной связи каждый атом связан со своим ближайшим (в данном направлении) соседом парой электронов (приставка «ко» – два). Один валентный электрон одного атома и один электрон соседнего атома (один «свой» и один «чужой») вращаются (согласно представлениям классической физики) вокруг этих двух атомов по одной общей орбите. Общая орбита двух электронов соседних атомов реализует ковалентную связь этих атомов в данном направлении.

Рис. 1.2. Модель формирования молекулы водорода  $H_2$

а), б) образование молекулы при сближении двух атомов за счёт ковалентной связи, реализованной двумя валентными электронами;  
в) символическое изображение ковалентной связи двух атомов.



Молекула водорода  $H_2$ , самого распространённого элемента Вселенной, является наглядным примером того, как за счёт ковалентной связи формируется новая структура, более сложная, нежели сам атом. При

сближении атомов до расстояний  $R$ , на которых орбиты валентных электронов начинают перекрываться (рис. 1.2а), два атома водорода объединяются в молекулу (рис. 1.2б) за счёт ковалентной связи, символически изображенной на рис. 1.2в.

С помощью одного электрона от каждого из двух атомов, объединяемых ковалентной связью, формируется структура не сложнее двухатомной молекулы (например, водорода). Посредством двух электронов от каждого из двух атомов, объединяемых ковалентной связью, возможно формирование более сложной структуры, состоящей, например, из трёх атомов (рис. 1.3).

Когда в реализации ковалентных связей участвует каждый из трёх валентных электронов, возможно формирование ещё более сложной структуры. Например, двумерной поверхности, состоящей из правильных шестиугольных атомных структур, где каждый атом связан ковалентными связями с тремя ближайшими соседями (рис. 1.4). Такую структуру имеют нанотрубки<sup>4</sup>, образованные атомами углерода. Правда, из 4-х валентных электронов атома углерода в формировании ковалентных связей участвуют только 3. Четвертый – может быть свободным.

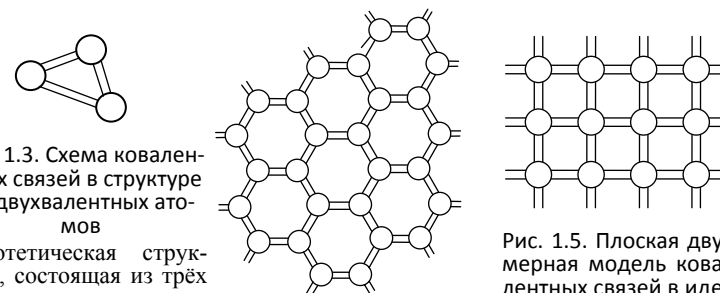


Рис. 1.3. Схема ковалентных связей в структуре из двухвалентных атомов

Гипотетическая структура, состоящая из трёх объединённых ковалентными связями, созданными двумя валентными электронами от каждого атома

Рис. 1.4. Схема ковалентных связей в структуре из трёхвалентных атомов

Рис. 1.5. Плоская двумерная модель ковалентных связей в идеальной кристаллической решётке типа алмаза, состоящей из четырёхвалентных атомов кремния

Объёмные твёрдотельные кристаллические решётки монокристаллов четырёхвалентного кремния и германия образуются за счёт ковалентных связей каждого атома с четырьмя ближайшими соседними атомами. Двумерная (плоская) модель связей в решётке четырёхвалентных атомов, показана на рис. 1.5. В такой решётке вокруг каждого атома находится 8 электронов. Консолидируясь в твёрдое тело, атомы дополняют свою внешнюю оболочку до 8 электронов, что соответствует полностью запол-

ненным (значит, устойчивым)  $s$ - и  $p$ -электронным уровням одиночного атома (рис. 1.6).

Связи реальных кристаллов имеют объёмную пространственную структуру. Плоская модель не имеет визуального сходства с реальной. Вместе с тем она правильно отражает главную особенность реальной решётки – структура связей в ней такова, что у каждого атома в кристалле имеется четыре симметрично расположенных ближайших соседа. Они размещены в вершинах правильного тетраэдра, в центре которого находится сам атом. Пространственная структура ковалентных связей атома германия, соответствующая плоской модели рис. 1.1, рис. 1.6, показана на рис. 1.7. Каждая из четырех связей осуществляется двумя электронами. Чёрными кружками изображены свои валентные электроны центрального атома. Светлыми кружками изображены четыре валентных электрона, которые принадлежат и соседним атомам.

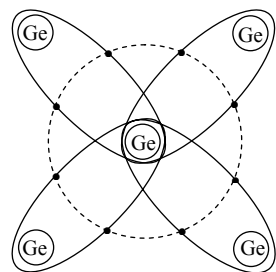


Рис. 1.6. Полное заполнение  $s$ - и  $p$ -электронных уровней при формировании идеальной кристаллической решётки типа алмаза

Характерная особенность ковалентной связи состоит в том, что электронные оболочки двух атомов частично перекрываются (например, на рис. 1.2а). Следствием такого взаимодействия является расщепление энергетического уровня электронов, формирующих ковалентную связь, на два подуровня – верхний и нижний. При этом выполняется общий принцип, согласно которому при сближении атомов полное число разрешённых состояний расщепляющегося уровня для каждого атома сохраняется. В качестве конкретного примера на рис. 1.8 показано расщепление  $1s$ -уровня атомов водорода по мере их сближения при образовании молекулы. Два атома, объединённые ковалентной связью, создали два подуровня разрешённых состояний; всего  $2 \times 2 = 4$  состояния для двух атомов. Состояния электронов на каждом подуровне различаются спинами.

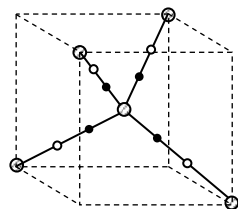


Рис. 1.7. Пространственная структура ковалентных связей атомов кремния и германия

Другая характерная особенность ковалентной связи заключается в том, что по мере увеличения числа атомов новые состояния появляются внутри крайних значений энергии расщеплённых уровней двух атомов в виде тонкой дополнительной структуры. В качестве примера на рис. 1.9

показана энергетическая диаграмма гипотетической (линейной) одномерной цепочки четырёх атомов, объединяемых ковалентными связями посредством электронов  $s$ -уровня при их сближении [1].

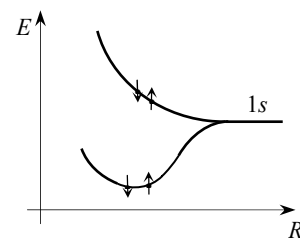


Рис. 1.8. Расщепление  $1s$ -уровня двух атомов водорода по мере их сближения и объединения в молекулу

Жирными стрелками показаны разрешённые состояния на каждом уровне.

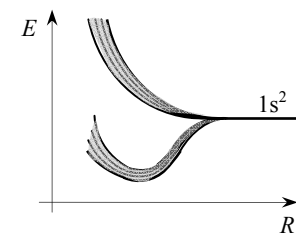


Рис. 1.9. Уровни энергии в зависимости от расстояния  $R$  между ядрами линейной цепочки нескольких атомов

Образование полос разрешённых состояний при ковалентной связи нескольких атомов.

Распространяя этот принцип на трёхмерные структуры, заключаем, что в твёрдом теле у электронов имеются не дискретные уровни энергии, как у двух отдельных атомов, объединённых ковалентной связью, а полосы разрешённых состояний. Отсюда следует, что при формировании твёрдого тела дискретные уровни электронов, реализующих ковалентные связи, расщепляются на две полосы разрешённых состояний, разделённых зоной запрещённых состояний (рис. 1.9).

### Модель энергетических зон

Элементы IV группы кремний и германий формируют кристаллическую решётку за счёт ковалентных связей каждого атома с 4-мя ближайшими соседями. Подобно уровням в молекуле водорода электронные уровни пары атомов каждой ковалентной связи расщепляются на два уровня, которые объединяются с электронными уровнями других атомов в две полосы разрешённых состояний – верхнюю и нижнюю с характерным минимумом.

На рис. 1.10 качественно показано, как энергетические уровни изолированных атомов кремния расщепляются в энергетические зоны разрешённых состояний при формировании твёрдого тела по мере сближения атомов до действительных расстояний между ближайшими соседями в кристалле кремния [2]. Изолированный атом кремния содержит во внешней оболочке два  $3s$ -электрона и два  $2p$ -электрона (таблица 1.2). При сближении атомов каждый из этих уровней расширяется в верхнюю и

нижнюю зону разрешенных состояний, которые при дальнейшем сближении атомов перекрываются.

При расстоянии, равном фактическому значению постоянной решетки кристалла кремния  $R_0$ , перекрывшиеся нижние зоны  $s$ - и  $p$ -уровней, образуют зону разрешенных состояний, которая называется **валентной зоной**. Перекрывшиеся верхние зоны  $s$ - и  $p$ -уровней образуют зону разрешенных состояний, которая называется **зоной проводимости**. Между ними находится **запрещенная зона**, шириной  $E_g = E_C - E_V$ .

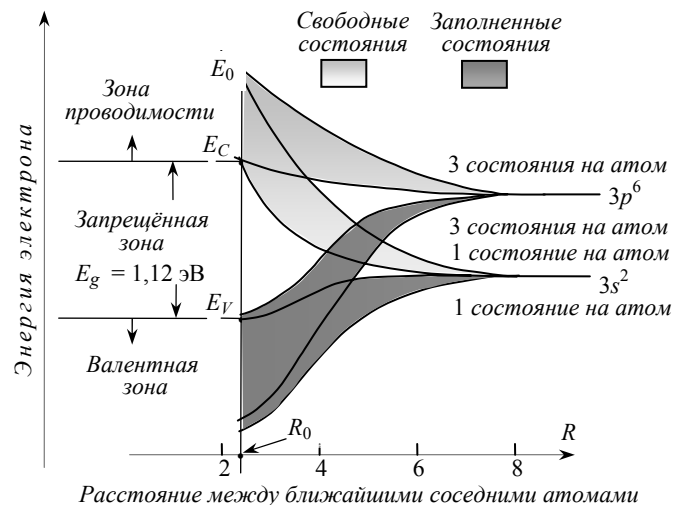


Рис. 1.10. Схема образования энергетических зон в кремнии при уменьшении расстояния между ближайшими соседними атомами

По достижении значения  $R_0$  зоны перекрываются, образуя зону проводимости, запрещенную зону шириной  $E_g = 1,12$  эВ и валентную зону. Германий имеет аналогичную схему энергетических зон, но  $E_g = 0,72$  эВ.

Поскольку сближение атомов оставляет неизменным полное число разрешенных состояний данного уровня, то в верхней и нижней зоне  $s$ -уровня на каждый атом приходится по одному разрешенному состоянию. Аналогично, в верхней и нижней зоне  $p$ -уровня на каждый атом приходится по три разрешенных состояния. Однако если зоны перекрываются, то уже невозможно отличить  $3s$ - от  $3p$ -состояний. Естественно поэтому считать, что в зоне проводимости и валентной зоне имеется по четыре разрешенных состояния на каждый атом.

На вертикали энергий для  $R_0$  различают уровень  $E_V$  – вершину валентной зоны<sup>5</sup>, уровень  $E_C$  – дно зоны проводимости<sup>6</sup>, уровень  $E_0$  – уро-

вень свободного электрона в вакууме (рис. 1.10). Глубина зоны проводимости  $E_0 - E_C = E_a$  называется **энергией электронного средства**.

Если ось пространственных координат  $x$  направить перпендикулярно рисунку 1.10, то получим **энергетическую зонную диаграмму** кремния/германия, показанную на рис. 1.11.

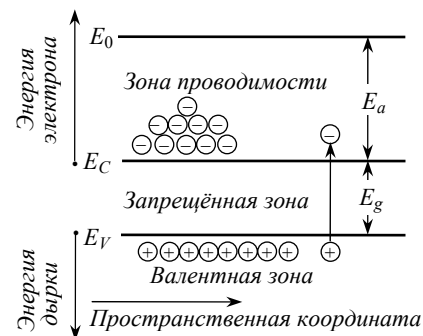


Рис. 1.11. Энергетическая зонная диаграмма

Зона проводимости, содержащая свободные электроны, и валентная зона, содержащая свободные дырки, разделены запрещенной зоной.

Разрыв ковалентной связи собственных атомов означает образование пары свободных носителей заряда и изображается как переход электрона из валентной зоны в зону проводимости.

Аналогично строятся энергетические зоны германия и других элементов IV группы. Однако ширина запрещенной зоны при температуре  $T_0 = 300^\circ\text{K}$  составляет величину порядка  $7\text{эВ} = 270\kappa T_0$  у алмаза,  $1,12\text{эВ} = 42\kappa T_0$  – у кремния,  $0,73\text{эВ} = 27\kappa T_0$  – у германия и около  $0,2\text{эВ} = 7\kappa T_0$  у серого олова, где  $\kappa$  – постоянная Больцмана,  $\kappa T_0 = 26\text{мВ} = 26 \cdot 10^{-3}$  эВ. Исходя из ширины запрещенной зоны, алмаз относят к изоляторам, кремний Si и германий Ge – к полупроводникам, олово – к металлам. Модель энергетических зон позволяет судить о состоянии носителей заряда в пространстве энергий.

В валентной зоне и в зоне проводимости на каждый атом приходится по четыре разрешенных состояния, а всего в атоме 4 валентных электрона. Значит, при низких температурах все 4 валентных электрона атома занимают энергетически более выгодные состояния с меньшей энергией в валентной зоне. Поэтому валентная зона полностью заполнена, а зона проводимости – пуста. Состояние заполненной валентной зоны и пустой зоны проводимости соответствует сохраненным (не разорванным) ковалентным связям, показанным на рис. 1.5. При таких условиях прохождение электрического тока проводимости в кристалле исключено. **Свободные носители заряда** (СНЗ) отсутствуют. Возможен только обмен электронами между атомами при суммарном импульсе, равном нулю.

Однако у всех элементов рассматриваемой группы, кроме алмаза, ширина запрещенной зоны невелика, поэтому при комнатной температуре ковалентные связи могут разрываться за счет энергии тепловых колеба-

ний решетки. Разрыв ковалентных связей в пространстве координат означает уход электрона с общей для двух атомов орбиты, формирующей ковалентную связь (рис. 1.12). В пространстве энергий – это уход электрона в зону проводимости с образованием пары свободных носителей заряда – положительной дырки в валентной зоне и отрицательного электрона в зоне проводимости. Действительно, ушедшие электроны оставляют пустыми уровни у потолка валентной зоны и заполняют свободные состояния у дна зоны проводимости. Пустой (не занятый электроном) уровень в валентной зоне называется дыркой. *Дырка – это отсутствие электрона в ковалентной связи двух соседних собственных атомов кристаллической решётки.* Вакантное место, которое образовалось в результате разрыва ковалентной связи и ухода электрона из валентной зоны, имеет положительный заряд. Оно ведёт себя в кристалле как виртуальная (мыслимая) элементарная частица<sup>7</sup>, имеющая реальный положительный заряд, равный по модулю заряду электрона. Состояние частично заполненной зоны проводимости и частично свободной валентной зоны (рис. 1.11) означает наличие разрывов ковалентных связей в модели кристалла, приведённой на рис. 1.5. Оно показано на рис. 1.12.

Разрыв ковалентной связи собственных атомов за счёт энергии тепловых колебаний решётки, т. е. уход электрона с общей орбиты, формирующей ковалентную связь, создаёт пару свободных носителей заряда – отрицательный электрон в зоне проводимости и положительную дырку в валентной зоне.

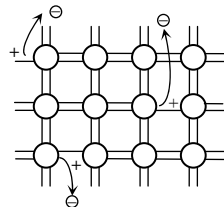


Рис. 1.12. Образование носителей заряда в собственном полупроводнике

Число атомов в кристалле велико, порядка  $10^{22} \text{ см}^{-3}$ , а глубина энергетических зон конечная, единицы эВ. Поэтому как в зоне проводимости, так и в валентной зоне образуется практически непрерывный спектр разрешённых состояний, состоящих из огромного числа дискретных уровней, разделённых незначительным зазором. Действительно, если принять, что  $E_a = 1 \text{ эВ}$ , объёмная плотность атомов равна  $10^{22} \text{ см}^{-3}$  и на каждый атом приходится по 4 разрешённых состояния, то энергетический интервал между разрешёнными состояниями составит  $1:(4 \cdot 10^{22}) = 2,5 \cdot 10^{-23} \text{ эВ}$ . Это на 20 порядков меньше характерного масштаба энергии тепловых колебаний решётки  $kT_0 = 26 \text{ эВ}$  для  $T_0 = 300 \text{ К}$ .

Наличие огромного количества разрешённых состояний, сопоставимого с числом атомов в кристалле, позволяет считать электроны зоны проводимости и дырки в валентной зоне *свободными носителями заряда* (СНЗ) в кристалле, т. к. они имеют возможность практически непрерывно изменять своё состояние в пространстве энергий и в пространстве коор-

динат полупроводника. Ведь все разрешённые состояния принадлежат конкретным атомам.

Посредством электрического поля, например, можно сформировать направленный перенос СНЗ и получить ток проводимости в полупроводнике. Электрическая проводимость будет иметь биполярный характер, поскольку осуществляется СНЗ двух знаков – положительными дырками и отрицательными электронами. Биполярная проводимость – необходимый признак *собственного* полупроводника, в котором нет примесей, и решётка содержит атомы только одной природы.

Разрыв ковалентной связи соседних собственных атомов решётки приводит к образованию (генерации) пары СНЗ разного знака (рис. 1.11, рис. 1.12, рис. 1.13б). *Условие электрической нейтральности* собственного полупроводника

$$(-q) \times n + (+q) \times p = 0,$$

выраженное через концентрацию носителей заряда, имеют вид

$$n = p \equiv n_i, \quad (1.1)$$

где  $n [\text{см}^{-3}]$  и  $p [\text{см}^{-3}]$  – концентрация электронов и дырок в зоне проводимости и валентной зоне соответственно,  $n_i$  – концентрация носителей собственного полупроводника<sup>8</sup>.

### Модель примесного полупроводника

Если собственный атом кристаллической решётки четырёхвалентного кремния или германия будет замещён пятивалентным атомом элемента V группы таблицы Менделеева, например, фосфором P, то четыре валентных электрона примесного атома будут задействованы в формировании ковалентных связей. Энергетический уровень пятого валентного электрона  $E_D$  будет находится в запрещённой зоне полупроводника, поскольку ни в валентной зоне, ни в зоне проводимости нельзя разместить ещё хотя бы один дополнительный уровень разрешённых состояний сверх положенных четырёх уровней на каждый атом (рис. 1.13). Не участвующий в формировании ковалентных связей пятый электрон слабо связан с ядром. Энергия его ионизации  $\Delta E_D = E_C - E_D$ , как правило, на один – два порядка меньше ширины запрещённой зоны. Например, энергия ионизации фосфора в кремнии составляет всего  $0,044 \text{ эВ} = 1,7kT_0$  при ширине запрещённой зоны  $E_g = 1,12 \text{ эВ} \approx 42kT_0$ . «Лишний» электрон легко отрывается от атома примеси, т.е. переходит с примесного уровня  $E_D$  в зону проводимости, где становится *свободным носителем заряда*. Такая примесь называется *донорной*, поскольку увеличивает концентрацию СНЗ-электронов в полупроводнике. Энергетический уровень электронов до-

норной примеси  $E_D$  находится в запрещённой зоне ниже дна зоны проводимости на величину энергии ионизации примеси (рис. 1.13в).

Ионизация атомов донорной примеси означает переход электрона с примесного уровня в зону проводимости. При этом образуется **свободный электрон** в зоне проводимости и **неподвижный положительный ион** донорной примеси, встроенный в кристаллическую решётку. Поскольку энергия ионизации донорной примеси невелика, то уже при комнатной температуре практически все атомы донорной примеси ионизированы. Ионизация атомов донорной примеси не исключает образование пар  $\text{CHЗ}$  за счёт перехода электронов из валентной зоны в зону проводимости. Ионизация собственных атомов создаёт концентрацию  $p$  [ $\text{см}^{-3}$ ] свободных дырок.

Условие электрической нейтральности примесного полупроводника, легированного донорной примесью, имеет вид:

$$n = p + N_D^+, \quad (1.2)$$

где  $N_D^+$  – концентрация встроенных в решётку положительных ионов донорной примеси.

Энергия ионизации примеси во много раз меньше энергии ионизации собственных атомов решётки. Поэтому в области рабочих температур в донорном полупроводнике  $n \gg p$ . Электроны являются **основными носителями заряда** (ОНЗ)<sup>9</sup>. Они образуются, в основном, за счёт ионизации атомов донорной примеси. Их концентрация может широко варьироваться путём изменения количества введённой примеси. Дырки являются **неосновными носителями заряда** (ННЗ). Они образуются за счёт разрыва ковалентных связей (ионизации) собственных атомов решётки.

Аналогично, если собственный атом кристаллической решетки четырёхвалентного кремния или германия замещён трёхвалентным атомом элемента III группы таблицы Менделеева, например, бором **В**, то три валентных электрона примесного атома будут задействованы в формировании ковалентных связей. Электрон, недостающий для формирования четвёртой ковалентной связи, привлекается за счёт разрыва ковалентной связи собственных атомов решётки в объёме кремния (рис. 1.13а).

При этом образуется **свободная дырка** в валентной зоне и **неподвижный отрицательный ион** атома акцепторной примеси, встроенный в кристаллическую решетку. Такая примесь называется **акцепторной** (от англ. ассерт – принимать), поскольку она «присоединяет» к себе электроны за счёт их перехода из валентной зоны на примесный уровень  $E_A$ . Энергетический уровень «присоединённых» электронов  $E_A$  находится в запрещённой зоне выше вершины валентной зоны на величину энергии ионизации акцепторной примеси (рис. 1.13а).

Условие электрической нейтральности примесного полупроводника, легированного акцепторной примесью, значит, обладающего в основном дырочной проводимостью, имеет вид:

$$p = n + N_A^-, \quad (1.3)$$

где  $N_A^-$  – концентрация встроенных в решётку отрицательно ионизированных атомов акцепторной примеси.

В полупроводнике, легированном акцепторной примесью, дырки являются **основными носителями заряда**, а электроны – **неосновные носители заряда**. Дырки образуются, в основном, за счёт ионизации атомов примеси, а ННЗ-электроны – за счёт разрыва ковалентных связей собственных атомов решётки.

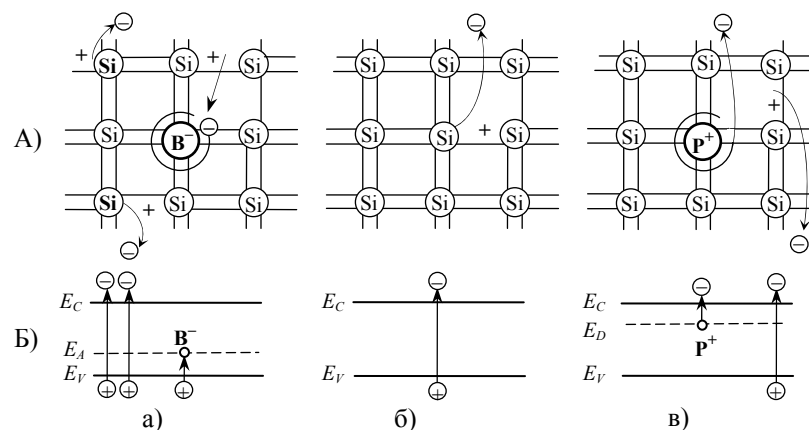


Рис. 1.13. Образование свободных носителей заряда в собственном и примесных полупроводниках за счёт тепловой генерации электронно-дырочных пар и ионизации атомов примеси

А) Модели кристаллических решёток. Б) Зонные диаграммы.

а) Модель **акцепторной примеси**. Трёхвалентный атом бора захватывает у собственного атома электрон, недостающий для формирования ковалентной связи. Образуется свободная дырка в валентной зоне и встроенный в кристаллическую решётку отрицательный ион  $\text{В}^-$ .  $E_A$  – акцепторный уровень.

б) Модель **собственного полупроводника**. Разрыв ковалентной связи собственных атомов решётки кремния создаёт пару свободных носителей заряда – положительную дырку в валентной зоне и отрицательный электрон в зоне проводимости.

в) Модель **донорной примеси**. Пятивалентный атом фосфора отдаёт незадействованный в ковалентных связях пятый электрон. Образуется свободный электрон в зоне проводимости и встроенный в кристаллическую решётку неподвижный положительный ион  $\text{P}^+$ .  $E_D$  – донорный уровень.

На зонных диаграммах примесных полупроводников показано также образование электронно-дырочных пар носителей заряда за счёт разрыва ковалентных связей (ионизации) собственных атомов решётки, которые обозначаются как переходы электронов из валентной зоны в зону проводимости.

В области рабочих температур примесных полупроводников основными являются «примесные» носители, созданные за счёт ионизации примеси. Неосновные носители – это «собственные» носители, созданные за счёт ионизации собственных атомов решётки.

#### Выводы

1. Элементы IV группы таблицы Менделеева Si и Ge формируют кристаллическую решётку за счёт ковалентных связей каждого атома с четырьмя ближайшими соседями. В каждом данном направлении ковалентная связь двух соседних атомов реализуется двумя валентными электронами  $s$ - и  $p$ -электронных уровней – одним «своим» и одним «чужим», которые (согласно представлениям классической физики) вращаются вокруг этих атомов по одной общей орбите.

2. При формировании твёрдого тела  $s$ - и  $p$ -электронные уровни атомов, объединяемых ковалентными связями, расщепляются на две – верхнюю и нижнюю – полосы разрешённых состояний, представляющих собой совокупность множества (по числу атомов) дискретных энергетических уровней.

Нижние полосы совокупности  $s$ - и  $p$ -электронных уровней перекрываются и создают **валентную зону**, в основном, занятых (электронами) состояний полупроводника. Перекрывающиеся верхние полосы  $s$ - и  $p$ -уровней создают **зону проводимости**, в основном, свободных состояний. Между ними находится **запрещённая зона**.

На каждый атом в валентной зоне и зоне проводимости приходится по 4 разрешённых состояния (по 1 от  $s$ - и по 3 от  $p$ -подуровней). Число атомов велико, а глубина зон конечная. В каждой из зон формируется *практически непрерывный спектр* большого числа разрешённых состояний, разделённых незначительным дискретом, величина которого на много порядков меньше характерного масштаба энергии тепловых колебаний.

3. Свободные носители заряда в полупроводнике создаются при разрыве ковалентных связей (ионизации) собственных атомов и внедрённых в решётку трёх- или пятивалентных атомов примеси<sup>10</sup>.

Разрыв ковалентных связей собственных атомов решётки, т. е. уход электрона с общей орбиты, формирующей ковалентную связь, означает освобождение занятого энергетического уровня в валентной зоне и переход электрона в зону проводимости, где он становится СНЗ. Незанятое электроном вакантное место – дырка – ведёт себя в кристалле как свободная виртуальная частица, имеющая элементарный положительный заряд. Электрическая проводимость осуществляется свободными зарядами обоих знаков, поэтому имеет биполярный характер.

Пятивалентные атомы примеси, замещая четырёхвалентные атомы Si или Ge, имеют один «лишний» электрон, который не участвует в формировании ковалентных связей. Он легко отрывается от примесного атома и

переходит в зону проводимости. При этом наряду со свободным отрицательным электроном образуется внедрённый в решётку неподвижный положительный ион атома донорной примеси. Условие электрической нейтральности соблюдается.

При замещении четырёхвалентного собственного атома Si или Ge трёхвалентным атомом примеси для образования четырёх ковалентных связей недостаёт одного электрона. Недостающий электрон восполняется за счёт разрыва ковалентной связи собственных атомов в объёме и присоединения электрона к атому примеси. Это означает переход электрона из валентной зоны на примесный уровень. Образуется свободная положительная дырка в валентной зоне и внедрённый в решётку неподвижный отрицательный ион атома акцепторной примеси. Условие электрической нейтральности сохраняется.

4. Носители в полупроводниках образуются путём ионизации примесных и собственных атомов. Для ионизации примеси требуется значительно меньше энергии, чем для ионизации собственных атомов. Поэтому большинство носителей заряда имеет «примесное» происхождение, и тип электрической проводимости определяется характером примеси. При донорной примеси преобладающими, т.е. **основными носителями заряда** являются электроны. Преобладает электронная проводимость. При акцепторной примеси ОНЗ становятся дырки. Преобладающей является дырочная проводимость.

5. Ионизация атомов, значит, и генерация СНЗ реализуется за счёт энергии тепловых колебаний решётки. Температурные и радиационные зависимости концентрации СНЗ являются принципиальной особенностью полупроводниковых приборов.

## 1.2. Свободные носители заряда

### Функция распределения Ферми–Дирака

Свободные носители заряда – электроны, каждый из которых независимо от остальных принимает в твёрдом теле то или иное состояние, в условиях теплового равновесия всё же имеют объективные меры, посредством которых описываются совокупные свойства всего ансамбля. К таким мерам относится **функция распределения**.

Функция распределения в статистических системах показывает относительное число членов ансамбля, обладающих данным свойством.

Конкретно в случае электронов она показывает, какую долю от общего числа свободных электронов составляют электроны с заданной энергией  $E$ . Как отмечалось максимальное число электронов, которые могут иметь данную энергию, значит, число разрешённых состояний в единице объёма, приходящихся на единичный интервал энергии



$n_{\max}(E)$  [Дж<sup>-1</sup> см<sup>-3</sup>], есть строго определённая для заданной энергии величина. Значит, функция распределения  $f_{\Phi-Д}(E)$  определяет отношение числа электронов  $n(E)$ , реально занимающих энергетический уровень  $E$ , к максимальному числу электронов, которые в принципе могут находиться на данном уровне:

$$f_{\Phi-Д}(E) = n(E)/n_{\max}(E). \quad (1.4)$$

Аналитическое выражение функции распределения электронов, которое независимо друг от друга предложили итальянский физик Энрико Ферми и английский физик Поль Дирак, имеет вид

$$f_{\Phi-Д}(E) = \frac{1}{1 + e^{(E-E_F)/kT}}, \quad (1.5)$$

где параметр  $E_F$  – уровень Ферми. Функция Ферми–Дирака, как элемент универсального математического аппарата, не зависит от свойств той или иной конкретной системы, а зависит лишь от температуры. Привязка к конкретной системе осуществляется через параметр  $E_F$ , который показывает, как нужно располагать функцию Ферми–Дирака относительно энергетических уровней конкретной системы (см. ниже рис. 1.20, рис. 1.21, рис. 2.2).

Функция распределения позволяет решать ряд важных задач, необходимых для анализа полупроводниковых приборов.

Во-первых, представляется возможным определять распределение свободных носителей заряда по энергиям. Распределение электронов в пределах зоны проводимости получаем из (1.4):

$$n(E) = n_{\max}(E) f_{\Phi-Д}(E) \left[ \text{см}^{-3} \text{Дж}^{-1} \right]. \quad (1.6a)$$

Если энергетический уровень заполнен не полностью, то  $n(E) < n_{\max}(E)$ .

Отсутствие электрона означает наличие дырки. Тогда число дырок на данном энергетическом уровне или распределение дырок в валентной зоне по энергиям есть разность

$$\begin{aligned} p(E) &= n_{\max}(E) - n(E) = n_{\max}(E) [1 - f_{\Phi-Д}(E)] \equiv \\ &\equiv n_{\max}(E) f_p(E), \end{aligned} \quad (1.6b)$$

где

$$f_p(E) = [1 - f_{\Phi-Д}(E)] = \frac{1}{1 + e^{-(E-E_F)/kT}} \quad (1.7)$$

– функция распределения дырок.

Во-вторых, пользуясь функцией распределения можно рассчитывать концентрации СНЗ, т.е. количество электронов или дырок в единице объёма, имеющих любое допустимое значение энергии в пределах зоны про-

водимости и валентной зоны соответственно. Концентрации электронов  $n$  и дырок  $p$  определяются интегралами

$$n = \int_{E_C} n(E) dE = \int_{E_C} n_{\max}(E) f_{\Phi-Д}(E) dE, \quad p = \int_{E_V} n_{\max}(E) f_p(E) dE, \quad (1.8)$$

где интегрирование следует ограничить областью энергий зоны проводимости (для электронов) и валентной зоны (для дырок).

Наконец, представляется возможным решать обратную задачу: зная концентрацию СНЗ, определять уровень Ферми  $E_F$ , и по положению уровня Ферми судить о свойствах полупроводника.

График функции распределения Ферми–Дирака для обозначенных температур приведён на рис. 1.14.

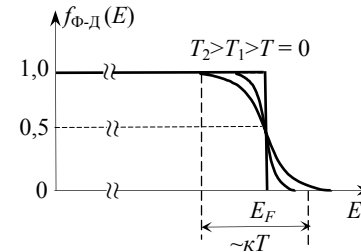


Рис. 1.14. Функция распределения Ферми–Дирака

• При температуре  $T = 0$  К он представляет собой прямоугольник, площадью  $E_F \times 1$ . При возрастании температуры освобождаемые площади под кривой распределения в области  $E < E_F$  равны возникающим площадям в области  $E > E_F$ . График функции содержит три характерные области:

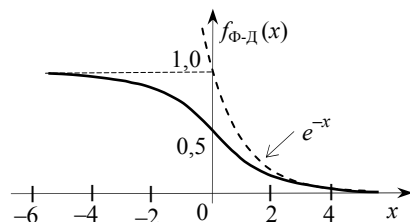
- область  $E < E_F$ , где  $f_{\Phi-Д}(E) = 1$ ;
- область в районе  $E \approx E_F$  протяжённостью в несколько  $kT$ , где функция испытывает резкий спад<sup>11</sup> и при  $E = E_F$  **независимо от температуры** равняется  $1/2$ ;
- область  $E > E_F$ , где «на хвосте» распределения при  $E - E_F > 2kT$  функцию Ферми–Дирака можно заменить функцией распределения Максвелла–Больцмана  $f_{М-Б}(E)$  (рис. 1.15). Действительно, пренебрегая единицей<sup>12</sup> в знаменателе (1.5), получим<sup>13</sup>

$$f_{\Phi-Д}(E) \approx e^{-(E-E_F)/kT} = \text{const} \times e^{-E/kT} = f_{М-Б}(E). \quad (1.9)$$

• Соотношение (1.4) есть число электронов, приходящихся на одно разрешённое состояние. Очевидно, что всегда  $n(E)/n_{\max}(E) \leq 1$ . Поскольку  $f_{\Phi-Д}(E) \leq 1$ , то **функцию Ферми–Дирака можно рассматривать как вероятность того, что энергетический уровень занят**. Тогда число электронов на данном уровне (1.6a) равно произведению максимальной плотности разрешённых состояний на вероятность того, что уровень занят. Функция распределения Ферми–Дирака определяет плотность вероятности занятости состояний с энергией от  $E$  до  $E + dE$ .

• При низких температурах функция Ферми–Дирака равна единице практически вплоть до энергии  $E \approx E_F$ , после чего она резко падает. Из вероятностного смысла функции Ферми–Дирака следует, что состояния с энергиями ниже уровня Ферми заняты (вероятность равна 1), а состояния с более высокими энергиями свободны. При повышении температуры определённая часть электронов переходит на более высокие энергетические уровни, значительно (по сравнению с  $\kappa T$ ) отстоящие от уровня Ферми, и в области «хвоста» подчиняются статистике Максвелла–Больцмана (1.9) (рис. 1.15). Этот факт является важным результатом, поскольку именно на этих уровнях находятся носители, формирующие токи в полупроводниковых приборах.

Рис. 1.15. Сравнение функции Ферми–Дирака с экспонентой  
На хвосте распределения при  $x = [(E - E_F)/\kappa T] \geq (2 \div 3)$  функция распределения Ферми–Дирака совпадает с функцией распределения Максвелла–Больцмана.



• Состояния занятого или свободного уровня являются противоположными событиями, сумма вероятностей которых равна 1 как вероятность достоверного события. Поэтому разность  $1 - f_{\text{Ф-Д}}(E)$ , дополняющая функцию Ферми–Дирака до единицы, есть вероятность того, что уровень занят дыркой (соотношение (1.7)).

• Распределение дырок в валентной зоне (1.7) выражается такой же формулой, что и распределение электронов (1.5) в зоне проводимости, но с противоположным знаком показателя экспоненты. Это даёт основание направлять ось отсчёта энергии дырок противоположно оси отсчёта энергии электронов (рис. 1.11, рис. 1.16). Тогда во встречно-вложенной системе координат распределение электронов и дырок изображается одной и той же кривой (рис. 1.16).

• С точки зрения теории вероятности уровень Ферми определяется как энергетический уровень, вероятность заполнения которого равна точно половине.

С точки зрения термодинамики уровень (энергия) Ферми является (в приближении равенства эффективных масс электронов и дырок) химическим потенциалом<sup>14</sup> (в расчёте на один электрон). Химический потенциал используется, в частности, при анализе состояния равновесия в таких гетерогенных (разнородных) системах как, например, жидкость–пар.

Равенство химических потенциалов служит критерием фазового равновесия.

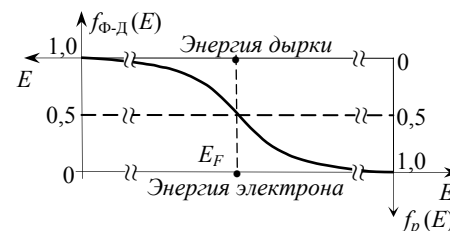


Рис. 1.16. Функция Ферми–Дирака во встречно-вложенной системе координат

Оси отсчёта энергии и ординат функций распределения электронов и дырок направлены в противоположные стороны.

Тогда, исходя из термодинамического смысла уровня Ферми, условием электронно-дырочного равновесия в разнородных системах (металлов, полупроводников) является равенство их уровней Ферми во всех частях системы. Единство и постоянство уровня Ферми является необходимым и достаточным условием равновесия электронно-дырочных систем.

Сам уровень Ферми определяется из условия, что полное число электронов в кристалле (системе) должно оставаться неизменным вне зависимости от их распределения по энергетическим уровням (ранее упомянутая обратная задача).

### Максимальная плотность разрешённых состояний

Для вычисления концентрации СНЗ в полупроводниках (соотношения (1.8)) необходимо кроме функции Ферми–Дирака располагать аналитическим выражением максимального числа электронов, способных иметь данную энергию, т. е. знать зависимость плотности разрешённых состояний  $n_{\text{max}}(E)$  от энергии. Для вывода этого соотношения определим сначала полное число электронов, энергия которых не превышает некоторого значения  $E$ .

В классической механике электрон считается частицей, не имеющей пространственного объёма, состояние которой во времени и пространстве точно определено координатами  $x, y, z$  и составляющими импульса  $p_x, p_y, p_z$ . Значит, состояние электрона будет задаваться (безразмерной) материальной точкой в шестимерном пространстве с взаимно перпендикулярными осями  $x, y, z, p_x, p_y, p_z$ . Это пространство называется **фазовым**. Полный объём фазового пространства  $V_{\text{фаз}}$  равен произведению объёмов пространства координат  $V_{xyz}$  и пространства импульсов  $V_{p_x p_y p_z}$

$$V_{\text{фаз}} = V_{xyz} \times V_{p_x p_y p_z}.$$

Объём изотропного пространства импульсов, соответствующий кинетической энергии

$$E = p^2/2m \Rightarrow p = \sqrt{2mE}, \quad (1.10)$$

где  $m$  – масса электрона, есть объём сферы радиуса  $p$ , равный  $V_{p_x p_y p_z} = (4/3)\pi p^3$ . Тогда

$$V_{\text{фаз}} = V_{xyz} \times (4/3)\pi p^3. \quad (1.11)$$

Согласно законам классической физики каждая точка этого пространства вполне характеризует состояние (координаты, скорость, направление движения) электрона, энергия которого не превышает величину  $E$  и может изменяться непрерывно.

Однако электрон не является классической частицей. При переходе из одного состояния в другое его энергия изменяется дискретным образом даже, если этот дискрет очень мал, как, например, в твёрдом теле. Если состояние изменяется с дискретом  $\Delta E$ , то, очевидно, что максимальное число возможных состояний электрона с энергией  $E$  не превышает отношения  $E/\Delta E$ . Поскольку энергия выражается через импульс и координату, то число состояний в пространстве импульсов и координат не превышает отношений  $p_x/\Delta p_x$  и  $x/\Delta x$  соответственно, где  $\Delta p_x$ ,  $\Delta x$  – дискреты изменения импульса и координаты, обязанные дискрету изменения энергии.

Из дискретности энергии логично следует принципиальная невозможность измерения физических величин с точностью, превышающей значения, обусловленные наличием соответствующих дискретов. Аналогично, применение линейки, проградуированной в миллиметрах, не позволяет измерять расстояния с точностью, превышающей значение дискрета, равного 1 мм.

Таким образом, при измерении дискретных физических величин (определении состояния объектов микромира) всегда имеет место принципиальная неопределённость, не связанная с погрешностями применяемых методов и используемых приборов.

Впервые к такому заключению пришёл немецкий физик Вернер Гейзенберг, который предложил принять эту принципиально неустранимую неопределённость в качестве специфического физического закона. Согласно этому закону, известному сейчас как соотношение неопределённости Гейзенберга, при одновременном определении координаты и импульса имеет место неопределённость измерения  $\Delta x$  и  $\Delta p_x$ , такая что

$$\Delta x \Delta p_x \geq h, \quad (1.12)$$

где  $h$  – постоянная Планка, посредством которой определяется минимальный дискрет (квант) энергии, равный  $h\nu$ ,  $\nu$  – частота излучения. Неопределённость выражается через произведение, что соответствует пред-

ставлению фазового пространства произведением пространства импульсов и координат<sup>15</sup>.

Отсюда следует, что для трёхмерного движения неопределённость составит величину, порядка  $\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z \geq h^3$ . Это означает, что объём, который занимает электрон в фазовом пространстве, всегда конечный, не меньше размера элементарной ячейки  $h^3$ . Учитывая, что в элементарной ячейке фазового пространства могут находиться 2 электрона с противоположными спинами<sup>16</sup>, приходим к выводу, что объём фазового пространства может содержать максимум  $2 \times (V_{\text{фаз}}/h^3)$  электронов. Тогда, используя (1.11) определим, что полное число электронов в единичном объёме координатного пространства  $N(E)[\text{см}^{-3}]$ , энергия которых не превышает  $E$ , будет равно

$$N(E) = 2V_{\text{фаз}}/(V_{xyz}h^3) = 2(4/3)\pi(p^3/h^3). \quad (1.13)$$

Максимальная плотность разрешённых состояний, т. е. число электронов в единице объёма с энергией  $E$ , приходящихся на единичный интервал энергии, по определению есть

$$n_{\text{max}}(E) = \frac{N(E+dE) - N(E)}{dE} = \frac{dN(E)}{dE} = \frac{dN}{dp} \frac{dp}{dE} = 4\pi(2m/h^2)^{3/2} \sqrt{E} \equiv \text{const} \times \sqrt{E} [\text{см}^{-3} \text{Дж}^{-1}]. \quad (1.14)$$

При вычислении сложной производной использованы соотношения (1.10), (1.13). В силу (1.10) здесь  $E$  – кинетическая энергия.

Отметим, что «отступления» от классической физики касались, в основном, обоснования размера элементарной ячейки фазового пространства через соотношение неопределённости Гейзенберга, которое по существу является формулировкой принципа исключения Паули на языке классической физики. Полученное соотношение (1.14) предполагает использование кинетической энергии классической частицы.

## Концентрация свободных носителей заряда

### Равновесный полупроводник

Концентрация СНЗ не содержащего примесей равновесного полупроводника определяется соотношениями (1.8). Вычисление интегралов можно упростить, если, учитывая особенности проходящих физических процессов, распространить интегрирование на бесконечные пределы. Такой шаг оправдан, поскольку почти все свободные электроны компактно располагаются на дне зоны проводимости, а дырки – у вершины валент-

ной зоны. Положив верхний предел интегрирования равным бесконечности, мы, не совершая существенной ошибки и сохраняя физический смысл результата, сможем получить более изящные аналитические выражения.

Вычисление интегралов (1.8) значительно упрощается также благодаря тому, что функцию Ферми–Дирака можно заменить функцией распределения Максвелла–Больцмана. Замена возможна, потому что для практически важных случаев *невырожденного* полупроводника уровень Ферми находится в запрещённой зоне ниже дна зоны проводимости, по крайней мере, на несколько  $\kappa T$ , а основная масса электронов – в зоне проводимости на расстояниях  $E - E_F > (2 \div 3)\kappa T$ . В этом случае в знаменателе уравнения (1.5) можно пренебречь единицей. Распределением электронов становится выражение (1.9), где полная энергия есть сумма потенциальной энергии  $E_C$  и кинетической энергии  $E_K$

$$E = E_C + E_K, \quad (1.15)$$

а максимальная плотность разрешённых состояний (1.14) выражается через кинетическую энергию  $E \equiv E_K$ . Тогда, введя обозначение  $x = E_K/\kappa T$ , используя (1.9), (1.14) и  $m = m_n$ , представим соотношение (1.8) в следующем виде:

$$n = \int_{E_C}^{E_0} n_{\max}(E) f_{\Phi-D}(E) dE \approx \int_{E_C}^{\infty} n_{\max}(E) f_{\Phi-D}(E) dE = \\ = \int_0^{\infty} n_{\max}(E_K) f_{M-B}(E_C + E_K) dE_K = 4\pi (2m_n/h^2)^{3/2} e^{-\frac{E_C - E_F}{\kappa T}} \int_0^{\infty} x^{1/2} e^{-x} dx,$$

где  $E_0$  – уровень свободного электрона в вакууме (рис. 1.10, рис. 1.11),  $m_n$  – эффективная масса электрона, посредством которой учитываются особенности движения в твёрдом теле. Воспользовавшись табличным интегралом  $\int_0^{\infty} x^{1/2} \exp(-x) dx = \sqrt{\pi}/2$ , окончательно получаем:

$$n = N_C e^{-\frac{E_C - E_F}{\kappa T}}, \quad (1.16)$$

где  $N_C = 2(2\pi m_n \kappa T/h^2)^{3/2} = 2,5 \cdot 10^{19} (m_n/m)^{3/2} \times (T/300)^{3/2} [\text{см}^{-3}]$

имеет смысл эффективной плотности состояний в зоне проводимости.

Определение концентрации дырок в приближении распределения Максвелла–Больцмана сводится к вычислению интеграла (1.8) для (1.7)  $f_p(E) \approx \exp[(E - E_F)/\kappa T]$ , где полная энергия дырки равна  $E = E_V - E_K$  и (как отмечалось ранее) отсчитывается от вершины валентной зоны  $E_V$

«вниз» в сторону отрицательных значений (рис. 1.11),  $p_{\max}(E_K)$  определяется уравнением (1.14) при эффективной массе дырки  $m = m_p$ , посредством которой учитываются особенности движения в твёрдом теле. Тогда

$$p = \int_{-\infty}^{E_V} p_{\max}(E) f_p(E) dE \approx \int_0^{\infty} p_{\max}(E_K) f_{M-B}(E_V - E_K) dE_K = \\ = 4\pi (2m_p/h^2)^{3/2} e^{-\frac{E_F - E_V}{\kappa T}} \int_0^{\infty} x^{1/2} e^{-x} dx, \quad \text{где } x = E_K/\kappa T. \quad \text{Окончательно} \\ p = N_V e^{-\frac{E_F - E_V}{\kappa T}}, \quad (1.17)$$

где  $N_V = 2(2\pi m_p \kappa T/h^2)^{3/2} = 2,5 \cdot 10^{19} (m_p/m)^{3/2} \times (T/300)^{3/2} [\text{см}^{-3}]$  – эффективная плотность состояний в валентной зоне.

Отметим, что полученные соотношения задают количество носителей заряда в единице объёма, но не закон их распределения по энергиям.

Таким образом, концентрация СНЗ полупроводника при данной температуре однозначно определяется положением уровня Ферми. Всякое изменение положения уровня Ферми приводит к экспоненциальному изменению концентрации носителей и обратно – изменение концентрации будь то за счёт температуры, легирования примесями, засветки и т. д. вызывает изменение положения уровня Ферми.

Количество носителей заряда определяется процессами *тепловой генерации* за счёт ионизации собственных и примесных атомов, а также обратными процессами *рекомбинации* электронов из зоны проводимости и дырок из валентной зоны (рис. 1.13, и ниже рис. 1.26, рис. 1.28). Процессы генерации и рекомбинации свободных носителей происходят непрерывно и параллельно. Равновесное состояние есть результат динамического равновесия этих процессов. При этом, однако, произведение концентраций электронов и дырок остаётся постоянным (правда, зависящим от температуры), *равным квадрату собственной концентрации* полупроводника. Действительно, учитывая, что в собственном полупроводнике носители образуются парами и  $n = p \equiv n_i^2$ , после почленного перемножения выражений (1.16) и (1.17), получим

$$np = n_i^2 = \text{const} = \\ = N_C N_V e^{-E_g/\kappa T} = 2,31 \cdot 10^{31} (m_n m_p/m^2)^{3/2} \times T^3 e^{-E_g/\kappa T}. \quad (1.18a)$$

где  $n_i$  – *собственная концентрация*. Такова особенность равновесного состояния полупроводников. Нижняя строчка этих равенств, раскрывающая произведение, зависит только от температуры и ширины запрещённой

ной зоны. Поэтому соотношение (1.18а) остаётся справедливым как для собственного, так и для примесного полупроводника независимо от концентрации носителей заряда и примесей<sup>17</sup>. Единственное налагаемое ограничение состоит в том, чтобы не нарушались условия, при которых распределение носителей может определяться статистикой Максвелла–Больцмана (1.9). Единичей в знаменателе формулы (1.5), можно пренебречь, если уровень Ферми находится в запрещённой зоне не ближе  $(2 \div 3)\kappa T$  от границ разрешённых зон. Для комнатной температуры это порядка  $(50 \div 80) \cdot 10^{-3}$  эВ при ширине запрещённой зоны  $(0,7 \div 1,43)$  эВ у промышленно используемых полупроводников.

Постоянство произведения концентраций носителей означает, что увеличение концентрации одних носителей с необходимостью сопровождается уменьшением концентрации носителей заряда другого знака.

По аналогии с известным законом химических реакций уравнение (1.18а) названо **законом действующих масс**. Закон справедлив, когда концентрация носителей собственных и примесных полупроводников определяется только температурой. В неравновесном состоянии концентрации носителей зависят от внешних воздействий, поэтому  $np \neq n_i^2$ .

Таким образом, равновесное состояние в полупроводнике наступает при одновременном соблюдении двух взаимосвязанных физических условий:

- ✓ условия **динамического равновесия** (1.18) или закона действующих масс и
- ✓ условия электрической нейтральности (1.1), либо (1.2), (1.3).

Равновесное состояние полупроводника можно образно характеризовать как состояние электрического и динамического равновесия.

### Собственный полупроводник

Из (1.18а) следует, что собственная концентрация, как параметр полупроводникового материала, при данной температуре зависит только от ширины запрещённой зоны полупроводника:

$$n_i = \sqrt{N_C N_V} e^{-E_g/2\kappa T} = \text{const} \times T^{3/2} e^{-E_g/2\kappa T}. \quad (1.18б)$$

Из (1.18а) видно также, что температурная (в данном случае экспоненциальная) зависимость параметров является принципиальной особенностью полупроводников. Стенная зависимость создаёт эффекты второго порядка по сравнению с экспоненциальной.

В собственном полупроводнике  $n = p$ . Положение уровня Ферми собственного полупроводника можно определить из равенства соотношений (1.16) и (1.17):

$$E_F = \frac{E_C + E_V}{2} + \frac{\kappa T}{2} \ln \frac{N_C}{N_V} = \frac{E_C + E_V}{2} + \frac{3}{4} \kappa T \ln \frac{m_p}{m_n} \approx \frac{E_C + E_V}{2}$$

Поскольку  $m_n \approx m_p$  (Таблица 1.3), приходим к выводу, что в собственном полупроводнике уровень Ферми находится примерно в середине запрещённой зоны. Действительно, если  $E_i$  – энергетический уровень середины запрещённой зоны, то  $E_C = E_i + \frac{1}{2}E_g$ ,  $E_V = E_i - \frac{1}{2}E_g$ . Тогда

$$\frac{1}{2}(E_C + E_V) = E_i \approx E_F. \quad (1.19)$$

Уровень  $E_i$  принято называть **собственным уровнем** полупроводника<sup>18</sup>.

Учитывая (1.19), из (1.16), (1.17) для собственной концентрации носителей получаем

$$n_i = N_C e^{-(E_C - E_i)/\kappa T} = N_V e^{-(E_i - E_V)/\kappa T} = p_i.$$

Собственный уровень и собственная концентрация являются параметрами материала.

Добавляя  $\pm E_i$  в показатели экспонент соотношений (1.16), (1.17), выразим концентрацию носителей через параметры полупроводника:

$$n = n_i e^{(E_F - E_i)/\kappa T}, \quad (1.20а)$$

$$p = n_i e^{-(E_F - E_i)/\kappa T}. \quad (1.20б)$$

В таком представлении концентрация носителей зависит только от относительного расстояния уровня Ферми до середины запрещённой зоны полупроводника. Поэтому полученные соотношения справедливы как для собственных, так и для примесных полупроводников. Отсюда

$$E_F = E_i + \kappa T \ln(n/n_i), \quad (1.21а)$$

$$E_F = E_i - \kappa T \ln(p/p_i). \quad (1.21б)$$

### Квазиуровни Ферми, произведение неравновесных концентраций

В состоянии равновесия как собственный, так и примесный полупроводник будут иметь единый для электронов и дырок уровень Ферми. Однако в неравновесном состоянии электронно-дырочная система носителей заряда полупроводника не может быть описана единым уровнем Ферми. Действительно, например, нагревание собственного полупроводника или поглощение излучения с энергией квантов  $h\nu \geq E_g$  приводит к увеличению числа разрывов ковалентных связей и повышает концентрацию как электронов, так и дырок. Но при увеличении концентрации электронов уровень Ферми согласно (1.21а) должен подниматься вверх, а при увеличении концентрации дырок он же согласно (1.21б) должен опускаться вниз. Выход (как это сделал Шокли) состоит в том, чтобы обобщить соот-

## Примесный полупроводник

ношения статистики на неравновесные состояния, если вместо единого Ферми формально ввести квазиуровень Ферми  $E_{Fn}$  для электронов и отдельный квазиуровень Ферми  $E_{Fp}$  для дырок. Тогда уже через квазиуровни Ферми соотношения (1.16), (1.17), (1.20) для неравновесных концентраций будут иметь такой же вид, как и в случае равновесия:

$$n = n_0 \pm \Delta n = N_C e^{-(E_C - E_{Fn})/\kappa T} = n_i e^{(E_{Fn} - E_i)/\kappa T}, \quad (1.22a)$$

$$p = p_0 \pm \Delta p = N_V e^{-(E_{Fp} - E_V)/\kappa T} = n_i e^{-(E_{Fp} - E_i)/\kappa T}, \quad (1.22б)$$

где через  $n_0, p_0$  обозначены равновесные концентрации, которые должны удовлетворять соотношению (1.18a). Но теперь произведение концентраций  $np \neq n_i^2$ . Для неравновесного состояния, используя (1.22), получаем

$$np = N_C N_V e^{-E_g/\kappa T} e^{(E_{Fn} - E_{Fp})/\kappa T} = n_0 p_0 e^{(E_{Fn} - E_{Fp})/\kappa T} = n_i^2 e^{(E_{Fn} - E_{Fp})/\kappa T}. \quad (1.23)$$

В равновесном состоянии уровень Ферми единый  $E_{Fn} = E_{Fp} = E_F$ . Соотношения (1.18a), (1.23) совпадают. Единство и постоянство  $dE_F/dx = 0$  уровня Ферми является необходимым и достаточным условием состояния равновесия. Неравновесное состояние и, следовательно, конечная разность квазиуровней Ферми возникает вследствие внешних факторов, например, напряжения, приложенного к полупроводниковому прибору, или инжекции в полупроводник носителей заряда.

Представления о параметрах некоторых широко используемых в электронике собственных полупроводников даёт таблица 1.3.

Т а б л и ц а 1.3

	Ge	Si	GaAs	InSb
$E_g$ , эВ	0,72	1,12	1,43	0,18
$E_a$ , эВ	4,0	4,05	4,07	4,59
$n_i$ , см <sup>-3</sup>	$2,4 \cdot 10^{13}$	$1,5 \cdot 10^{10}$	$2 \cdot 10^6$	$2 \cdot 10^{16}$
$N_C$ , см <sup>-3</sup>	$1,04 \cdot 10^{19}$	$2,8 \cdot 10^{19}$	$4,7 \cdot 10^{17}$	$4,2 \cdot 10^{16}$
$N_V$ , см <sup>-3</sup>	$6,1 \cdot 10^{18}$	$1,02 \cdot 10^{19}$	$7,0 \cdot 10^{17}$	$7,3 \cdot 10^{18}$
$m_n$	0,22 <i>m</i>	0,33 <i>m</i>	0,072 <i>m</i>	
$m_p$	0,31 <i>m</i>	0,56 <i>m</i>	0,5 <i>m</i>	
<i>m</i> – масса изолированного электрона, $E_a$ – электронное сродство.				

При определении концентрации ОНЗ в примесном (для определённости электронном) полупроводнике, необходимо исходить из того, что в равновесном состоянии одновременно должны выполняться два физических условия:

- условие электрической нейтральности (1.2)  $n_{n0} = p_{p0} + N_D$
- и закон действующих масс (1.18) –  $p_{p0} n_{n0} = n_i^2$ .

Индексы  $n$  и  $p$  в формулах обозначают электронный и дырочный тип проводимости соответственно, а индекс 0 показывает, что рассматривается равновесная концентрация полупроводника.

Подставив  $p_{p0} = n_i^2/n_{n0}$  в (1.2), получим квадратичное уравнение  $n_{n0}^2 - n_{n0} N_D - n_i^2 = 0$ . Из решения этого уравнения<sup>19</sup>

$$n_{n0} = \frac{1}{2} N_D \left[ 1 + \sqrt{1 + (2n_i/N_D)^2} \right] \approx \frac{1}{2} N_D \left[ 1 + 1 + \frac{1}{2} (2n_i/N_D)^2 \right]$$

закljučаем, что в области температур<sup>20</sup>, где  $n_i \ll N_D$ , концентрация ОНЗ электронного полупроводника равна концентрации доноров<sup>21</sup>

$$n_{n0} \approx N_D. \quad (1.24)$$

Из решения аналогичного уравнения для дырочного полупроводника определим, что в области температур, где  $n_i \ll N_A$ , концентрация ОНЗ-дырок равна концентрации акцепторов

$$p_{p0} \approx N_A. \quad (1.25)$$

Из (1.21), учитывая (1.24), (1.25), определим положение уровня Ферми в электронном и дырочном полупроводнике соответственно:

$$E_{Fn} = E_i + \kappa T \ln(n_{n0}/n_i) = E_i + \kappa T \ln(N_D/n_i), \quad (1.26a)$$

$$E_{Fp} = E_i - \kappa T \ln(p_{p0}/p_i) = E_i - \kappa T \ln(N_A/p_i). \quad (1.26б)$$

Значит, чем выше степень легирования, тем ближе уровень Ферми ко дну зоны проводимости электронного полупроводника или к вершине валентной зоны дырочного полупроводника.

В реальных условиях соотношения (1.24), (1.25) практически всегда выполняются. С учётом закона действующих масс (1.18) это означает, что чем выше концентрация примеси, тем выше концентрация ОНЗ и ниже концентрация ННЗ:

$$n_{n0} p_{p0} \approx N_D p_{p0} = n_i^2 \Rightarrow p_{p0} \approx n_i^2/N_D \ll n_{n0}, \quad (1.27a)$$

$$p_{p0} n_{n0} \approx N_A n_{n0} = n_i^2 \Rightarrow n_{n0} \approx n_i^2/N_A \ll p_{p0}. \quad (1.27б)$$

Например, для кремния, легированного донорной примесью с концентрацией  $N_D = 10^{15}$  см<sup>-3</sup> при  $n_i = 10^{10}$  см<sup>-3</sup> концентрация ННЗ-дырок равна

$p_{n0} = n_i^2 / N_D = 10^5 \text{ см}^{-3}$ , что на 5 порядков ниже собственной концентрации кремния и 10 порядков ниже концентрации ОНЗ-электронов, равной  $n_{n0} \approx N_D$ . Ясно что электрическая проводимость будет, в основном, электронной.

В кубическом сантиметре твёрдого тела содержится приблизительно  $10^{22}$  атомов. Рассмотренный пример показывает, что внедрение только одного атома примеси на  $10^{22}/10^{15} = 10^7$  собственных атомов превращает биполярную проводимость собственного полупроводника практически в монополярную проводимость примесного. Тип проводимости определяется примесью. Высокая чувствительность свойств полупроводника ко всякого рода примесям и дефектам выдвигает жёсткие требования к технологии производства.

### Температурная зависимость концентрации носителей

Полученные в предыдущем разделе соотношения и сделанные выводы основаны на предположении, что концентрация примеси существенно превышает собственную концентрацию полупроводника  $n_i \ll (N_D, N_A)$ .

Однако сама собственная концентрация (1.18б) экспоненциально зависит от температуры. Поэтому естественно возникает вопрос о характере температурной зависимости концентрации носителей тока в примесном полупроводнике, ибо она определяет температурную зависимость параметров полупроводниковых приборов.

В примесном полупроводнике свободные носители заряда образуются за счёт ионизации как примесных, так и собственных атомов (рис. 1.13). Однако для ионизации собственных атомов и перевода электрона из валентной зоны в зону проводимости требуется энергия, равная ширине запрещённой зоны. В то время как для ионизации примесных атомов и, например, перевода электрона с примесного уровня в зону проводимости требуется многократно меньшая энергия. Поэтому при каждой данной температуре вклад этих процессов в концентрацию носителей различен и зависит от температуры.

Экспериментальные зависимости концентрации электронов от температуры в кремнии и германии, легированных донорной примесью, приведены на рис. 1.17. Температурные зависимости имеют три характерные области.

В области низких температур средняя энергия тепловых колебаний решётки мала по сравнению с энергией ионизации донорной примеси. Доноры ионизированы лишь частично<sup>22</sup>. Концентрация свободных электронов незначительна, но экспоненциально растёт с увеличением температуры по мере ионизации доноров. Основную роль играют переходы электронов в зону проводимости с примесных уровней.

С повышением температуры средняя энергия фононов сравнивается с энергией ионизации доноров, оставаясь, однако, значительно меньше ширины запрещённой зоны. В этой области температур практически все атомы донорной примеси ионизированы, и их электроны находятся в зоне проводимости. Вместе с тем средняя энергия тепловых колебаний ещё недостаточна для того, чтобы перебрасывать электроны из валентной зоны в зону проводимости и повышать концентрацию носителей за счёт

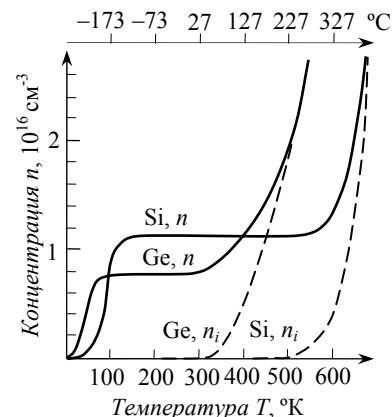


Рис. 1.17. Температурные зависимости концентрации электронов в кремнии, легированном мышьяком ( $As, N_D = 1,15 \cdot 10^{16} \text{ см}^{-3}$ ), и в германии, легированном мышьяком ( $As, N_D = 7,5 \cdot 10^{15} \text{ см}^{-3}$ )

Экспоненциальные участки кривых слева соответствуют области ионизации примеси, горизонтальные участки — области примесного истощения и примесной проводимости, экспоненциальные участки кривых справа — области собственной электропроводимости. Штриховые кривые описывают температурные зависимости собственных концентраций Si и Ge [5].

переходов «зона–зона». С этого момента и до температур, при которых энергия тепловых колебаний практически сравнивается с шириной запрещённой зоны, дальнейшее повышение температуры не приводит к заметному увеличению концентрации электронов. Такое состояние полупроводника называется состоянием **примесного истощения** в том смысле, что примесные уровни «истощились», отдав свои электроны в зону проводимости.

В области этих температур, протяжённостью в несколько сотен градусов абсолютной шкалы, концентрация ОНЗ-электронов практически не зависит от температуры (рис. 1.17) и равна концентрации доноров

$$n_i \ll n = N_D^+ \approx N_D.$$

Выполняются условия, использованные ранее для расчёта концентрации носителей (1.24), (1.25) в примесном полупроводнике. Остающаяся некоторая температурная зависимость концентрации обязана фактам ионизации собственных атомов решётки. Однако она незначительна, т. к. собственная концентрация  $n_i \ll n$ .

Состояние примесного истощения, важное своей температурной стабильностью концентрации основных носителей заряда, наступает тем раньше, чем меньше концентрация примеси и ниже энергия её ионизации.

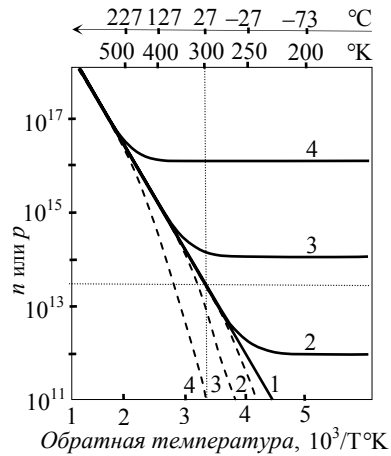
В зависимости от характера примеси нижняя граница температур (начальная температура) области примесного истощения может состав-

Рис. 1.18. Температурная зависимость концентрации электронов и дырок для Ge

Сплошные кривые – ОНЗ-электроны, штриховые кривые – ОНЗ-дырки.

- 1)  $N_D - N_A = 0$ ;
- 2)  $N_D - N_A = 10^{12} \text{ см}^{-3}$ ;
- 3)  $N_D - N_A = 10^{14} \text{ см}^{-3}$ ;
- 4)  $N_D - N_A = 10^{16} \text{ см}^{-3}$ .

В состоянии примесного истощения концентрация ОНЗ-электронов не зависит от температуры. Концентрация ОНЗ-дырок экспоненциально возрастает.



лять несколько десятков градусов абсолютной шкалы, т. е. сотни градусов отрицательных температур шкалы Цельсия<sup>23</sup>.

В области более высоких температур средняя энергия фононов начинает превышать ширину запрещённой зоны. Концентрация ОНЗ экспоненциально растёт. Ионизация собственных атомов создаёт пару ОНЗ  $n_i = p_i$ . Примесный полупроводник приобретает свойства собственного (рис. 1.18).

Верхняя граница – максимальная (конечная) температура состояния примесного истощения  $T_{\max}$  – определяется как температура, при которой собственная концентрация сравнивается с концентрацией легирующей примеси  $n_i = N_D$ . На рис. 1.19 приведены зависимости  $T_{\max}$  от степени легирования  $N$  для Ge ( $E_g = 0,72$  эВ), Si ( $E_g = 1,12$  эВ) и GaAs ( $E_g = 1,43$  эВ).

Область температур примесного истощения заканчивается, и состояние собственной проводимости наступает тем позже, чем выше концентрация примеси и шире запрещённая зона (рис. 1.17, рис. 1.19). Примесная

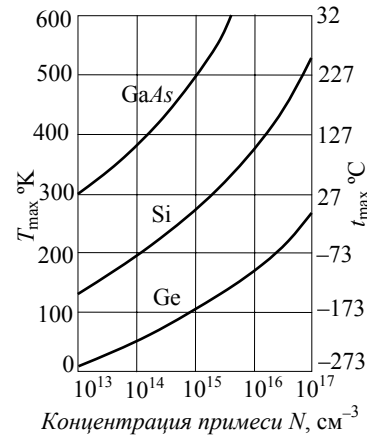


Рис. 1.19. Зависимость верхней границы примесного истощения  $T_{\max}$  от концентрации примеси для GaAs, Si и Ge

проводимость уступает место биполярной собственной проводимости.

### Положение уровня Ферми и тип проводимости

Электрическая проводимость зависит от концентрации носителей. Последняя однозначно определяется положением уровня Ферми. Тип проводимости зависит от соотношения концентраций электронов и дырок. Для определения зависимости положения уровня Ферми равновесного полупроводника от соотношения концентраций носителей разделим почленно выражения (1.16) и (1.17):

$$p/n = (N_V/N_C) e^{(E_C + E_V - 2E_F)/kT}$$

Отсюда, пренебрегая неравенством эффективных масс носителей и учитывая (1.19), определим:

$$E_F = \frac{1}{2}(E_C + E_V) + \frac{1}{2}kT \ln(n/p) + \frac{3}{4}kT \ln(m_p/m_n) \approx E_i + \frac{1}{2}kT \ln(n/p). \quad (1.28)$$

Значит, когда  $n = p$ ,  $\ln(n/p) = 0$ ,  $E_F \approx E_i$ , полупроводник собственный.

Если  $n > p$ , то  $E_F > E_i$ , преобладает электронная проводимость, полупроводник электронный. Если  $n < p$ , то  $E_F < E_i$ , преобладает дырочная проводимость, полупроводник дырочный. Таким образом, по положению уровня Ферми можно судить о преобладающей концентрации носителей и типе проводимости полупроводника (рис. 1.20).

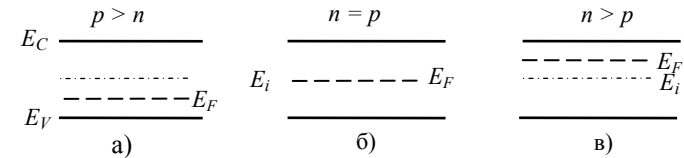


Рис. 1.20. Положение уровня Ферми и тип проводимости  
а)  $E_F \ll E_i$ , проводимость дырочная; б)  $E_F = E_i$ , проводимость собственная; в)  $E_F \gg E_i$ , проводимость электронная. Штриховые линии – уровень Ферми  $E_F$ , штрихпунктирные линии – собственный уровень  $E_i$ .

Из (1.28), следует, что при изменении концентрации уровень Ферми **всегда** движется в сторону зоны с большей концентрацией носителей заряда. Этим он похож на точку опоры равновесного коромысла, которая всегда приближается к более нагруженному плечу (рис. 1.21).

У невырожденного полупроводника уровень Ферми находится в запрещённой зоне. Однако следуя этой логике, уровень Ферми сильно леги-



рованных полупроводников может находиться в зоне разрешённых состояний.

Сильно легированные полупроводники, у которых уровень Ферми находится в разрешённой зоне, называются *вырожденными*. Далее в главе 2 мы увидим, что на таких полупроводниках формируются туннельные и обращённые диоды.

### Изгиб энергетических зон и электрическое поле

Значение энергии отличается от потенциала множителем:  $E = -q\varphi$ , где  $q > 0$  –

элементарный электрический заряд,  $\varphi$  – потенциал. Отсюда

$$\frac{dE}{dx} = -q \frac{d\varphi}{dx} = q\mathcal{E}_x \Rightarrow \mathcal{E}_x = \frac{1}{q} \frac{dE}{dx} \equiv \frac{1}{q} \nabla E,$$

где  $\mathcal{E}_x$  – проекция вектора напряжённости электрического поля на ось  $x$  пространственных координат. Для определённости удобно пользоваться собственным уровнем Ферми полупроводника. Уровень  $E_i$  однозначно характеризует наклон энергетических зон, значит, величину и направление вектора напряжённости электрического поля:

$$\bar{\mathcal{E}} = \frac{1}{q} \nabla E_i = \frac{1}{q} \overline{\text{grad} E_i} \quad (1.29)$$

Направление вектора напряжённости электрического поля совпадает с направлением градиента (возрастания) энергии собственного уровня полупроводника. Тангенс угла наклона пропорционален модулю напряжённости поля, а знак определяет направление вектора напряжённости электрического поля. При положительном знаке вектор напряжённости совпадает с положительным направлением оси координат, при отрицательном – противоположен.

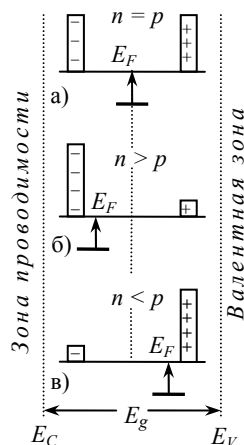
Из (1.29) следует, что всякий наклон энергетических зон свидетельствует о наличии в этой области электрического поля и обратно, область электрического поля сопровождается изгибом энергетических зон.

Формально полученное соотношение (1.29) согласуется с направлением физических процессов, происходящих в полупроводнике, помещён-

Рис. 1.21. Положение уровня Ферми

а) в собственном, б) в электронном, в) в дырочном невырожденном полупроводнике.

Подобно точке опоры равновесного коромысла уровень Ферми всегда движется в сторону зоны с большей концентрацией носителей заряда.



ном в электрическое поле. На рис. 1.22 показаны зонные диаграммы (А) и графики напряжённости электрического поля (Б) в поверхностном слое полупроводника<sup>24</sup>, когда внешнее поле направлено перпендикулярно его поверхности  $x = 0$ . Электрическое поле, проникающее на некоторую глубину в поверхностный слой изолированного, в частности, собственного полупроводника, вызывает перераспределение СНЗ. Электроны притягиваются к поверхности, дырки оттесняются в объём (рис. 1.22а). Согласно (1.20а,б) повышение концентрации электронов и уменьшение концентрации дырок возможно, если  $E_i$  уменьшается (снижается). Значит, зоны в области существования электрического поля должны изогнуться «вниз» относительно положения уровня  $E_i$  в объёме, где поле отсутствует. В электрическом поле противоположного направления повышение концентрации дырок и уменьшение концентрации электронов в поверхностном слое вызывает изгиб зон «вверх» (рис. 1.22б).

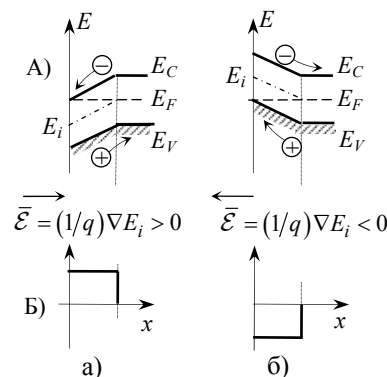


Рис. 1.22. Изгиб зон в электрическом поле

А) – энергетические зонные диаграммы.

Б) – напряжённость электрического поля.

а) вектор напряжённости электрического поля направлен по оси  $x$ ,

б) противоположное направление вектора напряжённости электрического поля.

### Заряд поверхности

В теории и технике полупроводников поверхность занимает особое место, ибо она является естественной неоднородностью, прерывающей пространственную периодичность кристаллической решётки. Разрыв кристаллической решетки оставляет на поверхности несформировавшиеся (разорванные) ковалентные связи, которые создают в запрещённой зоне полупроводника большое количество разрешённых состояний, играющих роль ловушек. Плотность этих состояний соизмерима с плотностью атомов на свободной поверхности кристалла, что составляет величину порядка  $10^{15} \text{ см}^{-2}$ . Захватывая или теряя заряд свободных носителей, поверхностные состояния формируют естественный заряд поверхности полупроводника  $Q_{SS}$  [Кл/см<sup>2</sup>], и заряжают поверхность. Плотность поверхностных зарядов, отнесённая к заряду электрона  $N_{SS} = Q_{SS}/q$ , находится в пределах

$10^{10} \div 10^{11} \text{ см}^{-2}$  для кремния и составляет величину порядка  $10^{13} \text{ см}^{-2}$  для арсенида галлия. Знак заряда поверхности, как правило, совпадает со знаком заряда основных носителей.

Возможный механизм формирования заряда поверхности показан на рис. 1.23. В электронном полупроводнике атомы донорной примеси отдают пятый электрон не в зону проводимости (как в объёме), а на разрешённые поверхностные состояния в запрещённой зоне (рис. 1.23а). Уход электронов с донорных уровней на поверхностные состояния  $n$ -полупроводника заряжает поверхность отрицательно  $Q_{SS} < 0$  и создаёт в прилегающей области положительный объёмный заряд обнажённых ионов донорной примеси  $Q_{\text{ОПЗ}}^+ = qN_D^+w > 0$ , где  $N_D$  – концентрация донорной примеси,  $w$  – размер области ОПЗ. Созданное поверхностным зарядом электрическое поле (направленное из объёма к поверхности) вытесняет ОНЗ-электроны из приповерхностного слоя в объём и согласно (1.29) (1.20а) вызывает изначальный изгиб энергетических зон «вверх» (рис. 1.22б, рис. 1.23а).

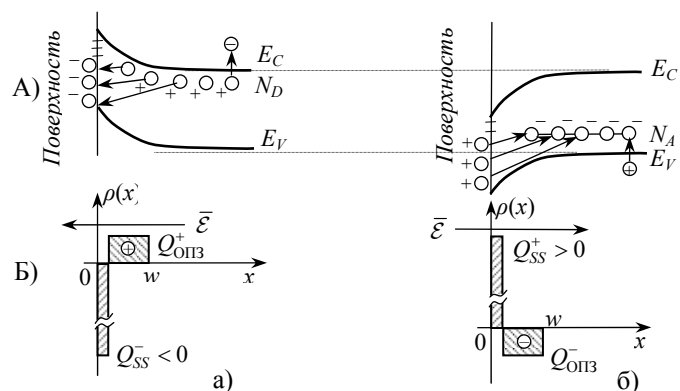


Рис. 1.23. Модель образования заряда поверхности

А) Зонные диаграммы; Б) Заряды в поверхностном слое;

а) электронный полупроводник, приход электронов на поверхностные состояния, заряд поверхности  $Q_{SS} < 0$ ; б) дырочный полупроводник, уход электронов с поверхностных состояний, заряд поверхности  $Q_{SS} > 0$ .

Аналогично, уход электронов с поверхностных состояний на уровни акцепторной примеси  $p$ -полупроводника заряжает поверхность положительно и создаёт в прилегающей области отрицательный объёмный заряд ионов акцепторной примеси  $Q_{\text{ОПЗ}}^- = qN_A^-w < 0$ , где  $N_A$  – концентрация акцепторной примеси,  $w$  – размер ОПЗ. Созданное электрическое поле на-

правлено от поверхности в объём (рис. 1.22а, рис. 23б). Зоны изгибаются вниз, что соответствует соотношениям (1.29), (1.20).

### Распределение концентрации носителей заряда по энергиям

Распределения концентраций носителей заряда по энергиям (1.6а), (1.6б) выражаются через произведение максимальной плотности разрешённых состояний (1.14) на вероятность занятия энергетического уровня (1.5) или (1.7). Процедура графического перемножения этих кривых пояснена на рис. 1.24 на примере собственного полупроводника.

На рис. 1.24а во встречно-вложенной системе координат, которая здесь в отличие от рис. 1.16 имеет вертикальную ориентацию, изображены графики используемых функций. Функция Ферми–Дирака изображена сплошной жирной кривой и расположена таким образом, чтобы кривая пересекала уровень Ферми  $E_i$  на ординате, равной 0,5. Графики функций (1.14) максимальной плотности разрешённых состояний электронов  $n_{\text{max}}(E)$  и дырок  $p_{\text{max}}(E)$ , расположенные соответственно в зоне проводимости и валентной зоне, изображены жирными штриховыми кривыми. Предполагается, что положение осей ординат этих кривых совпадает с энергетическими уровнями  $E_C$ ,  $E_V$  соответственно, а их направления – противоположны и совпадают с направлениями осей ординат соответствующих функций распределений  $n(E)$  и  $p(E)$ . Искомые графики распределения концентрации носителей заряда по энергиям (результат перемножения) заштрихованы.

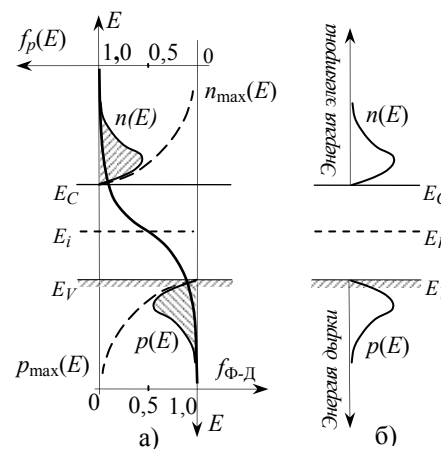


Рис. 1.24. К процедуре графического определения распределения концентрации носителей в собственном полупроводнике

а) графики во встречно-вложенной системе координат функций Ферми–Дирака (жирная кривая), максимальной плотности разрешённых состояний (штриховая кривая), распределения электронов и дырок (заштриховано),

б) распределение концентрации носителей заряда по энергиям.

На рис. 1.24б приведены результаты графических перемножений в общепринятом виде. Оси ординат графиков направлены одинаково слева направо вдоль уровней  $E_C$ ,  $E_V$ . Кривые распределений электронов и дырок идентичны и ограничивают равновеликие площади, что отражает равенство концентраций электронов и дырок  $n = p = n_i$  в собственном полупроводнике. Носители зарядов распределены в разрешённых зонах неравномерно. Основное количество свободных электронов находится выше дна зоны проводимости в пределах области, занимающей несколько  $kT$ . Энергия электронов отсчитывается «вверх». Поэтому указанное распределение соответствует минимуму их потенциальной энергии. Основное количество дырок сосредоточено в пределах области, занимающей несколько  $kT$  ниже вершины валентной зоны. Энергия дырок отсчитывается «вниз», поэтому минимум их потенциальной энергии соответствует вершине валентной зоны. Максимум распределений находится в пределах единиц  $kT$  от границ разрешённых зон.

По аналогичной процедуре построены кривые распределения концентрации носителей заряда в примесных полупроводниках, приведённые на рис. 1.25.

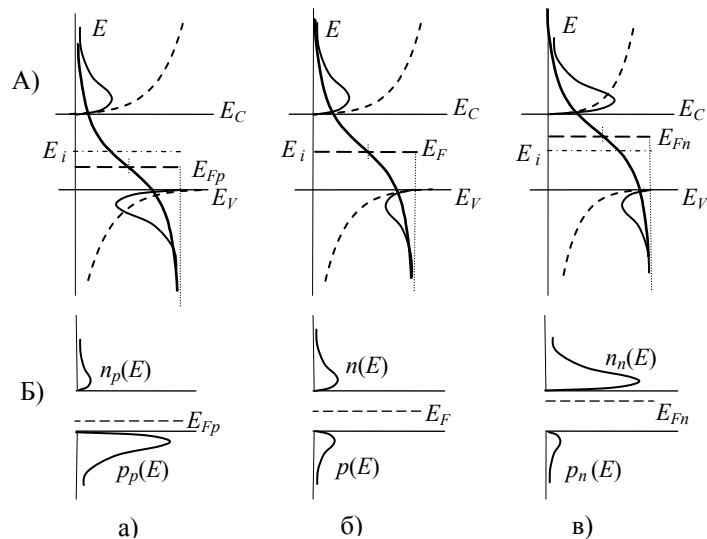


Рис. 1.25. Распределение концентраций носителей заряда по энергиям в дырочном (а), собственном (б), электронном (в) полупроводниках

А) графики функций, используемых для построения распределений.  
 Б) распределение концентраций носителей заряда по энергиям.  
 $p_p(E)$ ,  $n_n(E)$  – распределение ОНЗ,  $p_n(E)$ ,  $n_p(E)$  – распределение ННЗ.

На рисунках 1.25А показаны графики используемых функций. Тонким пунктиром обозначен также прямоугольный график функции Ферми–Дирака для температуры  $T = 0$  К. Его горизонтальная грань является уровнем Ферми соответствующего полупроводника. Кривые функций Ферми–Дирака для  $T > 0$  пересекают этот график ровно по уровню 0,5 при любом положении уровня Ферми в запрещённой зоне.

Уровень Ферми  $E_{Fp}$  дырочного полупроводника располагается ниже середины запрещённой зоны. Следствием становится смещение графика функции Ферми–Дирака вниз, уменьшение площади под кривой распределения  $n_p(E)$  ННЗ-электронов в зоне проводимости и соответствующее увеличение площади под кривой распределения  $p_p(E)$  ОНЗ-дырок в валентной зоне, поскольку в дырочном полупроводнике  $p_{p0} \gg n_{p0}$ .

Уровень Ферми  $E_{Fn}$  электронного полупроводника располагается выше середины запрещённой зоны. График функции Ферми–Дирака смещается вверх. Площадь под кривой распределения  $n_n(E)$  ОНЗ-электронов в зоне проводимости увеличивается, а площадь под кривой распределения  $p_n(E)$  ННЗ-дырок валентной зоне соответственно уменьшается, поскольку в электронном полупроводнике  $n_{n0} \gg p_{n0}$ .

Для удобства сравнения на рис. 1.25б приведено также распределение СНЗ в собственном полупроводнике, в ином масштабе повторяющее рис. 1.24. Здесь площади под кривыми распределения одинаковы, поскольку в собственном полупроводнике  $n = p = n_i$ . По идее площади под кривыми распределений примесных полупроводников должны быть такими, чтобы выполнялось условие (1.18а). Однако в силу естественных причин и ради наглядности рисунка это условие не соблюдено.

### 1.3. Рекомбинация носителей заряда

От динамического равновесия между процессом генерации носителей и обратным процессом их рекомбинации зависит концентрация свободных носителей заряда в полупроводнике.

На энергетической диаграмме акт генерации интерпретируется как переход электрона из валентной зоны в зону проводимости и обозначается стрелкой (рис. 1.26). Энергия, необходимая для разрыва валентной связи и переброса электрона в зону проводимости должна быть равна, по крайней мере, ширине запрещённой зоны

Рекомбинация есть обратный процесс перехода электрона из зоны проводимости в валентную зону. Обозначается стрелкой противоположного направления.

В электронно-дырочных генерационно-рекомбинационных процессах, как и при взаимодействиях других элементарных частиц, должны

выполняться законы сохранения энергии и импульса. Из закона сохранения энергии следует, что рекомбинация сопровождается выделением энергии, в частности, излучением. Такая рекомбинация, когда свободный электрон из зоны проводимости рекомбинирует со свободной дыркой из валентной зоны в одном элементарном акте, называется **прямой рекомбинацией зона-зона** в частности, **прямой излучательной рекомбинацией** (рис. 1.26а). Однако вероятность прямой излучательной рекомбинации у широкозонных полупроводников мала.

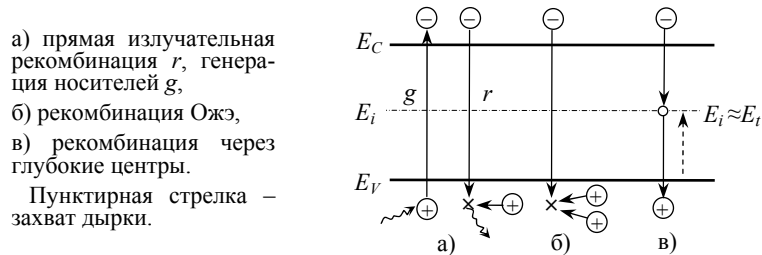


Рис. 1.26. Механизмы рекомбинации носителей заряда

Необходимость выполнения закона сохранения импульса накладывает жёсткие ограничения на характер взаимодействия частиц и возможность акта рекомбинации. Прямая излучательная рекомбинация осложняется тем, что импульс фотона  $p_{\text{фТ}} = h\nu/c$  очень мал. Отсюда следует, что рекомбинация возможна, если электрон и дырка имеют приблизительно одинаковые и встречно направленные импульсы. (рис. 1.27).

Гораздо более вероятными являются механизмы рекомбинации с участием третьего тела, за счёт которого облегчается выполнение законов сохранения. Таковыми могут быть ещё один свободный носитель (рис. 1.26б, рекомбинация Ожэ) или центры рекомбинации (ловушки) в запрещённой зоне. Роль ловушек могут играть, например, примесные атомы, либо различные структурные дефекты решётки (рис. 1.26в, рис. 1.28).

Наиболее распространёнными являются **непрямые механизмы рекомбинации**, через (ловушки) центры рекомбинации (уровни разрешённых состояний), расположенные в запрещённой зоне. Переход электрона из зоны проводимости в валентную зону и рекомбинация с дыркой происходит двумя этапами (рис. 1.28в). Сначала свободный электрон захватывается ловушкой – переходит из зоны проводимости на пустой энергетический уровень ловушки. Затем происходит захват дырки из валентной зоны на заполненный электроном центр рекомбинации и сам

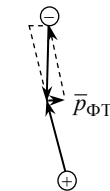


Рис. 1.27. Суммирование импульсов

акт рекомбинации. Свободная дырка рекомбинирует со связанным (третьим участником) электроном. Именно это облегчает выполнение законов сохранения. Последний этап эквивалентен эмиссии (переходу) электрона с уровня ловушки в валентную зону.

Ловушки мелкого залегания, расположенные в запрещённой зоне около дна зоны проводимости или вершины валентной зоны, захватывают носитель заряда одного знака (рис. 1.28а,б). Однако из-за значительного энергетического интервала вероятность захвата носителя заряда противоположного знака и, следовательно, вероятность рекомбинации невелика. Возвращая через некоторое время захваченный носитель в зону, такие ловушки существенно влияют на время жизни.

Глубокие ловушки (центры рекомбинации), расположенные примерно в середине запрещённой зоны, в силу своей равноудалённости обладают примерно одинаковой вероятностью захвата носителей заряда обоих знаков (рис. 1.28в). Вероятность рекомбинации максимальна.

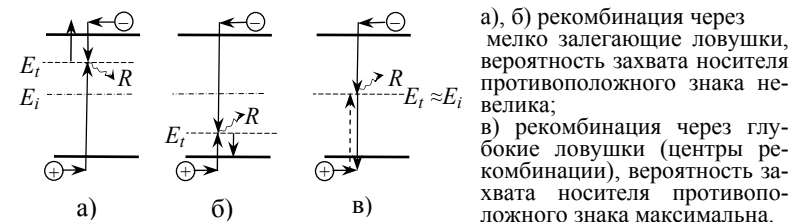


Рис. 1.28. Рекомбинация через ловушки и центры рекомбинации

Мерой генерационно-рекомбинационных процессов является число актов рекомбинации в единице объёма за единицу времени  $R$  [ $\text{см}^{-3}\text{с}^{-1}$ ]. Эту величину принято называть **скоростью рекомбинации**. Поскольку процессы генерации и рекомбинации происходят параллельно, и рекомбинируют носители парами, формула для результирующей скорости рекомбинации представляет собой разность скорости собственно процесса рекомбинации и процесса генерации:  $R = r - g$ , где  $r$  – истинная скорость рекомбинации,  $g$  – скорость генерации, обусловленные естественным внутренним (не внешним) термодинамическим процессам.

Согласно теории Шокли-Рида-Холла результирующая скорость рекомбинации через ловушки определяется соотношением<sup>25</sup>

$$R = r - g = \frac{pn - n_0 p_0}{(n + n_1)\tau_p + (p + p_1)\tau_n} = \frac{pn - n_i^2}{(n + n_1)\tau_p + (p + p_1)\tau_n}, \quad (1.30)$$

где  $n_i^2 = n_0 p_0$  – равновесное значение произведения концентраций (1.18а),  $p_1, n_1$  – концентрации дырок и электронов в случае, когда уровень

Ферми совпадает с уровнем ловушек (или центров рекомбинации), через которые осуществляются генерационно-рекомбинационные переходы носителей,  $\tau_p$  – время жизни дырок в материале  $n$ -типа,  $\tau_n$  – время жизни электронов в материале  $p$ -типа<sup>26</sup>.

Если  $E_t$  – энергетический уровень ловушек<sup>27</sup>, то согласно (1.20)

$$n_1 = n_i \exp[(E_t - E_i)/\kappa T], \quad p_1 = p_i \exp[-(E_t - E_i)/\kappa T]. \quad \text{Тогда}$$

$$R = r - g = \frac{pn - n_i^2}{\left[ n + n_i \exp\left(\frac{E_t - E_i}{\kappa T}\right) \right] \tau_p + \left[ p + p_i \exp\left(-\frac{E_t - E_i}{\kappa T}\right) \right] \tau_n}. \quad (1.31)$$

Полагая, что в силу малой концентрации ловушек времена жизни носителей одинаковы  $\tau_p = \tau_n \equiv \tau$ , отсюда получим

$$R = \frac{pn - n_i^2}{\tau \left[ p + n + 2n_i \operatorname{ch}\left(\frac{E_t - E_i}{\kappa T}\right) \right]}. \quad (1.32)$$

Другим важным параметром полупроводника, органически связанным с рекомбинацией, является **время жизни**. С точки зрения эксплуатационных характеристик время жизни относится к числу основных факторов, влияющих на инерционность полупроводниковых приборов.

Установим зависимость между временем жизни носителей заряда и скоростью их рекомбинации. Для определённости рассмотрим электронный полупроводник, у которого  $n_{n0} \gg p_{n0}$ . Пусть  $n_n = n_{n0} + \Delta n_n$ ,  $p_n = p_{n0} + \Delta p_n$  и  $\Delta n_n = \Delta p_n$ . Тогда, подставив эти значения в (1.30), получим

$$R = \frac{(n_{n0} + p_{n0} + \Delta p_n) \Delta p_n}{(n_{n0} + \Delta n_n + n_1) \tau_p + (p_{n0} + \Delta p_n + p_1) \tau_n}.$$

Отсюда для низкого уровня инжекции  $(\Delta p_n / n_{n0}) \ll 1$  определим

$$R = \Delta p_n / \tau_p = (p_n - p_{n0}) / \tau_p = R_p \Rightarrow \quad (1.33a)$$

$$\Rightarrow \Delta p_n = R_p \tau_p, \quad (1.33б)$$

где  $R_p$  – скорость рекомбинации ННЗ-дырок в электронном полупроводнике. В приближении  $\Delta n_p / p_{p0} \ll 1$  аналогичные соотношения нетрудно получить для дырочного полупроводника:

$$R = \Delta n_p / \tau_n = (n_p - n_{p0}) / \tau_n = R_n \Rightarrow \quad (1.34a)$$

$$\Rightarrow \Delta n_p = R_n \tau_n, \quad (1.34б)$$

где  $R_n$  – скорость рекомбинации ННЗ-электронов.

Таким образом, в случае непрямых механизмов рекомбинации время жизни, скорость рекомбинации и невысокая избыточная концентрация неосновных носителей связаны линейными зависимостями (1.33), (1.34). Нередко поэтому непрямую рекомбинацию называют **линейной рекомбинацией**, в отличие от прямой межзонной рекомбинации, где указанные зависимости не являются линейными.

Эксперименты свидетельствуют о том, что основную роль в процессах естественной рекомбинации играют непрямые механизмы. В частности, если бы в кремнии излучательная рекомбинация была основным процессом, то время жизни электронно-дырочных пар составило бы величину порядка 3-х часов. Тогда как наблюдаемое максимальное время жизни в чистом кремнии при комнатной температуре на шесть порядков меньше (не превышают 3-х миллисекунд). С другой стороны время жизни очень чувствительно к содержанию примесей некоторых металлов, играющих роль глубоких центров рекомбинации. Например, внедрение 1 атома золота на  $10^7$  атомов германия (концентрация примеси  $\approx 10^{15} \text{ см}^{-3}$ ) снижает время жизни на 6 порядков с  $10^{-3} \text{ с}$  до  $10^{-8} \div 10^{-9} \text{ с}$ .

До сих пор наше рассмотрение касалось объёмной рекомбинации. Скорость **поверхностной рекомбинации** принято характеризовать несколько иным образом.

Поверхность, нарушая периодическую структуру объёмной кристаллической решётки, представляет собой значительную неоднородность, что создаёт большое количество разрешённых состояний в запрещённой зоне поверхностного слоя полупроводника. Наличие разрешённых состояний повышает скорость рекомбинации на поверхности и вызывает диффузионный поток необходимых для рекомбинации носителей, направленный из объёма к поверхности. Величина потока будет тем больше, чем больше актов рекомбинации, значит, чем выше скорость натекающего потока. Темп рекомбинации на поверхности принято характеризовать **скоростью поверхностной рекомбинации**, которая и есть скорость потока, расходуемого на рекомбинацию носителей.

Если  $\Delta p$  – избыточная концентрация расходуемых на рекомбинацию носителей, то (см. (1.36б)) за единицу времени на единичной площади рекомбинируют  $\Pi_p = S_p \Delta p \left[ \text{см}^{-2} \text{с}^{-1} \right]$  носителей заряда, где

$$S_p = \Pi_p / \Delta p \left[ \text{см/с} \right] \quad (1.35)$$

– скорость потока, являющаяся в данном случае скоростью поверхностной рекомбинации. В частности, на совершенном омическом контакте металл–полупроводник носители заряда не скапливаются,  $(\Delta p, \Delta n) = 0$ , поэтому  $S \rightarrow \infty$ .

## 1.4. Электрические токи в полупроводниках

В природе существует электрический ток проводимости и электрический ток смещения. Ток проводимости формируется за счёт направленного переноса электрических зарядов (электронов и дырок). Ток смещения создаётся изменением (во времени) напряжённости электрического поля<sup>28</sup>.

Явления направленного переноса (частиц, зарядов, жидкости, энергии, газа, фотонов, ...) описываются через понятие «потока». Поток характеризуется направлением, средней скоростью, плотностью.

Плотность потока частиц  $\Pi$  [см<sup>-2</sup>с<sup>-1</sup>] есть число частиц, пересекающих единичную площадь, ортогональную потоку, за единицу времени. За единицу времени частица (поток) проходит путь, численно равный средней скорости. Поэтому независимо от механизмов переноса и природы частиц плотность потока равна произведению концентрации на среднюю скорость частиц.

Поток зарядов создаёт электрический ток проводимости. Очевидно, что плотность тока проводимости  $J$  [А/см<sup>2</sup>] есть произведение элементарного заряда на плотность потока зарядов  $J = q\Pi$  [А см<sup>-2</sup>]. Плотность дырочного тока  $J_p = q\Pi_p$ , плотность электронного тока  $J_n = (-q) \times \Pi_n = -q\Pi_n$ . Знак характеризует направление тока относительно положительного направления пространственной координаты.

В полупроводниках направленный перенос свободных зарядов осуществляется за счёт двух механизмов:

- **дрейфа** в электрическом поле под действием разности потенциалов пропорционально градиенту потенциала, т.е. напряжённости электрического поля, и

- **диффузии** под действием разности концентраций пропорционально градиенту концентрации носителей зарядов (закон Фике).

Соответственно различают дрейфовый и диффузионный токи проводимости. Однако, независимо от механизма переноса плотности электронного и дырочного потоков и токов проводимости выражаются через произведение концентрации на среднюю скорость  $U_n, U_p$  диффузии или дрейфа соответствующих носителей:

$$\Pi_n = nU_n, \quad J_n = -qnU_n, \quad (1.36a)$$

$$\Pi_p = pU_p, \quad J_p = qpU_p. \quad (1.36b)$$

### Диффузионный ток

Диффузионный перенос зарядов осуществляется из того места, где их концентрация выше, в то место, где их концентрация ниже, т.е. в сто-

рону, противоположную направлению градиента концентрации. Поэтому в математические выражения диффузионного потока градиент концентрации всегда входит со знаком минус.

Основным законом диффузии в неподвижной среде является закон Фике, согласно которому плотность диффузионного потока пропорциональна градиенту концентрации. Коэффициентом пропорциональности служит коэффициент диффузии  $D$  [см<sup>2</sup>/с].

Плотности электронных и дырочных диффузионных потоков и токов будут равны соответственно:

$$\Pi_{p \text{ диф}} = D_p \times (-\nabla p) = -D_p \text{grad } p, \quad J_{p \text{ диф}} = q\Pi_{p \text{ диф}} = -qD_p \nabla p, \quad (1.37)$$

$$\Pi_{n \text{ диф}} = D_n \times (-\nabla n) = -D_n \text{grad } n, \quad J_{n \text{ диф}} = (-q) \times (-D_n \nabla n) = qD_n \nabla n. \quad (1.38)$$

Диффузионные токи имеют противоположные направления. Хотя под действием градиента электроны и дырки движутся в одном и том же направлении, из-за разных знаков зарядов их диффузионные токи противоположны.

Таким образом,

- как электронный, так и дырочный диффузионный поток всегда направлен из того места, где концентрация выше, в то место, где концентрация ниже,

- дырочный диффузионный ток всегда совпадает с направлением диффузии,

- электронный диффузионный ток всегда противоположен направлению диффузии (из-за отрицательного заряда) (рис. 1.29).

Анализируя диффузионные токи, мы полагали, что по умолчанию диффузионные **потоки** направлены одинаково. В этом случае

электронные и дырочные диффузионные **токи** имеют противоположные направления. Значит, при биполярной проводимости результирующий диффузионный ток является разностью составляющих и может быть равен нулю. Однако если диффузионные потоки направлены противоположно, то диффузионные токи электронов и дырок складываются:

Диффузионные токи характерны для полупроводников и отсутствуют в металлах. Причина в том, что в металлах, где много высокопо-

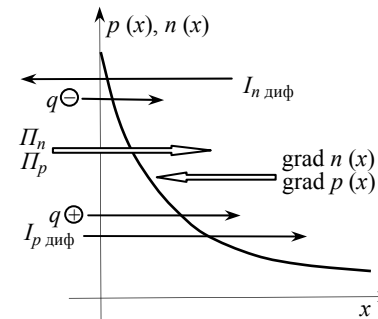


Рис. 1.29. Диффузионные потоки и токи для приведённого распределения электронов и дырок

Диффузионные потоки электронов и дырок направлены одинаково, токи – противоположно.

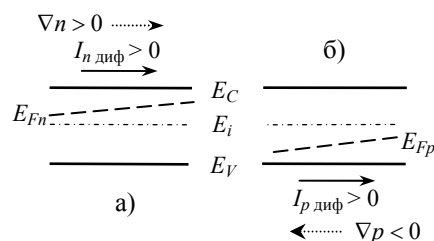
движных свободных электронов, нельзя создать пространственно разнесённые области с различающейся плотностью свободных зарядов. В полупроводниках СНЗ на много порядков меньше. Концентрации электронов и дырок могут изменяться в широких пределах при постоянной суммарной концентрации зарядов.

**Пример 1.1.** Нарисовать энергетическую зонную диаграмму электронного/дырочного полупроводника, по которому проходит диффузионный ток заданного направления.

Пусть направление электронного тока  $I_n > 0$  совпадает с положительным направлением оси  $x$  (слева направо). По условию ток диффузионный. Значит, электрическое поле отсутствует и согласно (1.29)  $\nabla E_i = 0$   $E_i = \text{const}$ . Следовательно, энергетические зоны горизонтальны.

Согласно (1.38) направление электронного тока совпадает с направлением градиента концентрации. Значит, градиент направлен по оси  $x$ . Отсюда следует, что концентрация электронов должна увеличиваться слева направо. В этом же направлении должен повышаться уровень Ферми, поскольку согласно (1.20)  $n = n_i \exp[(E_F - E_i)/kT]$  и увеличение концентрации электронов при  $E_i = \text{const}$  должно сопровождаться повышением уровня Ферми. Значит, зонная диаграмма электронного полупроводника должна иметь вид, показанный на рис. (1.30а).

Рис. 1.30. Зонные диаграммы полупроводников при прохождении диффузионного тока  
а) электронный полупроводник,  
б) дырочный полупроводник.



Аналогично нетрудно показать, что энергетическая диаграмма дырочного полупроводника должна иметь вид, приведённый на рис. (1.30б).

### Дрейфовый ток.

В электрическом поле дырки движутся (дрейфуют) по направлению вектора напряжённости. По физическому смыслу напряжённость есть сила, действующая на положительный единичный заряд. Отрицательные электроны движутся в электрическом поле в противоположном направлении.

Воздействие электрического поля на электрон в вакууме приводит его в состояние равномерно-ускоренного прямолинейного движения. Иной характер движения наблюдается в кристалле. Особенности дрейфа в полупроводнике обусловлены тем, что носители заряда, ускоряясь в электрическом поле, приобретают дополнительную энергию и в процессе движения отдают её решётке. Двигаясь в кристалле, электрон испытывает столкновения с колеблющимися узлами кристаллической решётки, дефектами периодической структуры, рассеяние на ионизированных атомах примеси и т.д. Из-за этого при сохранении среднего направления дрейфа траектория электрона приобретает вид кусочно-ломанной кривой, в узлах которой электрон практически теряет свою энергию, а затем опять ускоряется (рис. 1.31). Тем не менее, в умеренных полях средняя скорость дрейфа прямо пропорциональна напряжённости электрического поля

$$v_{др} = \mu \mathcal{E}. \quad (1.39)$$

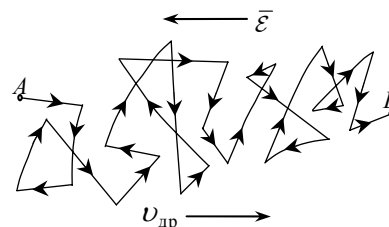


Рис. 1.31. Характер движения электрона в электрическом поле кристалла

Коэффициентом пропорциональности является *подвижность* электронов  $\mu_n$  [см<sup>2</sup>/В·с] или дырок  $\mu_p$ . Численно подвижность равна скорости дрейфа, которую приобретает электрон в единичном поле  $\mathcal{E} = 1$  [В/см].

Для определения плотности дрейфового потока дырок выделим в потоке дырок трубку тока, представляющую собой прямой круглый цилиндр с площадью основания 1 см<sup>2</sup>, образующая которого совпадает с направлением (вектора напряжённости электрического поля) скорости потока (рис. 1.32).

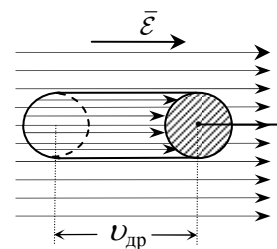


Рис. 1.32 Трубка тока в потоке дырок

За секунду выделенное штриховкой основание цилиндра пересекут все дырки, отстоящие от основания не далее расстояния, численно равного скорости дрейфа  $v_{p,др}$  (пути, пройденному за единицу времени), т.е. находящиеся в объёме, равном  $v_{p,др} \times 1$ . Значит, плотность потока прошедших дырок равна произведению концентрации дырок на объём выделенной трубки тока

$\Pi_p = p v_{p,др} = p \mu_p \mathcal{E}$ . Аналогично, плотность потока электронов будет

равна  $\Pi_n = -n\upsilon_{n\text{др}} = -n\mu_n\mathcal{E}$ . Знак минус учитывает противоположное направление дрейфовой скорости электронов.

Проведённый вывод является обоснованием ранее сделанного утверждения о том, что независимо от механизмов переноса плотность потока частиц равна произведению концентрации на среднюю скорость потока. Тогда плотности дырочных и электронных дрейфовых токов будут равны соответственно:

$$J_{p\text{др}} = q\Pi_{p\text{др}} = q\rho\mu_p\mathcal{E}, \quad (1.40)$$

$$J_{n\text{др}} = (-q) \times (-\Pi_{n\text{др}}) = q\Pi_{n\text{др}} = qn\mu_n\mathcal{E}. \quad (1.41)$$

Эти соотношения раскрывают выражения (1.36) для дрейфового тока.

Таким образом,

о дрейфовые электронный и дырочный токи всегда имеют одинаковое направление, совпадающее с направлением вектора напряжённости электрического поля (потока дырок),

о дырочные дрейфовые потоки всегда направлены по напряжённости электрического поля, а электронные – против,

о дырочный дрейфовый ток всегда совпадает с направлением дрейфа дырок, а электронный ток противоположен направлению дрейфа электронов (из-за отрицательного заряда электрона, рис. 1.33).

В случае биполярной проводимости электронная и дырочная составляющие дрейфового тока суммируются:

$$I_{\text{др}} = S(J_{p\text{др}} + J_{n\text{др}}) = S(q\rho\mu_p + qn\mu_n)\mathcal{E} \equiv S\sigma_0\mathcal{E}, \quad (1.42)$$

где  $\sigma_0 = (q\rho\mu_p + qn\mu_n) \equiv (\sigma_p + \sigma_n)[\text{Ом} \cdot \text{см}]^{-1}$  (1.43)

**удельная проводимость** полупроводника – проводимость  $1 \text{ см}^3$  (единицы объёма) полупроводникового материала,  $S$  – площадь поперечного сечения.

**Пример 1.2.** Нарисовать энергетическую зонную диаграмму электронного/дырочного полупроводника, по которому протекает дрейфовый ток заданного направления.

Пусть направление электронного тока (справа налево) противоположно направлению оси  $x$ , т. е.  $I_n < 0$ . По условию ток дрейфовый. Значит, из (1.41) следует, что направление тока совпадает с направлением элек-

трического поля и вектор напряжённости  $\vec{\mathcal{E}} < 0$ . Согласно (1.29) область электрического поля сопровождается наклоном энергетических зон, а напряжённость поля пропорциональна градиенту (тангенсу угла наклона а) энергетических зон. Но отрицательная производная означает, что

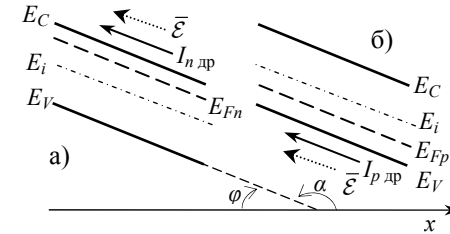


Рис. 1.34. Зонные диаграммы полупроводников при прохождении дрейфового отрицательного тока а) электронный полупроводник, б) дырочный полупроводник.

$\text{tg}\alpha > \pi/2$ . Отсюда следует, что как собственный уровень электронного полупроводника, так и энергетические зоны будут наклонены к оси  $x$  под отрицательным (отсчитываемым по часовой стрелке) углом  $\varphi = 180^\circ - \alpha$ . Энергетическая диаграмма электронного полупроводника должна иметь вид, показанный на рис. (1.34а).

Дрейфовые токи электронов и дырок совпадают по направлению. Значит, наклоны энергетических зон дырочного и электронного полупроводников тоже должны совпадать (рис. 1.34б). Однако уровень Ферми  $E_{Fp}$  должен располагаться ниже собственного уровня  $E_i$  (см. рис. 1.20 и соотношение (1.28)).

### Уравнения токов

При наличии как электрического поля, так и градиента концентрации носителей заряда ток проводимости будет содержать дрейфовую и диффузионную составляющие. При этом полный ток будет включать также ток смещения, если электрическое поле переменное.

С учётом соотношений (1.38), (1.41) плотность электронной составляющей тока проводимости есть сумма

$$J_n = J_{n\text{др}} + J_{n\text{диф}} = qn\mu_n\mathcal{E} + qD_n\text{grad}n. \quad (1.44)$$

Плотность дырочной составляющей с учётом (1.37), (1.40) равна

$$J_p = J_{p\text{др}} + J_{p\text{диф}} = q\rho\mu_p\mathcal{E} - qD_p\text{grad}p. \quad (1.45)$$

Полный ток равен сумме токов проводимости и смещения

$$I = S \left( J_n + J_p + \varepsilon_{\text{ПП-к}} \varepsilon_0 \frac{\partial \mathcal{E}}{\partial t} \right),$$



где  $\epsilon_0$  – электрическая постоянная, зависящая от выбора системы единиц,  $\epsilon_{\text{ПП-к}}$  – относительная диэлектрическая проницаемость полупроводника.

Из (1.44), (1.45) следует, что управление дрейфовыми токами сводится к управлению напряжённостью (потенциалом) электрического поля. В частности, ниже будет показано, что управлять дрейфовым током при данном напряжении в конкретном полупроводниковом образце можно также посредством изменения его геометрических размеров.

Управление диффузионным током сводится к управлению градиентами концентраций носителей заряда. Определение градиентов требует знания пространственного распределения концентрации носителей. Таким образом, хотя оба тока являются токами проводимости, способы управления существенно различаются. Правда, в конечном счёте, управление обоими токами производится с помощью напряжения, поскольку градиенты концентрации также зависят от приложенного напряжения.

### Температурная зависимость тока, температурный коэффициент

Как видно из (1.42), физическими причинами температурной зависимости дрейфового тока могут быть температурные изменения концентрации носителей заряда  $n(T)$ ,  $p(T)$  и подвижности  $\mu(T)$ . Концентрация основных носителей в области температур примесного истощения практически не зависит от температуры (рис. 1.17). В этой области температурная зависимость тока определяется температурными изменениями подвижности

$$\mu(T) = \text{const} \times T^{-3/2}$$

С повышением температуры подвижность снижается (рис. 1.35) [4].

Физическая природа температурной зависимости подвижности основных и неосновных носителей одинакова. С ростом температуры увеличивается амплитуда и частота тепловых колебаний решётки. Растёт число актов рассеяния носителей. Подвижность электронов и дырок, значит, и дрейфовый ток уменьшается

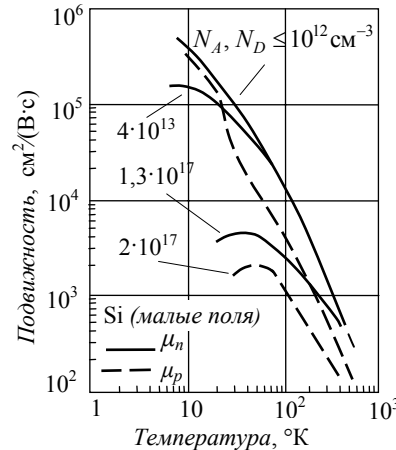


Рис. 1.35. Температурная зависимость подвижности носителей заряда в кремнии  
 Параметр – концентрация примесей.

Однако концентрация неосновных носителей (1.27) экспоненциально резко увеличивается при возрастании температуры, поскольку прямо пропорциональна квадрату собственной концентрации (1.18). Действительно, из (1.27), (1.18) следует, что

$$p_{n0}(T) \approx n_i^2 / N_D = (N_C N_V / N_D) e^{-E_g / \kappa T},$$

$$n_{p0}(T) \approx n_i^2 / N_A = (N_C N_V / N_A) e^{-E_g / \kappa T}.$$

Экспоненциальная зависимость собственной концентрации от температуры (рис. 1.36) приводит к сильной температурной зависимости тока ННЗ и влияет на температурную стабильность приборов, работа которых основана на ННЗ.

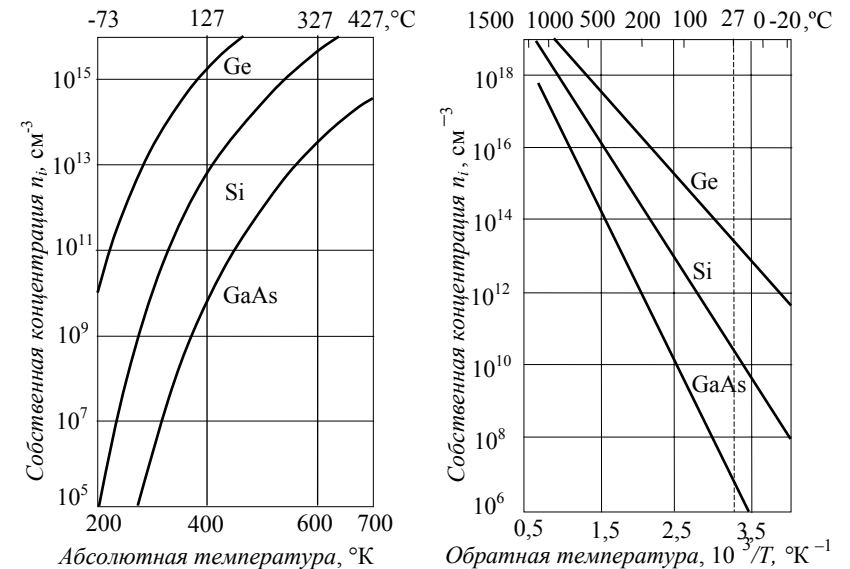


Рис. 1.36. Температурная зависимость собственной концентрации в Ge, Si и GaAs

Собственная концентрация экспоненциально растёт при повышении температуры и уменьшении ширины запрещенной зоны.

Такая многопараметрическая (концентрация, подвижность) и к тому же разнонаправленная температурная зависимость тока дополняется определяющим влиянием температурных зависимостей происходящих физических процессов. В следующих главах мы увидим, что, например, дрейфовый ток основных носителей полупроводникового резистора (при постоянном напряжении) с повышением температуры уменьшается. В то же время ток основных носителей диода Шоттки на контакте металл-

полупроводник экспоненциально возрастает. Ток полевого транзистора с индуцированным каналом при возрастании температуры уменьшается так же, как у резистора, а ток биполярного транзистора увеличивается.

Величину и направление температурных изменений параметров полупроводниковых приборов принято оценивать посредством **температурных коэффициентов**. Температурный коэффициент определяется как производная от температурной зависимости соответствующего параметра. Численно он равен изменению параметра при изменении температуры на один градус шкалы Цельсия или Кельвина. Например, температурный коэффициент прямого тока (ТКПТ) есть производная

$$\text{ТКПТ} = \left. \frac{dI(T)}{dT} \right|_{U=\text{const}} \approx \frac{\Delta I}{\Delta T} \Big|_{\Delta T=1^\circ\text{C}} = \Delta I \left[ \text{A}/^\circ\text{C} \right]. \quad (1.46a)$$

**Абсолютный температурный коэффициент** – размерная величина, в данном случае равная изменению тока в амперах при изменении температуры на  $1^\circ\text{C}$ .

Согласно общематематическому смыслу произвольной модуль температурного коэффициента характеризует величину и скорость изменения параметра при изменении температуры. Чем больше модуль, тем выше крутизна (угол наклона касательной к температурной зависимости), сильнее и быстрее изменяется параметр. И наоборот, небольшой температурный коэффициент свидетельствует о незначительных и медленных температурных изменениях параметра. Знак температурного коэффициента характеризует направление изменения параметра. Положительный знак свидетельствует об увеличении параметра при увеличении температуры. Отрицательный – наоборот, об уменьшении параметра при возрастании температуры.

**Относительный температурный коэффициент** характеризует относительное (относительно текущего значения параметра) изменение параметра при изменении температуры на один градус. Например, относительный ТКПТ есть отношение

$$\frac{\text{ТКПТ}}{I(T)} = \left. \frac{dI/I}{dT} \right|_{U=\text{const}} \approx \frac{\Delta I/I}{\Delta T} \Big|_{\Delta T=1^\circ\text{C}} = \frac{\Delta I}{I(T)} \left[ ^\circ\text{C}^{-1} \right]. \quad (1.46b)$$

Относительные изменения нередко выражают в процентах. Тогда относительный ТКПТ, равный  $(\Delta I/I) \times 100 [\% \cdot ^\circ\text{C}^{-1}]$ , указывается в процентах изменения параметра в диапазоне рабочих температур. Уточнения «абсолютный», «относительный» обычно опускают. О температурном коэффициенте можно судить по размерности.

Относительный температурный коэффициент равен логарифмической производной, т. е. производной от логарифма температурной зависимости. Действительно, например, производная от логарифма тока равна

$$\frac{d \ln I(T)}{dT} = \frac{1}{I} \frac{dI(T)}{dT} = \frac{\text{ТКПТ}}{I} \left[ ^\circ\text{C}^{-1} \right]. \quad (1.46в)$$

### Сопротивление полупроводника

Для определения сопротивления полупроводника используем соотношение (1.42), представляющее собой дифференциальный (локальный) закон Ома. Если к полупроводниковому образцу с размерами  $h \times b \times l$  постоянного поперечного сечения  $S = h \times b$  приложено напряжение  $U$  (рис. 1.37), то дрейфовый ток  $I(x)$  в произвольном сечении  $0 \leq x \leq l$  будет равен

$$I(x) = SJ(x) = S\sigma_0 \mathcal{E}(x) = \sigma_0 bh \mathcal{E}(x) = -\sigma_0 bh \frac{d\varphi}{dx},$$

где напряжённость поля выражена через потенциал  $\varphi$ . Отсюда для тока через весь образец получим

$$I(x)dx = -\sigma_0 bh d\varphi \Rightarrow I = \int_0^l I(x)dx = -\sigma_0 bh \int_{\varphi(0)}^{\varphi(l)} d\varphi.$$

По условию **непрерывности тока** проводимости<sup>29</sup>

$$\text{div}J(x) = \frac{dJ(x)}{dx} = 0 \Rightarrow J = \text{const} \quad (1.47)$$

ток не зависит от координаты. Тогда, продолжая интегрирование, имеем:

$$\begin{aligned} I \times l = -\sigma_0 bh [\varphi(l) - \varphi(0)] &\Rightarrow I = -\sigma_0 (bh/l) (-U) = \sigma_0 (bh/l) U \equiv \sigma U \equiv \\ &\equiv U/R, \text{ где } \sigma = \sigma_0 (bh/l) \text{ [Ом}^{-1}\text{]} - \text{проводимость} \\ &\text{образца указанных размеров,} \\ R = \sigma^{-1} = \sigma_0^{-1} (l/bh) &\equiv \rho_0 (l/bh) \text{ [Ом]} - \text{сопротивление} \\ &\text{полупроводникового образца,} \\ \rho_0 = \sigma_0^{-1} \text{ [Ом} \cdot \text{см]} &- \text{удельное сопротивление} \\ &\text{полупроводника, } \varphi(l) = -U, \varphi(0) = 0. \end{aligned}$$

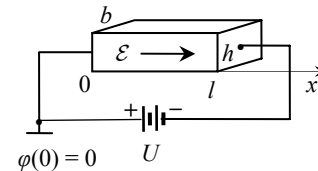


Рис. 1.37. Определение сопротивления полупроводникового образца

Законом Ома выражается прямая пропорциональность между напряжением (напряжённостью поля) и током. Из проведённого вывода следует, что прямая пропорциональность соблюдается до тех пор, пока дрейфовая скорость пропорциональна напряжённости поля (1.39).

Характер зависимость дрейфовой скорости от напряжённости поля определяется физическими механизмами, посредством которых носители заряда передают решётке избыточную энергию, приобретённую ими в электрическом поле. Фактическое сопротивление дрейфовому потоку носителей в полупроводнике определяется тем, насколько часто они теряют свою энергию, испытывая столкновения с узлами кристаллической ре-

шётки, дефектами периодической структуры, рассеяние на ионизированных атомах примеси<sup>30</sup> и т.д. Спецификой дрейфа обусловлен также характер зависимости самой дрейфовой скорости от напряжённости поля.

Эксперименты показывают, что дрейфовая скорость в кремнии и германии прямо пропорциональна напряжённости внешнего электрического поля вплоть до полей порядка  $5 \div 15$  кВ/см (рис. 1.38). В линейной области коэффициент пропорциональности (подвижность) не зависит от электрического поля. Рассеяние носителей происходит, в основном, на акустических (тепловых) колебаниях решётки. Однако в более сильных полях поток отбирает большую энергию от поля. Реализуются более энергоёмкие механизмы взаимодействия с решёткой, включая возбуждение оптических колебаний и процессы ударной ионизации атомов. При этом подвижность уменьшается. Рост дрейфовой скорости замедляется вплоть до **насыщения скорости**, когда дрейфовая скорость перестаёт зависеть от напряжённости поля. Хотя напряжённость поля (напряжение) растёт, электрический ток при насыщении дрейфовой скорости не изменяется, потому что подвижность снижается и произведение  $\mu\mathcal{E} = v_s = \text{const}$  остаётся постоянным. Дрейфовая скорость насыщения большинства полупроводников составляет величину порядка  $10^7$  см/с.

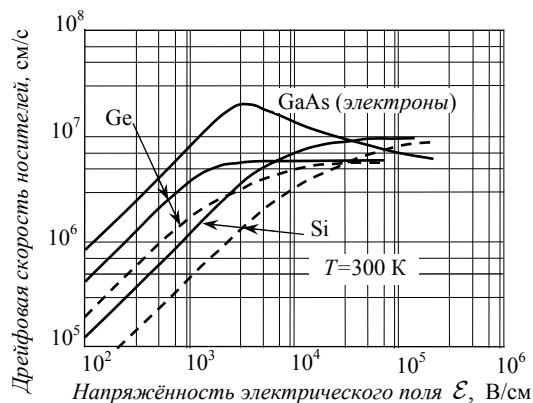


Рис. 1.38. Зависимость дрейфовой скорости от напряжённости электрического поля в германии, кремнии и арсениде галлия

Сплошные кривые – электроны, штриховая кривая – дырки [4].

Явление насыщения дрейфовой скорости используются при конструировании полупроводниковых приборов. В частности, в полевых транзисторах принципиальным является возможность получить режим насыщения тока за счёт насыщения дрейфовой скорости. Насыщение дрейфовой скорости является одной из причин независимости обратного тока идеализированного  $p-n$ -перехода от напряжения.

Обсуждаемые параметры некоторых собственных полупроводников при комнатной температуре  $T = 300$  К приведены в таблице 1.4.

Т а б л и ц а 1.4

	$\rho_0, \text{Ом}\cdot\text{см}$	$D_p, \text{см}^2/\text{с}$	$D_n, \text{см}^2/\text{с}$	$\mu_n, \text{см}^2/\text{В}\cdot\text{с}$	$\mu_p, \text{см}^2/\text{В}\cdot\text{с}$
Ge	45	47	99	3800	1800
Si	$2,3 \cdot 10^5$	13	34	1300	500
GaAs	$6,4 \cdot 10^7$			8500	450

### Соотношение Эйнштейна

Как при диффузии, так и при дрейфе, процесс направленного переноса носителей заряда в твёрдом теле сопровождается одними и теми же физическими явлениями. Среда, в которой происходит движение носителей, одинаково воздействует на движущийся поток независимо от причин, вызывающих перенос. Поэтому основные параметры, характеризующие дрейф и диффузию – подвижность и коэффициент диффузии – должны быть связаны между собой.

Связь между коэффициентом диффузии и подвижностью устанавливается **соотношением Эйнштейна**.

В неявном виде это соотношение содержат уравнения (1.44), (1.45), ибо в равновесном состоянии токи проводимости должны быть равны нулю. Физически это означает, что диффузионные токи, возникшие за счёт градиента концентрации носителей заряда, должны компенсироваться дрейфовыми токами за счёт возникающего поля. И наоборот.

Определим напряжённость возникающего электрического поля, если, например,  $\text{grad } n \neq 0$ . Используя соотношение (1.16), получаем:

$$\text{grad } n = \frac{dn}{dE_F} \frac{dE_F}{dx} = \frac{N_C e^{-(E_C - E_F)/kT}}{kT} \frac{dE_F}{dx} \Big|_{E_F=q\phi} = \left( \frac{n}{kT/q} \right) \frac{d\phi}{dx} = -\frac{n}{U_T} \mathcal{E},$$

где  $U_T = kT/q$  – **температурный потенциал**, равный 26 мВ для  $T = 300\text{K}$  ( $27^\circ\text{C}$ ). Отсюда  $\mathcal{E} = -(U_T/n)\text{grad } n$ .

Проведённые выкладки основаны на том, что концентрация  $\text{CH}_3$  однозначно определяется положением уровня Ферми. Правомерно поэтому выражать градиент концентрации носителей через сложную производную.

Подставив найденное значение напряжённости поля в (1.44) и приравняв ток нулю, определим:

$$D = (kT/q)\mu = U_T \mu. \quad (1.48a)$$

$$(D/\mu) = U_T, \quad (1.48б)$$

Соотношение Эйнштейна устанавливает прямую пропорциональность между коэффициентом диффузии и подвижностью. Коэффициентом пропорциональности служит температурный потенциал.

Это соотношение справедливо для невырожденных полупроводников, т. е. при относительно небольших концентрациях примесей, когда коэффициент диффузии и подвижность не зависят от концентрации носителей заряда, а уровень Ферми находится в запрещённой зоне.

## 1.5. Уравнения непрерывности

Известное из раздела «Электричество» курса общей физики уравнение непрерывности, которое в дифференциальной форме выражается как

$$\operatorname{div} J = - \frac{d\rho}{dt},$$

где  $\rho$  [Кл/см<sup>3</sup>] – объёмная плотность заряда, является математическим выражением постулата сохранения электричества (заряда). При анализе полупроводниковых приборов уравнение непрерывности используется в иной более специфической форме, имеющей тот же физический смысл, но выраженный через сохранение числа носителей заряда.

Поскольку  $\rho_p = qp$ ,  $\rho_n = -qn$ , отсюда получаем

$$\frac{\partial p}{\partial t} = - \frac{1}{q} \operatorname{div} J_p, \quad \frac{\partial n}{\partial t} = \frac{1}{q} \operatorname{div} J_n, \quad (1.49)$$

где  $J_p, J_n$  – плотности дырочных и электронных токов проводимости (1.45), (1.44) соответственно. В такой записи уравнения непрерывности выражают темп изменения концентрации носителей через изменения соответствующих потоков.

Однако в полупроводниках изменения концентрации носителей могут быть вызваны также генерацией, обусловленной внешним воздействием, которая неизбежно сопровождается естественной рекомбинацией. Тогда, учитывая в (1.49) темп объёмной внешней генерации носителей  $g$  [см<sup>-3</sup>с<sup>-1</sup>], например, за счёт поглощения оптического излучения и темп их естественной рекомбинации  $R$  [см<sup>-3</sup>с<sup>-1</sup>], приходим к следующему виду уравнений непрерывности:

$$\frac{\partial p}{\partial t} = g_p - R_p - 1/q \operatorname{div} J_p, \quad \frac{\partial n}{\partial t} = g_n - R_n + 1/q \operatorname{div} J_n. \quad (1.50)$$

Уравнения непрерывности являются следствием закона сохранения заряда. Закон сохранения заряда касается самих зарядов, а уравнения непрерывности – темпов их изменений во времени и в пространстве. Факти-

чески уравнения непрерывности выражают условия непрерывности потоков зарядов.

### Вывод уравнений

Для независимого вывода уравнения непрерывности, например, дырок выделим в пространстве дырок произвольный неподвижный объём, протяжённостью  $\Delta x$  с площадью поперечного сечения  $\Delta S$  и размером  $\Delta V = \Delta x \Delta S$  (рис. 1.39). Возможные изменения концентрации дырок в выделенном объёме могут быть обусловлены следующими причинами:

- внешней генерацией  $g_p$  [см<sup>-3</sup>с<sup>-1</sup>];

- естественной рекомбинацией, скорость которой есть разность

$$R_p = (r_p - g_T) [\text{см}^{-3}\text{с}^{-1}]$$

между скоростью естественной рекомбинации  $r_p$  и генерации  $g_T$ , обусловленными тепловыми переходами;

- изменениями потока

$$\Pi_p(x) - \Pi_p(x + \Delta x) = -\Delta \Pi_p.$$

Изменения концентрации  $\Delta p$  в единичном объёме за единицу времени пропорциональны разности

$$\Delta p \sim [g_p - (R_p + \Delta \Pi_p)].$$

Тогда за время  $\Delta t$  во всём выделенном объёме  $\Delta V = \Delta x \Delta S$  изменения концентрации составят величину

$$\Delta p \Delta x \Delta S = g_p \Delta x \Delta S \Delta t - R_p \Delta x \Delta S \Delta t - \Delta \Pi_p \Delta S \Delta t. \quad (1.51)$$

Разделив обе части этого уравнения на  $\Delta x \Delta S \Delta t$ , и перейдя к пределу при  $\Delta x, \Delta S, \Delta t \rightarrow 0$ , получим соотношение для темпа изменения концентрации дырок в локальном физически бесконечно малом объёме

$$\frac{\partial p}{\partial t} = g_p - R_p - \frac{d\Pi_p}{dx} = g_p - R_p - 1/q \frac{dJ_p}{dx} \equiv g_p - R_p - 1/q \nabla_x J_p,$$

которое и является дифференциальным уравнением непрерывности для дырок. Производная от плотности тока рассматривается здесь как составляющая дивергенции. Аналогично выводится уравнение непрерывности для электронов. Частная производная в левой части уравнений применена для того, чтобы подчеркнуть неподвижность рассматриваемого объёма.

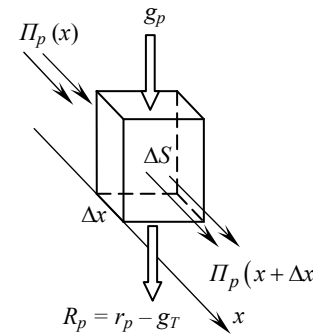


Рис. 1.39. Составляющие изменения концентрации дырок

Для вывода уравнения непрерывности в виде (1.50), явно содержащем дивергенцию тока, учтём, что изменения потока вектора тока  $\Delta I_p \Delta S = (1/q) \Delta J_p \Delta S$  в пределах некоторого объёма равны потоку вектора через замкнутую поверхность, ограничивающую этот объём,

$$\Delta I_p \Delta S = \frac{1}{q} \Delta J_p \Delta S = \frac{1}{q} \oint_S \bar{J}_p d\bar{S}. \quad (1.52)$$

Интегрирование проводится по замкнутой поверхности  $S$ , окружающей выделенный объём. Под знаком интеграла стоит скалярное произведение вектора плотности тока  $\bar{J}_p$  и векторного элемента поверхности  $d\bar{S}$ . Тогда, подставив (1.52) в (1.51), разделив обе части уравнения (1.51) на  $\Delta V \Delta t$ , и перейдя к пределу при  $\Delta V, \Delta t \rightarrow 0$ , получим:

$$\frac{\partial p}{\partial t} = g_p - R_p - \frac{1}{q} \lim_{\Delta V \rightarrow 0} \left( \oint_S \bar{J}_p d\bar{S} / \Delta V \right) = g_p - R_p - \frac{1}{q} \operatorname{div} J_p.$$

Учтено, что (по физическому смыслу) дивергенция есть предел отношения потока вектора через замкнутую поверхность к объёму, ограниченному этой поверхностью, при объёме, стремящемся к нулю (поток вектора из точки) [5].

Уравнения непрерывности (1.50) показывают, в силу каких причин изменяется концентрация носителей заряда полупроводника во времени и в пространстве. Концентрация носителей может изменяться из-за нетепловой генерации при внешних воздействиях, например, при засветке или ударной ионизации (первое слагаемое). В зависимости от знака второго слагаемого концентрация может изменяться из-за преобладания рекомбинации, либо тепловой генерации носителей. Наконец, концентрация может изменяться за счёт не нулевой дивергенции тока, т. е. изменений потока свободных зарядов, например, из-за наличия поля, неподвижных зарядов или градиента (неравномерной) концентрации носителей в рассматриваемом физически бесконечно малом объёме (третье слагаемое). Уравнения упрощаются при отсутствии какой-либо из названных причин.

Когда концентрация неосновных носителей существенно меньше равновесной концентрации основных носителей заряда

$$n_p \ll p_{p0}, p_n \ll n_{n0},$$

реализуется линейная рекомбинация через ловушки и центры рекомбинации. Скорости линейной рекомбинации определяются соотношениями (1.33), (1.34). В этом случае развёрнутые одномерные уравнения (1.50) для ННЗ имеют следующий вид

$$\frac{\partial p_n}{\partial t} = g_p - \frac{p_n - p_{n0}}{\tau_p} - p_n \mu_p \frac{\partial \mathcal{E}}{\partial x} - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2}, \quad (1.53)$$

$$\frac{\partial n_p}{\partial t} = g_n - \frac{n_p - n_{p0}}{\tau_n} + p_n \mu_n \frac{\partial \mathcal{E}}{\partial x} + \mu_n \mathcal{E} \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2}. \quad (1.54)$$

Вид уравнений непрерывности сохраняется также и для приращений концентраций  $\Delta p_n = p_n - p_{n0}$ ,  $\Delta n_p = n_p - n_{p0}$ , поскольку производные от равновесных концентраций равны нулю.

## 1.6. Задачи на инжекцию и экстракцию неосновных носителей заряда

При анализе физических процессов в полупроводниковых приборах возникают задачи, связанные с инжекцией неосновных носителей заряда. **Инжекция** – это поступление носителей заряда в области полупроводника, где они являются неосновными.

**Уровень инжекции**  $\delta$  определяется величиной отношения концентрации неравновесных неосновных носителей заряда к равновесной концентрации основных

$$\delta = \begin{cases} n_p / p_{p0} \text{ или } \Delta n_p / p_{p0}, & p\text{-полупроводник, } p_{p0} > n_{p0}, \\ p_n / n_{n0} \text{ или } \Delta p_n / n_{n0}, & n\text{-полупроводник, } n_{n0} > p_{n0} \end{cases} \quad (1.55)$$

При низком уровне инжекции  $\delta \ll 1$ . При высоком –  $\delta \gg 1$ .

Инжекция повышает концентрацию ННЗ в полупроводнике, нарушает условие термодинамического равновесия (1.18) и стимулирует рекомбинацию. Высокий уровень инжекции может вызвать изменение электрофизических параметров полупроводника, появление электрического поля, изменить характер рекомбинации.

В настоящем разделе с помощью уравнений непрерывности проведено подробное решение и физический анализ некоторых задач, результаты которых используются в дальнейшем при исследовании физических процессов в полупроводниковых приборах.

**Задача 1.1.** *Определить распределение концентрации дырок в электронном полупроводниковом образце, если на одном его торце  $x = 0$  поддерживается стационарное значение избыточной неравновесной концентрации дырок  $\Delta p_n(0) = p_n(0) - p_{n0}$ . Концентрацию ННЗ на другом конце образца считать равновесной. Рассмотреть случаи полубесконечного образца и образца конечных размеров  $w$ .*

Поддержание стационарной неравновесной концентрации ННЗ-дырок на торце означает их стационарное поступление в образец. Избыточная концентрация на одном торце создаёт в образце стационарный

диффузионный поток инжектированных дырок. Физическое содержание задачи заключается в определении пространственного распределения (зависимости от пространственной координаты) концентрации инжектированных ННЗ для образца заданной длины, если на другом конце образца поддерживается равновесная концентрация. Такая задача может возникнуть, например, когда один торец полупроводникового образца освещается коротковолновым оптическим излучением, постоянно создающим на торце тонкий слой фотогенерированных дырок, а другой торец заземлён через омический контакт металл–полупроводник. Тогда эта задача становится частью анализа физических процессов в фотоэлектрическом приборе. Ниже мы увидим, что в полупроводниковых приборах неравновесная концентрация возникает, когда к прибору приложено внешнее напряжение. Тогда результаты решения такой задачи позволят анализировать диффузионные токи проводимости в приборах.

Математическая сторона широкого круга физических задач, связанных с анализом процессов в полупроводниковых приборах, сводится к решению стационарных уравнений непрерывности (1.53), (1.54) для приращений концентраций  $\Delta p_n(x)$ ,  $\Delta n_p(x)$  в отсутствие внешней генерации и электрического поля, т. е. при  $\partial \Delta p_n / \partial t = 0$ ,  $\partial \Delta n_p / \partial t = 0$ ,  $g = 0$ ,  $\vec{E} = 0$ . Граничные условия определяются заданными физическими условиями задачи.

### Полубесконечный образец

Конкретно для задачи 1.1 в случае полубесконечного образца граничные условия уравнения (1.53) принимают следующий вид:

$$\Delta p_n(x)|_{x=0} = \Delta p_n(0), \quad \Delta p(x)|_{x=\infty} = 0.$$

В отсутствие поля и внешней генерации уравнение упрощается до однородного уравнения второго порядка с постоянными коэффициентами:

$$D_p \frac{d^2 \Delta p_n}{dx^2} - \frac{\Delta p_n}{\tau_p} = 0, \quad \text{или} \quad L_p^2 \frac{d^2 \Delta p_n}{dx^2} - \Delta p_n = 0, \quad (1.56)$$

$$\text{где} \quad L_p^2 = D_p \tau_p \Rightarrow L_p = \sqrt{D_p \tau_p} \quad (1.57)$$

**диффузионная длина** неосновных носителей заряда – дырок в электронном полупроводнике.

Для решения в виде  $\Delta p_n(x) = C \exp(\lambda x)$ , где  $C$  – константа, характеристическое уравнение  $(L_p \lambda)^2 - 1 = 0$  дифференциального уравнения (1.56) имеет корни  $\lambda_{1,2} = \pm(1/L_p)$ . Общее решение

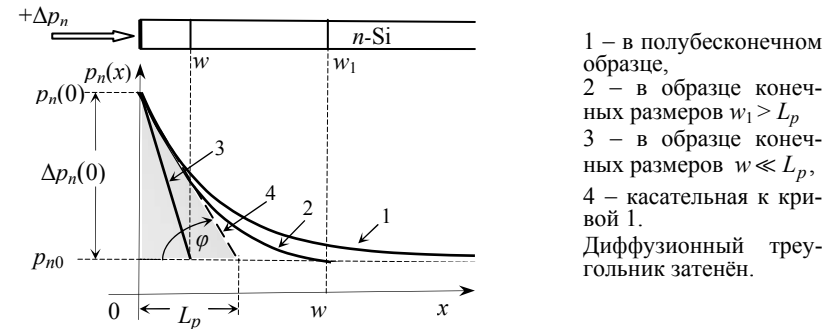
$$\Delta p_n(x) = C_1 e^{-x/L_p} + C_2 e^{x/L_p} \quad (1.58)$$

для постоянных  $C_1 = \Delta p_n(0)$ ,  $C_2 = 0$ , удовлетворяющих заданным граничным условиям, будет равно

$$\Delta p_n(x) = \Delta p_n(0) e^{-x/L_p} \Rightarrow p_n(x) = p_{n0} + \Delta p_n(0) e^{-x/L_p}. \quad (1.59)$$

Инжектированные носители распределены в полупроводнике по экспоненциальному закону с характерной постоянной  $L_p$ , равной диффузионной длине неосновных носителей заряда (рис. 1.40 кривая 1).

Физический анализ полученного решения позволяет определить смысл диффузионной длины и ответить на вопрос, куда деваются инжектированные дырки, стационарно поступающие в полупроводник.



- 1 – в полубесконечном образце,
  - 2 – в образце конечных размеров  $w_1 > L_p$
  - 3 – в образце конечных размеров  $w \ll L_p$ ,
  - 4 – касательная к кривой 1.
- Диффузионный треугольник затенён.

Рис. 1.40. Стационарное распределение концентрации инжектированных неосновных носителей заряда

**Диффузионная длина** – это расстояние, на котором избыточная концентрация ННЗ уменьшается (изменяется) в  $e = 2,72$  раз. Действительно, положив в (1.59)  $x = L_p$ , получим  $[\Delta p_n(x)/\Delta p_n(0)]_{x=L_p} = 1/e = 0,37$ .

Диффузионная длина (1.57) является характерным масштабом процесса диффузии. На графике диффузионная длина определяется точкой пересечения касательной (штриховая линия) к кривой распределения ННЗ в плоскости инъекции  $x = 0$  с уровнем равновесной концентрации  $p_{n0}$ .

Уменьшение концентрации дырок по мере распространения потока происходит за счёт рекомбинации, потому что в каждой точке кривой 1 рис. 1.40 скорость потока дырок  $U_{\text{диф}}$  (производная от потока по координате) равна скорости их рекомбинации (1.33). Это следует из вида самого уравнения (1.56). Действительно, преобразуя слагаемое, содержащее вторую производную, получаем<sup>31</sup>

$$D_p \frac{d^2 \Delta p_n(x)}{dx^2} = \frac{d}{dx} \left[ D_p \frac{dp_n(x)}{dx} \right] = \frac{d}{dx} \left[ D_p \text{grad } p_n(x) \right] = -\frac{d}{dx} \left[ \Pi_p(x) \right] = U_{\text{диф}}.$$

Тогда 
$$\frac{d}{dx} [ \Pi_p(x) ] - \Delta p_n / \tau_p = v_{\text{диф}} - R_p = 0 \Rightarrow v_{\text{диф}} = R_p.$$

Равенство скоростей диффузионного потока и скорости рекомбинации является физическим условием непрерывности потока дырок.

Однако для рекомбинации дырок необходим стационарный приток электронов. Поэтому рассматривая физическую сторону задачи, мы должны (в качестве варианта) домыслить также существование потока электронов. Из непрерывности потока дырок в условиях происходящей рекомбинации следует, что к каждой точке кривой распределения ННЗ-дырок должен подходить поток ОНЗ-

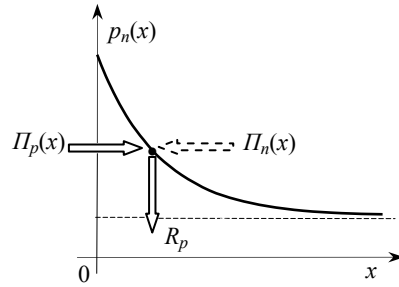


Рис. 1.41. Рекомбинирующие потоки при стационарной инжекции неосновных носителей заряда

электронов, равный потоку дырок  $\Pi_p(x) = \Pi_n(x)$  (рис. 1.41).

Распределение плотности потока дырок по координате и, в частности, плотность первоначально инжектированного потока через торец  $x = 0$  можно определить из полученного решения (1.59):

$$\Pi_p(x) = D_p \nabla p_n(x) = D_p \left[ \Delta p_n(0) / L_p \right] e^{-x/L_p} \Big|_{x=0} = D_p \left[ \Delta p_n(0) / L_p \right]. \quad (1.60)$$

Отношение  $\Delta p_n(0) / L_p$  является модулем градиента концентрации, который характеризует величину инжектированного потока. Из рис. 1.40 видно, что  $\Delta p_n(0) / L_p = \text{tg} \varphi$ , где  $\varphi$  угол наклона касательной (штриховая прямая 4) к кривой пространственного распределения ННЗ в затенённом треугольнике. Этот треугольник принято называть «диффузионным треугольником».

Таким образом, в отсутствие электрического поля инжекция ННЗ в полупроводниковый образец создаёт в нём экспоненциально-распределённый диффузионный поток ННЗ. Величина потока прямо пропорциональна тангенсу угла наклона касательной к кривой распределения инжектированных носителей. В плоскости инжекции тангенс угла наклона  $\text{tg} \varphi \sim \Delta p_n(0) / L_p$  определяется из «диффузионного треугольника», катетами которого являются приращения концентраций и диффузионные длины ННЗ, а гипотенузами – отрезки касательных. Диффузионный поток имеет максимальную величину в плоскости инжекции и при распространении экспоненциально затухает с постоянной  $L_p$  за счёт неизбежно происходящей рекомбинации.

### Образец конечных размеров

Распределение концентрации дырок, инжектированных в образец конечных размеров  $w$ , есть решение уравнения (1.56) для граничных условий

$$\Delta p_n(x)|_{x=0} = \Delta p_n(0), \quad \Delta p_n(x)|_{x=w} = 0. \quad (1.61)$$

Из (1.58) определяем постоянные  $C_1, C_2$ , удовлетворяющие заданным граничным условиям (1.61):

$$C_1 = \frac{\Delta p_n(0) e^{w/L_p}}{e^{w/L_p} - e^{-w/L_p}} = \frac{1/2 \Delta p_n(0) e^{w/L_p}}{\text{sh}(w/L_p)}, \quad C_2 = -\frac{1/2 \Delta p_n(0) e^{-w/L_p}}{\text{sh}(w/L_p)}.$$

Тогда решение уравнения (1.56) для найденных постоянных имеет вид

$$\Delta p_n(x) = \frac{\Delta p_n(0)}{\text{sh}(w/L_p)} \frac{\text{sh}(w-x)}{L_p}. \quad (1.62)$$

В частности, если образец короткий с размером  $w \ll L_p$  то, разлагая гиперболические функции в ряд Тейлора и ограничиваясь линейным приближением для  $\text{sh}z \approx z$ , из (1.62) получим линейное распределение

$$\Delta p_n(x) \approx \Delta p_n(0) [1 - (x/w)], \quad (1.63)$$

обозначенное номером 3 на рис. 1.40.

Таким образом, распределение инжектированных ННЗ имеет экспоненциальный характер и выражается через гиперболические функции (1.62) (рис. 1.40, кривые 1, 2). Однако, если размер образца много меньше диффузионной длины ННЗ, распределение является линейным (кривая 3).

Знание пространственного распределения носителей позволяет определять потоки инжектированных зарядов. Необходимый для этого градиент концентрации в плоскости инжекции  $x = 0$  находим из (1.62):

$$\text{grad } p_n(x) = - \left[ \Delta p_n(0) / L_p \text{sh}(w/L_p) \right] \text{ch} \frac{w-x}{L_p} \Big|_{x=0} = \quad (1.64)$$

$$= - \frac{\Delta p_n(0)}{L_p} \frac{w}{L_p} \approx \begin{cases} -\Delta p_n(0) / w, & \text{когда } w \ll L_p, \\ -\Delta p_n(0) / L_p, & \text{когда } w \gg L_p, w \rightarrow \infty. \end{cases} \quad (1.65)$$

Учтено, что  $\text{cth}(w/L_p) \approx \begin{cases} 1, & \dots \text{если } w \gg L_p \text{ или } w \rightarrow \infty, \\ L_p/w, & \text{если } w \ll L_p. \end{cases}$

Плотность первоначально инжектированного потока равна

$$\Pi_p(x) = -D_p \nabla p(x)|_{x=0} = \begin{cases} \Delta p_n(0) [D_p/w] & \text{для } w \ll L_p, \\ \Delta p_n(0) [D_p/L_p] & \text{для } w \gg L_p \text{ и } w \rightarrow \infty. \end{cases} \quad (1.67)$$

Отметим, если  $w < L_p$ , диффузионная длина в формулах заменяется фактическим размером, что естественно.

Для инжектированного потока справедливо общее правило, выраженное, в частности, соотношениями (1.36), согласно которым плотность потока равна произведению скорости потока на концентрацию. Действительно, из определения диффузионной длины (1.57) следует, что  $L_p = (D_p/L_p)\tau_p = (L_p/\tau_p)\tau_p \equiv v_{\text{диф}}\tau_p$ . Отношения

$$D_p/L_p = L_p/\tau_p = v_{\text{диф}}, \quad (1.68)$$

имеющие размерности скорости, уместно считать разными выражениями средней скорости диффузии или скорости диффузионного потока. Тогда выражения (1.66), (1.67) приобретают вид, совпадающий с (1.36):

$$\begin{aligned} \Pi_p(x)|_{x=0} &= \Delta p_n(0) [D_p/L_p] = \Delta p_n(0) [L_p/\tau_p] = \Delta p_n(0) v_{\text{диф}}, \text{ или} \\ \Pi_p(x)|_{x=0} &= p_n [D_p/L_p] = p_n [L_p/\tau_p] = p_n v_{\text{диф}}. \end{aligned} \quad (1.69)$$

Из соотношения (1.68) видно, что диффузионная длина  $L$  есть среднее расстояние, которое проходит носитель за время жизни  $\tau$  до рекомбинации в объёме. Таков ещё один смысл диффузионной длины.

Результаты решения для образца конечных размеров позволяют определить коэффициент переноса инжектированных носителей с одного конца полупроводникового образца на другой. Статический коэффициент переноса  $A_T$  есть отношение потоков на торцах образца

$$A_T = \frac{\Pi_p(w)}{\Pi_p(0)} = \frac{\nabla p_n(x)|_{x=w}}{\nabla p_n(x)|_{x=0}} = \frac{1}{\text{ch}(w/L_p)} \leq 1. \quad (1.70)$$

В частности, для полубесконечного или длинного образца, когда  $w \gg L_p$ ,  $\text{ch}(w/L_p) \rightarrow \infty$ ,  $A_T = 0$ . Все инжектированные ННЗ прорекомбинируют прежде, чем достигнут другого торца.

Для короткого образца, когда  $w \ll L_p$ ,

$$A_T = \text{ch}^{-1}(w/L_p) = \text{sch}(w/L_p)|_{w \ll L_p} \approx 1 - \frac{1}{2}(w/L_p)^2. \quad (1.71)$$

Например, если  $w = 0,1L_p$ ,  $A_T \approx 0,995$ . Диффузионный поток неосновных носителей заряда, инжектированных в полупроводник, размеры которого существенно меньше диффузионной длины, доходит до противоположного торца практически без потерь на рекомбинацию в объёме:

$$A_T \rightarrow 1 \Rightarrow \Pi_p(w)|_{w \ll L_p} = A_T \Pi_p(0) \approx \Pi_p(0). \quad (1.72)$$

## Экстракция (вытягивание) неосновных носителей

**Задача 1.2.** Определить распределение концентрации дырок в электронном полупроводниковом образце, если на одном его торце  $x=0$  поддерживается нулевая концентрация ННЗ, а на другом – равновесная. Рассмотреть случай полубесконечного образца.

В равновесном примесном полупроводнике ННЗ существуют всегда. Нулевая концентрация ННЗ на торце означает их экстракцию – стационарное вытягивание из образца. Физическая задача состоит в том, чтобы определить пространственное распределение ННЗ в полупроводниковом образце для условий их стационарного вытягивания с одного торца при равновесной концентрации на другом.

Математическая сторона задачи сводится к решению уравнения (1.56) для граничных условий

$$p_n(x)|_{x=0} = 0 \Rightarrow \Delta p_n(x)|_{x=0} = -p_{n0}, \quad \Delta p_n(x)|_{x=\infty} = 0, \quad (1.73)$$

отражающих заданные физические условия в полубесконечном образце. Общее решение (1.58) однородного уравнения (1.56) при постоянных  $C_1 = -p_{n0}$ ,  $C_2 = 0$ , удовлетворяющих граничным условиям (1.73), будет равно

$$\Delta p_n(x) = -p_{n0} e^{-x/L_p} \Rightarrow p_n(x) = p_{n0} (1 - e^{-x/L_p}). \quad (1.74)$$

Вытягивание ННЗ создаёт градиент концентрации, вызывающий диффузионный поток дырок из объёма в сторону торца с пониженной концентрацией:

$$\Pi_p(x) = -D_p \nabla p_n(x) =$$

$$-p_{n0} (D_p/L_p) e^{-x/L_p} \quad (\text{рис. 1.42}).$$

Минус в выражении потока свидетельствует о том, что его направление противоположно направлению оси  $x$ . Наибольшее значение поток имеет в плоскости экстракции при  $x=0$ . По мере удаления вглубь полупроводника поток уменьшается и концентрация восстанавливается. В частности, при  $x=L_p$  согласно определению диффузионной длины

$$\Pi_p(x)|_{x=L_p} / \Pi_p(x)|_{x=0} = e^{-1} = 0,37.$$

Однако стационарное вытягивание ННЗ нарушает электрическую нейтральность полупроводника. Поэтому рассматривая физическую сто-

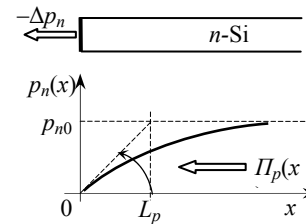


Рис. 1. 42. Пространственное распределение неосновных носителей при вытягивании из торца полупроводникового образца



рону задачи, мы вправе (в качестве варианта) домыслить также существование противоположно направленного потока электронов, обеспечивающего электрическую нейтральность за счёт ухода ОНЗ из образца (рис. 1.42 пунктирная стрелка).

### Двусторонняя инжекция\экстракция неосновных носителей

**Задача 1.3.** *Определить распределение концентрации дырок в электронном полупроводниковом образце, если на одном его торце  $x=0$  поддерживается стационарное значение избыточной неравновесной концентрации дырок, равное  $\Delta p_n(0) = p_n(0) - p_{n0}$ , а на другом – стационарное значение избыточной неравновесной концентрации дырок, равное  $\Delta p_n(w) = p_n(w) - p_{n0}$ . Рассмотреть случай образца конечных размеров  $w \ll L_p$ .*

Физическим содержанием задачи является определение распределения концентрации ННЗ при двусторонней инжекции в полупроводниковый образец конечных размеров. Такая задача возникает, например, при инжекции/экстракции неосновных носителей в базу биполярного транзистора со стороны эмиттера и коллектора. Чтобы определить распределение инжектированных носителей, необходимо решить уравнение (1.56) при следующих граничных условиях

$$\Delta p_n(x)|_{x=0} = \Delta p_n(0), \quad \Delta p_n(x)|_{x=w} = \Delta p_n(w), \quad (1.75)$$

соответствующих заданному состоянию полупроводника.

Общее решение (1.58) однородного уравнения (1.56) при постоянных

$$C_1 = -\frac{\Delta p_n(w) - \Delta p_n(0)e^{w/L_p}}{2\text{sh}(w/L_p)}, \quad C_2 = \frac{\Delta p_n(w) - \Delta p_n(0)e^{-w/L_p}}{2\text{sh}(w/L_p)},$$

удовлетворяющих граничным условиям (1.75), будет равно

$$\Delta p_n(x) = \Delta p_n(0) \frac{\text{sh}[(w-x)/L_p]}{\text{sh}(w/L_p)} + \Delta p_n(w) \frac{\text{sh}(x/L_p)}{\text{sh}(w/L_p)}. \quad (1.76)$$

Первое слагаемое в (1.76) является частью распределения, управляемой с торца  $x=0$ , путём изменения величины задаваемого значения неравновесной концентрации  $\Delta p_n(0)$ . Управление вторым слагаемым производится с противоположного торца  $x=w$  задаваемым значением неравновесной концентрации  $\Delta p_n(w)$ .

Двусторонняя инжекция создаёт в образце встречно-направленные диффузионные потоки, величина которых уменьшается из-за рекомбинации (при  $A_T \neq 1$ ). Действительно, поток на торце  $x=0$  равен разности

$\Pi_p(x)|_{x=0} = D_p \nabla(\Delta p_n(x))|_{x=0} = (D_p/L_p) \text{cth}(w/L_p) \cdot [\Delta p_n(0) - A_T \Delta p_n(w)]$  потока, инжектированного с торца  $x=0$  (уменьшаемое), и потока, перенесённого с торца  $x=w$  (вычитаемое), уменьшенного за счёт рекомбинации. Аналогично поток на торце  $x=w$  есть разность

$\Pi_p(x)|_{x=w} = D_p \nabla(\Delta p_n(x))|_{x=w} = (D_p/L_p) \text{cth}(w/L_p) \cdot [A_T \Delta p_n(0) - \Delta p_n(w)]$  потока, перенесённого с торца  $x=0$  (уменьшаемое), уменьшенного за счёт рекомбинации, и потока, инжектированного на торце  $x=w$  (вычитаемое).

Для короткого образца с размером  $w \ll L_p$ , ограничиваясь первым членом разложения  $\text{cth}z \approx 1/z$ , откуда получим:

$$\Pi_p(x)|_{x=0} = -(D_p/w) [\Delta p_n(0) - A_T \Delta p_n(w)], \quad (1.77)$$

$$\Pi_p(x)|_{x=w} = (D_p/w) [A_T \Delta p_n(0) - \Delta p_n(w)], \quad (1.78)$$

где коэффициент переноса  $A_T$  определяется соотношением (1.71). На противоположном торце каждый из встречно инжектированных потоков уменьшается в  $A_T$  раз из-за рекомбинации в объёме.

**Задача 1.3-1.** В частности, если, например, на одном торце  $x=0$  полупроводникового образца, размер которого  $w \ll L_p$ , реализуется инжекция ННЗ, а на другом конце  $x=w$  происходит их вытягивание, то  $\Delta p_n(x)|_{x=0} = \Delta p_n(0)$ ,  $\Delta p_n(x)|_{x=w} = -p_{n0} \Rightarrow p_n(w) = 0$ . В приближении  $(1 - A_T) \rightarrow 0$  соотношения (1.77), (1.78) принимают следующий вид<sup>33</sup>:

$$\Pi_p(x)|_{x=0} = \frac{D_p}{w} [p_n(0) - p_{n0}(1 - A_T)] \approx D_p \frac{p_n(0)}{w} \equiv D_p \text{tg} \varphi_{|x=0},$$

$$\begin{aligned} \Pi_p(x)|_{x=w} &= \frac{D_p}{w} [A_T p_n(0) + p_{n0}(1 - A_T)] \approx \\ &\approx D_p \frac{A_T p_n(0)}{w} \equiv D_p \text{tg} \varphi_{|x=w} = A_T D_p \text{tg} \varphi_{|x=0}. \end{aligned}$$

В квадратных скобках этих соотношений стоят значения суммарной концентрации на торцах. Это позволяет выразить потоки зарядов через тангенсы углов наклона касательных (градиенты распределений) на торцах и получить наглядный результат рекомбинации. Из-за рекомбинации  $\Pi_p(x)|_{x=w} < \Pi_p(x)|_{x=0}$ , поэтому  $\varphi_{|x=l} < \varphi_{|x=0}$ , и распределение концентрации ННЗ нелинейное (рис. 1.43).

В линейном приближении, полагая, что гиперболические синусы равны их аргументам, из (1.76) получим линейное распределение ННЗ

$$p_n(x) = \left(1 - \frac{x}{w}\right) \Delta p_n(0) + \frac{x}{w} \Delta p_n(w) + p_{n0}, \quad (1.79)$$

По условиям задачи на торце  $x = w$  происходит вытягивание носителей,  $p_n(w) = 0$ , и (1.79) принимает вид линейного распределения

$$p_n(x) = (1 - x/w) \Delta p_n(0) - (x/w) p_{n0} + p_{n0}, \quad (1.80)$$

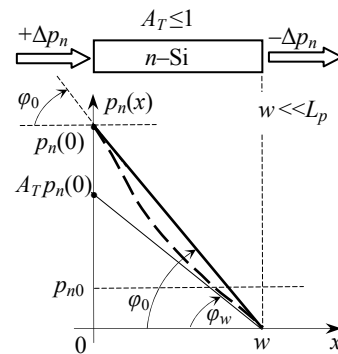
обеспечивающего односторонний поток зарядов через весь образец без потерь на рекомбинацию (рис. 1.43). Действительно, в этом случае  $\text{grad } p_n(x) = -p_n(0)/w$  не зависит от  $x$ , что и обеспечивает постоянство потока. Значит, при линейном распределении ННЗ рекомбинация в объёме полупроводника отсутствует. В главе 4 мы увидим, что такие

Рис. 1.43. Распределение концентрации неосновных носителей при одностороннем потоке через узкий образец

Линейное распределение при отсутствии рекомбинации (сплошная прямая).

Нелинейное распределение при учёте рекомбинации в объёме (штриховая кривая).

Градиент концентрации на торце  $x = 0$  больше градиента концентрации на торце  $x = w$  из-за рекомбинации в объёме,  $\varphi_0 > \varphi_w$ .



процессы происходят в узкой базе биполярного транзистора в активном режиме работы при  $A_T = 1$ .

### Контрольные вопросы

1. Что такое дырка с точки зрения структуры кристаллической решётки собственного и примесного полупроводников, состава свободных носителей заряда?
2. Напишите выражение электронной конфигурации атома германия.
3. Почему в зонных моделях полупроводников дырка находится в валентной зоне, а электрон в зоне проводимости?
4. Каковы механизмы образования СНЗ в кремниевых и германиевых полупроводниках?
5. Что такое донорная/акцепторная примесь? Почему она так называется?
6. В чём различие механизмов формирования СНЗ в полупроводниковых материалах элементов IV группы Si и Ge и углеродных нанотрубках или графенах?
7. Каковы условия электрической нейтральности собственных и примесных полупроводников?

8. Каковы физические причины принципиально существующей температурной зависимости электрических свойств и параметров полупроводниковых приборов?

9. Каков общезначимый смысл функции распределения в статистических системах?

10. Почему считается, что функция распределения Ферми–Дирака показывает вероятность занятости энергетического уровня?

11. Какие основные задачи теории полупроводниковых приборов решаются с помощью функции распределения Ферми–Дирака?

12. Каковы свойства уровня Ферми?

13. Как изменяется положение уровня (квазиуровня) Ферми при изменении концентрации носителей заряда?

14. Каков смысл закона действующих масс для равновесных и неравновесных состояний?

15. Что такое состояние примесного истощения? Охарактеризуйте температурную зависимость концентрации носителей заряда в примесном полупроводнике.

16. Почему в примесных полупроводниках в принципе всегда существуют неосновные носители зарядов? Как можно изменять их концентрацию?

17. Как зависит соотношение основных и неосновных носителей от количества легирующей примеси?

18. Каковы электрические и динамические условия равновесного состояния полупроводника?

19. О чём свидетельствует наклон/изгиб энергетических зон полупроводника?

20. Почему поверхность полупроводника имеет заряд? Оцените его плотность и знак?

21. Какое направление имеют электронные и дырочные диффузионные токи, если соответствующие потоки совпадают или противоположны по направлению?

22. Как можно управлять диффузионным и дрейфовым токами полупроводника? Чем различается управление дрейфовым и диффузионным токами?

23. Как изменятся диффузионный и дрейфовый токи проводимости и сама проводимость при изменении концентрации носителей, например, в 2 раза?

24. Как ведёт себя сопротивление полупроводникового резистора при изменении температуры?

25. Что такое абсолютный, относительный температурный коэффициент?

26. Каков физический смысл соотношения Эйнштейна?

27. Что является необходимым условием соблюдением закона Ома в полупроводнике? Каковы возможные причины его нарушения?

28. Каков физический смысл уравнения непрерывности?

29. Дайте определения диффузионной длины, коэффициента диффузии, времени жизни, подвижности.

30. Каково характерное распределение диффузионного потока инжектированных неосновных носителей по длине полупроводникового образца разных размеров?

31. Что такое коэффициент переноса инжектированных ННЗ?

32. Охарактеризуйте двустороннюю/одностороннюю инжекцию ННЗ в короткий полупроводниковый образец.

## ПОЛУПРОВОДИКОВЫЕ ДИОДЫ НА ЭЛЕКТРОННО-ДЫРОЧНЫХ ПЕРЕХОДАХ

Полупроводниковые диоды, характеристики которых определяются свойствами электронно-дырочного перехода, играют важную роль в современной электронике и как класс самостоятельных приборов, и как функциональная часть других классов полупроводниковых приборов и интегральных схем. Полупроводниковый диод относится к числу базовых элементов твердотельной электроники.

Под электронно-дырочным или  $p-n$ -переходом мы понимаем гальванический контакт полупроводников электронной и дырочной проводимости, содержащий область объёмного пространственного заряда (ОПЗ), или собственно  $p-n$ -переход, и прилегающие к нему нейтральные  $n$ - и  $p$ -области. Соответствующий полупроводниковый диод отличается разве что наличием омических контактов, которые предполагаются по умолчанию.

Электронно-дырочные переходы привлекли серьёзное внимание после того, как в 1938 г. Д. И. Давыдов рассмотрел явления выпрямления переменного тока  $p-n$ -переходом и появления фотоэдс<sup>34</sup>. Математическая теория  $p-n$ -перехода была заложена Шокли (W. Shockley<sup>35</sup>) в 1949 г.

Туннельный диод создан в 1958 г японским физиком Л. Есаки (L. Esaki<sup>36</sup>), исследовавшим сплавные переходы сильно легированного германия. Своим названием диод обязан туннельному эффекту, посредством которого носители заряда преодолевают потенциальный барьер.

Анализ свойств электронно-дырочного перехода проведён на модели идеального  $p-n$ -перехода. Физические процессы рассмотрены на основе идеи диффузионно-дрейфового равновесия. Проведён физический вывод вольтамперной характеристики диода. Разобраны механизмы формирования прямого и обратного токов. Аналитическим и физическим способами введены ёмкости, а также схемы замещения диода по переменному току. Проанализированы температурные зависимости вольтамперной характеристики, тока и напряжения диода. Учтено влияние ряда физических факторов, не рассматриваемых моделью идеального  $p-n$ -перехода, на вид вольтамперной характеристики диода. Туннельный и обращённый диоды рассмотрены на качественном уровне.

### 2.1. Физические основы работы идеализированного $p-n$ -перехода

#### Формирование, диффузионно-дрейфовое равновесие

При гальваническом контакте электронного и дырочного полупроводников электроны из  $n$ -полупроводника, где они основные носители

ник. Логика физических процессов, направленных на восстановление равновесия, приводит к образованию пространственного объёмного заряда.

Основной модели формирования выпрямляющего контакта металл-полупроводник в настоящем приложении является идея об образовании отрицательного объёмного заряда в полупроводнике за счёт непосредственной ионизации атомов акцепторной примеси электронами, пришедшими не из валентной зоны объёма полупроводника, как обычно, а из металла. Приход электронов обусловлен меньшей работой выхода металла.

Выпрямляющий контакт металл- $p$ -полупроводник может быть реализован, если работа выхода металла меньше работы выхода полупроводника  $A_M < A_{ПП-к}$  (рис. 3.2). При гальваническом контакте электроны металла за счёт термоэлектронной эмиссии переходят в полупроводник. Перешедшие электроны ионизируют 3-х валентные атомы акцепторной примеси (заполняют недостающие связи, формируют ковалентные связи) в области контакта. В объёме полупроводника ионизация атомов акцепторной примеси происходит за счёт разрыва ковалентных связей собственных атомов решетки и образования дырки, т.е. за счёт перехода электронов из зоны проводимости на примесный уровень акцепторной примеси. В приповерхностной области контакта ионизация происходит за счёт электронов, пришедших из металла. При этом образуется ОПЗ отрицательно ионизированных атомов акцепторной примеси в полупроводнике и положительный заряд поверхности металла. Индуцированное объёмным зарядом электрическое поле в полупроводнике, направленное из металла в полупроводник, способствует увеличению эмиссии электронов из полупроводника в металл, выравнивает уровни Ферми. Равновесие наступит, когда в системе установится единый для металла и полупроводника уровень Ферми. При этом на контакте зоны изгибаются «вниз» на величину, равную разности уровней Ферми:  $qU_0 = E_{Fn} - E_{Fm}$ , где  $U_0$  – контактная разность потенциалов. Изгиб зон образует потенциальный барьер  $qU_0$  для ОНЗ-дырок, переходящих из полупроводника в металл, и потенциальный барьер  $q\phi_{Vp}$  для обратного перехода дырок из металла в полупроводник (рис. 3.2б).

<sup>1</sup> По предложению Друде.

<sup>2</sup> В 1874 г.

<sup>3</sup> В настоящее время таблица Менделеева содержит почти вдвое больше – 116 элементов.

<sup>4</sup> Нанотрубки – продукты нанотехнологии. Термин «нанотехнология» введён в 1974г японским физиком Норё Танигути для описания процессов построения но-

вых объектов и материалов при помощи манипуляций с отдельными атомами. Нанометр –  $10^{-9}$  м.

<sup>5</sup> Индекс от англ. Valance.

<sup>6</sup> Индекс от англ. Conductance

<sup>7</sup> В физике в таком случае говорят о «квазичастице».

<sup>8</sup> Часто обозначения физических величин, относящихся к собственному полупроводнику, имеют индекс  $i$  – от англ. intrinsic – присущий, собственный.

<sup>9</sup> Носители называются *основными носителями заряда* (ОНЗ), если их концентрация больше концентрации собственных носителей заряда  $n_i$  при данной температуре. В противном случае их называют неосновными носителями заряда (ННЗ).

<sup>10</sup> Примером иного механизма формирования СНЗ являются углеродные нанотрубки и графен, где СНЗ могут создаваться за счёт структуры материала (см. рис. 1.4).

<sup>11</sup> При комнатной температуре  $T=T_0 = 300$  К интервал  $kT_0 = 4,14 \cdot 10^{-21}$  Дж =  $= 0,026$  эВ;  $1$  эВ  $\approx 38 kT_0$ .

<sup>12</sup> Для ориентировки:  $e^2 = (2,72)^2 \approx 7,4$ ;  $e^{2,3} \approx 10$ ;  $e^3 \approx 20$ ;  $e^{4,6} \approx 10^2$

<sup>13</sup> В такой записи функция Максвелла–Больцмана (в отличие от функции Ферми–Дирака) показывает, что относительное число частиц с энергией выше некоторого значения экспоненциально падает при увеличении их энергии.

<sup>14</sup> Химический (термодинамический) потенциал определяет приращение энергии системы частиц при увеличении числа частиц на единицу. Поэтому энергия Ферми определяется общим числом частиц.

<sup>15</sup> Отметим, что если взять другие сопряженные физические величины, например, энергию и время или амплитуду и фазу сигнала, то получим аналогичные неравенства, предел которых вытекает из соотношения (1.12).

<sup>16</sup> Иначе: на одном энергетическом уровне могут находиться два электрона с противоположными спинами.

<sup>17</sup> Подчёркивая значимость этого равенства, его называют иногда *уравнением полупроводника*.

<sup>18</sup> См. сноску №8.

<sup>19</sup> Использовано линейное приближение ряда Тейлора  $(1+x)^{1/2} \approx 1 + (1/2)x + \dots$ . Решение с отрицательным знаком перед радикалом отброшено, т. к. должно быть  $n_p > 1$ .

<sup>20</sup> Эта область температур соответствует состоянию примесного истощения (см. основной текст ниже).

<sup>21</sup> По умолчанию предполагается, что речь идёт об ионизированных атомах примеси или, что все атомы примеси ионизированы.

<sup>22</sup> Отсюда очевидно, что применительно к характеристикам полупроводниковых приборов понятия низких (малых) и высоких (больших) температур связаны с концентрацией и природой примеси, а не только с показаниями градусника.

<sup>23</sup> Напомним, что  $T^\circ\text{K} = t^\circ\text{C} + 273,15^\circ\text{C}$ .

<sup>24</sup> В предположении кусочно-ломаного графика энергетических зон, использованного для того, чтобы чётко обозначить границы существования поля.

<sup>25</sup> С выводом формулы Шокли–Рида–Холла можно познакомиться, например, в книгах: *Смит Р.* Полупроводники. М.: Мир, 1982; *Бонч-Бруевич В.Л., Калашиников С.Г.* Физика полупроводников. М.: Наука, 1977.

<sup>26</sup> В равновесном состоянии  $r = g$ . Отсюда получаем условие термодинамического равновесия  $np = n_i^2$  (1.18a).

<sup>27</sup> Индекс  $t$  от англ. trap – ловушка.

<sup>28</sup> Понятие тока смещения ввёл шотландский физик Д.К. Максвелл. Формальной причиной послужила необходимость согласования собственных уравнений электродинамики с уравнениями для постоянного тока.

<sup>29</sup> Напомним, в школьном курсе физики этот закон формулируется так: «Сила тока во всех участках последовательной цепи одинакова».

<sup>30</sup> Отметим, что в этом одна из причин температурной зависимости подвижности, электрического сопротивления полупроводника и, в конечном счёте, дрейфового тока.

<sup>31</sup> Отрицательный знак перед производной от потока свидетельствует о снижении скорости потока в направлении  $x > 0$ .

<sup>32</sup> Это равенство справедливо только в том случае, если по осям координат используются одинаковые масштабы. Здесь и в дальнейшем по умолчанию предполагается, что у такого рода равенств названное условие выполнено.

<sup>33</sup> См. сноску №32.

<sup>34</sup> *Давыдов Б.*, – ЖТФ, **5**, 7987 (1938).

<sup>35</sup> Нобелевская премия в 1956 г совместно с J. В. Bardeen и W. Brattain за исследование полупроводников и открытие транзисторного эффекта.

<sup>36</sup> Нобелевская премия в 1973 г. совместно с ----- за экспериментальные открытия эффекта туннелирования в полупроводниках и сверхпроводниках.

<sup>37</sup> Другие физические модели формирования ОПЗ и внутреннего электрического поля см. в Приложении 2.1.

<sup>38</sup> Для получения энергетической зонной диаграммы  $p$ – $n$ -перехода зонная диаграмма  $n$ -области как единое целое опускается вниз до совмещения уровней Ферми. Затем энергетические уровни зон соединяются плавными кривыми.

<sup>39</sup> Объёмный пространственный заряд иногда называют *двойным электрическим слоем*.

<sup>40</sup> Доказательство соотношений (2.2а), (2.2б) см. в Приложении 2.2.

<sup>41</sup> Можно показать также, что  $qU_0 = E_g - \kappa T \ln(N_V N_C / N_A N_D)$ . Значит,  $qU_0 \rightarrow E_g$  при  $(N_A, N_D) \rightarrow (N_V, N_C)$  с увеличением степени легирования. Например, для Si-перехода при изменении  $N_D = (10^{14} \div 10^{17}) \text{см}^{-3}$   $U_0 = (0,8 \div 1) \text{В}$ .

<sup>42</sup> Область ОПЗ содержит участок с собственной, значит, минимальной концентрацией СНЗ  $p = n = n_i$  (рис. 2.6). Поэтому ОПЗ является наиболее высокоомной