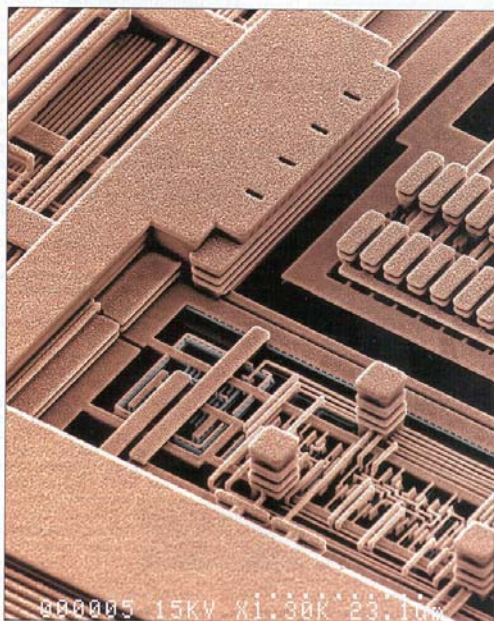


ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
МОСКОВСКИЙ ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

Г.И. ЗЕБРЕВ

ФИЗИЧЕСКИЕ ОСНОВЫ КРЕМНИЕВОЙ НАНОЭЛЕКТРОНИКИ

Рекомендовано УМО «Ядерные физика и технологии»
в качестве учебного пособия для студентов
высших учебных заведений



Москва 2008

УДК 121.382(075)+620.3(075)

ББК 32.85я7

З - 47

Зебрев Г.И. Физические основы кремниевой наноэлектроники: Учебное пособие. — М.: МИФИ, 2008. — 288 с.

Книга посвящена описанию основных физических принципов, структур и методов моделирования, а также тенденций развития современной и перспективной кремниевой наноэлектроники с технологическими нормами < 100 нм.

Предназначена для преподавателей и студентов, специализирующихся по направлениям микро- и наноэлектроники, электроники, электронных измерительных систем. Может быть использована в учебном процессе при подготовке учебных курсов «Физические основы наноэлектроники», «Наноэлектронные технологии», «Физика микроэлектронных структур».

Пособие подготовлено в рамках Инновационной образовательной программы

Рецензент: д-р техн. наук, профессор О.Р. Мочалкина

ISBN 978-5-7262-0954-8

© Московский инженерно-физический институт
(государственный университет), 2008

Предисловие

Данная книга предназначена, в первую очередь, для студентов, специализирующихся в области микро- и нанoeлектроники. Для лучшего восприятия изложение ведется по возможности в замкнутом виде от простого к сложному. Тем не менее, предполагается наличие у читателей базовых знаний в размере курса общей физики и элементарных знаний физики полупроводников и полупроводниковых приборов.

Предметом данной книги является кремниевая нанoeлектроника, которая понимается как электроника на основе наноразмерных МОП транзисторов, изготовленных на основе кремния. Даже в таком усеченном виде нанoeлектроника представляет собой необъятную область. Поэтому отбор материала для относительно небольшой книги определяется интересами автора и его субъективными представлениями о том минимуме, который должен знать современный специалист в области микро- и нанoeлектроники.

Книга организована следующим образом. В главе 1 кратко излагаются некоторые положения элементарные результаты физической кинетики и квантовой механики. В главе 2 описываются основные физические проблемы и принципы развития индустрии нанoeлектроники. В главе 3 дается описание структур «металлоокисел-полупроводник». Четвертая глава посвящена уже физике работы МОП транзистора – основе современной нанoeлектроники. В пятой главе рассказывается об физических и технологических эффектах, влияющих на характеристики МОП транзисторов. В главе 6 отдельно описываются эффекты, связанные с сильными электрическими полями в транзисторах. В главе 7 дается продвинутая физическая модель МОП транзистора, основанная на аналитическом решении уравнения непрерывности для плотности тока в канале. Восьмая глава посвящена описанию транзисторов на основе технологий «кремний-на-изоляторе» (КНИ) – наиболее быстроразвивающейся отрасли кремниевой нанoeлектроники. Здесь нужно отметить, что на русском языке практически отсутствует учебная литература, посвященная КНИ технологиям. В главе 9 впервые подробно описываются модели КНИ транзисторов. Десятая глава посвящена фундаментальной технологической проблеме токов утечек в нанoeлектронных структурах. Наконец, в главе 11 дается опи-

сание некоторых физических эффектов, возникающих в наноразмерных транзисторных структурах.

К сожалению, ограниченный объем книги и недостаток времени для подготовки текста привели к тому, что многие важные вопросы либо остались незатронутыми, либо упомянуты вскользь.

Пособие подготовлено на основе курса лекций «Физические основы нанoeлектроники», читаемого автором на протяжении ряда лет для студентов кафедры микро- и нанoeлектроники и Высшего физического колледжа Московского инженерно-физического института (государственного университета) (МИФИ). Некоторые подходы, описанные в книге, основаны на собственных результатах автора. Особенно это касается глав 7, 9 и, отчасти, 3 и 11. При подготовке текста автор опирался на стиль, обозначения и подбор материалов, принятые в современной литературе [П1-11]. Предполагается, что читатель может также воспользоваться более доступной хорошей литературой на русском языке более старых изданий [П12-20].

Для каждой главы приведена небольшая библиография, которая не претендует на полноту, но может помочь потенциальному читателю сориентироваться в потоке современной литературы на эту тему.

Автор выражает благодарность и признательность В.С. Першенкову, Р.Г. Усейнову за внимание к работе, всем коллегам по кафедре микро- и нанoeлектроники МИФИ за благожелательность, отдельно – О.Р. Мочалкиной – за то, что она взяла на себя труд прочесть книгу в рукописи и сделать неоценимые замечания, своим студентам и аспирантам – Максиму Горбунову, Валере Шункову, Мише Федоренко, Свете Белобородовой, Алексею Рубанцеву, и особенно, – Лизе Филипенковой, – за проявленный интерес и помощь в оформлении книги; П.Н. Осипенко из НИИСИ РАН и В.С. Анашину из НИИ КП – за сотрудничество и финансовую поддержку, а также своей семье, – Марине, Ане и Ивану Филипповичу – за терпение и чувство юмора.

1. Базисные физические уравнения

1.1. Предмет нанoeлектроники

Термины «нанoeлектроника» и «нанoeлектронные технологии» используются часто в двояком смысле. С одной стороны, под нанoeлектроникой понимают продукт эволюционного развития микроэлектронной транзисторной (главным образом, КМОП) технологии на основе кремния в сторону дальнейшей миниатюризации и увеличения степени интеграции. С другой стороны, часто под нанoeлектроникой подразумеваются приборы, основанные на принципах, материалах и конфигурациях, отличных от стандартных КМОП технологий. В данной книге будем иметь в виду, главным образом, первый аспект, понимая под «нанoeлектроникой» продукт эволюционного развития «микроэлектроники». Исторически термин «микроэлектроника» привязан к характерным размерам базового элемента – транзистора. Первые транзисторы имели вполне макроскопические размеры порядка 10 микронов. За почти полвека развития размеры микроэлектронных компонентов уменьшились приблизительно в 200 раз. До самого последнего времени прогресс в области освоения наноразмеров носил чисто эволюционный характер, основанный на идеях технологического скейлинга, т.е. масштабной миниатюризации. Поэтому граница между «субмикронной» и «нанoeлектронной» технологией, часто проводимая на размерах ~ 100 нанометров (0.1 мкм), носит несколько условный характер: удобней оперировать нанометрами, а не долями микрона. В указанном аспекте мировая микроэлектронная индустрия уже давно находится в «нанoeлектронной» области, – в конце 2007 года Intel анонсировало начало производства схем с технологической нормой (характерным масштабом технологии) 45 нм.

Следует подчеркнуть, что на масштабах десятков нанометров характерные размеры элементов сравниваются с некоторыми физическими размерами (например, длина экранирования, длина пробега электрона, длина волны де Бройля), что предполагает появление новых физических эффектов и некоторых фундаментальных физических ограничений на работу таких приборов. В этом проявляется особенность «нанoeлектроники» по сравнению с «микроэлектроникой», описывающейся макроскопическими законами классической

физики. Дело в том, что на протяжении почти полувека все развитие традиционной микроэлектроники осуществлялось в эволюционной форме, главным образом, усилиями технологов, схемотехников и материаловедов, практически без участия физиков. Переход к наноразмерной электронике означает переход широкого класса относительно новых физических задач и эффектов из чисто научной в инженерную область.

Что же касается внедрения в индустрию приборов на новых физических принципах или даже на новых материалах, то это представляется делом достаточно отдаленного будущего (не ранее чем через 20-30 лет). Несомненно, что некоторые уже давно известные и недавно открытые материалы (например, графен – моноатомная 2-мерная решетка углерода с уникальными электронными и механическими свойствами) займут значительную нишу в будущей электронике. Тем не менее, полное вытеснение кремния как основного материала электроники представляется столь же маловероятным событием, как вытеснение железа в качестве основного конструкционного материала.

В этой книге речь пойдет о наноразмерных транзисторных структурах, изготовленных на базе кремния.

1.2. Пространственные масштабы нанoeлектроники

Согласно определению, транзисторы как элементы современных (2008) цифровых схем с размерами 45-90 нм относятся к разряду нанoeлектронных, или, в более общем плане, наноразмерных объектов. Структуры таких размеров занимают промежуточное место между макроскопическими и микроскопическими объектами. Характерные масштабы микроскопических объектов ограничиваются сверху приблизительно размерами атомов или молекул, не превышающими, как правило, 10^{-7} см = 10 ангстрем = 1 нм.

Для иллюстрации промежуточного положения современной транзисторной структуры между макроскопическим и микроскопическим миром приведем следующую иерархию масштабов:

- диаметр человеческого волоса ~ 20..100 мкм;
- частичка табачного дыма ~ 5 мкм;
- биологическая клетка ~ 1.7 мкм;
- разрешение оптического микроскопа ~ 0.2 мкм;

- длина волны фотона фиолетового цвета ~ 0.41 мкм;
- длина волны фотона красного цвета ~ 0.68 мкм;
- минимальная толщина сусального золота ~ 0.125 мкм;
- типичная бактерия ~ 1 мкм;
- вирус гриппа ~ 100 нм;
- размеры современных МОП транзисторов – $45 \dots 90$ нм;
- спираль ДНК ~ 2 нм в диаметре (так же, как и углеродная нанотрубка (carbon nanotube, CNT));
- длина де Бройля волны свободного электрона в кремнии (≤ 10 нм);
- постоянная кристаллической решетки $\sim 0.3 \dots 0.6$ нм;
- разрешение электронного микроскопа ~ 0.2 нм;
- атом $\sim 0.1 \dots 0.4$ нм.

Для описания микроскопических объектов требуется квантовая механика. Макроскопические объекты с микронными (микрон = = микрометр = 1000 нм) размерами и более, к которым можно отнести те же МОП-транзисторы двадцатилетней давности, описываются законами макроскопической (классической) физики. Масштабы, занимающие промежуточное положение между микроскопическими и макроскопическими размерами, часто называют мезоскопическими. На таких масштабах начинает проявляться ряд новых эффектов, не свойственных макроскопической физике. В первую очередь, это квантовомеханические эффекты: туннелирование, размерное квантование, квантовая интерференция. Туннелирование в стандартной транзисторной КМОП технологии играет роль только паразитного эффекта, обуславливая основной механизм утечек в наноэлектронных структурах. Квантовая (волновая) интерференция не проявляет себя заметным образом при работе современных кремниевых транзисторов. А вот размерное квантование играет фундаментальную роль не только на уровне эффектов влияния, но и на уровне формирования самой структуры.

Как известно, электрон в полупроводнике заключен в глубокую потенциальную яму, образуемую потенциальными энергиями ионов решетки. Если размеры этой ямы в одном из направлений сравниваются с длиной волны де Бройля, то электрон по этому направлению следует рассматривать не как частицу, а как стоячую волну. При этом одна из размерностей трехмерной системы «схло-

пывается» и трехмерная система превращается в двумерную (т.н. 2D) систему. Образцом такой двумерной (точнее, квазидвумерной) 2D системы, играющей огромную роль в современной технологии, является канал МОП-транзистора (МОПТ). В этом смысле уже первые образцы МОП-транзисторов являлись наноразмерными структурами.

Далее, если сделать двумерную систему достаточно узкой, то стоячая волна образуется и по ширине канала. Проводник превращается в своеобразный волновод, в котором электрон может свободно распространяться только в одном направлении. В таких случаях говорят о квантовых проволоках или нитях (quantum wires), являющихся (квази-)одномерными (1D) системами. В настоящее время размеры полупроводниковой структуры можно сделать настолько малым, что стоячие волны образуются по всем направлениям, что дает основания говорить о «нульмерных» (0D) системах, или, по-другому, квантовых точках (quantum dots) или искусственных атомах. Следует подчеркнуть условность терминологии: как и в реальных атомах, электрон находится, конечно же, в трехмерном пространстве. Таким образом, главной особенностью т.н. низкоразмерных структур является образование стационарной интерференционной картинки (стоячей волны), по крайней мере, в одном из направлений.

1.3. Общая структура нанoeлектронных приборов

Цифровая техника может быть построена на основе устройств, позволяющих модулировать ток во внешней цепи.

Принципиальная схема для широкого класса таких устройств представлена на рис. 1.1. Оно состоит из контактов внешней цепи («источка» и «стока»), активной области («канала») и управляющего электрода («затвора»). Именно такую структуру имеет МОП-транзистор – основной элемент современной цифровой техники. В принципе, роль активной области («канала») может играть не только инверсионный слой в кремниевой подложке, но и квантовая точка, углеродная нанотрубка или даже отдельная органическая молекула.



Рис. 1.1. Принципиальная схема транзисторной структуры с управляющим затвором

Основой приборов такого типа является электростатическое управление положением энергетических уровней в активной области прибора («канале») и связанная с этим возможность модуляции (контролируемого изменения) тока между «истоком» и «стоком», желательно в максимальном диапазоне значений. Поэтому «каналы» изготавливаются из полупроводников, а не из металлов, проводимостью которых управлять очень сложно.

Напротив, контакты (сток и исток) должны быть изготовлены из хорошего проводника (металла, либо сильнолегированного полупроводника), и контактное сопротивление должно быть минимальным. Управляющий затвор должен иметь максимальную взаимную емкость с каналом и минимальную взаимную емкость с истоком и стоком. В идеальном приборе напряжение между стоком и истоком должно влиять только на ток, но не на проводимость канала.

1.4. Энергии и потенциалы

Полная энергия носителя (электрона или дырки), не взаимодействующего с окружающей средой, представляется в виде суммы потенциальной и кинетической энергии

$$E = \varepsilon_p + U(x). \quad (1.4.1)$$

Кинетическая энергия в кремнии в параболическом изотропном приближении записывается в виде

$$\varepsilon_p = \frac{p^2}{2m}, \quad (1.4.2)$$

где m – эффективная масса в кремнии, p – импульс электрона.

Потенциальная энергия определяется электростатическим потенциалом $\varphi(x)$, а ее знак определяется знаком заряда носителя

$$U = \pm q\varphi, \quad (1.4.3)$$

где знаки плюс (минус) соответствуют дырке (электрону).

Изменение потенциальной энергии удобно отслеживать по изменению дна зоны проводимости полупроводника $E_C(x)$. Максимальный потенциал соответствует минимальной потенциальной энергии для электронов и максимальной потенциальной энергии для дырок. Сила в потенциальном поле всегда направлена в сторону уменьшения потенциальной энергии

$$f_x = -\frac{\partial U}{\partial x} = -q \frac{\partial \varphi}{\partial x}. \quad (1.4.4)$$

Вспоминая, что x -проекция электрического поля определяется как $F = -\partial\varphi/\partial x$, получаем очевидный вывод, что электрон притягивается в сторону большего, а дырка – в сторону меньшего потенциала.

Если носитель (для определенности, электрон) ни с чем не взаимодействует, то его полная энергия сохраняется, хотя кинетическая и потенциальная энергия могут переходить друг в друга. В этом случае имеем простую механическую систему, которую иногда называют консервативной. Перенос носителей в потенциальном поле с сохранением его полной энергии называется баллистическим переносом (рис. 1.2). Такой процесс доминирует, когда длина канала прибора, т.е. расстояние между контактами становится сопоставимым со средней длиной пробега носителя без взаимодействия с окружающей средой. Длина пробега в кремнии при комнатной температуре составляет величину ~ 10 нм, что уже сопоставимо с размерами активных областей современных приборов. Вообще говоря, электрон взаимодействует с окружением, которое иногда называют термостатом, и его транспорт носит т.н. диффузный характер (т.е. с частыми столкновениями). Например, электрон может взаимодействовать с колебаниями решетки (фононами), получая и отдавая энергию и импульс.

В равновесии скорость получения и набора электроном энергии и импульса от фононов равны друг другу. В неравновесном случае электрон получает энергию и импульс от электрического поля и

передает их фононам. Таким образом, энергия электрического поля преобразуется в энергию колебаний решетки, т.е. тепло. Этот необратимый процесс происходит с возрастанием энтропии и называется диссипацией. Рассеяние с потерей импульса происходит, как правило, чаще, чем рассеяние с потерей энергии, поэтому т.н. время пробега по энергии τ_ϵ всегда больше времени пробега по импульсу $\tau_\epsilon \geq \tau$.

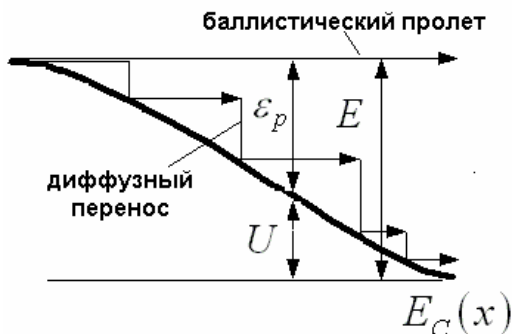


Рис. 1.2. Схема, иллюстрирующая диффузный (диссипативный) и баллистический перенос электрона

Например, в кремнии потеря энергии происходит за счет испускания т.н. оптических фононов с энергией ~ 0.06 эВ, а потеря импульса происходит за счет столкновений с т.н. акустическими фононами, при которых энергия электронов практически не меняется, но меняется направление импульса. Прежде чем электрон столкнется с оптическим фононом и изменит энергию, он успевает несколько раз столкнуться с акустическими фононами.

1.5. Что такое электрохимический потенциал?

Энергетическое распределение системы равновесных электронов дается функцией распределения Ферми-Дирака

$$f_0(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)}. \quad (1.5.1)$$

Этой же формулой можно иногда пользоваться и в неравновесном случае, вводя вместо уровня Ферми E_F неравновесное значение квазиуровня Ферми E_{Fn} (E_{Fp}), разное для электронов и дырок.

Если изменение потенциальной энергии в полупроводнике $U(x)$ (то же самое, что изменение, например, дна зоны проводимости $E_C(x)$) достаточно плавное, то в каждой точке трехмерного проводника можно определить локальное значение плотности электронов

$$\begin{aligned} n(\mathbf{r}) &= 2 \int \frac{dp_x dp_y dp_z}{(2\pi\hbar)^3} \left(1 + \exp\left(\frac{\varepsilon_p + U(\mathbf{r}) - E_F}{k_B T} \right) \right)^{-1} = \\ &= \int [d^3 p] \left(1 + \exp\left(\frac{\varepsilon_p - q\zeta(\mathbf{r})}{k_B T} \right) \right)^{-1}, \end{aligned} \quad (1.5.2)$$

где мы ввели сокращенное обозначение $[d^3 p] = 2dp_x dp_y dp_z / (2\pi\hbar)^3$ (двойка соответствует двум спинам на один уровень).

Из формулы (1.5.2) видно, что локальная концентрация носителей определяется энергетическим промежутком между локальным значением квазиуровня Ферми для этих носителей и краем соответствующей зоны

$$q\zeta(\mathbf{r}) \equiv E_{Fn} - U(\mathbf{r}). \quad (1.5.3)$$

Определенную таким образом величину называют химическим потенциалом для электронов. Для вырожденных полупроводников (когда $E_F > U = E_C$) химический потенциал положителен, для невырожденных ($E_F < E_C$) – отрицателен.

В физической литературе иногда эту же формулу представляют в виде

$$E_{Fn} \equiv \mu = q\zeta(\mathbf{r}) + U(\mathbf{r}) = q(\zeta(\mathbf{r}) - \varphi(\mathbf{r})), \quad (1.5.4)$$

вводя вместо понятия квазиуровня Ферми эквивалентное понятие электрохимического потенциала μ (не путать с подвижностью).

Электрохимический потенциал может изменяться двумя, в общем случае, независимыми способами:

- а) за счет изменения электростатического потенциала;
- б) за счет изменения концентрации носителей.

В равновесном случае изменения концентрации и электростатического потенциала связаны друг с другом через условие равенства нулю полного тока. В неравновесном случае электростатический потенциал φ и концентрация носителей (химический потенциал) являются, вообще говоря, независимыми величинами.

Во всех книгах по физике полупроводников легко найти, что для случая невырожденного полупроводника интеграл (1.5.2) дает следующее соотношение для локальной концентрации электронов:

$$n(x) = N_C \exp\left(-\frac{E_C(x) - E_{Fn}(x)}{k_B T}\right) = N_C \exp\left(-\frac{|\zeta(x)|}{\varphi_T}\right), \quad (1.5.4)$$

где N_C – эффективная плотность электронов в зоне проводимости ($\sim 10^{19} \text{ см}^{-3}$ в кремнии).

Роль электрохимического потенциала удобно проиллюстрировать на зонной диаграмме равновесного pn -перехода (рис. 1.3), где химический и электрический потенциал по толщине перехода меняются, но электрохимический потенциал остается постоянным.

$$|\mu(x)|/q = \varphi(x) + |\zeta(x)| = \varphi(x) + \varphi_T \ln \frac{N_C}{n(x)} = \text{const}, \quad (1.5.5)$$

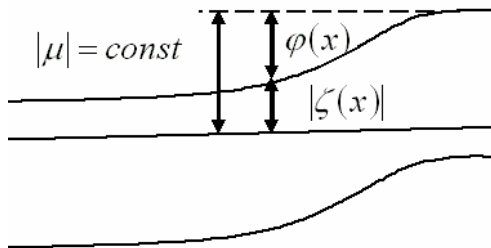


Рис. 1.3. Зонная диаграмма равновесного pn -перехода

Важно помнить, что вольтметр измеряет не разность электрических потенциалов, а разность электрохимических потенциалов! Например, разность электрических потенциалов в равновесном pn -переходе составляет величину барьера ($\sim 0.7 \dots 0.9 \text{ В}$), в то время как разность электрохимических потенциалов при отсутствии тока через переход строго равна нулю.

1.6. Элементарная кинетика

В невырожденных полупроводниках электроны движутся хаотично с тепловой скоростью $\sim \sqrt{k_B T/m} \sim 10^7$ см/с (для комнатных температур), причем средний вектор скорости системы электронов в равновесном полупроводнике равен нулю. Предположим, в момент времени $t=0$ мгновенно включили однородное электрическое поле F . Тогда уравнение для избыточной (по сравнению с нулевым электрическим полем) скорости электрона Δv имеет вид

$$m \frac{d \Delta v}{dt} = q F - \frac{m \Delta v}{\tau}, \quad (1.6.1)$$

где m – эффективная масса электрона, τ – время свободного пробега по импульсу ($\sim 10^{-12} \dots 10^{-14}$ с). Первое слагаемое в правой части уравнения (1.4.1) описывает набор импульса электрона за счет внешнего электрического поля; второе слагаемое описывает «торможение» электронов за счет передачи импульса фононам и примесям (рис. 1.4).

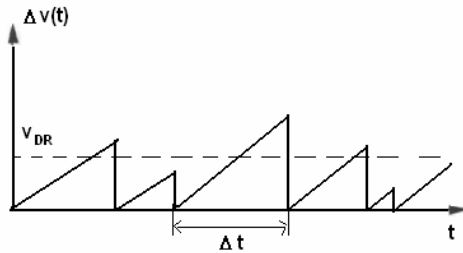


Рис. 1.4. Избыточная скорость носителя во внешнем поле как функция времени

За очень малые времена (порядка нескольких времен свободного пробега $\sim 10^{-12} \dots 10^{-13}$ с) устанавливается стационарное состояние ($d\Delta v/dt = 0$) с постоянной дрейфовой скоростью v_{DR} , когда набор импульса электронов компенсируется сбросом импульса за счет процессов рассеяния

$$v_{DR} = \frac{q}{m} \tau F = \mu F, \quad (1.6.2)$$

где введена подвижность носителей $\mu \equiv q\tau/m$.

Плотность дрейфового тока j_{DR} и удельная проводимость системы σ определяются через дрейфовую скорость и плотность свободных носителей заряда n

$$j_{DR} = qn v_{DR} = q \mu n v_{DR} \equiv \sigma F . \quad (1.6.3)$$

Определенная таким образом удельная проводимость σ (обратное удельное сопротивление ρ) дается выражением

$$\sigma = \frac{1}{\rho} = \frac{q^2 \tau n}{m} . \quad (1.6.4)$$

Легко убедиться, что для однородных проводников соотношение (1.6.3) эквивалентно закону Ома:

$$j = \sigma F = F/\rho \quad \Leftrightarrow \quad I = V/R \quad (1.6.5)$$

и имеет одинаковый вид в системах разной мерности (3D, 2D и 1D). Но размерность проводимости для разного числа измерений имеет разный вид (ниже выражения записаны в системах единиц СИ и СГСЭ):

$$\begin{aligned} [j_{DR}]_{3D} &= \frac{\text{Кл}}{\text{см}^2 \times \text{с}} , \\ [\sigma]_{3D} &= \frac{1}{\text{Ом} \times \text{см}} \text{ (СИ)}, \quad [\sigma]_{3D} = \frac{1}{\text{с}} \text{ (СГСЭ)} \\ [j_{DR}]_{2D} &= \frac{\text{Кл}}{\text{см} \times \text{с}} , \\ [\sigma]_{2D} &= \frac{1}{\text{Ом}} \text{ (СИ)}, \quad [\sigma]_{2D} = \frac{\text{см}}{\text{с}} \text{ (СГСЭ)}, \\ [j_{DR}]_{1D} &= \frac{\text{Кл}}{\text{с}} , \\ [\sigma]_{1D} &= \frac{\text{см}}{\text{Ом}} \text{ (СИ)} , \quad [\sigma]_{1D} = \frac{\text{см}^2}{\text{с}} \text{ (СГСЭ)}. \end{aligned}$$

Отметим, что для трехмерной системы проводимость в СГСЭ имеет размерность обратного времени, для двумерной – скорости, а для одномерной – коэффициента диффузии. Важно, что для корректного определения удельной проводимости, размеры системы должны быть таковы, чтобы в ней происходило несколько актов рассеяния, т.е. чтобы длина свободного пробега ℓ была много

меньше размеров образца $\ell \ll L$. Длина свободного пробега определяется временем свободного пробега носителя τ и скоростью Ферми v_F для вырожденного или тепловой скоростью v_{th} для невырожденного электронного газа $\ell = v_{F/th}\tau$.

Как уже говорилось, размерность проводника определяется соотношением длины пробега и характерных размеров проводника. Например, если для толщины d и ширины W проводника мы имеем неравенства $W \gg \ell$ и $d \ll \ell$, то такой проводник можно считать двумерным; если $W, d < \ell$ – то одномерным.

1.7. Диффузионно-дрейфовый ток

Диффузионный ток j_{DIFF} обусловлен градиентом концентрации

$$j_{DIFF} = -qD \frac{dn}{dx}, \quad (1.7.1)$$

где D – коэффициент диффузии с размерностью $[cm^2/c]$.

Плотность полного тока в полупроводниках всегда можно представить суммой плотностей дрейфового j_{DR} и диффузионного тока j_{DIFF} :

$$j = j_{DR} + j_{DIFF} = \sigma F - qD \frac{dn}{dx}. \quad (1.7.2)$$

Формально это соотношение (записанное для носителей с условно положительным зарядом) справедливо и для металлов, но в металлах очень сложно создать сколько-нибудь заметный градиент концентрации свободных носителей, что обусловлено их огромной концентрацией ($\sim 10^{23} cm^{-3}$). В полупроводниках с типичным диапазоном концентраций носителей $10^{15} \dots 10^{20} cm^{-3}$ градиент концентраций можно относительно легко создать за счет неоднородного легирования и/или внешней инжекции.

В равновесии плотность тока равна нулю $j = 0$ и диффузионные и дрейфовые потоки носителей строго уравниваются друг друга независимо для электронов и дырок. Приравнявая плотность тока (1.7.2) нулю, получаем связь между подвижностью μ и коэффициентом диффузии D :

$$\frac{D}{\mu} = n \frac{F dx}{dn} = \frac{n}{|dn/d\phi|}. \quad (1.7.3)$$

Эта связь называется соотношением Эйнштейна и в форме (1.7.3) оно остается справедливым как для вырожденных, так и для невырожденных систем, т.е как для металлов, так и для полупроводников.

Для невырожденных полупроводников справедлива статистика Больцмана $dn/d\phi = qn/k_B T$, что непосредственно следует из (1.5.4), и соотношение Эйнштейна (1.7.3) принимает простой вид

$$\frac{D}{\mu} = \frac{k_B T}{q} \equiv \varphi_T, \quad (1.7.4)$$

где определяется т.н. тепловой потенциал ($\varphi_T \cong 0.026$ В для комнатной температуры $T = 300$ К).

1.8. Уравнение Больцмана

Для простоты обозначений будем рассматривать одномерное приближение, поскольку обобщение на многомерный случай, как правило, не представляет труда. В неравновесном случае формула (1.5.1), вообще говоря, перестает быть справедливой. Для определения неравновесной функции распределения необходимо решать кинетическое уравнение Больцмана, которое в можно записать в виде ($q > 0$):

$$\frac{\partial f}{\partial t} + v_x \frac{\partial f}{\partial x} - qF \frac{\partial f}{\partial p_x} = -\frac{f - f_0}{\tau}. \quad (1.8.1)$$

Правая часть уравнения Больцмана называется интегралом столкновений и характеризует, скорость с которой любое отклонение от равновесия в ФР распределения возвращается в равновесное состояние, характеризуемое функцией распределения Ферми-Дирака f_0 (см.1.5.1). Скорость этой релаксации в очень грубом приближении (т.н. приближение времени релаксации) характеризуется временем свободного пробега τ .

Все интересующие нас величины (плотности носителей и плотности токов) можем получить согласно общим правилам:

плотность тока:

$$j_x(x) = q \int [d^d p] v_x f(x, t) \equiv q \langle v_x f \rangle, \quad (1.8.2)$$

и концентрация носителей:

$$n(x,t) = \int [d^d p] f_p(x,t) \equiv \langle f \rangle. \quad (1.8.3)$$

Плотность тока и концентрацию можно в принципе измерить, и к тому же они зависят только от координат и времени, и не зависят от импульсов; поэтому с ними проще иметь дело, чем с функциями распределений.

Рассмотрим равновесный, стационарный, но неоднородный случай. Неоднородная система типична для приборов, что может быть обусловлено, например, неоднородным легированием. В равновесном случае правая часть уравнения Больцмана (т.е. интеграл столкновения) равна нулю. Стационарность (независимость от времени) соответствует условию $\partial f / \partial t = 0$.

Тогда уравнение Больцмана для электронов принимает вид

$$v_x \frac{\partial f}{\partial x} - qF \frac{\partial f}{\partial p_x} = 0. \quad (1.8.4)$$

Нетрудно проверить, что решением этого уравнения является *любая* функция *полной* энергии

$$\tilde{f}(E) = \tilde{f}(\varepsilon_p + U(x)) = \tilde{f}(\varepsilon_p - q\varphi(x))$$

$$v_x \frac{\partial \tilde{f}}{\partial x} - qF \frac{\partial \tilde{f}}{\partial p_x} = \frac{d\tilde{f}}{d\varepsilon} \left(v_x \frac{\partial}{\partial x} (-q\varphi(x)) - qF \frac{\partial \varepsilon_p}{\partial p_x} \right) = 0, \quad (1.8.5)$$

где мы воспользовались тем, что $F = -\partial\varphi/\partial x$ и $v_x = \partial\varepsilon_p/\partial p_x$.

Это означает, что распределение зарядов в равновесной структуре определяется распределением электрического потенциала и для его нахождения необходимо решать уравнение Пуассона совместно с выражением типа (1.5.1). Для нахождения положения постоянного электрохимического потенциала в равновесной структуре достаточно воспользоваться условием электрической нейтральности системы в локальной (как в объеме), либо в глобальной (как слоистой структуре МОПТ) форме.

1.9. Уравнение непрерывности

Уравнение Больцмана в общем виде решить почти невозможно. На практике вместо уравнения Больцмана используют его более простые физические следствия, переходя от описания с функциями распределений к описанию с плотностями токов и концентрациями. Для того чтобы избавиться от импульсной зависимости, проинтегрируем каждое слагаемое уравнения Больцмана (1.8.2), используя (1.8.2), (1.8.3):

$$\left\langle \frac{\partial f}{\partial t} \right\rangle = \frac{\partial}{\partial t} \langle f \rangle = \frac{\partial n}{\partial t}, \quad (1.9.1)$$

$$\left\langle v_x \frac{\partial f}{\partial x} \right\rangle = \frac{\partial}{\partial x} \langle v_x f \rangle = \frac{\partial J_x}{\partial x}, \quad (1.9.2)$$

где J_x – плотность потока частиц,

$$qF_x \int [dp] \frac{\partial f}{\partial p_x} \propto \int_{-\infty}^{+\infty} dp_x \frac{\partial f}{\partial p_x} = f(p_x \rightarrow +\infty) - f(p_x \rightarrow -\infty) = 0, \quad (1.9.3)$$

$$\langle f - f_0 \rangle = \langle f \rangle - \langle f_0 \rangle = 0. \quad (1.9.4)$$

Таким образом, т.н. нулевой момент уравнения Больцмана дает уравнение непрерывности, характеризующее сохранение количества свободных носителей (для краткости опускаем рассмотрение процессов рекомбинации и генерации):

$$\frac{\partial n}{\partial t} + \frac{\partial J_x}{\partial x} = 0. \quad (1.9.5)$$

1.10. Уравнение баланса импульсов и диффузионно-дрейфовое приближение

Домножим каждое слагаемое уравнения Больцмана (1.8.1) на проекцию скорости (или импульса, что эквивалентно в рамках используемого нами приближения) и проинтегрируем по импульсам

$$\left\langle v_x \frac{\partial f}{\partial t} \right\rangle = \frac{\partial}{\partial t} \langle v_x f \rangle = \frac{\partial}{\partial t} J_x, \quad (1.10.1)$$

$$\left\langle v_x^2 \frac{\partial f}{\partial x} \right\rangle = \frac{\partial}{\partial x} \langle v_x^2 f \rangle, \quad (1.10.2)$$

$$\left\langle -qF_x \frac{p_x}{m} \frac{\partial f}{\partial p_x} \right\rangle = -\frac{qF_x}{m} \int p_x \frac{\partial f}{\partial p_x} [dp] = \frac{qF_x}{m} \int f [dp] = \frac{qF_x}{m} n, \quad (1.10.3)$$

$$-\frac{\langle f - f_0 \rangle}{\tau} = \frac{\langle v_x f \rangle}{\tau} - \overbrace{\frac{\langle v_x f_0 \rangle}{\tau}}^{=0} = -\frac{J_x}{\tau}. \quad (1.10.4)$$

Таким образом, получаем т.н. уравнение баланса импульсов

$$\frac{\partial J_x}{\partial t} + \frac{\partial}{\partial x} \langle v_x^2 f \rangle + \frac{qF_x}{m} n = -\frac{J_x}{\tau}. \quad (1.10.5)$$

Рассмотрим подробнее второе слагаемое в (1.10.5):

$$\langle v_x^2 f \rangle = \frac{\langle v_x^2 f \rangle}{\langle f \rangle} \langle f \rangle \equiv \langle \langle v_x^2 \rangle \rangle n = \frac{nk_B T}{m}. \quad (1.10.6)$$

Последнее равенство в (1.10.6) обусловлено тем, что в пространстве с размерностью d имеем

$$\langle \langle v_x^2 \rangle \rangle = \frac{\langle \langle v_x^2 \rangle \rangle + \langle \langle v_y^2 \rangle \rangle + \langle \langle v_z^2 \rangle \rangle}{d} = \frac{\langle \langle v^2 \rangle \rangle}{d} = \frac{k_B T}{m}, \quad (1.10.7)$$

поскольку все направления скоростей равноправны и для бoльцмановской статистики справедлив известный закон равнораспределения для d степеней свободы

$$\frac{m \langle \langle v^2 \rangle \rangle}{2} = \frac{d}{2} k_B T. \quad (1.10.8)$$

Если вместо плотности потока электронов ввести плотность тока электронов $j_x = -qJ_x$, то в стационарном случае уравнение (1.10.5) превращается в выражение для диффузионно-дрейфового тока для электронов

$$j_x = \overbrace{\left(\frac{q^2 \tau}{m} \right)}^{q\mu_n} n F_x + q \frac{\partial}{\partial x} (nk_B T \tau) = \sigma F_x + q \frac{\partial}{\partial x} (D_n n), \quad (1.10.9)$$

где подвижность и коэффициент диффузии связаны соотношением Эйнштейна для невырожденных носителей:

$$\mu_n = \frac{q}{m} \tau, \quad D_n = \frac{k_B T}{m} \tau, \quad \frac{D_n}{\mu_n} = \frac{k_B T}{q}. \quad (1.10.10)$$

Таким образом, мы показали, что для стационарной задачи уравнение баланса импульса (1.10.5) эквивалентно выражению для диффузионно-дрейфового тока (1.7.2), а знак плюс в (1.10.9) соответствует тому, что поток электронов направлен противоположно направлению электрического поля.

Распределение локальных температур (если это существенно и такую величину можно ввести) можно получить из уравнения баланса энергий. Это уравнение можно получить, домножая каждое слагаемое уравнения Больцмана на кинетическую энергию и интегрируя по импульсам.

1.11. Электрон как волна и длина когерентности

Из квантовой механики известно, что электроны могут интерферировать, так же как и свет. На рис. 1.2 представлена схема т.н. двухщелевого эксперимента. Электроны из источника S падают с очень малой интенсивностью на непроницаемый экран с двумя щелями. За экраном на небольшом расстоянии L расположена фотопластинка, позволяющая визуализировать распространение электронов, проходящих через щели. Если закрыть одну из щелей, то пик прошедших электронов окажется под открытой щелью. С точки зрения классической физики естественно было бы предположить, что эксперимент с двумя открытыми щелями даст два пика, соответствующим двум экспериментам с одной закрытой щелью (эти пики обозначены на рисунке штриховой линией).

В реальности эксперимент с двумя открытыми щелями может приводить к интерференционной картинке распределения интенсивности с пиком в центре между щелями, очень похожей на ту, что наблюдается при прохождении света (фотонов) через подобную структуру.

Интерференция пропадает за счет случайного взаимодействия электронов с окружающей средой (атомы, фононы, фотоны). Процессы потери когерентности иногда называются декогеренизацией (или декогеренцией). Эти процессы можно характеризовать временем и длиной когерентности. Время когерентности (или время потери фазы) τ_φ определяется, главным образом, неупругими процессами (т.е. происходящими с изменением энергии) взаимодействия электронов с окружающей средой.

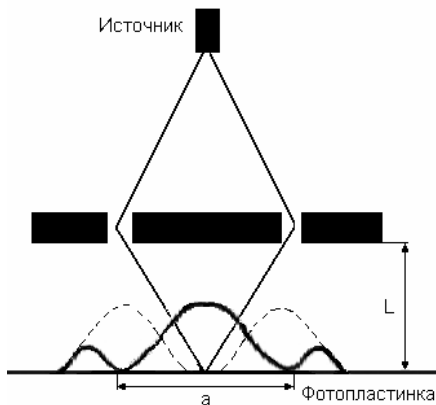


Рис. 1.2. Схема двухщелевого эксперимента

Длина когерентности L_ϕ – это расстояние, преодолеваемое электроном за время τ_ϕ . Если L на рис. 1.2 достаточно велико $L \gg L_\phi$, то интерференционная картинка не наблюдается. Строго говоря, именно соотношение длины когерентности (а не длины свободного пробега) и размеров проводника определяет его размерность.

1.12. Математическое описание волн

Волновой процесс можно характеризовать, например, синусоидой (или косинусоидой) с длиной волны λ , частотой ν (рис. 1.3):

$$\psi(x, t) = A \cos\left(\frac{2\pi x}{\lambda} - 2\pi \nu t\right) \equiv A \cos(kx - \omega t), \quad (1.12.1)$$

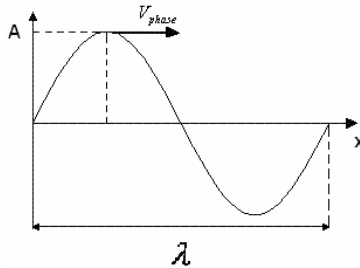


Рис. 1.3. Синусоида бегущей волны с амплитудой A и длиной волны λ , где введено обозначение $k = 2\pi/\lambda$ – волновой вектор, $\omega = 2\pi\nu$ – циклическая частота.

Положение максимума x_m любого из горбов синусоиды в (1.12.1) движется вправо постоянной (т.н. фазовой скоростью) $V_{phase} = \omega/k = \nu\lambda$:

$$k x_m - \omega t = 0 \Rightarrow x_m = (\omega/k)t \equiv V_{phase} t. \quad (1.12.2)$$

Волну, бегущую влево, можно описывать функцией $\cos(kx + \omega t)$ (или $\sin(kx + \omega t)$). Линейная сумма двух или множества независимых волн называется суперпозицией. Волны, бегущие навстречу друг другу с равной амплитудой A и подходящей фазой, образуют стоячую волну, в которой пучности (горбы) и узлы не движутся:

$$A \cos(kx - \omega t) + A \cos(kx + \omega t) = 2A \cos(kx) \cos(\omega t). \quad (1.12.3)$$

Часто удобно пользоваться не синусами и косинусами, а комплекснозначными экспонентами, основанными на известных формулах Муавра $\exp(i\varphi) = \cos(\varphi) + i \sin(\varphi)$,

$$A \exp(kx - \omega t) + A \exp(-kx - \omega t) = 2A \cos(kx) \exp(-i\omega t), \quad (1.12.4)$$

где не зависящее от времени положение узлов и пучностей характеризует стационарный характер стоячей волны.

1.13. Уравнение Шредингера и волновая функция

Согласно постулатам квантовой механики движение частиц описывается волной, причем частота ω соответствует энергии, а длина волны и волновой вектор k – импульсу:

$$\varepsilon = h\nu = \hbar\omega, \quad (1.13.1)$$

$$\lambda = \frac{h}{p} \Rightarrow p = \frac{h}{2\pi} \frac{2\pi}{\lambda} \equiv \hbar k, \quad (1.13.2)$$

где введена постоянная Планка

$$\hbar \equiv h/2\pi = 1.05 \times 10^{-34} \text{ Дж}\cdot\text{с} = 6.6 \times 10^{-16} \text{ эВ}\cdot\text{с}. \quad (1.13.3)$$

Реальный электрон можно представить как сумму (линейную суперпозицию) волн с разными λ и k (волновой пакет). Каждой длине волны соответствует своя энергия, и соответствующая связь называется законом дисперсии. Например, закон дисперсии в кремнии хорошо описывается т.н. параболическим приближением с эффективной массой m :

$$\varepsilon_p = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2}{2m} \left(\frac{2\pi}{\lambda} \right)^2. \quad (1.13.4)$$

Классическая скорость электрона определяется не фазовой, а групповой скоростью (скорость перемещения максимума огибающей кривой волнового пакета):

$$v = \frac{d\omega}{dk} = \frac{d\varepsilon_p}{dp} = \frac{p}{m}. \quad (1.13.5)$$

Важно подчеркнуть, что для электронов в материалах закон дисперсии отражает структуру кристаллической решетки данного материала и может иметь совершенно различный вид. Например, в монослое углерода (т.н. графене) закон дисперсии вообще не содержит эффективной массы, но содержит константу скорости: $\varepsilon(p) = \hbar v_0 k = v_0 p$. Такой закон дисперсии больше похож на закон дисперсии для фотона, где роль «скорости света» играет некая характерная скорость носителя в графене, постоянна по абсолютной величине $v = d\varepsilon/dp = v_0$ ($\cong 10^8$ см/с) и никак не связанная с величиной импульса носителя!

Волновые процессы квантовой механики описываются комплексной функцией координаты и времени, которая называется

волновой функцией $\psi(x,t)$, представляемых в виде суперпозиции комплексных волн типа (1.12.4). Физический смысл волновой функции определяется постулатом Борна, который говорит о том, что величина

$$|\psi(x,t)|^2 dx = \psi^*(x,t)\psi(x,t)dx \quad (1.13.5)$$

дает вероятность того, что в момент времени t электрон находится в точке $x \div x + dx$.

Волновая функция является результатом решения уравнения Шредингера. Структуру уравнения Шредингера можно понять (но не вывести), рассматривая выражение для полной энергии электрона в системах, где она сохраняется (1.4.1).

Согласно принципу соответствия, энергии и импульсу соответствуют операторы, действующие на комплекснозначную волновую функцию $\psi(x,t)$:

$$E \rightarrow i\hbar \frac{\partial}{\partial t}, \quad p_x \rightarrow -i\hbar \frac{\partial}{\partial x}, \quad p_y \rightarrow -i\hbar \frac{\partial}{\partial y}, \quad p_z \rightarrow -i\hbar \frac{\partial}{\partial z}. \quad (1.13.6)$$

Тогда из (1.4.1) «получаем» одномерное уравнение Шредингера:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) + U\psi. \quad (1.13.7)$$

Это дифференциальное уравнение требует, конечно, задания начального и граничных условий.

1.14. Стационарное уравнение Шредингера

Для простоты сначала рассмотрим одномерную задачу. Если полная энергия электронов сохраняется ($E = \text{const}$), то уравнение Шредингера (1.13.7) с помощью замены переменных $\Psi(x,t) = \psi(x)\exp(-iEt/\hbar)$ приводится к более простому стационарному:

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x). \quad (1.14.1)$$

Эта задача на собственные значения $H\psi = E\psi$ оператора энергии (т.н. гамильтониана)

$$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x), \quad (1.14.2)$$

где энергия – это собственные значения, волновая функции – собственные функции гамильтониана.

В свободном пространстве потенциальная энергия равна нулю $U(x) = 0$, и уравнение Шредингера принимает очень простой вид

$$-\frac{\hbar^2}{2m}\psi''(x) = E\psi(x) \quad (1.14.3)$$

с собственными функциями $\exp(\pm ikx)$ и собственным значением, в том смысле, что подставляя такие собственные функции в (1.14.3), мы получаем собственные значения в форме закона дисперсии

$$E = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}. \quad (1.14.5)$$

Общее решение уравнения Шредингера в свободном пространстве имеет вид

$$\begin{aligned} \psi(x, t) &= c_1 \exp\left(ikx - i\frac{\hbar^2 k^2}{2m}t\right) + c_2 \exp\left(-ikx - i\frac{\hbar^2 k^2}{2m}t\right) = \\ &= (c_1 \exp(ikx) + c_1 \exp(-ikx)) \exp(-iEt/\hbar), \end{aligned} \quad (1.14.6)$$

где два слагаемых интерпретируются как описывающие частицу с одной энергией, но с противоположными направлениями импульса (первое слагаемое соответствует волне, бегущей вправо, а второе – волне, бегущей влево). Если $c_1 = c_2$, то волновая функция (1.14.6) описывает стоячую волну, как в случае (1.12.5).

Важно отметить, что в квантовой механике любая линейная суперпозиция волновых функций

$$\psi(x, t) = c_1\psi_1(x, t) + c_2\psi_2(x, t) \quad (1.14.7)$$

может описывать какую-то физическую ситуацию (принцип суперпозиции), а выбор базисного набора собственных функций определяется характером физической системы и соображениями удобства.

1.15. Электрон в бесконечно глубокой яме

Для бесконечного пространства энергия (1.14.5) и волновой вектор (1.14.6) являются непрерывной функцией. Это означает, что энергия и импульс (по направлению и абсолютной величине) могут принимать любые значения. Математически такая ситуация возникает из-за того, что формально бесконечное пространство не накладывает на решение граничных условий.

Электрон в твердом теле можно грубо представить как объект, находящийся в трехмерной потенциальной яме с глубиной порядка работы выхода (обычно, несколько эВ). Для любых наноразмерных структур ключевую роль играет наличие границ. Граничные условия, накладываемые на волновую функцию, приводят к тому, что импульс (и, соответственно, энергия) может принимать только дискретные значения, либо значения в некотором диапазоне. В этом случае говорят об эффектах размерного квантования в квантовых ямах.

Рассмотрим модель электрона в одномерной бесконечно глубокой яме, где $U(x)=0$ при $0 < x < L$ и $U(0)=U(L)=\infty$. Общее решение (1.14.3) в такой яме можно записать в виде

$$\psi(x) = a \sin(kx) + b \cos(kx), \quad (1.15.1)$$

и эта задача отличается от случая бесконечного пространства только накладываемыми граничными условиями

$$\psi(x=0) = \psi(x=L) = 0, \quad (1.15.2)$$

причем строгое равенство нулю имеет место для формально бесконечной глубины потенциальной ямы. Из граничных условий (1.15.2) очевидно, что $b = 0$ и волновой вектор (и длина волны) может принимать только дискретные значения, определяемые условием

$$kL = \pi n, \quad n = 1, 2, 3, \dots, \quad k_n = \frac{\pi}{L} n. \quad (1.15.3)$$

Вспоминая определение волнового вектора (1.13.2), получаем, что (1.15.3) эквивалентно тому, что на длине ямы может уместиться только целое число полуволин

$$n = \frac{L}{\lambda_n/2}. \quad (1.15.4)$$

Тогда волновые функции и собственные значения энергии для разного значения n (т. н. квантового числа) будут иметь вид

$$\Psi_n(x) = a \sin\left(\frac{\pi x}{L} n\right),$$

$$\varepsilon_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2}{2m} \frac{\pi^2}{L^2} n^2 \equiv \varepsilon_1 n^2, \quad (1.15.5)$$

а энергия основного состояния (т.е. состояния с минимальной энергией) электрона в яме будет соответствовать волновой функции с одной полуволной на длину ямы ($n=1$) со значением в кремнии

$$\varepsilon_1 = \frac{\hbar^2}{2m} \frac{\pi^2}{L^2} \cong \frac{m_0}{m} \left(\frac{10 \text{ нм}}{L}\right)^2 3.8 \text{ мэВ}, \quad (1.15.6)$$

где m_0 – масса электрона в пустоте.

Константу a (1.15.5) можно определить с помощью условия нормировки волновой функции:

$$1 = a^2 \int_0^{\overbrace{L}^{=L/2}} \sin^2\left(\frac{\pi x}{L} n\right) dx \quad \Rightarrow \quad a = (2/L)^{1/2}.$$

Тогда состояние электрона на n -ом уровне описывается волновой функцией

$$\Psi_n(x, t) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{\pi x}{L} n\right) \exp\left(-\frac{\varepsilon_n t}{\hbar}\right). \quad (1.15.7)$$

В трехмерном случае для потенциальной ямы с размерами $L_x \times L_y \times L_z$ уравнение Шредингера имеет вид

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(x) = E \psi(x). \quad (1.15.8)$$

Волновые функции электронов трехмерной задачи с нулевыми граничными условиями записываются с помощью тройки независимых целочисленных квантовых чисел n_x, n_y, n_z (ср. (1.15.7) и (1.15.5)):

$$\Psi_{n_x n_y n_z}(x, t) = \left(\frac{8}{L_x L_y L_z} \right)^{1/2} \times \\ \times \sin\left(\frac{\pi x}{L_x} n_x \right) \sin\left(\frac{\pi y}{L_y} n_y \right) \sin\left(\frac{\pi z}{L_z} n_z \right) \exp\left(-\frac{\varepsilon_{n_x n_y n_z} t}{\hbar} \right), \quad (1.15.9)$$

с квантованными значениями энергии

$$\varepsilon_{n_x n_y n_z} = \frac{\hbar^2 \pi^2}{2m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right). \quad (1.15.10)$$

Если расстояние между ближайшими энергетическими уровнями оказывается меньше $k_B T$, то уровни «размываются», и размерное квантование становится несущественным. Например, кубик металла с объемом $\Omega = 1 \text{ км}^3$ слишком велик, чтобы там было заметно влияние размерного квантования. Действительно, концентрация электронов в металлах $N_{at} \sim 10^{22} \text{ см}^{-3}$, и металл можно представить как потенциальную яму глубиной $\sim 5 \text{ эВ}$, причем на каждом уровне находится два электрона с разными спинами. Тогда среднее расстояние между уровнями $\sim 5 \text{ эВ} / N_{at} \Omega \sim 10^{-10} \text{ эВ} \ll \ll k_B T$.

1.16. Плотность дискретного и непрерывного спектра двумерной системы

Рассмотрим задачу о количестве разрешенных состояний двумерной системы с энергией, меньшей некоторого значения ε для наноразмерных образцов (т.е. с учетом квантования).

Для простоты рассмотрим квадратную систему $L_x = L_y = L$:

$$E_{n_x, n_y} = \frac{\hbar^2 \pi^2}{2mL^2} (n_x^2 + n_y^2) \equiv \varepsilon_0 \times (n_x^2 + n_y^2). \quad (1.16.1)$$

Параметры n_x и n_y , принимающие целочисленные положительные значения, имеют смысл количества полувольт, укладываемых на соответствующей грани прямоугольной двумерной системы. Состояние электрона с самой низкой энергией описывается квантовыми числами

$$(n_x, n_y) = (1, 1), \quad E_{11} = 2\varepsilon_0.$$

Далее имеем два состояния (1, 2) и (2, 1) с одинаковой энергией $(n_x, n_y) = (1, 2)$ и $(n_x, n_y) = (2, 1)$, $E_{12} = E_{21} = 5\varepsilon_0$.

Равенство энергий для разных состояний называется вырождением по энергиям. В данном случае оно носит случайный характер, поскольку обусловлено квадратностью системы. Следующим по величине энергии будет невырожденный уровень $E_{22} = 8\varepsilon_0$ с энергией ниже, чем у двух вырожденных уровней $E_{13} = E_{31} = 10\varepsilon_0$ (рис.1.4).

В этом случае легко подсчитать количество состояний с энергией, меньшей некоторой заданной энергии ε . Действительно, удвоенное (из-за наличия спина) количество квадратиков в четверти круга с радиусом $k_\varepsilon = \sqrt{2m\varepsilon/\hbar^2}$ и площадью $\pi k_\varepsilon^2/4$ можно приблизительно оценить как

$$Z_{2D}(\varepsilon) = 2 \frac{\pi k_\varepsilon^2/4}{(\pi/L)^2} = L^2 \frac{k_\varepsilon^2}{2\pi^2} = L^2 \frac{m}{\pi \hbar^2} \varepsilon. \quad (1.16.2)$$

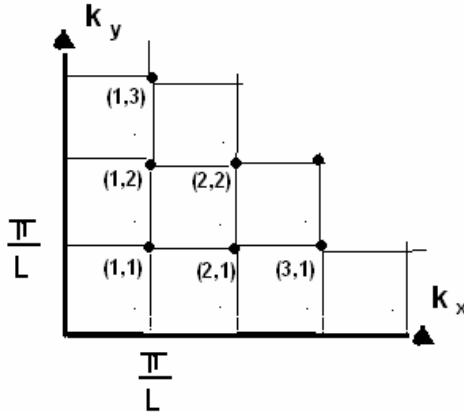


Рис. 1.4. Дискретный спектр двумерного (2D) проводника квадратной формы

Отсюда плотность состояний на единицу энергии и на единицу площади двумерной системы

$$g_{2D} = \frac{1}{L^2} \frac{dZ_{2D}}{d\varepsilon} = \frac{m}{\pi \hbar^2}. \quad (1.16.3)$$

1.17. Энергетическая плотность состояний

Чем больше размеры потенциальной ямы, тем меньше энергетическое расстояние между ближайшими уровнями. Например, при увеличении размеров L_x и L_y энергетическое расстояние между уровнями при изменении n_x и n_y уменьшается, и дискретный спектр по этим направлениям постепенно превращается в непрерывный:

$$\varepsilon_{n_x n_y n_z} \xrightarrow{L_x, L_y \rightarrow \infty} \varepsilon_{p_x p_y n_z} \cong \frac{p_x^2 + p_y^2}{2m} + \frac{\hbar^2 \pi^2}{2m} \frac{n_z^2}{L_z^2}. \quad (1.17.1)$$

В предельном случае мы имеем непрерывный спектр по плоскости $L_x \times L_y$ и дискретный спектр по поперечному направлению z . При этом все электроны разбиваются на группы с разным значением квантового числа n_z (т.е. количества полуволн, укладываемых-

ся на ширине ямы L_z). Такие группы образуют т.н. подзоны, каждая из которых имеет собственное дно, соответствующее разным значениям n_z (рис. 1.5).

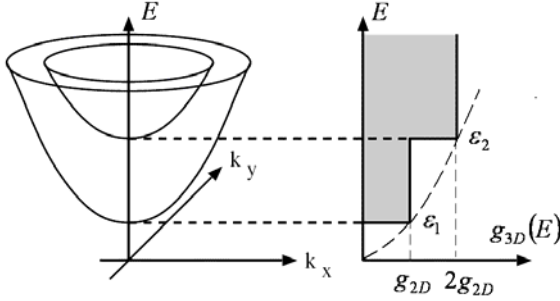


Рис. 1.5. Энергетический спектр и плотность состояний трехмерного проводника как суперпозиция плотностей состояний двумерных подзон

Результат (1.16.3) для непрерывного двумерного спектра можно получить другим способом:

$$2 \frac{dp_x dp_y}{(2\pi\hbar)^2} = 2 \frac{2\pi p dp}{(2\pi\hbar)^2} = \frac{m}{\pi\hbar^2} d\varepsilon \equiv g_{2D} d\varepsilon. \quad (1.17.1)$$

Концентрация электронов для двумерного электронного газа вычисляется как

$$\begin{aligned} n_{2D} &= \int f(p) [d^2 p] \cong \int_{\varepsilon_1}^{E_F} \frac{g_{2D}(\varepsilon)}{1 + \exp\left(\frac{\varepsilon - E_F}{k_B T}\right)} d\varepsilon = \\ &= g_{2D} k_B T \ln \left(1 + \exp\left(\frac{E_F - \varepsilon_1}{k_B T}\right) \right), \end{aligned} \quad (1.17.2)$$

где $\varepsilon_F = E_F - \varepsilon_1$ – положение уровня Ферми, отсчитанного от дна нижней подзоны.

Для вырожденного случая ($E_F - \varepsilon_1 \gg k_B T$) имеем

$$n_{2D} = g_{2D} \varepsilon_F = \frac{m \varepsilon_F}{\pi \hbar^2} = \frac{p_F^2}{2\pi \hbar^2}, \quad (1.17.3)$$

и p_F – энергия и импульс Ферми для вырожденной системы.

Для невырожденного случая ($E_F - \varepsilon_1 < 0$) получаем из (1.17.2) бoльцмановскую статистику для двумерного случая (ср. (1.5.4))

$$n_{2D} = g_{2D} k_B T \exp\left(-\frac{|E_F - \varepsilon_1|}{k_B T}\right). \quad (1.17.4)$$

Действуя аналогично (1.17.1), для одномерного проводника получаем

$$g_{1D}(\varepsilon) = \frac{2}{\pi \hbar v(\varepsilon)}, \quad v(\varepsilon) = \sqrt{2\varepsilon/m}, \quad (1.17.5)$$

а для трехмерного случая воспроизводим обычное выражение (рис.1.6)

$$g_{3D}(\varepsilon) = \frac{\sqrt{m\varepsilon}}{\pi^2 \hbar^3}. \quad (1.17.5)$$

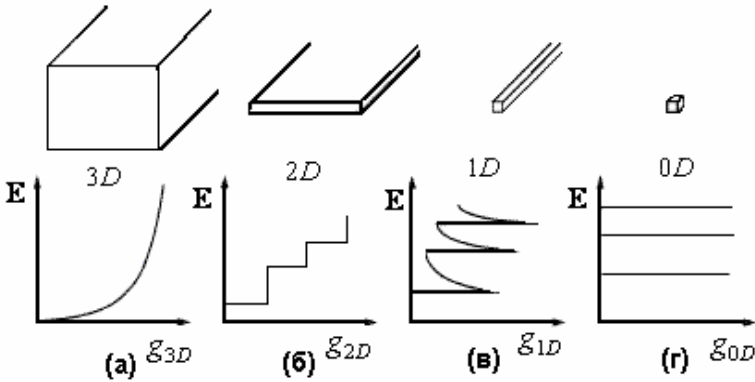


Рис. 1.6. Энергетические плотности состояний $g(\varepsilon)$ для (а) объемного полупроводника (3D); (б) квантовой ямы или инверсионного слоя МОПТ (2D); (в) квантовой проволоки (нити) (1D); (г) квантовой точки (0D)

Энергетический спектр квантовой точки напоминает дискретный спектр атомов, поэтому квантовые точки иногда называют искусственными атомами.

1.18. Подбарьерное туннелирование

Одним из важнейших практических следствий волновой природы электрона является возможность т.н. подбарьерного туннелирования, т.е. проникновения через классически запрещенную область, где высота потенциального барьера больше полной энергии электрона (рис. 1.7).

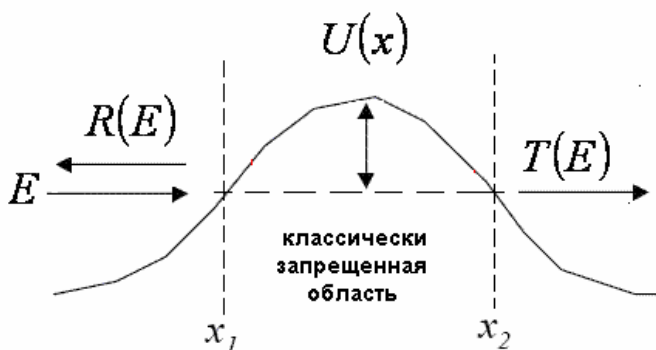


Рис. 1.7. Схема, иллюстрирующая возможность подбарьерного туннелирования. Вероятность прохождения барьера $T(E)$ и отражения от него $R(E)$ связаны соотношением $R(E) + T(E) = 1$

Вероятность туннелирования $T(E)$ зависит от энергии электрона и формы барьера, и ее можно рассчитать по формуле

$$T(E) = \exp\left[-\frac{2}{\hbar} \int_{x_1}^{x_2} (2m(U(x) - E))^{1/2} dx\right]. \quad (1.18.1)$$

В качестве практического примера можно рассмотреть туннелирование через барьер треугольной формы (т.н. туннелирование Фаулера-Нордгейма), играющее важную роль в МОП-структурах и вызывающее ток через изолирующий окисел между затвором и подложкой кремния (см. гл. 10).

Форма потенциального барьера задается его высотой U_B (относительно уровня Ферми в затворе) и наклоном линейного участка, имеющего смысл электрического поля F (рис. 1.8)

$$U(x) = U_B - qFx. \quad (1.18.2)$$

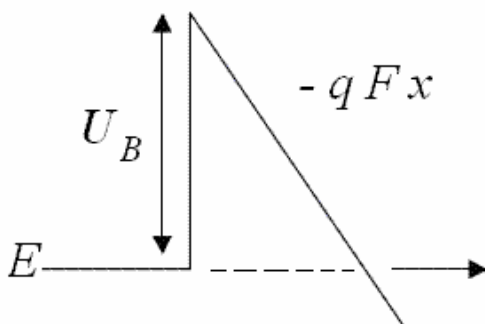


Рис. 1.8. Схема туннелирования Фаулера-Нордгейма

Высота барьера обычно равна работе выхода материала затвора (несколько электрон-вольт). Соответственно, вероятность туннелирования Фаулера-Нордгейма равна

$$\begin{aligned}
 T(E) &= \exp \left[-\frac{2}{\hbar} \int_0^{U_B/qF} (2m(U_B - qFx))^{1/2} dx \right] = \\
 &= \exp \left[-\frac{4}{3} \frac{(2mU_B)^{1/2}}{\hbar} \frac{U_B}{qF} \right]. \quad (1.18.3)
 \end{aligned}$$

2. Основные принципы КМОП-технологии

2.1. Цифровая техника и логические вентили

Современная цифровая схемотехника основана на бинарной логике и обработке сигналов двух уровней – логического нуля и логической единицы. Хорошо различимые устойчивые логические состояния должны быть соотнесены с хорошо различимыми устойчивыми физическими состояниями, в качестве которых можно использовать, например, степень зарядки конденсатора. Заряженный конденсатор и высокое напряжение на его клеммах соответствуют логической единице, разряженный конденсатор и низкое (нулевое) напряжение – логическому нулю. Хранение, передача и обработка сигналов возможны только при наличии эффективной системы коммутации, т.е. возможности легко запирать конденсатор, заряжать и разряжать его с помощью управляющих сигналов. Прибор, осуществляющий такую функцию, должен под воздействием управляющего сигнала менять свою проводимость, по меньшей мере, на много порядков.

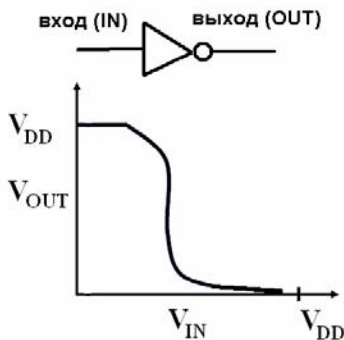


Рис. 2.1. Условное графическое обозначение и передаточная характеристика инвертора (V_{DD} – стандартное обозначение напряжения питания)

Обработка информации осуществляется с помощью логических блоков, именуемых вентилями, простейшим из которых является инвертор, осуществляющий логическую операцию отрицания (логическая операция НЕ). Инвертор имеет один вход и один выход и

преобразует высокий сигнал (потенциал) на входе V_{IN} в низкий на выходе V_{OUT} и наоборот (рис. 2.1).

Принципиальным является то, что вентиль должен иметь сильную нелинейность передаточной характеристики, т.е. зависимости $V_{OUT}(V_{IN})$. Это необходимо для того, чтобы обеспечить хорошее отношение сигнал/шум даже при передаче сигнала через последовательный ряд вентилях. Внешняя помеха подавляется, обеспечивая тем самым хорошую помехоустойчивость, если коэффициент усиления $|dV_{OUT}/dV_{IN}|$ практически везде равен нулю, за исключением бесконечно узкой в идеале области переключения, где этот коэффициент стремится к бесконечности. Поэтому идеальная передаточная характеристика инвертора должна иметь вид ступеньки. Таким образом, хороший вентиль и логику, а также хорошую память можно построить только на основе приборов, обладающих сильной нелинейностью.

Почти идеальным прибором, сочетающим в себе свойства конденсатора, способного хранить информацию, и коммутационного ключа с нелинейными характеристиками, является полевой транзистор со структурой МОП (металл-окисел-полупроводник, МОПТ). Международный термин — MOSFET (metal-oxide-semiconductor field effect transistor).

2.2. Интегральные схемы и планарная технология

Вообще говоря, транзистором (от английского transfer resistor – передаточное сопротивление) называют трехэлектродный полупроводниковый электронный прибор, в котором ток в цепи двух электродов управляется третьим электродом. Управление током в выходной цепи осуществляется за счет изменения входного тока (в базе биполярного транзистора), либо входного напряжения (на затворе полевого транзистора). Небольшое изменение входных величин может приводить к существенно большему изменению выходного напряжения и тока. Резкая нелинейная зависимость выходного напряжения от входного параметра дает возможность построения цифровых элементов (вентилей), работающих с битами информации, на чем основана цифровая техника (логика в составе микропроцессоров и микроконтроллеров, память и т.д.).

Биполярный транзистор был изобретен и изготовлен впервые в 1947 г. (Шокли, Бардин и Браттейн). Идея физически более простого полевого транзистора появилась гораздо раньше. Полевой транзистор был запатентован в 1930 г. (J.E. Lilienfeld), а впервые изготовлен – только в 1960 г. (D. Kahng, M.M. Attala). Ключевым пунктом оказалась возможность термического выращивания на пластине кремния собственного окисла с химически стабильной границей раздела Si-SiO₂ очень высокого качества. Именно эта возможность (в некотором смысле, совсем не обязательная) обеспечила бурное развитие производства приборов на основе структур «металл-окисел-полупроводник» (МОП).

Другой ключевой технологической идеей, определившей развитие современной индустрии, явилась идея планарной интегральной технологии. Ее суть состоит в том, что с помощью специальных методов оптической и электронной литографии и ряда последовательных технологических операций на одной кремниевой подложке одновременно изготавливается схема практически любой степени сложности. Биполярная интегральная схема была впервые изготовлена в 1958 г. Джеком Килби (Jack Kilbey), получившим за это Нобелевскую премию 2000 г., и независимо Робертом Нойсом (Robert Noyce), который впоследствии стал одним из основателей фирмы Intel.

По историческим причинам в цифровой технике вначале доминировали биполярные технологии (рис. 2.2). Разработка первого микропроцессора Intel 4004 в 1971 г. ознаменовала начало эры доминирования цифровых МОП технологий. Исторически первой МОП технологией была технология на *p*-канальных МОП транзисторах. Ее сменила технология на *n*-канальных транзисторах, которую с начала 80-ых гг. XX в. начала вытеснять комбинированная технология, использующая комплементарные (взаимодополняющие) комбинации *n*-канальных и *p*-канальных МОП транзисторов, известная как КМОП (CMOS) технология.

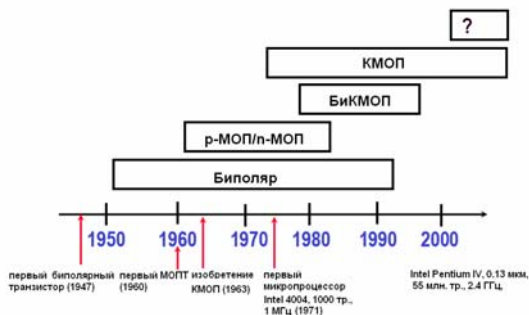


Рис. 2.2. Развитие доминирующих ветвей микроэлектронной индустрии

Эта технология является основой современной микро- и наноэлектроники и, как предполагается, останется таковой в обозримой перспективе.

2.3. МОП транзистор и КМОП технология

Принцип действия МОП транзистора основан на модуляции тока с помощью электростатического эффекта поля (рис. 2.3). Положительный заряд на металлическом (или сильнолегированном поликремниевом) затворе (gate, G), образующийся там при приложении положительного потенциала, индуцирует в кремниевой подложке для определенности *p*-типа равное количество отрицательного заряда.

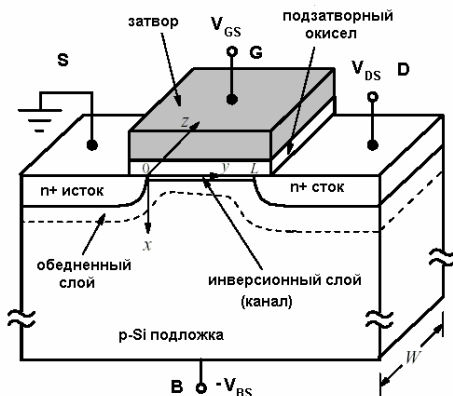


Рис. 2.3. Сечение *n*-канального МОП транзистора (*n*-МОПТ)

Последний состоит из слоя отрицательно заряженных акцепторов (обедненного основными носителями – дырками) и очень тонкого слоя подвижных электронов (инверсионный слой неосновных носителей), образующего проводящий канал между истоком и стоком. Ток в цепи транзистора I_D (ток стока) течет между электродами стока (drain, D) и истока (source, S) под воздействием электрического смещения, приложенного между ними $V_{DS} = V_D - V_S$.

Проводимостью этого канала можно управлять, изменяя напряжение на затворе, приложенное, как правило, относительно заземленного истока $V_{GS} = V_G - V_S$. Током в канале можно частично управлять и с помощью электрического смещения на подложке, прикладывая полярность напряжения $V_{BS} = V_B - V_S < 0$ так, чтобы *pn*-переход исток-подложка смещался в обратную сторону.

Типичный вид вольт-амперных характеристик *n*-канальных МОПТ показан на рис. 2.4.

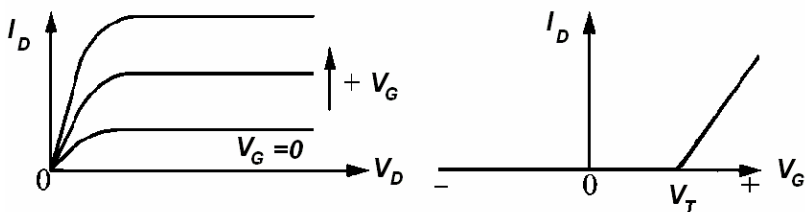


Рис. 2.4. Вольт-амперные характеристики *n*-МОПТ: (а) выходные ВАХ $I_D(V_{DS})$ при фиксированных V_{GS} ; (б) передаточные ВАХ $I_D(V_{GS})$ при фиксированных V_{DS}

Электронный ток в цепи стока *n*-МОПТ в закрытом состоянии I_{OFF} , т.е. при затворных напряжениях, меньших некоего порогового значения $V_{GS} < V_T$, очень мал ($\sim 10^{-11} \dots 10^{-8}$ А). В открытом состоянии ($V_{GS} > V_T$) ток составляет $I_{ON} \sim 1$ мА, что обеспечивает динамический диапазон изменения тока I_{ON}/I_{OFF} около 5-8 порядков. Транзистор с *p*-каналом на *n*-Si подложке (*p*-МОПТ) имеет дырочную проводимость канала, и его проводимость уменьшается при увеличении положительного напряжения на затворе. Эта дополни-

тельность типов проводимости дает возможность построить т.н. комплементарную МОП (КМОП) логику (CMOS).

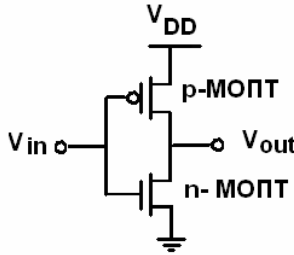


Рис. 2.5. Электрическая схема КМОП инвертора

На рис. 2.5 представлена схематическая реализация логической функции НЕ (инвертора) на основе КМОП технологии, использующей одновременно транзисторы двух типов. Увеличивая напряжение на входе (затворах транзисторов обоих типов), мы одновременно увеличиваем проводимость *n*-МОПТ и уменьшаем проводимость *p*-МОПТ. Благодаря этому передаточная характеристика инвертора оказывается в наибольшей степени близкой к идеальной ступеньке, что обеспечивает малое энергопотребление и доминирование схем КМОП технологии над схемами *n/p*-МОП и биполярной технологии.

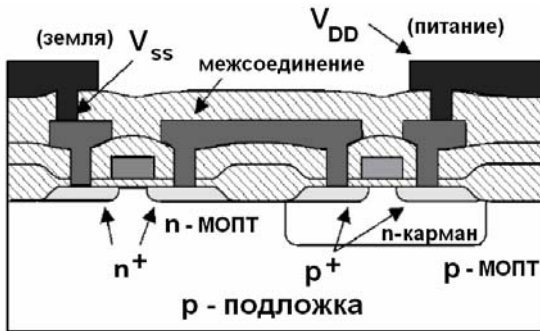


Рис. 2.6. Сечение объемной КМОП технологии с *n*-карманом

Технологическая реализация интегрального КМОП инвертора показана на рис. 2.6.

2.4. Закон Мура

Анализ эмпирических тенденций первых 5 лет развития интегральной технологии позволил Гордону Муру (рис. 2.7) сформулировать в 1965 г. некоторую закономерность, впоследствии получившую название закона Мура.



Рис. 2.7. Гордон Мур, один из основателей Intel

Первоначальная формулировка закона звучала следующим образом: «Количество транзисторов на чипе (интегральной схеме) удваивается каждые 12 месяцев». Несмотря на некоторый произвол в формулировке и обосновании, закон Мура имеет определенную физическую основу. Его физическое содержание состоит в идее возможности размерного скейлинга – масштабного уменьшения геометрических размеров микроселектронных компонентов с сохранением электрических и улучшением функциональных характеристик отдельных приборов и всей схемы в целом.

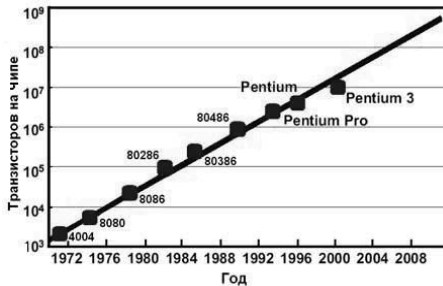


Рис. 2.8. Зависимость степени интеграции от времени

Именно возможность геометрического скейлинга микроселектронных структур явилась технологическим и экономическим основанием, обеспечившим победное шествие закона Мура на протяжении уже почти полувека (рис. 2.8).

Увеличение количества элементов на одном кристалле (чипе) (то же самое, повышение степени интеграции) достигается, главным образом, за счет уменьшения т.н. технологической нормы, и, в некоторой степени, за счет роста площади чипа. Площадь одного

чипа (кристалла), на котором может располагаться микропроцессор или схема памяти, имеет тенденцию к росту и в настоящее время составляет приблизительно 2 см^2 . В настоящее время (2008 г.) на одном чипе располагается порядка миллиарда транзисторов. Ожидается, что к 2016 г. количество транзисторов на одном чипе будет составлять ~ 10 млрд.

2.5. Технологическая (проектная) норма

Основным параметром технологии является минимальная технологическая (топологическая, проектная) норма (technology node, feature size). Говоря о проектной норме, обычно имеют в виду минимальный для данной технологии характеристический размер. Длина затвора МОП транзистора приблизительно в 1.5...2 раза меньше технологической нормы (рис. 2.9). Длина канала, как правило, оказывается еще меньше длины затвора.

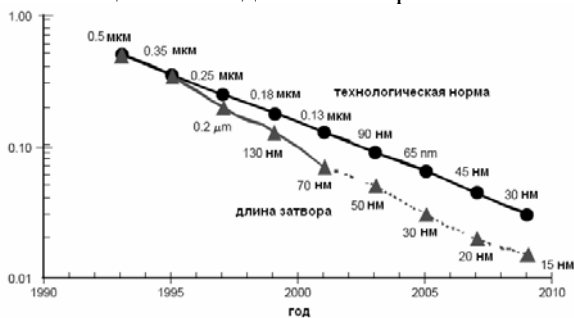


Рис. 2.9. Технологическая норма и длина затвора для технологий разных поколений

Изменение технологической нормы идет по поколениям, с приблизительно одинаковым масштабным множителем $L \rightarrow L/\sqrt{2} \sim \sim 0.7L$. Соответствующий ряд параметров технологических норм исторически имеет следующий вид:

0.5 мкм..0.35..0.25..0.18..0.13 ..(130 нм)..90..60..45..32 нм...

Переход к следующему поколению соответствует приблизительно увеличению степени интеграции в 2 раза. Очевидно, что закон, декларирующий экспоненциальный рост количества транзисторов на одной интегральной схеме, не может быть справедливым в течение неограниченного времени. Ожидается (рис. 2.10), что ко-

нец «эры закона Мура», т.е. когда технологическая норма современной кремниевой технологии достигнет своего минимума $\sim 5 \dots 10$ нм, определяемого возможностями литографии и фундаментальными физическими ограничениями, наступит не ранее 2020-2030 гг.

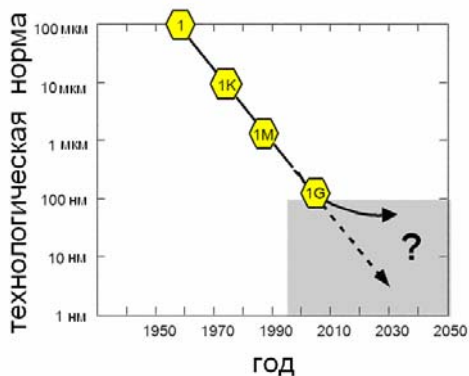


Рис. 2.10. Технологическая норма как функция времени

Все основные геометрические параметры технологии уменьшаются вместе с технологической нормой. Это касается длины, ширины канала транзистора, а также толщины подзатворного окисла (рис. 2.11).

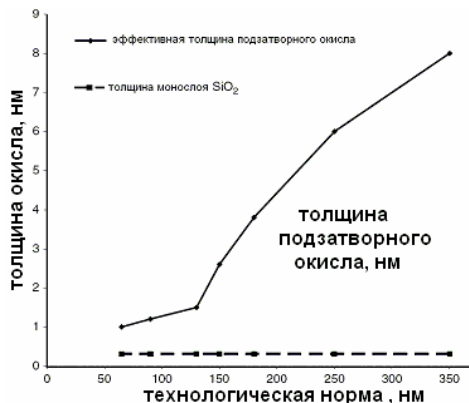


Рис. 2.11. Зависимость толщины подзатворного окисла от технологической нормы

Зависимость геометрического размера от нормы носит приблизительно пропорциональный характер. Например, многолетняя технологическая тенденция фирмы Intel состоит в том, что толщина подзатворного окисла составляет $\sim 1/45$ от технологической нормы. Эта традиция не может быть выдержана в дальнейшем, поскольку в современных транзисторах толщина подзатворного окисла уже составляет 3-5 постоянных решетки SiO_2 .

2.6. Тактовая частота

Важнейшей характеристикой микропроцессора является его быстродействие. Максимальное быстродействие интегральных схем определяется системной, т.н. тактовой частотой (clock). В последние 30 лет тактовая частота микропроцессоров увеличивалась в среднем приблизительно в 2 раза за 1 год (рис. 2.12).

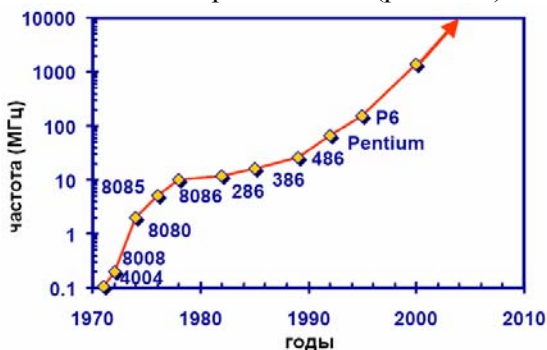


Рис. 2.12. Зависимость тактовой частоты от года выпуска

Тактовая частота определяется скоростью переключения отдельных цифровых элементов и, в конечном итоге, в очень сильной степени зависит от геометрического фактора и, следовательно, от технологической нормы (рис. 2.13).

До настоящего времени тактовая частота процессора приблизительно удваивается в каждом следующем поколении. При технологической норме ~ 10 нм тактовая частота ожидается на уровне 100 ГГц (10^{11} Гц).

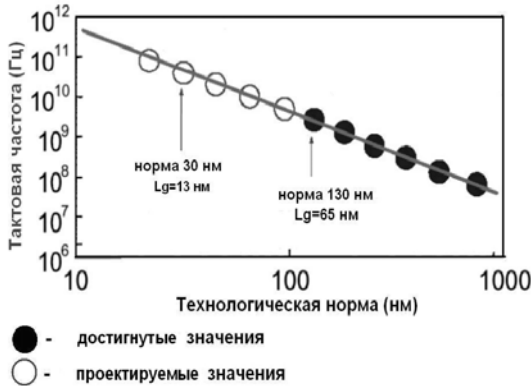


Рис. 2.13. Зависимость тактовой частоты от технологической нормы

Геометрический ресурс для дальнейшего увеличения тактовой частоты практически исчерпан, поскольку существует фундаментальное ограничение, связанное с конечностью скорости распространения электрического сигнала. Например, для металлической линии на диэлектрической подложке с проницаемостью ϵ_i (≈ 4 для SiO_2) скорость распространения сигнала не может превышать $c/\sqrt{\epsilon_i}$, где $c = 3 \times 10^{10}$ см/с – скорость света в вакууме.

2.7. Основные проблемы миниатюризации

Основными параметрами микропроцессоров являются: длина затвора L , толщина подзатворного окисла d_{ox} , тактовая частота f , степень интеграции, плотность потока энергии, рассеиваемой в тепло на единицу площади чипа, P . Основными проблемами при увеличении степени интеграции и уменьшении размеров активных областей приборов являются увеличение токов утечек и проблема тепловыделения.

В табл. 2.1 представлен прогноз Международной ассоциации производителей микроэлектронных компонентов ITRS (International Roadmap for Semiconductors), публикуемый с 1994 г. и регулярно обновляемый на сайте www.itrs.net. Из этого прогноза видно, что роль токов утечки резко возрастает с уменьшением технологической нормы.

Таблица 2.1

Эволюция параметров микропроцессоров

Год Норма	1999 180нм	2001 130нм	2004 90нм	2008 60нм	2011 40нм	2014 30нм
Длина за- твора L_G (нм)	120	90	70	45	32	22
d_{ox} (ангстре- мы)	19- 25	15- 19	12- 15	8-12*	6-8*	5-6*
Ток утечки при 100 °С (нА/мкм)	7	10	16	40*	80*	160*

*— прогнозируемые размеры и величины

Анализ проблемы тепловыделения проведен в следующих пунктах. О токах утечки более подробно оговаривается в главе 9.

2.8. Анализ проблемы тепловыделения

Фундаментальной проблемой при увеличении степени интеграции является проблема тепловыделения. Первый ламповый компьютер ENIAC (1944 г.) имел довольно высокую тактовую частоту (~100 кГц), весил 20 т, и его 18 000 вакуумных ламп выделяли тепловую мощность ~ 150 кВт. Изобретение твердотельных слаботочных транзисторов и переход на интегральную технологию на какое-то время сняли проблему тепловыделения, но рост степени интеграции вновь поставил ее в ряд важнейших.

Переключение цифровых элементов – это в конечном итоге зарядка и разрядка эффективных конденсаторов, при которых, как правило, выделяется тепло. На рис. 2.14 показана упрощенная модель элемента памяти, позволяющая проанализировать его работу. Она состоит из конденсатора с полной эффективной емкостью C_{tot} , [Ф], на котором может храниться информационный заряд. Этот конденсатор может заряжаться с помощью ключа через эффективное сопротивление R с помощью источника питания V_{DD} [В] и разряжаться с помощью контакта «земля». Как обычно, высокий потенциал (V_{DD}), запираемый с помощью ключа на узле М, будет соответствовать логической единице, нулевой потенциал («земля») – логическому нулю.

Основными процедурами обработки сигнала условно являются:
 WRITE – зарядка конденсатора через сопротивление до напряжения питания V_{DD} ;

READ – считывание сигнала через вольтметр;

ERASE – уничтожение сигнала через заземление.

Потенциальная энергия конденсатора с емкостью C , заряженного до разницы потенциалов V_{DD} ,

$$E_{pot} = \int_0^{V_{DD}} Q(V') dV' = \int_0^{V_{DD}} C_{tot} V' dV' = C_{tot} V_{DD}^2 / 2 = Q^2 / 2 C_{tot}. \quad (2.8.1)$$

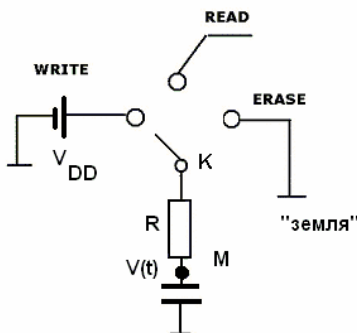


Рис. 2.14. Прототип цифровой ячейки памяти

Конденсатор заряжается через сопротивление R . Каждый электрон при переходе через резистор отдает в тепло энергию, равную разности потенциалов на выводах сопротивления $q(V_{DD} - V')$. Тогда энергия, перешедшая в тепло при повышении потенциала в точке M от 0 до V_{DD} (энергия диссипации E_{dissip}), равна

$$\begin{aligned} E_{dissipat} &= \int_0^{V_{DD}} (V_{DD} - V') I(V') dt = \\ &= \int_0^{V_{DD}} (V_{DD} - V') dQ(V') = \int_0^{V_{DD}} C_{tot} (V_{DD} - V') dV' = \\ &= C_{tot} V_{DD}^2 / 2 \equiv Q^2 / 2 C_{tot}. \end{aligned} \quad (2.8.2)$$

Полная энергия, необходимая для зарядки конденсатора ($Q \equiv C_{tot} V_{DD}$),

$$E_S = E_{pot} + E_{dissipat} = C_{tot} V_{DD}^2 = Q^2 / C_{tot}. \quad (2.8.3)$$

Таким образом, при зарядке описанного типа необратимо теряется (переходит в тепло) ровно половина от полной затраченной энергии $E_{dissip} = (1/2)E_S$. Другая половина накапливается в форме энергии электрического поля конденсатора и, в принципе, может быть использована. К сожалению, современная архитектура микропроцессоров такова, что пока никто не заботится об использовании накопленной потенциальной энергии конденсаторов, и в следующих тактах она неизбежно теряется (переходит в тепло).

Итак, в цикле WRITE часть энергии $CV_{DD}^2/2$ теряется в сопротивлении; такая же энергия $CV_{DD}^2/2$ накапливается в форме потенциальной энергии конденсатора. В цикле READ энергия не затрачивается, по крайней мере, в идеале. И, наконец, в цикле ERASE потенциальная энергия $CV_{DD}^2/2$ переходит в тепло. Таким образом, за один полный активный такт затрачивается и уходит в тепло CV_{DD}^2 энергии. На практике перезарядка элементов происходит не каждый такт, т.е. не каждый такт является активным. Определим активность схемы $0 < \alpha < 1$ как среднюю долю элементов, перезаряженных за один такт (нет активности – нет потребления).

Тогда, если тактовая частота f [Гц], то $\alpha CV_{DD}^2 f$ [Вт] – мощность, рассеиваемая в одном элементе. Если N_G – плотность интеграции, т.е. количество затворов на см^2 чипа, то тепловая мощность теплового потока динамического энергопотребления (т.е., происходящего при переключениях) можно грубо оценить по формуле

$$P = \alpha N_G C_{tot} V_{DD}^2 f \text{ [Вт/см}^2\text{]}. \quad (2.8.4)$$

Рост степени интеграции и тактовой частоты привел к тому, что мощность теплового потока от поверхности микропроцессора современного компьютера угрожающе растет (табл. 2.3).

Таблица 2.2

Характерные мощности тепловых потоков

Intel 486	$P = 5 \dots 8 \text{ Вт/см}^2$
Сковородка (1 кВт, площадь 1 дм ²)	$P \cong 10 \text{ Вт/см}^2$
Поверхность Intel Pentium 4	$P \cong 30 \text{ Вт/см}^2$
Поверхность ядерного реактора АЭС	$P \cong 300 \text{ Вт/см}^2$
Поверхность Солнца	$P = 10^4 \text{ Вт/см}^2$

В 2016 г. прогноз ITRS дает поток тепла 93 Вт/см^2 с поверхности микросхемы. Оценки показывают, что максимально возможный отвод тепла соответствует нескольким сотням Вт/см^2 (при водяном охлаждении).

2.9. Проблема отвода тепла

Таким образом, энергия переключения $C_{tot} V_{DD}^2$ является фундаментальным параметром технологии, определяющим энергопотребление. Очевидным путем решения проблемы минимизации E_S является (а) снижение полной емкости C_{tot} , что в конечном итоге достигается за счет уменьшения размеров элемента; (б) снижение напряжения питания V_{DD} .

В результате энергия переключения непрерывно уменьшается. Это уменьшение наглядно продемонстрировано в табл. 2.3, где энергия переключения представлена в единицах $k_B T$ ($\sim 0.025 \text{ эВ}$ для комнатных температур).

Таблица 2.3

Энергия переключения в единицах kT

1995 г.	$\sim 5 \times 10^5 kT$
2005 г.	$\sim 10^4 kT$
2015 г.	$\sim 10^3 kT$ (прогноз)

Вместе с тем, уменьшение размеров элементов приводит к повышению степени интеграции, и мощность потока тепла при этом только увеличивается. При этом температура микросхемы при функционировании все время должна оставаться постоянной (пусть и повышенной), тепло нужно отводить (подложка, поток воздуха или даже воды). В настоящее время используется воздуш-

ное охлаждение с помощью вентиляторов (кулеров). В перспективе необходимым будет использование водяного охлаждения.

При этом очевидно, что максимально возможная отводимая мощность теплового потока P_{cool} определяет максимально возможную производительность $[N] = 1/(\text{см}^2 \text{ с})$, т.е. количество операций за единицу времени на единице площади

$$N = P_{cool} / C_{tot} V_{DD}^2. \quad (2.9.1)$$

Отводимая в подложку тепловая мощность пропорциональна разности температур кристалла и теплоотводящей подложки ΔT . Тогда выделяемая тепловая мощность, отводимая в холодильник, оценивается формулой

$$P_{cool} = K \Delta T, \quad (2.9.2)$$

где K – коэффициент теплопередачи.

Коэффициент теплопередачи зависит от способа отвода тепла, и его значения представлены в табл. 2.4.

Таблица 2.4

Коэффициенты теплопередачи

K	Вт/(см ² К)
Воздух (естественная конвекция)	0.001
Воздух (обдувание)	0.01
Вода (без циркуляции)	0.1
Вода (с циркуляцией)	1
Вода (кипящая)	10

При отводе тепла в подложку становится важной ее коэффициент. В GaAs этот коэффициент в три раза больше, а в алмазе – в 14 раз больше, чем в кремнии.

2.10. Проблема диссипации тепла и обратимости вычисления

Любое вычисление является в конечном итоге физическим процессом. Например, рассеивание тепла при вычислении тесно связано с необратимостью вычисления. Дело в том, что стандартные логические вентили (например, И-НЕ (NAND), ИЛИ-НЕ (NOR)) имеют два входа и один выход (рис. 2.15).



Рис. 2.15. Условная схема логического вентиля с двумя входами и одним выходом

Для таких вентилях после акта вычисления не представляется возможным по состоянию выхода восстановить состояние входа. В таких случаях говорят, что имеет место логическая необратимость. Действительно, на входе имеем 2 бита информации, на выходе – 1 бит. Таким образом, количество информации при одном акте вычисления уменьшается на 1 бит.

$$\delta I = \ln 1 - \ln 2 = -\ln 2. \quad (2.10.1)$$

По словам известного специалиста Р. Ландауэра: «Современные компьютеры скорее не производят информацию, а уничтожают ее».

Согласно термодинамике, любая потеря информации эквивалентна увеличению энтропии S

$$\delta S = -k_B \delta I = k_B \ln 2, \quad (2.10.2)$$

что на физическом уровне равносильно выделению тепла

$$\delta Q = T \delta S = k_B T \ln 2. \quad (2.10.3)$$

Согласно первоначальной идее Ландауэра (1957), эта величина соответствует минимально возможному тепловыделению при акте вычисления. Как выяснилось впоследствии, это утверждение оказалось неправильным. В принципиальном плане вычисления могут проходить без диссипации энергии и, следовательно, обратимым образом. Примером обратимых процессов без диссипации энергии являются процессы без трения. При достаточно медленном движении нет трения и нет тепловыделения. Процедуру вычислений можно организовать так, чтобы она происходила практически без выделения тепла, но, к сожалению, этого можно достичь только за счет уменьшения быстродействия.

2.11. Адиабатическая логика

Адиабатичность означает отсутствие обмена или выделения тепла. Адиабатической логикой называется такая организация процесса вычисления, при которой диссипация энергии (переход в тепло) минимизирована. Простейшим примером такой организации является адиабатическая зарядка конденсатора.

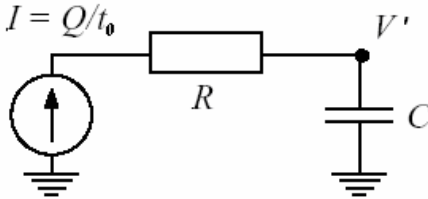


Рис. 2.16. Схема зарядки информационного конденсатора с помощью источника тока

Простейшим способом зарядки конденсатора является зарядка с помощью источника постоянных напряжений, энергия диссипации при этом равна $C_{tot} V_{DD}^2 / 2$ (см. п. 2.8). С другой стороны, зарядку информационного конденсатора можно проводить через источник тока, поддерживая в цепи постоянный ток $I = Q/t_0$ в течение заданного времени t_0 (рис. 2.16). Тогда энергию диссипации можно оценить следующим образом:

$$E_{diss} = I^2 R t_0 = (Q/t_0)^2 R t_0 = C_{tot} V_{DD}^2 \frac{R C_{tot}}{t_0}. \quad (2.11.1)$$

При достаточно маленьком токе зарядки ($R C_{tot} \ll t_0$) диссипация может быть существенно снижена (но только за счет уменьшения быстродействия).

Одним из возможных принципов адиабатической логики является недопущение открытия МОП транзисторов, когда между стоком и истоком приложено большое напряжение. Ценой этого решения является усложнение архитектуры и относительное уменьшение быстродействия. Ясно, что до тех пор, пока остаются возможности для геометрического скейлинга, у производителей нет значимого стимула для существенного усложнения архитектуры.

2.12. Оценка максимального быстродействия

Максимальная тактовая частота и быстродействие определяется минимальной временной задержкой функционирования одного транзистора. Минимальная внутренняя задержка транзистора в значительной степени определяется временем пролета носителей через канал длиной L . При малых электрических полях в канале время пролета определяется дрейфовой скоростью носителей $v_d = \mu E$.

Современные транзисторы в цифровых схемах работают в режиме насыщения дрейфовой скорости, когда скорость носителя оказывается порядка тепловой скорости v_{th} ($\sim 10^7$ см/с в Si при $T = 300$ К). Максимальная скорость носителей в Si достигается при электрических полях в канале $\sim 2 \times V/\mu\text{км}$.

Таким образом, максимально возможное быстродействие определяется временем пролета носителя через канал

$$t_F = L / v_{th}. \quad (2.12.1)$$

Но за это время можно перезарядить только заряд, находящийся непосредственно в канале под затвором $Q_S = C_G(V_{DD} - V_T)$. Поэтому максимальный рабочий ток транзистора I_{\max} (drive current) определяется только зарядом под затвором

$$I_{\max} = \frac{Q_S}{t_F} = \frac{C_G(V_{DD} - V_T)}{t_F}, \quad (2.12.2)$$

где

$$C_G = A_G (\epsilon_i / d_{ox}) \quad (2.12.3)$$

– полная емкость затвора, выраженная через полную площадь канала A_G , толщину подзатворного изолятора с диэлектрической проницаемостью ϵ_i .

Быстродействие (время задержки) определяется быстротой перезарядки полной емкости транзистора, которая складывается из емкости затвора и паразитной емкости

$$C_{tot} = C_G + C_{par} \equiv C_G(1 + \eta) \quad \eta \equiv C_{par} / C_G, \quad (2.12.4)$$

где паразитная емкость $C_{par} = C_J + C_{wire}$ складывается из емкости *pn*-переходов истока-стока и емкости межсоединений. Паразитная емкость уменьшает быстродействие и должна быть минимизирована

на. Полный заряд (информационный + паразитный) выражается через полную емкость и напряжение питания $C_{tot}V_{DD}$. Тогда можно сделать следующие оценки.

Минимальное время переключения (рассасывания информационного заряда)

$$t_{d,\min} \cong 2 \frac{C_{tot}V_{DD}}{I_{\max}} = 2 \frac{C_{tot}V_{DD} t_F}{C_G (V_{DD} - V_T)} = 2 \frac{(1+\eta)V_{DD} L}{(V_{DD} - V_T) v_{th}}. \quad (2.12.5)$$

Минимальное время определяет максимальную частоту f_{\max} :

$$f_{\max} \sim \frac{1}{t_{d,\min}} = \frac{(1 - V_T/V_{DD}) v_{th}}{2(1+\eta) L}. \quad (2.12.6)$$

В качестве примера рассмотрим стандартную существующую технологию 90 нм, для которой примем $L = 0.1$ мкм, $V_T/V_{DD} \sim 1/3$; $\eta \cong 8$. Это дает оценку $f_{\max} = 30$ ГГц, что приблизительно на порядок превышает реальную тактовую частоту, связанную с недооценкой паразитных емкостей металлических межсоединений (проводов). Тем не менее, из этой оценки видны меры для увеличения частоты:

- уменьшение длины канала;
- уменьшение порогового напряжения (приводит к увеличению подпороговых токов утечек);
- уменьшение паразитных емкостей (емкость межсоединений трудно уменьшить).

2.13. Проблемы при миниатюризации межсоединений

Проблема межсоединений в настоящее время, а особенно в перспективе, является одной из самых острых. В современных схемах общая длина линий металлизации 5-7 км (до 10 км) на кристалл (≤ 1 мм на один транзистор). Для обеспечения коммутации всех элементов необходимо 5-8 уровней металлизации (вплоть до 10) (рис. 2.17).

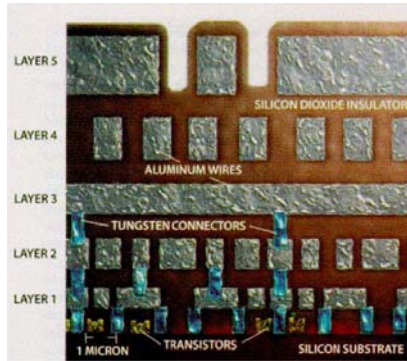


Рис. 2.17. Сечение чипа с линиями металлизации на 5 уровнях

Типичная площадь сечения межсоединений $\sim 1 \text{ мкм}^2$ и меньше. Типичная плотность тока в проводе $\sim 10^5 \text{ А/см}^2$. Межсоединения вводят паразитные сопротивления и емкости. Их роль возрастает при миниатюризации активных областей. Поэтому временные задержки распространения сигнала в современных схемах определяются не столько активными транзисторами, сколько линиями межсоединений.

Скорость распространения электрического сигнала в проводнике на изоляторе с проницаемостью ϵ_i замедляется за счет перезарядки распределенной емкости и оказывается заметно меньшей скорости света в пустоте c . $v = c / \sqrt{\epsilon_i}$.

Оценим характеристики линии металлизации на диэлектрической подложке с проницаемостью ϵ_i (рис. 2.18).

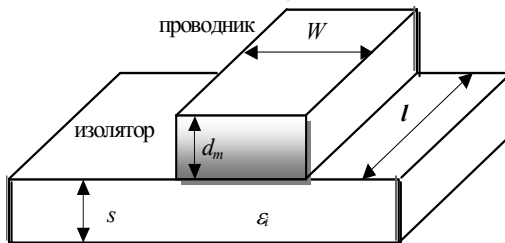


Рис. 2.18. Линия металлизации на диэлектрической подложке

Характерное время перезарядки проводника длиной ℓ можно оценить через геометрические размеры и удельное электрическое сопротивление материала провода ρ :

$$RC \sim \left(\rho \frac{\ell}{W d_m} \right) \left(\frac{\varepsilon_0 \varepsilon_i}{s} W \ell \right) = \frac{\rho \varepsilon_0 \varepsilon_i \ell^2}{s d_m} = \tau_r \frac{\ell^2}{s d_m}, \quad (2.13.1)$$

где $\tau_r = \varepsilon_i \varepsilon_0 \rho$ – постоянная диэлектрической релаксации. Для увеличения скорости передачи сигнала по межсоединению необходимо уменьшать его удельное сопротивление. Именно по этой причине (табл. 2.5), а также по причине более высокой теплопроводности на смену алюминию в качестве основного материала для межсоединений приходит медь.

Таблица 2.5

Сравнение удельного сопротивления меди и алюминия

Материал	Удельное сопротивление (мкОм×см)
Медь (Cu)	1.68
Алюминий(Al)	2.66

Из формулы (2.13.1) видно, что миниатюризация линии межсоединения приводит только к уменьшению ее быстродействия. Длина линий не может быть уменьшена, а при уменьшении толщины проводника и изолятора в k раз время перезарядки линии увеличится в k^2 раз! Емкость линии на единицу длины составляет $\sim 100 \dots 300$ фФ/см (фемтофарада = 10^{-15} фарада) и увеличивается с уменьшением технологической нормы.

По этой причине, роль задержки в линиях разводки неуклонно растёт:

до 90-х гг. доминировали задержки в самом транзисторе;

в настоящее время (60...90 нм) времена задержки в приборах составляют ~ 1 пс; 1 мм линии ~ 6 пс;

в 2015 (22 нм) ожидаются значения 0.1 пс в приборе и ~ 30 пс на 1 мм линии межсоединения.

Очевидным способом уменьшения RC задержки является использование изоляторов с низким значением диэлектрической проницаемости (low-K insulators) изолирующих слоев, на которую положена линия, что позволяет увеличить эффективную скорость распространения сигнала по линии. В идеале желательно, чтобы

проницаемость приближалась к единице. В настоящее время изолирующие слои многоуровневой разводки имеют проницаемость от 3.5 до 2.6. В 10-летней перспективе ожидается снижение этой величины до уровня < 2 . Для этого используются различные полимеры (органические и неорганические) и модифицированные окислы. Например, в технологии Intel Pentium, 0.18 мкм, 6 Al слоев металлизации, используется флюоринированный оксид кремния SiOF с диэлектрической проницаемостью ~ 3.1 .

2.14. Принципы скейлинга

Анализ, проведенный Деннардом (Dennard) в начале 70-х гг. XX в., привел к неожиданному выводу: уменьшение размеров приборов улучшает почти все характеристики схем, как функциональные, так и экономические. Этот вывод дал зеленый свет к продолжающейся по сию пору технологической гонке на основе идей масштабной миниатюризации – скейлинга.

Основная идея скейлинга – уменьшение геометрических размеров приборов с сохранением некоторых функциональных и параметрических инвариантов (рис. 2.19).

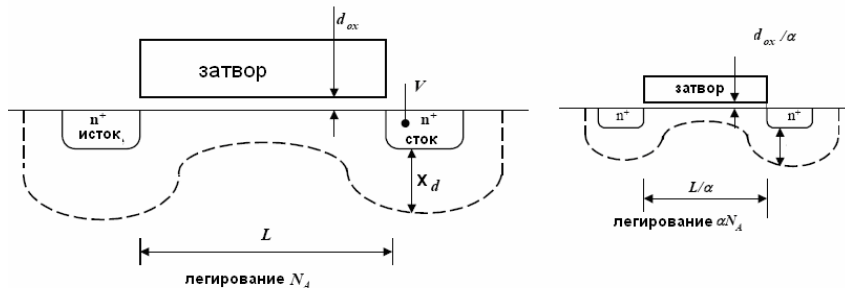


Рис. 2.19. Принципы масштабирования (скейлинга)

В частности, необходимо, чтобы при скейлинге не изменялись электрические характеристики ВАХ транзисторов. Для этого необходимо оставлять постоянными некоторые параметрические инварианты. Один из возможных таких инвариантов – электрические поля в транзисторе (табл. 2.6). Для того чтобы электрические поля внутри приборов оставались приблизительно постоянными, необходимо уменьшать напряжения питания.

Численной характеристикой скейлинга является безразмерный масштабный фактор α . Наблюдаемая численная характеристика скейлинга $\alpha \cong 5$ за 10 лет (здесь $\alpha > 1$). Часто обсуждаются и другие стратегии «обобщенного скейлинга», учитывающие разные масштабные факторы для параметров разных типов. Например, при увеличении степени интеграции характерные значения электрического поля в транзисторе увеличиваются. Это связано с тем, что напряжение питания уменьшается медленней, чем характерные размеры, например, толщина подзатворного окисла.

Таблица 2.6

Принципы скейлинга с постоянным электрическим полем

1	Длина/ ширина/ толщина окисла	$L \rightarrow L/\alpha; Z \rightarrow Z/\alpha; d_o \rightarrow d_o/\alpha$
2	Напряжение питания	$V_{DD} \rightarrow V_{DD}/\alpha$
3	Полная емкость затвора	$C_G = \epsilon_i Z L / d_o \rightarrow C_G/\alpha$
4	Крутизна $\beta = \mu(Z/L)(\epsilon_i/d_{ox})$	$\beta \rightarrow \alpha\beta$
5	Легирования подложки	$N_B \rightarrow \alpha N_B$
6	Максимальный ток $I_{max} \cong \beta V_{DD}^2$	$I_{max} \rightarrow I_{max}/\alpha$
7	Плотность интеграции	$N_G \rightarrow N_G \alpha^2$
8	Энергия переключения	$E_S = C_{tot} V_{DD}^2 \rightarrow C_{tot}/\alpha^3$
9	Время переключения	$t_d = (1+\eta)(L/v_i) V_{DD}/(V_{DD}-V_T) \rightarrow t_d/\alpha$
10	Частота [Гц]	$f_{max} \rightarrow \alpha f_{max}$
11	Мощность на ячейку [Вт]	$P \sim C_{tot} V_{DD}^2 f \rightarrow P/\alpha^2$
12	Плотность потока энергии [Вт/см ²]	$P/A \rightarrow (P/A) = \text{const}!$
13	Плотность потока энергии на чип	Растет как степень интеграции (количество ячеек на кристалле) (~5%/год)

Увеличение электрических полей в приборе иллюстрируется табл. 2.7.

Изменение напряжения питания, толщины подзатворного окисла и поля в этом окисле для разных технологических норм

Технологическая норма	Напряжение питания	Толщина окисла	Поле в окисле
2 мкм	5 В	35 нм	1.4 МВ/см
1.2 мкм	5 В	25 нм	2.0 МВ/см
0.8 мкм	5 В	18 нм	2.8 МВ/см
0.5 мкм	3.3 В	12 нм	2.8 МВ/см
0.35 мкм	3.3 В	10 нм	3.3 МВ/см
0.25 мкм	2.5 В	7 нм	3.6 МВ/см

На практике невозможно выдержать последовательно любую из этих стратегий. Поэтому разработчики часто пользуются эмпирическими правилами скейлинга, основанными на анализе фактических тенденций и технологических традиций в своей фирме.

Обычно длина канала L используется как технологический индикатор скейлинга. Все параметры вычисляют, исходя из заданного L , по эмпирическим формулам:

- $d_{ox} \cong \max (21 \times L^{0.77}, 14 \times L^{0.55})$, где d_{ox} (нм), L (мкм);
- $V_{DD} \cong \max (5 \times L^{0.75})$, где V_{DD} (В);
- $N_B \cong 4 \times 10^{16} L^{-1.6}$; N_B (см⁻³).

При этом стратегия скейлинга для разных областей, например, микропроцессора, может быть различна. Это связано, в частности, с тем, что структура микропроцессора неоднородна. В нем есть «упорядоченные» области (например, память) и «неупорядоченные» участки (например, логика).

2.15. Компромиссы миниатюризации

В процессе проектирования типичной является ситуация, когда улучшение одних функциональных характеристик (например, быстродействия) негативно сказывается на надежности приборов и некоторых других функциональных характеристиках. Например, высокое быстродействие требует высоких значений электрических полей, что приводит к ряду нежелательных эффектов.

Во-первых, растут токи утечки (через затворный изолятор и между стоком и истоком, стоком и подложкой). Во-вторых, возникает эффект износа подзатворного окисла, что приводит к резкому уве-

личению токов утечек через подзатворный изолятор и к его пробой (глава 10). В-третьих, высокие электрические поля приводят к снижению времени жизни приборов из-за деградации горячими носителями, возникающими в канале в сильном поле между стоком и истоком (глава 6).

Другая проблема состоит в дилемме – высокое быстродействие или низкое динамическое энергопотребление. Как мы уже указывали, энергия переключения $E_S \sim C_{tot} V_{DD}^2$ является константой данной технологии, характеризующей среднее рассеиваемое тепло за один активный такт. Очевидно, что существует дополнительность между мощностью и задержкой

$$P \times t_d \sim C_{tot} V_{DD}^2,$$

где время переключения выражается через максимальный рабочий ток I_{max} транзистора

$$t_d \sim C_{tot} V_{DD} / I_{max}.$$

Из этих соотношений ясно, что для того чтобы уменьшить время переключения (т.е. увеличить тактовую частоту и быстродействие), нужно увеличить максимальные токи и, соответственно, энергопотребление. Таким образом, при геометрической миниатюризации всегда возникает необходимость искать оптимизационные компромиссы между улучшением одних функциональных характеристик и ухудшением других.

2.16. Ограничения скейлинга

Пределы скейлинга диктуются постоянством фундаментальных и технологических характеристик материала, большинство из которых не поддаются масштабированию. Это, например:

- ширина запрещенной зоны кремния (ограничивает снизу пороговое напряжение, ширину обедненной области);
- эффективная масса носителей (ограничивает сверху подвижность, скорость насыщения и быстродействие);
- внутренние характеристики материалов – например, поле в подзатворном окисле SiO_2 не может превышать максимальное напряжение пробоя окисла ($\sim 10^7$ В/см).

Другой причиной ограничения скейлинга являются некоторые фундаментальные законы, лежащие в основе работы прибора. На-

пример, подпороговые токи утечки между стоком и истоком определяются больцмановской статистикой термоактивационного перехода носителей между стоком и истоком и, вообще говоря, не зависят от геометрических размеров приборов. Паразитные туннельные токи при скейлинге даже возрастают (глава 10).

Кроме того, при малых размерах начинает работать геометрический фактор, связанный с ухудшением электростатического контроля заряда в канале зарядом на затворе. Это является следствием нарушения планарности и проявлением трехмерности уравнения Пуассона, что приводит к появлению паразитных короткоканальных эффектов (глава 5).

Возрастание последовательных паразитных сопротивлений истоков/стоков (S/D) – уменьшение максимального тока. Возможное решение состоит в использовании металлических контактов в качестве стоков и истоков (глава 5).

Обеднение затвора приводит к падению потенциала на затворе и выражается в уменьшении эффективной затворной емкости. Возможное решение состоит в использовании металлических затворов вместо поликремниевых (глава 3).

Уменьшение размеров и использование нетрадиционных изоляторов уменьшает подвижность. Возможное решение состоит в использовании для увеличения подвижности Si/Ge сплавов и слоев напряженного кремния.

3. Структуры металл - окисел - полупроводник

3.1. Контактная разность потенциалов в МОП структуре

Фундамент современной микроэлектронной индустрии составляет приборы на основе планарных многослойных структур, среди которых главное место занимает структура металл-окисел-полупроводник (МОП).

Зонные диаграммы структуры металл-окисел-полупроводник изображены на рис. 3.1.

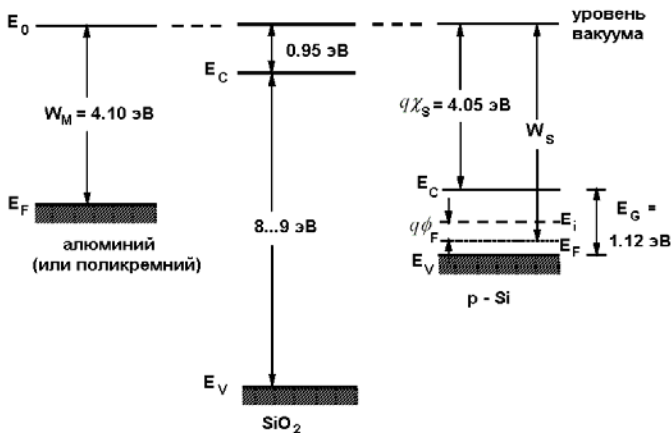


Рис. 3.1. Зонные диаграммы алюминия, SiO₂ и кремния p-типа с концентрацией легирующей примеси N_A

Здесь E₀ – уровень вакуума, E_F – уровни Ферми в металле, χ_S (Si) (4.05 эВ) – средство к электрону (electron affinity) кремния, χ_i (SiO₂) (0.95 эВ) – средство к электрону SiO₂.

Работа выхода полупроводника определяется средством к электрону, шириной запрещенной зоны и уровнем легирования

$$W_S = \chi_S + E_G / 2q + \phi_F. \quad (3.1.1)$$

Контактная разность потенциалов есть разность работ выхода в затворе и полупроводнике

$$q\phi_{MS} \equiv W_M - W_S = W_M - q\chi_S - E_G / 2 - q\phi_F. \quad (3.1.2)$$

Работа выхода алюминия W_M (4.1 эВ), но в современных приборах затворы изготавливаются чаще всего из сильнолегированного

поликремния. Уровень Ферми затвора из сильнолегированного n^+ -поликремния находится вблизи его края зоны проводимости, т.е. $W_S \cong \chi_S$. Поэтому в этом случае

$$\begin{aligned} q\varphi_{MS} = W_M - W_S &\cong -E_G / 2 - q\varphi_F = \\ &= -0.56 \text{ эВ} - q\varphi_T \ln(N_A/n_i) \cong -0.9 \text{ В}. \end{aligned} \quad (3.1.3)$$

В этих формулах N_A – легирование кремниевой подложки, затвор легирован гораздо сильнее $\sim 10^{20} \text{ см}^{-3}$. Для затвора на основе p^+ -поликремния имеем

$$\begin{aligned} W_M &= q\chi_S + E_G, \\ q\varphi_{MS} &= E_G / 2 - q\varphi_F. \end{aligned}$$

Металлические затворы менее технологичны, но как мы уже отмечали, во многих отношениях лучше поликристаллических. Как правило, работа выхода в металле меньше, чем в полупроводнике; т.о. уровень Ферми в металле выше (рис. 3.2). Электронов в затворе больше, и при замыкании цепи они переходят в полупроводник, образуя в металле слой положительного заряда.

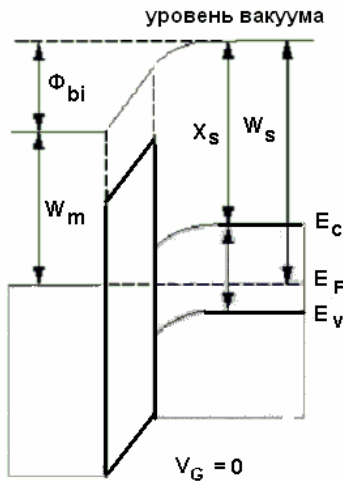


Рис. 3.2. Зонная диаграмма МОП структуры в равновесии

Уровень Ферми в затворе снижается и выравнивается с уровнем Ферми в подложке. Это есть условие термодинамического равнове-

сия, при котором поток электронов из затвора в кремний в точно-
сти уравнивается потоком электронов из кремния в затвор.

Для обеспечения выравнивания уровней Ферми края зон крем-
ния искривляются, что эквивалентно появлению электрического
поля и падения потенциала в полупроводнике и окисле. На затворе
и в полупроводнике индуцируются заряды, алгебраическая сумма
которых в равновесии в точности равна нулю.

Для подложек p -типа, как правило, $\varphi_{MS} < 0$; на затворе при этом
индуцируется положительный заряд, что эквивалентно наличию
встроенного внутреннего положительного напряжения на затворе

$$\phi_{BI} = -\varphi_{MS} = W_S - W_M = \chi_S + E_G / 2 + \varphi_F - W_M > 0.$$

3.2. Электростатика плоских слоев заряда

Электростатика в средах описывается в общем случае трехмер-
ным уравнением Пуассона

$$\Delta\varphi = -\rho(\varphi)/\varepsilon\varepsilon_0, \quad (3.2.1)$$

где Δ – трехмерный лапласиан и в правой части стоит объемная
плотность заряда, самосогласованным и, как правило, нелинейным
образом зависящая от потенциала. Нелинейный характер такого
уравнения практически исключает в общем случае его простое ана-
литическое рассмотрение.

Тем не менее, в микроэлектронике мы чаще всего имеем дело с
планарными структурами, в которых поперечные размеры много
больше полной толщины. Для таких структур можно использовать
одномерное уравнение Пуассона, полагая, что электрические поля
в плоскости структуры много меньше полей поперек структуры
(выбираем для этого направления ось x)

$$\overbrace{\frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2}}{\approx 0} - \frac{\partial^2 \varphi}{\partial x^2} \approx \frac{dE(x)}{dx} = \frac{\rho(x)}{\varepsilon \varepsilon_0}. \quad (3.2.2)$$

В этом случае объемную плотность заряда $\rho(x)$ и электриче-
ское поле $E(x)$ можно считать функцией только одной координаты.
Уравнение Пуассона превращается при этом в обыкновенное
дифференциальное (хотя и нелинейное) уравнение. На практике не
возникает необходимости детального решения уравнения Пуассона.
При расчете частот оказываются важными только некоторые

интегральные характеристики, например, падение потенциала на некотором слое или нескольких разнородных слоях. Поэтому при анализе уравнения Пуассона используются интегральные подходы, например, метод моментов.

Для начала важно подчеркнуть, что алгебраическая сумма поверхностной плотности зарядов во всех слоях планарной структуры равна нулю (локальная электронеутральность):

$$\int \rho(x) dx = 0. \quad (3.2.4)$$

На границах двух сред следует использовать условие непрерывности электрической индукции

$$\varepsilon_1 E_1 = \varepsilon_2 E_2. \quad (3.2.5)$$

Например, поля в окисле и полупроводнике связаны соотношением

$$E_{ox} = (\varepsilon_S / \varepsilon_i) E_S \cong 3E_S. \quad (3.2.6)$$

Вместо дифференциального уравнения Пуассона на практике удобнее использовать его интегральные моменты. Если домножить обе стороны (3.2.2) на x^n и проинтегрировать по всей толщине структуры, то это называется n -м моментом уравнения Пуассона. Например, проинтегрировать нулевой момент уравнения Пуассона в слое от a до b с постоянной проницаемостью законом Гаусса

$$E(b) - E(a) = \frac{1}{\varepsilon \varepsilon_0} \int_a^b \rho(x) dx. \quad (3.2.7)$$

Это есть не что иное, как закон Гаусса, т.е. интегральная форма уравнения Пуассона.

Первый (т.н. дипольный) момент уравнения Пуассона по толщине планарной структуры определяется интегралом

$$\int_a^b x \frac{dE}{dx} dx = \int_a^b x \frac{\rho(x)}{\varepsilon \varepsilon_0} dx,$$

который после интегрирования по частям приводится к виду

$$\varphi(a) - \varphi(b) = b E(b) - a E(a) - \frac{1}{\varepsilon \varepsilon_0} \int_a^b x \rho(x) dx. \quad (3.2.8)$$

Эта общая формула оказывается очень полезной при анализе одномерной электростатики многослойных планарных структур. Отметим, что условие полной планарной электронеутральности

обеспечивает независимость расчетов от выбора начала системы координат.

Необходимо помнить, что в системе СИ относительная диэлектрическая проницаемость среды во всех формулах всегда соседствует с диэлектрической проницаемостью вакуума ϵ_0 , являющейся размерным переводным коэффициентом. Поэтому в дальнейшем мы часто будем использовать сокращенное обозначение типа ϵ , подразумевая под этим размерное значение $\epsilon\epsilon_0$.

3.3. Электростатика МОП структуры с однородно - легированной подложкой

Для определенности будем всегда рассматривать подложку p -типа с концентрацией легирующей примеси N_A . При приложении положительного смещения на затвор на нем образуется поверхностный слой положительного заряда с поверхностной плотностью N_G . Этот заряд компенсируется отрицательным зарядом в полупроводнике, который складывается из заряда электронов в инверсионном слое n_s и заряда акцепторов. В приближении обедненного слоя мы считаем, что имеется однородный слой отрицательно заряженных акцепторов толщиной x_d (начало координат при $x=0$) и очень тонкий слой электронов инверсионного слоя (канала)

$$\rho(x) = -qN_A(x) - qn_s \delta(x-0), \quad (3.3.1)$$

где объемная плотность электронов в инверсионном слое представлена в виде произведения известной δ -функции Дирака и поверхностной плотности электронов в инверсионном слое n_s . Аргумент δ -функции чуть сдвинут в положительную сторону так, чтобы выполнялось основное свойство функции Дирака $\int_0^\infty \delta(x-0)dx = 1$,

откуда следует $\int_0^\infty n(x)dx = n_s$.

Пользуясь общим выражением (3.2.8), получаем полное падение потенциала на слое с неоднородным легированием от нейтральной подложки (при $x = x_d$) до границы раздела кремния и подзатворного окисла (при $x = 0$)

$$\varphi(x=0) - \varphi(x=x_d) \equiv \varphi_S = \frac{1}{\epsilon_S} \int_0^{x_d} q N_A(x) x dx. \quad (3.3.2)$$

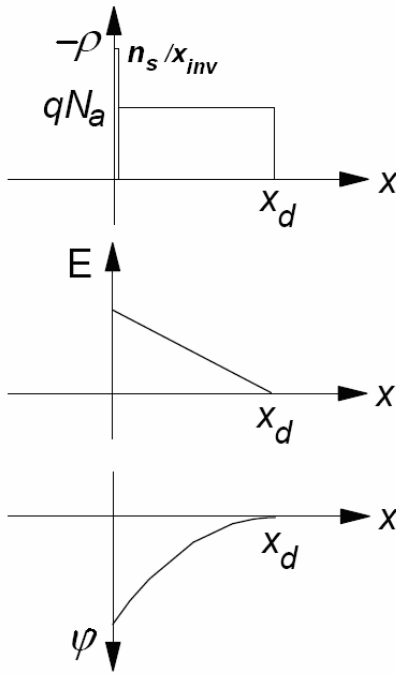


Рис. 3.3. Распределение плотности заряда, электрического поля и потенциала в n -Si подложке МОП структуры

Если подложка однородно легирована ($N_A = \text{const}$), то получаем (рис. 3.3)

$$E(x) \cong E_S \left(1 - \frac{x}{x_d} \right), \quad (3.3.3)$$

$$E_S = \frac{q}{\varepsilon_S} (n_s + N_A x_d), \quad (3.3.4)$$

$$\varphi(x) \cong \varphi_S \left(1 - \frac{x}{x_d} \right)^2, \quad (3.3.5)$$

$$\varphi_S \cong q N_A x_d^2 / 2 \varepsilon_S, \quad (3.3.6)$$

где E_S – электрическое поле точно на границе раздела Si-SiO₂ со стороны кремния.

Ширина (x_d) и плотность заряда обедненной области на единицу площади для однородно-легированной подложки определяются условием электронейтральности и выражаются в приближении обедненного слоя простыми зависимостями от поверхностного потенциала

$$x_d \cong (2\varepsilon_S \varphi_S / qN_A)^{1/2},$$

$$Q_d(\varphi_S) = qN_A x_d = \sqrt{2\varepsilon_S qN_A \varphi_S}. \quad (3.3.7)$$

Отсюда легко получить удельную (на единицу площади) емкость обедненной области

$$C_D = \frac{dQ_d}{d\varphi_S} = \left(\frac{\varepsilon_S q N_A}{2\varphi_S} \right)^{1/2} = \frac{\varepsilon_S}{x_d(\varphi_S)}. \quad (3.3.8)$$

НВ-1. Мы получили, что полное падение потенциала на обедненной области не зависит от плотности электронов в инверсионном слое! Это связано с тем, что мы пренебрегли толщиной инверсионного слоя по сравнению с толщиной обедненной области. Это приближение вполне оправдано для больших затворных напряжений ($V_G \gg 1$ В), поскольку падение потенциала на толщине инверсионного слоя составляет всего несколько φ_T (50...60 мВ). Тем не менее, в нанoeлектронных приборах с низким напряжением питания падение напряжения на инверсионном слое необходимо учитывать.

НВ-2. Из закона Гаусса следует, что электрическое поле в полупроводнике на границе раздела *правее* тонкого инверсионного слоя определяется только зарядом акцепторов

$$E(x_1 \geq x_{inv}) \cong (q/\varepsilon_S) N_A x_d.$$

Из закона Гаусса получаем, что поперечное электрическое поле точно на границе раздела:

$$E_S = E(x=0) = \frac{q}{\varepsilon_S} (n_S + N_A x_d). \quad (3.3.9)$$

Таким образом, на очень малой толщине (~ 10 нм) инверсионного слоя электрическое поле испытывает скачок, изменяясь в несколько раз. Поэтому часто пользуются эффективным (фактически, средним) значением прижимающего электрического поля в инверсионном слое

$$E_{eff} = \frac{q}{\epsilon_S} (N_A x_d + 0.5 n_S). \quad (3.3.10)$$

3.4. Падение потенциалов в неоднородно-легированном полупроводнике

Для контроля порогового напряжения и смягчения короткоканальных эффектов широко применяется неоднородное легирование подложки. Будем рассматривать упрощенную модель неоднородного легирования, когда от 0 до x_1 уровень легирования составляет постоянную величину N_1 , а при $x > x_1$ концентрация легирующей примеси – N_2 (рис. 3.4). Величина x_1 является технологическим параметром, а полная толщина обедненной области x_d определяется условием электронейтральности.

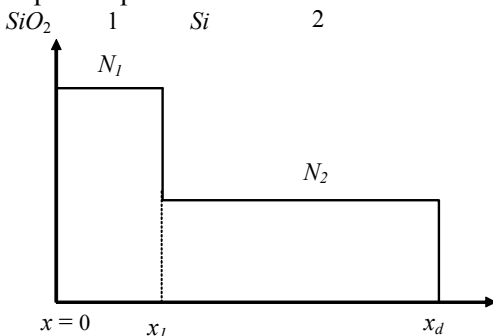


Рис. 3.4. Идеализированный профиль легирования и распределение ионизированной примеси в подложке

Если толщина обедненной области $x_d < x_1$, то это соответствует случаю однородного легирования. Нас будет интересовать случай $x_d > x_1$. В соответствии с общей формулой (3.2.8) имеем падение потенциала на слое 1 ($a=0, b=x_1$)

$$\Delta\varphi_1 = E_1 x_1 + \frac{qN_1 x_1^2}{2\varepsilon_S},$$

и падение потенциала на слое 2 ($a=x_1, b=x_d$)

$$\Delta\varphi_2 = -E_1 x_1 + \frac{qN_2 (x_d^2 - x_1^2)}{2\varepsilon_S}.$$

Электрическое поле в точке x_1 равно $E_1 = qN_2 (x_d - x_1) / \varepsilon_S$,

а полное падение потенциала на обедненном слое

$$\varphi_S = \Delta\varphi_1 + \Delta\varphi_2 = \frac{qN_1 x_1^2}{2\varepsilon_S} + \frac{qN_2 (x_d^2 - x_1^2)}{2\varepsilon_S}. \quad (3.4.1)$$

Отсюда находим толщину обедненного слоя

$$x_d(\varphi_S) = \left(\frac{2\varepsilon_S \varphi_S}{qN_2} - \frac{(N_1 - N_2)x_1^2}{N_2} \right)^{1/2}. \quad (3.4.2)$$

Таким образом, получаем все интересующие нас величины как функции, зависящие от поверхностного потенциала и технологических параметров. Например, полный заряд обедненной области выражается

$$Q_d(\varphi_S) = qN_1 x_1 + qN_2 (x_d(\varphi_S) - x_1). \quad (3.4.3)$$

Дифференцируя это выражение по потенциалу, можно получить емкость обедненной области. Оно будет иметь вид как в (3.3.8), где обедненная область определяется формулой (3.4.2).

3.5. Учет напряжения, приложенного к затвору

Рассмотрим ситуацию, когда относительно подложки к затвору приложено положительное напряжение V_{GB} . Ток через структуру

блокируется подзатворным изолятором, но уровни Ферми в затворе и подложке отличаются:

$$qV_{GB} = E_F(\text{gate}) - E_F(\text{Si}). \quad (3.5.1)$$

NB. Запомните, что повышение потенциала для электрона означает понижение его полной энергии.

Рассмотрим два первых момента уравнения Пуассона. Поверхностная плотность положительного заряда на затворе должна быть в точности равна плотности отрицательного заряда в полупроводнике (электроны инверсионного слоя и ионизированные акцепторы)

$$N_G \equiv n_S + N_A x_d. \quad (3.5.2)$$

NB. В общем (трехмерном) случае нужно приравнять не плотности, а полные заряды на электродах (затвор, подложка, исток, сток).

По закону Гаусса в окисле у границы раздела с металлом

$$E_{ox} = qN_G / \varepsilon_i. \quad (3.5.3)$$

Если заряд в окисле отсутствует или расположен в тонком слое у границы раздела с кремнием (что является типичной ситуацией), то полное падение потенциала в окисле $V_{ox} = E_{ox} d_{ox}$. Полное падение потенциалов на толщине структур складывается из падений потенциала в окисле V_{ox} и полупроводнике φ_S :

$$\begin{aligned} V_{GB} &= \varphi_{MS} + \varphi_S + E_{ox} d_{ox} = \\ &= \varphi_{MS} + \varphi_S + \frac{qN_G}{\varepsilon_i} d_{ox} = \varphi_{MS} + \varphi_S + \frac{qN_G}{C_o}, \end{aligned} \quad (3.5.4)$$

где $C_o = \varepsilon_i / d_{ox}$ — удельная емкость оксида. Выше было показано, что отрицательная контактная разность потенциалов $\varphi_{MS} = -\phi_{BI}$ между материалами затвора и кремния эквивалентна положительному смещению на затворе

$$\begin{aligned} V_{GB} + \phi_{BI} &= \varphi_S + E_{ox} d_{ox} = \varphi_S + \frac{N_G}{\varepsilon_i} d_{ox} \cong \\ &\cong \underbrace{\frac{qN_A x_d^2(\varphi_S)}{2\varepsilon_S}}_{\varphi_S} + \overbrace{\frac{qN_A x_d(\varphi_S)}{C_o} + \frac{qn_S(\varphi_S)}{C_o}}^{E_{ox} d}. \end{aligned} \quad (3.5.5)$$

Обычно в окисле есть положительный заряд, захваченный на дефекты вблизи раздела Si-SiO₂ (< 1 - 3 нм) и непосредственно на границе раздела. Дефекты такого рода могут обмениваться носителями с кремниевой подложкой, меняя свое зарядовое состояние. Заполнение и опустошение уровней дефектов контролируется положением их уровней относительно приповерхностного уровня Ферми. Поэтому полный заряд на поверхностных и приповерхностных дефектах также зависит от поверхностного потенциала.

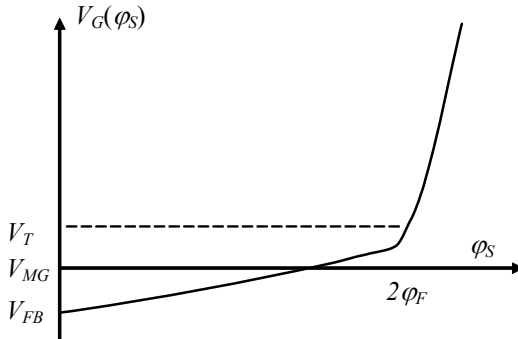


Рис. 3.5. Зависимость напряжения на затворе от поверхностного потенциала

При наличии таких заряженных дефектов имеем $N_G = n_S + N_A x_d - N_t(\varphi_S)$, и соответственно, зависимость затворного напряжения от поверхностного потенциала приобретает вид (см. также рис. 3.5)

$$V_{GB}(\varphi_S) = \varphi_{MS} + \varphi_S + \frac{q}{C_o} (n_S(\varphi_S) + N_A x_D(\varphi_S) - N_t(\varphi_S)). \quad (3.5.6)$$

Обратите внимание, что в этом базовом уравнении все слагаемые в правой части являются функцией поверхностного потенциала. Схематически зависимость затворного напряжения представлена на рис. 3.5. При малых напряжениях на затворе, когда в кремнии еще не сформировался инверсионный слой электронов, зависимость V_{GB} от поверхностного потенциала имеет линейный и сублинейный характер. Начало формирования инверсионного слоя (канала) сопровождается резким ростом функции $V_{GB}(\varphi_S)$, обу-

словленным экспоненциальной зависимостью плотности электронов в канале от поверхностного потенциала.

3.6. Характерные затворные напряжения

В нейтральной кремниевой p -подложке с объемной плотностью акцепторов N_A равновесные концентрации невырожденных дырок и электронов выражаются формулами

$$p \cong N_A, \quad n \cong \frac{n_i^2}{N_A} = N_A \exp\left(-2 \frac{\varphi_F}{\varphi_T}\right), \quad (3.6.1)$$

так что выполняется закон действующих масс

$$np = n_i^2, \quad (3.6.2)$$

где $n_i \cong 10^{10} \text{ см}^{-3}$ (при $T = 300 \text{ К}$) – собственная концентрация кремния, $\varphi_F \equiv \varphi_T \ln(N_A/n_i)$ – потенциал Ферми в объеме, характеризующий положение уровня Ферми относительно середины запрещенной зоны кремния.

Прикладывая положительное напряжение к затвору, мы увеличиваем потенциал в объеме p -кремния и на границе раздела (см. (3.3.5)). При этом концентрация электронов экспоненциальным образом увеличивается, а дырок – уменьшается. В частности, для объемной концентрации электронов и дырок на границе раздела с окислом имеем

$$n(\varphi_S) \cong (n_i^2 / N_A) \exp(\varphi_S / \varphi_T), \quad (3.6.3)$$

$$p(\varphi_S) \cong N_A \exp(-\varphi_S / \varphi_T).$$

При некотором значении поверхностного потенциала концентрации электронов и дырок на границе раздела сравниваются $n(\varphi_S, x=0) = p(\varphi_S, x=0)$

$$\varphi_{MG} = \varphi_F \equiv \varphi_T \ln(N_A/n_i). \quad (3.6.4)$$

Такой потенциал называется потенциалом середины зоны (midgap potential), поскольку уровень Ферми при этом находится в середине зоны. Соответствующее напряжение (midgap voltage) называется напряжением середины зоны

$$V_{MG} \equiv V_G(\varphi_F) = \varphi_{MS} + \varphi_F + (qN_A x_d(\varphi_F) - Q_i(\varphi_F)) / C_O. \quad (3.6.5)$$

Напряжение, при котором поверхностный потенциал в кремнии равен нулю, называют напряжением «плоских зон» V_{FB} (flatband

voltage). Полный заряд в кремнии при этом равен нулю, и края зон кремния не изогнуты (условие «плоских зон»). Подставляя $\varphi_S = 0$ в (3.5.6), получаем

$$V_{FB} \equiv V_G(0) = \varphi_{MS} - Q_i(\varphi_S = 0)/C_O < 0. \quad (3.6.6)$$

Необходимо подчеркнуть, что из-за наличия заряда в окисле, условие «плоских зон» в кремнии не означает «плоские зоны» (т.е. отсутствие электрического поля в окисле) в окисле.

3.7. Пороговое напряжение

При некотором потенциале плотность *неосновных* носителей (электронов) на *границе раздела* сравнивается с плотностью *основных* носителей (дырок) в *подложке*: $n(\varphi_S) = N_A$. Этот потенциал соответствует началу режима, который называется режимом сильной инверсии

$$\varphi_S(inv) = 2\varphi_F \equiv 2\varphi_T \ln(N_A/n_i). \quad (3.7.1)$$

Соответствующее напряжение пороговое напряжения (threshold voltage, SPICE параметр V_{TO})

$$V_T \equiv V_G(2\varphi_F) = \varphi_{MS} + 2\varphi_F + (qN_A x_d(2\varphi_F) - Q_i(2\varphi_F))/C_O. \quad (3.7.2)$$

При дальнейшем увеличении напряжения на затворе поверхностный потенциал практически перестает расти (см. рис. 3.5). Соответственно, перестает расти толщина обедненной области, насыщаясь на своем максимальном значении

$$x_{d,max} \equiv (4\epsilon_S \varphi_F / qN_A)^{1/2} = (4\epsilon_S \varphi_T \ln(N_A/n_i) / qN_A)^{1/2}. \quad (3.7.3)$$

Это очень важная величина, поскольку она определяет, например, минимально возможную длину канала. Наконец, замедление изменения поверхностного потенциала (фактически, уровня Ферми на границе раздела) приводит к резкому уменьшению перезарядки приповерхностных дефектов в окисле. Это означает, что в формуле (3.5.6) в режиме сильной инверсии существенно изменяется только слагаемое, связанное с экспоненциальной функцией электронов в канале. С другой стороны, это слагаемое практически равно нулю при $\varphi_S < 2\varphi_F$. Это дает возможность использовать в надпороговом режиме ($\varphi_S > 2\varphi_F \Leftrightarrow V_G > V_T$) замечательное приближение, находящееся в очень хорошем согласии с экспериментальными данными,

$$q n_S \cong C_O (V_G - V_G |_{\varphi_S=2\varphi_F}) = C_O (V_G - V_T). \quad (3.7.4)$$

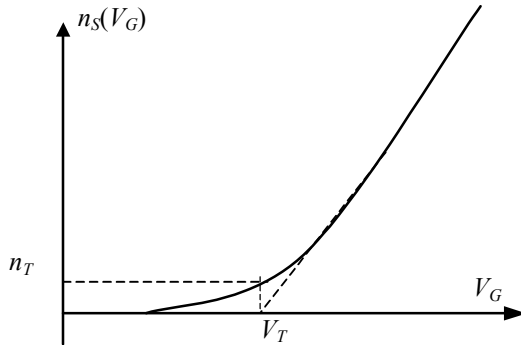


Рис. 3.6. Схематическая зависимость плотности носителей в канале от напряжения на затворе

Таким образом, при напряжении, превышающем пороговое хотя бы на несколько φ_T , плотность носителей в канале начинает расти практически линейно (см. рис. 3.5). Использование этого приближения позволяет обходить многие сложные расчетные проблемы. Например, точный расчет вида зависимости $n_S(\varphi_S)$ требует, вообще говоря, квантостатистического рассмотрения (учет размерного квантования, вырождения и заполнения подуровней). На практике введение феноменологического порогового напряжения и использование формулы (3.7.4) позволяют избегать все эти проблемы. Это является одной из причин того, что в современных системах компактных моделей МОП транзисторов (типа BSIM4) практически отсутствуют квантовые эффекты. В современной 70 нм технологии имеем следующие характерные параметры $N_A \cong 4.6 \times 10^{18} \text{ см}^{-3}$, $\varphi_F \cong 0.52 \text{ В}$, $V_{FB} \cong -0.56 - \varphi_F \cong 1.08 \text{ В}$, $V_T \cong 0.38 \text{ В}$ (ITRS 2005).

3.8. Полный заряд в полупроводнике при заданном поверхностном потенциале

В практических расчетах обычно ограничиваются рассмотрением приближения невырожденного электронного газа с бoльцмановской статистикой. Это приближение работает в режиме обеднения и слабой инверсии ($\varphi_S < 2\varphi_F$) и вполне удовлетворительно описывает ситуацию сильной инверсии ($\varphi_S > 2\varphi_F$), соответствующее вырожденным носителям.

Одномерное уравнение Пуассона для объема полупроводника p -типа

$$\frac{d^2\varphi(x)}{dx^2} = -\frac{q}{\varepsilon_S} (p(x) - n(x) - N_A). \quad (3.8.1)$$

Нас интересует случай обеднения и инверсии, когда $\varphi_S > \varphi_F$, где с учетом (3.6.1) имеем следующее приближение для правой части уравнения Пуассона:

$$\begin{aligned} \rho(x) = q(p(x) - n(x) - N_A) &= q N_A \left(\exp\left(-\frac{\varphi}{\varphi_T}\right) - \exp\left(\frac{\varphi - 2\varphi_F}{\varphi_T}\right) - 1 \right) \cong \\ &\cong -q N_A \left(\exp\left(\frac{\varphi - 2\varphi_F}{\varphi_T}\right) + 1 \right), \end{aligned} \quad (3.8.2)$$

где мы пренебрегли концентрацией дырок.

В равновесии (в отсутствии тока вдоль структуры) в правой части уравнения Пуассона вся зависимость от координаты присутствует в зависимости потенциала от этой координаты $\rho(x) = \rho(\varphi(x))$ (это справедливо только при отсутствии тока, поскольку при наличии тока появляется *явная* зависимость электрохимического потенциала и плотности носителей от координаты).

Это дает возможность использовать для решения нелинейного уравнения (3.8.1) следующий математический трюк. Представим (3.8.1) в эквивалентной форме (3.8.3) с использованием переменной $E = d\varphi/dy$, имеющей смысл электрического поля, и проинтегрируем обе стороны уравнения от границы раздела кремния с изолятором, где электрическое поле равно искомому E_S , а потенциал равен по-

верхностному потенциалу φ_S , до границы обедненного слоя, где электрическое поле и потенциал равны нулю:

$$\left(\underbrace{\frac{d\varphi}{dx}}_{E(x)} \right) d \left(\underbrace{\frac{d\varphi}{dx}}_{E(x)} \right) \Big|_{E_S}^0 = -\frac{q}{\varepsilon_S} \rho(\varphi) d\varphi = \frac{qN_A}{\varepsilon_S} \left(1 + \exp\left(\frac{\varphi - 2\varphi_F}{\varphi_T} \right) \right) d\varphi \Big|_{\varphi_S}^0. \quad (3.8.3)$$

Отсюда получаем точную (в рамках использованных приближений) связь электрического поля на границе раздела и поверхностного потенциала

$$\frac{1}{2} E_S^2 = \frac{q N_A}{\varepsilon_S} \varphi_T \left(\frac{\varphi_S}{\varphi_T} + \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T} \right) \right). \quad (3.8.4)$$

Удобно ввести обозначение для безразмерной функции

$$F(\varphi_S) = \left(\frac{\varphi_S}{\varphi_T} + \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T} \right) \right)^{1/2}, \quad (3.8.5)$$

и для характерной длины задачи, которая называется дебаевской длиной,

$$L_D \equiv \sqrt{\frac{2 \varepsilon_S k T}{q N_A}}. \quad (3.8.6)$$

Тогда полная поверхностная концентрация заряда в полупроводнике для заданного потенциала

$$Q_S \equiv q N_S = \varepsilon_S E_S = q N_A L_D F(\varphi_S). \quad (3.8.7)$$

Электрическое поле на границе раздела в кремнии

$$E_S = \frac{\varphi_T}{L_D} F(\varphi_S). \quad (3.8.8)$$

Предельные случаи сильной инверсии и глубокого обеднения

$$N_S(\varphi_S) \cong \begin{cases} N_A L_D \exp\left(\frac{\varphi_S - 2\varphi_F}{2\varphi_T} \right), & \varphi_S > 2\varphi_F; \\ N_A L_D \left(\frac{\varphi_S}{\varphi_T} \right)^{1/2} = N_A x_d(\varphi_S), & 2\varphi_F > \varphi_S > \varphi_F. \end{cases} \quad (3.8.9)$$

В первом случае мы пренебрегли вторым слагаемым в $F(\varphi_S)$, ответственным за заряд неподвижных акцепторов, во втором – зарядом инверсионного слоя (канала).

3.9. Плотность электронов в канале как функция поверхностного потенциала

Поверхностная плотность электронов в канале есть разность между полным зарядом в полупроводнике и зарядом обедненного слоя

$$n_S = N_S - N_A x_d = N_A L_D \left[\left(\frac{\varphi_S}{\varphi_T} + \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T} \right) \right)^{1/2} - \left(\frac{\varphi_S}{\varphi_T} \right)^{1/2} \right]. \quad (3.9.1)$$

Это приближение часто называют приближением зарядовых плоскостей (charge sheet approximation). Воспользовавшись алгебраическим тождеством $\sqrt{a} - \sqrt{b} \equiv (a - b) / (\sqrt{a} + \sqrt{b})$, получаем выражение для поверхностной плотности заряда в канале как функции поверхностного потенциала

$$n_S(\varphi_S) = N_A L_D \frac{\exp((\varphi_S - 2\varphi_F)/\varphi_T)}{\left[\varphi_S/\varphi_T + \exp((\varphi_S - 2\varphi_F)/\varphi_T) \right]^{1/2} + \left[\varphi_S/\varphi_T \right]^{1/2}}. \quad (3.9.2)$$

Эта классическая формула справедлива от глубокого обеднения до сильной инверсии. Плотность электронов в канале в предельных случаях сильной инверсии и глубокого обеднения записывается в виде

$$qn_S \cong \begin{cases} C_D \varphi_l \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T} \right), & 2\varphi_F > \varphi_S > \varphi_F, \\ qN_A L_D \exp\left(\frac{\varphi_S - 2\varphi_F}{2\varphi_T} \right), & \varphi_S > 2\varphi_F. \end{cases} \quad (3.9.3)$$

где емкость обедненной области C_D определяется формулой (3.3.8). Формула (3.9.2) может быть использована для расчета зависимости $V_G(\varphi_S)$.

NB. Приближенная эмпирическая формула (3.7.4) дает $n_S(V_G = V_T) \cong 0$. На самом деле плотность носителей в канале при напряжении, равном пороговому, может быть достаточно большой (рис. 3.6):

$$\begin{aligned}
 q n_T = q n_S (\varphi_S = 2\varphi_F) &= q N_A L_D \frac{1}{[2\varphi_F/\varphi_T + 1]^{1/2} + [2\varphi_F/\varphi_T]^{1/2}} \cong \\
 &\cong \sqrt{\frac{q N_A \varepsilon_S k T}{4 \ln(N_A/n_i)}} \cong 10^{11} \text{ см}^{-2}. \quad (3.9.4)
 \end{aligned}$$

3.10. Тепловая толщина инверсионного слоя (канала)

Выражение (3.9.2) полезно получить иным способом, используя приближение эффективного прижимающего электрического поля E_{eff} . Действительно, потенциальную энергию электронов в очень тонком инверсионном слое можно записать в виде $U(x) = -qE_{eff}x$. Тогда распределение объемной концентрации невырожденных электронов вблизи границы раздела можно записать в виде

$$n(x) = n(x=0) \exp\left(-\frac{qE_{eff}x}{kT}\right) = \underbrace{\frac{n_i^2}{N_A} \exp\left(\frac{q\varphi_S}{kT}\right)}_{n(x=0)} \exp\left(-\frac{qE_{eff}x}{kT}\right). \quad (3.10.1)$$

Эта формула записана в приближении локальной объемной плотности носителей, которое для краткости будем называть квазиобъемным приближением. Формально говоря, оно неприменимо для таких квазидвумерных структур, какой является инверсионный слой МОП транзистора. Тем не менее, квазиобъемное приближение дает вполне разумные результаты. Поверхностная плотность электронов в инверсионном слое в этом приближении выражается формулой

$$n_S = \int_0^\infty n(x) dx = \frac{n_i^2}{N_A} \exp\left(\frac{q\varphi_S}{kT}\right) \frac{kT}{qE_{eff}} = N_A \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T}\right) \frac{\varphi_T}{E_{eff}}. \quad (3.10.2)$$

Мы видим, что поверхностная плотность электронов в канале представляется в виде произведения объемной концентрации на границе раздела на т.н. тепловую толщину

$$x_{inv} = \varphi_T / E_{eff}. \quad (3.10.3)$$

Физический смысл тепловой толщины состоит в том, что для невырожденного случая в слое такой толщины находится экспоненциально большая часть носителей канала.

Используя выражение для эффективного значения электрического поля (3.3.10) и (3.10.2), можно вновь получить формулу (3.9.2).

3.11. Зависимость эффективного прижимающего поля от затворного напряжения в надпороговом режиме

Эффективное прижимающее поле в канале является важной характеристикой, определяющей, в частности, значение подвижности носителей в канале в надпороговом режиме ($V_G > V_T$). Поэтому важно уметь вычислять эту величину как явную функцию напряжения на затворе. Для этого вспомним, что плотности зарядов в обедненном и инверсионном слоях в надпороговом режиме ($V_G > V_T$) можно оценить по формулам

$$qN_A x_d (2\varphi_F) = C_o(V_T - V_{FB} - 2\varphi_F) \text{ и } qn_s = C_o(V_G - V_T).$$

Для МОПТ с n^+ поли-Si затвором имеем приближенно $V_{FB} + 2\varphi_F \cong 0$. Тогда имеем

$$E_{eff} = \frac{q}{\varepsilon_S} (0.5 n_s + N_A x_d) \cong q \left(\frac{\varepsilon_{ox}}{\varepsilon_S} \right)^{\cong 1/3} \left(\frac{V_T - \overbrace{(V_{FB} + 2\varphi_F)}^{\cong 0}}{d_{ox}} + 0.5 \frac{V_G - V_T}{d_{ox}} \right),$$

и получаем оценку для эффективного прижимающего электрического поля в канале

$$E_{eff} \approx \frac{V_G + V_T}{6d_{ox}}. \quad (3.11.1)$$

Отметим, что прижимающее поле точно на границе раздела больше эффективного ($E_S > E_{eff}$)!

$$E_S = \frac{q}{\varepsilon_S} (N_A x_d (2\varphi_F) + n_s) = q \frac{\varepsilon_{ox}}{\varepsilon_S} \left(\frac{V_G - \overbrace{(V_{FB} + 2\varphi_F)}^{\cong 0}}{d_{ox}} \right) \approx \frac{V_G}{3d_{ox}}. \quad (3.11.2)$$

NB. Приведенные в этом пункте расчетные формулы справедливы для случая, когда зарядом, захваченным на дефектах в окисле и на границе раздела, можно пренебречь!

3.12. Контроль порогового напряжения за счет легирования подложки

Контроль порогового напряжения играет ключевую роль в разработке и изготовлении МОПТ. Пороговое напряжение для МОП транзисторов с разными типами подложек можно записать в форме

$$V_T = V_{FB} + 2\phi_F \pm \frac{Q_D}{C_O}, \quad (3.12.1)$$

где знак плюс соответствует подложке p -типа; минус – подложке n -типа. Повышение уровня легирования увеличивает пороговое напряжение для n -типа подложки и уменьшает для p -типа (рис. 3.7).

Из формулы (3.12.1) видно также, что подгонку порогового напряжения можно проводить технологически за счет контролируемого изменения:

- толщины подзатворного окисла (и удельной емкости C_O);
- уровня легирования подложки (добавление акцепторов в подложку увеличивает V_T , а добавление доноров уменьшает порог вне зависимости от типа подложки!);
- изменения контактной разности потенциалов за счет подбора материала затворов (например, за счет использования сплавов кремния с германием).

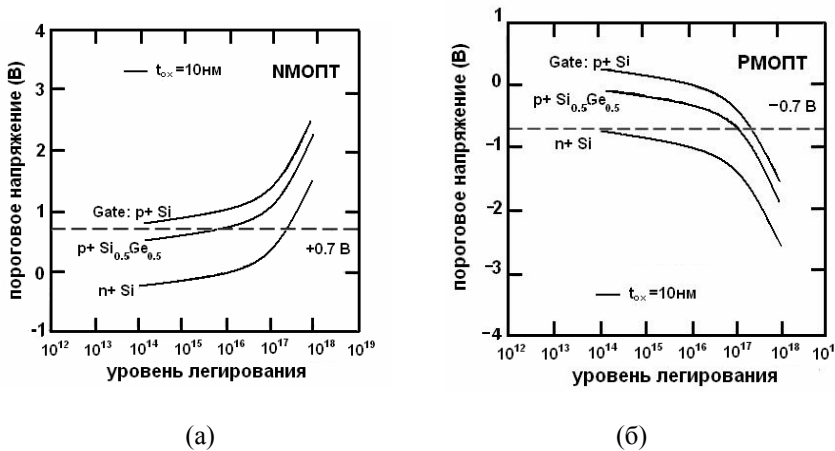


Рис. 3.7. Зависимости порогового напряжения от уровня легирования для подложек (а) n -типа и (б) p -типа

3.13. Регулирование порогового напряжения за счет работы выхода материала затвора

Идея такого регулирования состоит в возможности изменения контактной разности потенциалов между материалом подложки и затвора $\varphi_{MS} = W_G - W_S$.

Работа выхода в кремнии равна (рис. 3.1)

$$W_S = \chi_S + E_G/2q + \varphi_F. \quad (3.13.1)$$

Для сильнолегированного поликремниевых затвора n^+ - типа:

$$W_G \cong \chi_{Si}; \quad \varphi_{MS} \cong -0.9\text{В}.$$

Для поликремниевых затвора p^+ - типа:

$$W_G \cong \chi_{Si} + E_G/q; \quad \varphi_{MS} \cong 0\text{В}.$$

Как видим, пороговое напряжение для n^+ -затворов всегда меньше приблизительно на 0.9 В, чем в затворах на p^+ -поли-Si.

Для затворов на основе германиевых сплавов с поликремнием поли-Si_xGe_{1-x} имеем промежуточное значение работы выхода

$$W_G \cong \chi_{Si} + E_G/2q,$$

которое можно регулировать, меняя долю x германия в сплаве. Работа выхода в таких сплавах уменьшается с ростом доли германия.

3.14. Профили легирования

В старых технологиях, в которых уровень легирования подложки был относительно низким, актуальной являлась задача увеличения порогового напряжения за счет более сильного в области канала и спадающего профиля легирования вглубь подложки.

Увеличение степени легирования подложки вызывает негативные явления, а именно:

уменьшение подвижности носителей в канале;

усиление эффекта влияния подложки на пороговое напряжение (глава 4).

По мере уменьшения технологической нормы и увеличения степени легирования подложки использовался постоянный профиль легирования (рис. 3.8)

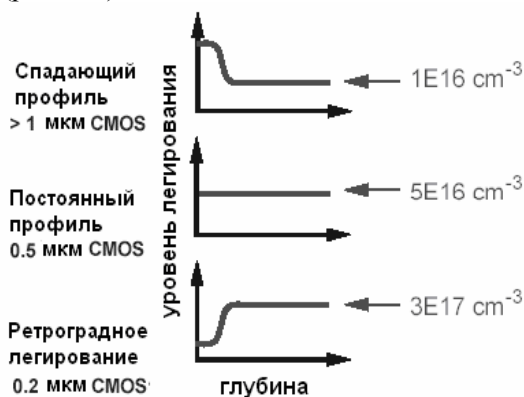


Рис. 3.8. Эволюция профилей легирования

Появилась проблема понижения порогового напряжения, которую решают с помощью повышающего профиля легирования или т.н. ретроградного легирования.

Другим важнейшим преимуществом ретроградного легирования является уменьшение интенсивности рассеяния носителей канала на атомах примеси и соответствующее увеличение подвижности и быстродействия транзистора.

3.15. Спадающий (HIGH-LOW) профиль

Процедура определения порогового напряжения при неоднородном легировании (рис. 3.9) проводится с использованием общих формул в п. 3.2 и 3.4. Сначала определяем поверхностный потенциал при заданной толщине обедненной области

$$\varphi_S = \Delta\varphi_1 + \Delta\varphi_2 = \frac{qN_1x_1^2}{2\varepsilon_S} + \frac{qN_2(x_d^2 - x_1^2)}{2\varepsilon_S}. \quad (3.15.1)$$

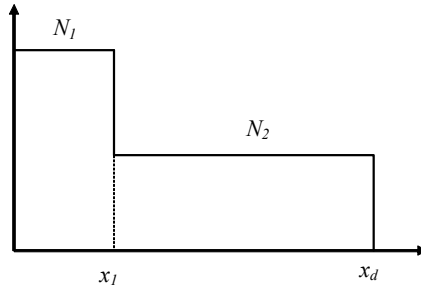


Рис. 3.9. Спадающий профиль легирования

Затем находим толщину обедненной области при потенциале $\varphi_S = 2\varphi_F$, соответствующую пороговому напряжению,

$$x_d(\varphi_S) = \left(\frac{2\varepsilon_S\varphi_S}{qN_2} - \frac{(N_1 - N_2)x_1^2}{N_2} \right)^{1/2}. \quad (3.15.2)$$

Далее вычисляем полный заряд обедненной области

$$Q_d = qN_1x_1 + qN_2(x_d - x_1) = qN_2x_d + q(N_1 - N_2)x_1. \quad (3.15.3)$$

И наконец, получаем выражение для порогового напряжения

$$V_T = V_{FB} + 2\varphi_F + \frac{Q_d}{C_O} = V_{FB} + 2\varphi_F + \frac{1}{C_O} \left(\sqrt{2\varepsilon_S q N_2 \left(2\varphi_F - \frac{q(N_1 - N_2)x_1^2}{2\varepsilon_S} \right)} + q(N_1 - N_2)x_1 \right). \quad (3.15.4)$$

Эти формулы справедливы, конечно, только для случая $x_d > x_1$. При обратном неравенстве следует использовать формулы для однородного легирования.

3.16. Нарастающий профиль (LOW-HIGH, ретроградное легирование)

Формулы, полученные в предыдущем пункте, справедливы и для случая ретроградного легирования $N_1 < N_2$ (рис. 3.10)

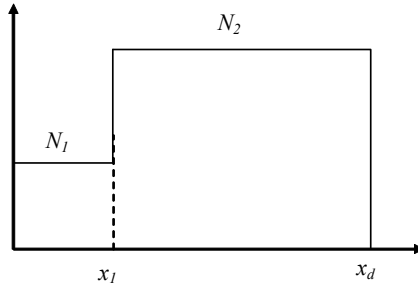


Рис. 3.10. Идеализированный профиль ретроградного легирования

Например, полагая $N_1 = 0$, получаем выражение для порогового напряжения

$$V_T \cong V_{FB} + 2\varphi_F + \frac{1}{C_O} \left(\sqrt{2\varepsilon_S q N_2 \left(2\varphi_F + \frac{q N_2 x_1^2}{2\varepsilon_S} \right)} - q N_2 x_1 \right). \quad (3.16.1)$$

На практике ретроградное легирование в современных транзисторах создается с помощью создания относительно сильнолегированного слоя в относительно слаболегированной подложке (рис. 3.11).

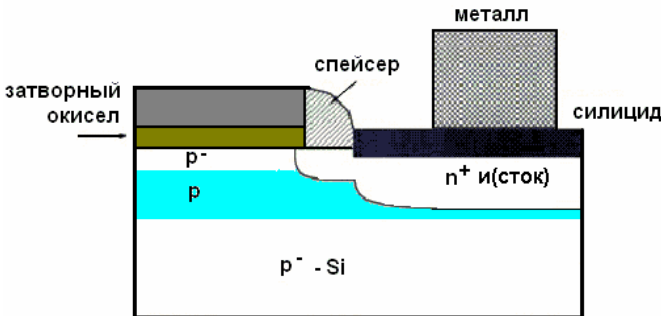


Рис. 3.11. Профиль легирования в современном транзисторе

При этом, сильнолегированный p -слой экранирует pn -переход стока, уменьшая его толщину и улучшая электростатическое качество транзистора.

3.17. Легирование дельта-слоем

Если имплантированный сильнолегированный слой является достаточно узким, то в этом случае говорят о легировании дельта-слоем. Профиль концентрации легирующей примеси при имплантации можно аппроксимировать гауссовским распределением

$$N_{\delta}(x) = \frac{N_I}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-x_c)^2}{2\sigma^2}\right), \quad (3.17.1)$$

где N_I – полная поверхностная концентрация примеси в слое с характерной толщиной σ , находящейся на расстоянии x_c от границы раздела. Если это слой тонкий ($\sigma \ll x_d$), то гауссовскую функцию можно приближенно заменить дельта-функцией Дирака $N_{\delta}(x) \cong N_I \delta(x - x_c)$. Тогда поверхностный потенциал приближенно выражается формулой

$$\varphi_S = \frac{q}{\varepsilon_S} \int_0^{x_d} x(N_{\delta}(x) + N_A) dx = \frac{q}{\varepsilon_S} N_I x_c + \frac{qN_A x_d^2}{2\varepsilon_S}, \quad (3.17.2)$$

где N_A – уровень легирования слаболегированной толщи подложки. Отсюда получаем толщину обедненной области в точке порога

$$x_d(2\varphi_F) = \left(\frac{4\varphi_F \varepsilon_S}{qN_A} - \frac{2N_I x_c}{N_A} \right)^{1/2}. \quad (3.17.3)$$

Используя это значение, получаем выражение для порогового напряжения при дельта-легировании

$$V_T = V_{FB} + 2\varphi_F + \frac{q}{C_O} (N_A x_d(2\varphi_F) + N_I). \quad (3.17.4)$$

Эта формула справедлива только для случая $x_c > x_d(2\varphi_F)$. При достаточно сильном легировании подложки и/или большом значении глубины залегания дельта-слоя x_c электрическое поле затвора и обедненная область не достигает дельта-слоя, и мы имеем случай однодального легирования.

Более интересным является случай, когда легированием подложки можно пренебречь. Электрическое поле в подложке при этом почти постоянно и все его силовые линии оканчиваются на примесях дельта-слоя (рис. 3.12). Высоколегированный слой при этом играет роль второго затвора.

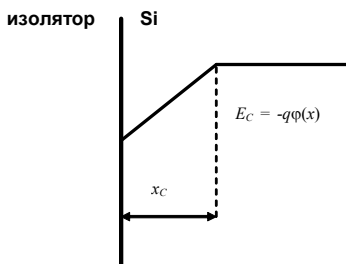


Рис. 3.12. Профиль потенциальной энергии при дельта-легировании

Существенным преимуществом дельта-легирования является уменьшение нежелательного влияния обратного смещения на подложке на пороговое напряжение (глава 4)

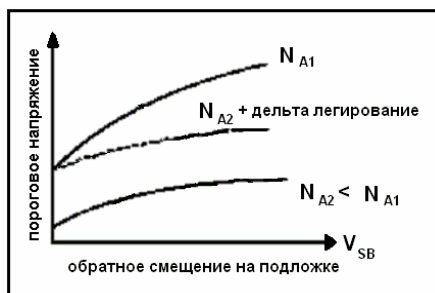


Рис. 3.13. Зависимость порогового напряжения от величины обратного смещения на подложке

Равномерное увеличение уровня легирования подложки увеличивает влияние подложки, а использование дельта-легирования подавляет этот эффект (рис. 3.13).

3.18. Заряженные ловушки на и вблизи границы раздела

Граничные ловушки (дефекты) расположены точно на границе раздела либо в окисле в пределах 1-2 нм от границы раздела. Эти

дефекты имеют разные зарядовые состояния и способны перезаряжаться, обмениваясь носителями (электронами и дырками) с кремниевой подложкой.

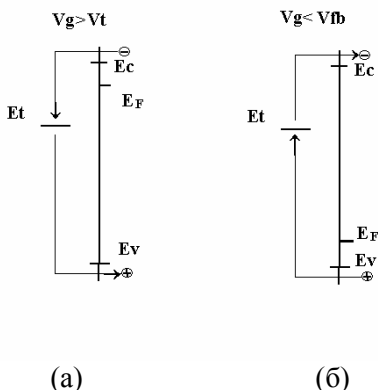


Рис. 3.14. Механизмы обмена носителей:
(а) заполнение, (б) опустошение

Эти перезаряжаемые дефекты (или ловушки) «чувствуют» положение уровня Ферми в кремнии. Если уровни ловушек E_t оказываются выше уровня Ферми – они заполняются электронами, если ниже – опустошаются (рис. 3.14).

Различают ловушки двух типов – акцепторные и донорные. Ловушки акцепторного типа в заполненном состоянии отрицательны ($0/-$), в незаполненном – нейтральны. Ловушки донорного типа положительно заряжены в пустом состоянии и нейтральны в заполненном ($+/0$). В любом случае при увеличении V_G уровень Ферми в кремнии опускается вниз, и ловушки начинают заполняться, т.е. они становятся более отрицательно заряженными (рис. 3.15).

Каждому потенциалу соответствует свой уровень Ферми на границе раздела и свое «равновесное» заполнение и соответствующая плотность условно положительного заряда $Q_t(\varphi_s) = q N_t(\varphi_s)$.

Для ловушек с малыми временами перезарядки быстро устанавливается равновесие с подложкой. Те ловушки, которые быстро обмениваются носителями с подложкой, называются «поверхностными состояниями» (interface traps, N_{it}). Те ловушки, которые не успевают обмениваться зарядом с подложкой за время измерения,

называются «фиксированным зарядом в окисле» (oxide traps, N_{ot}). Граница между ними условна и определяется временем развертки и температурой измерения.

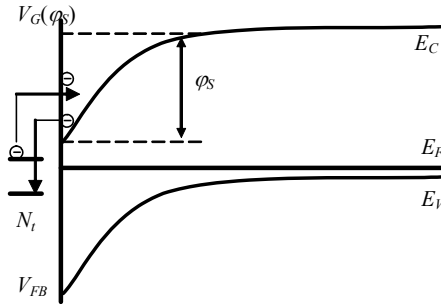


Рис. 3.15. Зонная диаграмма границы раздела Si-SiO₂

Когда имеется зависимость плотности заряда от потенциала, всегда можно ввести соответствующую емкость C_{it} :

$$C_{it} \equiv \frac{d}{d\varphi_S} (-qN_t(\varphi_S)). \quad (3.18.1)$$

В эту формулу входит выражение для заряда на всех ловушках (быстрых и медленных), но для конечных времен развертки t_s напряжения вклад в перезарядку дают только «поверхностные состояния» со временами перезарядки $\tau_r < t_s$. Именно поэтому эту величину называют емкостью поверхностных состояний. Она характеризует темп уменьшения положительного заряда с ростом поверхностного потенциала и имеет размерность удельной емкости $\text{Ф}/\text{см}^2$. С другой стороны, удельная емкость поверхностных состояний с точностью до размерного множителя есть просто энергетическая плотность поверхностных состояний D_{it} , т.е. имеет размерность $\text{см}^{-2} \text{эВ}^{-1}$. Легко убедиться, что эти величины связаны соотношением

$$C_{it} = q^2 D_{it}(\varphi_S). \quad (3.18.2)$$

Типичные емкости поверхностных состояний в современных транзисторах $\sim 10^{11} \text{ см}^{-2} \times \text{эВ}^{-2}$. При воздействии горячих носителей и ионизирующих излучений плотность поверхностных состояний может вырасти до значений $\sim 10^{12} \text{ см}^{-2} \text{эВ}^{-2}$.

3.19. Емкость инверсионного слоя

Аналогично можно ввести удельную емкость инверсионного слоя

$$C_{inv} \equiv q \frac{dn_S}{d\varphi_S}. \quad (3.19.1)$$

Используя зависимость $Q_S = q N_A L_D F(\varphi_S)$, полученную в п. 3.8, находим

$$C_{inv} = q \left(\frac{dn_S}{d\varphi_S} \right) = \frac{\varepsilon_S}{L_D} \frac{\exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T}\right)}{\left(\frac{\varphi_S}{\varphi_T} + \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T}\right)\right)^{1/2}}. \quad (3.19.2)$$

Отметим, что емкость инверсионного слоя экспоненциально мала в подпороговой области $V_G < V_T$, когда $\varphi_S < 2\varphi_S$.

Сравнивая выражения n_S и dn_S/φ_S , можно получить формулу для оценки емкости инверсионного слоя

$$C_{inv} \equiv \frac{q n_S}{2\varphi_T} \left(1 + \frac{N_A x_d}{n_S + N_A x_d} \right) = \begin{cases} \frac{q n_S}{2\varphi_T}, & V_G > V_T, \\ \frac{q n_S}{\varphi_T}, & V_G < V_T. \end{cases} \quad (3.19.3)$$

В надпороговом режиме ($V_G > V_T$) имеем следующее приближение для оценки удельной емкости инверсионного слоя:

$$C_{inv} \equiv \frac{C_O(V_G - V_T)}{2\varphi_T}. \quad (3.19.4)$$

Эти формулы справедливы формально только для невырожденного случая, но могут быть использованы для оценок и в вырожденных каналах.

3.20. Полная емкость МОП структуры

Полная дифференциальная емкость МОП структуры C_G по определению равна производной от плотности заряда на затворе по затворному напряжению V_G :

$$C_G \equiv q \left(\frac{dN_G}{dV_G} \right) = q \left(\frac{dN_G}{d\varphi_S} \right) / \left(\frac{dV_G}{d\varphi_S} \right). \quad (3.20.1)$$

Вспомогая соотношение (3.5.6), в котором мы пренебрегли падением потенциала в материале затвора,

$$V_G - \varphi_{MS} = \overbrace{\varphi_S}^{\text{полупроводник}} + \overbrace{q(n_S + N_A x_d - N_t)/C_O}^{V_{ox}}, \quad (3.20.2)$$

получаем

$$\begin{aligned} q \frac{dN_G}{d\varphi_S} &= q \left[\frac{dn_S}{d\varphi_S} + \frac{d(N_A x_d)}{d\varphi_S} + \frac{d}{d\varphi_S} (-N_t) \right] \equiv \\ &\equiv C_{inv} + C_D + C_{it}, \end{aligned} \quad (3.20.3)$$

$$\frac{dV_G}{d\varphi_S} = 1 + \frac{C_{inv} + C_D + C_{it}}{C_O} \equiv m. \quad (3.20.4)$$

Отсюда получаем формулу для полной емкости МОП структуры

$$\frac{1}{C_G} = \frac{1}{C_O} + \frac{1}{C_{inv} + C_D + C_{it}}. \quad (3.20.5)$$

Эта формула означает, что емкости инверсионного, обедненного слоя и поверхностных состояний соединены параллельно, и все они вместе последовательно соединены с емкостью окисла (рис. 3.16).

Эквивалентные схемы МОП структуры

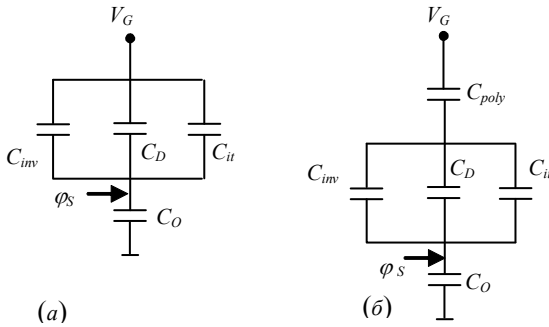


Рис. 3.16. Эквивалентная емкостная схема МОП структуры для режимов обеднения (а) и инверсии (б)

В режиме инверсии (надпороговом) емкость инверсионного слоя очень велика: $C_{inv} \gg C_D, C_{it}, C_O$, и поэтому емкость всей структуры приблизительно равна емкости окисла

$$C_G \cong C_O = \frac{\epsilon_i}{d_{ox}}. \quad (3.20.6)$$

Следует подчеркнуть, что это справедливо только для вольт-фарадных характеристик (ВФХ) МОП конденсаторов, измеряемых при относительно низких частотах, на которых успевает происходить образование инверсионного слоя. При высокочастотных измерениях в конденсаторах инверсионный слой за время цикла изменения малого сигнала на затворе не успевает сформироваться и составляющая емкости C_{inv} отсутствует (рис. 3.17).

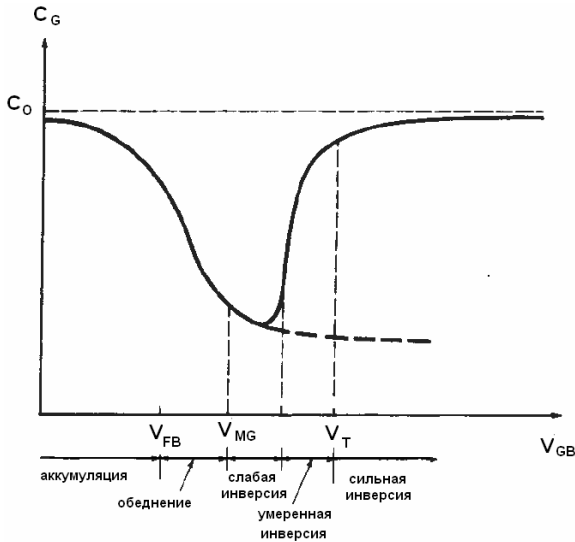


Рис. 3.17. Низкочастотная вольт-фарадная характеристика (ВФХ) МОП структуры с p -подложкой. Пунктирной линией показана высокочастотная ВФХ

ВФХ транзистора всегда имеет вид низкочастотной характеристики, поскольку из-за наличия электродов стока и истока образование инверсионного слоя происходит практически мгновенно (за время $\sim 10^{-11}$ с). В режиме обеднения емкость инверсионного слоя оказывается слишком малой по сравнению с емкостью обедненного слоя и поверхностных состояний, и поэтому, как и в инверсии, C_{inv} практически не влияет на вид вольт-фарадной характеристики МОП структуры. Используя теоретическое значение емкости плоских зон $C_{FB}^{-1} = C_O^{-1} + L_D/\epsilon_S$ и экспериментальную кривую (3.17), легко получить экспериментальное значение V_{FB} . Зная V_{FB} и контактную разность потенциалов, можно, используя (3.6.6), получить оценку для фиксированного заряда в окисле.

В режиме обогащения роль емкости инверсионного слоя основных носителей C_{inv} начинает играть емкость обогащенного слоя основных носителей $C_{acc} = |Q_{acc}|/2\phi_T$. Эта емкость очень быстро растет с увеличением отрицательного напряжения на затворе,

и поэтому, как и в инверсии, емкость всей структуры стремится к емкости окисла.

3.21. Учет влияния падения напряжения в затворе и инверсионном слое

Ранее мы предполагали, что на затворе есть заряд, но нет падения потенциала. Строго говоря, это справедливо только для металлических затворов, концентрация электронов в которых $\sim 10^{22} \text{ см}^{-3}$. В современной микроэлектронике в качестве затвора используют не металл, а сильнолегированный поликристаллический кремний с концентрацией легирования не более $(1...5) \times 10^{20} \text{ см}^{-3}$. Дополнительное падение потенциала в таком затворе может достигать значения десятков милливольт.

Падение потенциала в однородно-легированном поли-Si затворе V_{poly} оценивается по известной формуле (3.3.6) через толщину обедненной области в затворе x_{dG} и уровень легирования затвора N_{poly} :

$$V_{poly} = qN_{poly}x_{dG}^2 / 2\varepsilon_S. \quad (3.21.1)$$

Распределение падений напряжений (3.20.2) модифицируется следующим образом:

$$V_G = \varphi_{MS} + \varphi_S + \varphi_{inv} + V_{ox} + V_{poly}. \quad (3.21.2)$$

Тогда выражение для полной удельной емкости МОП структуры имеет вид

$$C_G \cong \left(\frac{1}{C_O} + \frac{1}{C_{poly}} + \frac{1}{C_{inv} + C_D + C_{it}} \right)^{-1}. \quad (3.21.3)$$

Учет падения потенциала в затворе приводит к появлению дополнительной емкости, включенной последовательно с емкостью окисла (рис. 3.16(б)). В режиме сильной инверсии (когда $C_{inv} \gg C_D, C_{it}$) эффективная емкость всей МОП структуры определяется формулой

$$C_G \cong \left(\frac{d_{ox}}{\varepsilon_i} + \frac{x_{dG}}{\varepsilon_S} + \frac{x_{inv}}{\varepsilon_S} \right)^{-1} \cong \frac{\varepsilon_i}{d_{ox} + (x_{dG} + x_{inv})(\varepsilon_i/\varepsilon_S)}. \quad (3.21.4)$$

Таким образом, падение потенциала в материале затвора эквивалентно тому, что изолятор становится толще; ITRS рекомендует

в этом случае говорить об эквивалентной электрической толщине окисла (equivalent oxide thickness electrical, EOT_{elec}). Если уровень легирования затвора $\sim 10^{20} \text{ см}^{-3}$, то толщина обедненного слоя в материале затвора оказывается $\sim 0.4..1.0 \text{ нм}$.

С учетом отношения диэлектрических проницаемостей эффективная толщина увеличивается на $0.15..0.3 \text{ нм}$. Это существенно, поскольку современные подзатворные изоляторы имеют толщину, сопоставимую с 1 нм . Уменьшение эффективной емкости подзатворного окисла крайне нежелательно и, поэтому, разрабатываются методы борьбы с эффектом затворного обеднения:

- использование металлического затвора, который сложнее поликремниевому с технологической точки зрения;
- увеличение степени легирования затвора, например, за счет дополнительной ионной имплантации, что также приводит к усложнению и удорожанию технологии.

3.22. Температурная зависимость порогового напряжения

Важно, что пороговое напряжение МОПТ зависит от температуры, что, главным образом, определяется температурной зависимостью положения уровня Ферми в объеме полупроводника φ_F и ширины запрещенной зоны кремния E_G . Другим источником температурной нестабильности порогового напряжения является наличие поверхностных состояний.

Контактная разность потенциалов для n^+ -поликремниевого затвора равна

$$\varphi_{MS} \cong -E_G / 2q - \varphi_F, \quad (3.22.1)$$

и поэтому формула для порогового напряжения принимает вид

$$V_T = -\frac{E_G}{2q} + \varphi_F + \frac{1}{C_O} (qN_A x (2\varphi_F) - qN_i (2\varphi_F)). \quad (3.22.2)$$

Тогда температурный коэффициент порогового напряжения определяется производной

$$\frac{dV_T}{dT} = -\frac{1}{2q} \frac{dE_G}{dT} + \frac{d\varphi_F}{dT} + \left(\frac{C_D(2\varphi_F) + q^2 D_{ii}(2\varphi_F)}{C_O} \right) \frac{d}{dT} (2\varphi_F). \quad (3.22.3)$$

Вспомогая определение для φ_F и собственной концентрации n_i

$$n_i = (N_C N_V)^{1/2} \exp(-E_G/2q\varphi_T), \quad (3.22.4)$$

получаем

$$\varphi_F = 2\varphi_T \ln\left(\frac{N_A}{n_i}\right) \cong E_G/2q - \varphi_T \ln \frac{(N_C N_V)^{1/2}}{N_A}, \quad (3.22.5)$$

$$\frac{d\varphi_F}{dT} = \frac{1}{2q} \frac{dE_G}{dT} - \frac{k_B}{q} \left(\ln\left(\frac{(N_C N_V)^{1/2}}{N_A}\right) + \frac{3}{2} \right). \quad (3.22.6)$$

Поскольку экспериментальное значение температурного коэффициента ширины запрещенной зоны кремния в области комнатных температур составляет $dE_G/dT \cong -0.27$ мэВ/К, то для уровней легирования $N_A \sim 10^{16}$ см⁻³ и малых плотностей поверхностных состояний ($D_{it} < 10^{11}$ см⁻² эВ⁻¹) типичные значения температурного коэффициента находятся в диапазоне -4.0...-0.5 мВ/К.

Важно подчеркнуть, что в соответствии с формулой (3.2.3) нестабильность плотности поверхностных состояний приводит к нестабильности температурного коэффициента порогового напряжения.

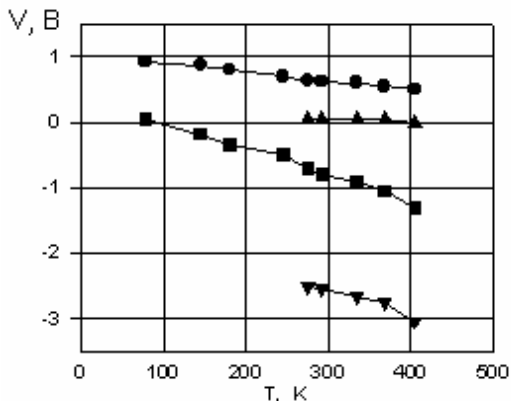


Рис. 3.18. Экспериментальные зависимости порогового напряжения V_T до (●) и после (■) облучения, а также напряжение середины зоны V_{MG} до (▲) и после (▼) облучения в n -МОП транзисторах как функции температуры измерения

Например, воздействие ионизирующего излучения приводит к накоплению поверхностных состояний и существенному усилению зависимости порогового напряжения от температуры (рис. 3.18).

4. Вольт-амперные характеристики МОПТ

4.1. Затворное напряжения как функция поверхностного потенциала в подпороговой области

В подпороговой области заряд в инверсионной области очень мал, а зависимости всех зарядов являются плавными функциями поверхностного потенциала

$$V_G - \varphi_{MS} \cong \varphi_S + q(N_A x_d - N_t)/C_O. \quad (4.1.1)$$

Это означает, что во всей подпороговой области изменение затворного напряжения приблизительно пропорционально изменению затворного потенциала

$$\Delta V_G \cong \left(\frac{dV_G}{d\varphi_S} \right) \Delta \varphi_S \equiv m \Delta \varphi_S, \quad (4.1.2)$$

где $\Delta V_G = V_{G1}(\varphi_{S1}) - V_{G2}(\varphi_{S2})$, $\Delta \varphi = \varphi_{S1} - \varphi_{S2}$.

Безразмерный коэффициент пропорциональности определяется емкостями поверхностных состояний C_{it} , обедненной области C_D и подзатворного окисла C_O :

$$\frac{dV_G}{d\varphi_S} \equiv m = 1 + \frac{C_D + C_{it}}{C_O}. \quad (4.1.3)$$

Вспоминая, что напряжению середины зоны V_{MG} и пороговому напряжению V_T соответствуют значения поверхностных потенциалов φ_F и $2\varphi_F$, можно записать приближенные равенства:

$$V_G - V_T = m(\varphi_S - 2\varphi_F), \quad V_T - V_{MG} = m\varphi_F. \quad (4.1.4)$$

Используя (4.1.4) и (3.9.3), получаем плотность заряда в инверсионном слое как функцию затворного напряжения

$$Q_{inv} \cong C_D \varphi_T \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T}\right) \cong C_D \varphi_T \exp\left(\frac{V_G - V_T}{m\varphi_T}\right). \quad (4.1.5)$$

Это выражение справедливо только в подпороговой области, когда $V_G < V_T$.

4.2. Плотность носителей в канале как функция затворного напряжения в форме интерполяции (BSIM3)

Как было показано выше, мы имеем хорошее описание зависимости плотности носителей в канале от поверхностного потенциала φ_S . На практике непосредственно задается не внутренний поверхностный потенциал, а напряжение на внешних электродах, в частности, на затворе. Формально, рассматривая нелинейную зависимость $V_G = f(\varphi_S)$ (см. (3.5.6)) как трансцендентное уравнение, мы можем численно решить его для каждого потенциала и получить искомую зависимость $\varphi_S = f(V_G)$. Другой способ состоит в составлении таблицы $\{V_G(\varphi_S), n_S(\varphi_S)\}$. Перебираем с нужным шагом значения поверхностного потенциала, получая тем самым табличное представление для $n_S(V_G)$.

Эти подходы, как правило, не используются, поскольку требуют чрезмерных затрат по времени счета и не соответствуют духу «компактного моделирования». Более эффективным, с точки зрения экономии компьютерного времени, является использование интерполяционной функции (выбор Беркли).

В разделе (3.9) было получено

$$qn_S(V_G) \cong \begin{cases} qN_A L_D \exp\left(\frac{\varphi_S(V_G) - 2\varphi_F}{2\varphi_T}\right) \cong C_O(V_G - V_T), & \varphi_S > 2\varphi_F, \\ C_D \varphi_T \exp\left(\frac{V_G - V_T}{m\varphi_T}\right), & 2\varphi_F > \varphi_S > \varphi_F. \end{cases} \quad (4.2.1)$$

Основываясь на этих двух предельных случаях, была предложена общая интерполяционная формула для плотности электронов (дырок) в канале Q_{inv} (Кл/см²) как функции затворного напряжения (BSIM3)

$$Q_{inv} = \frac{2 C_O m \varphi_T \ln \left[1 + \exp\left(\frac{V_{GS} - V_T}{2 m \varphi_T}\right) \right]}{1 + 2 \frac{C_O}{C_D} m \exp\left(-\frac{V_{GS} - V_T - V_{off}}{2 m \varphi_T}\right)}. \quad (4.2.2)$$

В надпороговых и подпороговых режимах работы МОПТ экспериментально наблюдается линейная и экспоненциальная зависимость плотности носителей в канале от затворного напряжения:

$$Q_{inv} \cong \begin{cases} C_o(V_{GS} - V_T), & V_G > V_T, \\ C_D \varphi_T \exp\left(\frac{V_{GS} - V_T - V_{off}}{m \varphi_T}\right), & V_G < V_T. \end{cases} \quad (4.2.3)$$

Здесь для лучшего согласования расчетного и экспериментального тока в BSIM3 введен не имеющий физического смысла подгочный коэффициент V_{off} (SPICE параметр VOFF).

4.3. Подпороговый размах напряжения (subthreshold swing)

Подпороговый ток МОПТ практически не зависит от напряжения на стоке, поскольку почти целиком состоит из диффузионной компоненты $I_D \sim D_n n_s / L$. Концентрация носителей в канале МОПТ в подпороговой области экспоненциально зависит от поверхностного потенциала и напряжения на затворе (см. (3.9.3)), что схематично отражено на рис. 4.1.

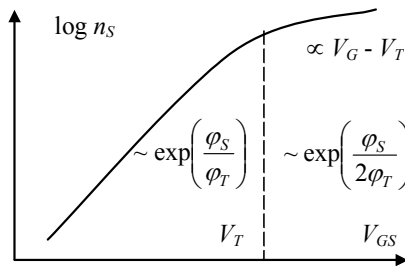


Рис. 4.1. Логарифмическая зависимость плотности носителей от напряжения на затворе для подпороговой и надпороговой области

Экспериментальная зависимость логарифма тока от напряжения на затворе в подпороговой области близка к прямой линии (рис. 4.2).

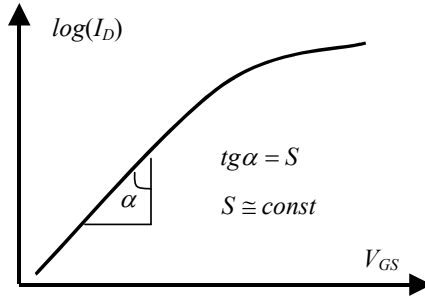


Рис. 4.2. Зависимость подпорогового тока МОПТ как функция затворного напряжения

При этом зависимость диффузионного тока МОПТ от затворного напряжения в подпороговой области определяется зависимостью плотности носителей в канале, и зависимость подпорогового тока от затворного напряжения определяется соответствующей зависимостью плотности носителей в канале

$$I_D \propto n_S \propto \exp\left(\frac{\varphi_S}{\varphi_T}\right). \quad (4.3.1)$$

Для характеристики наклона логарифма тока от напряжения на затворе используется логарифмический размах напряжения на декаду (порядок) тока (S-фактор)

$$S \equiv \frac{dV_G}{d(\log I_D)} = \ln 10 \frac{dV_G}{d(\ln n_S)} \cong \ln 10 n_S \frac{dV_G}{dn_S} = \ln 10 \left(\frac{n_S}{dn_S/d\varphi_S} \right) \left(\frac{dV_G}{d\varphi_S} \right). \quad (4.3.2)$$

Учитывая соотношение (3.9.3), в подпороговой области имеем

$$\frac{n_S}{dn_S/d\varphi_S} \cong 1, \quad (4.3.3)$$

и, вспоминая обозначение (4.1.3) для фактора неидеальности m (ideality factor), имеющего смысл безразмерного отношения скоростей изменения затворного напряжения и поверхностного потенциала, получаем выражение для размаха напряжения на декаду тока

$$S = \ln 10 \varphi_T \left[1 + \frac{C_D + C_{it}}{C_O} \right] \cong m \ln 10 \varphi_T. \quad (4.3.4)$$

Эта величина характеризует возможность быстрого открывания или закрывания транзистора, а также определяет динамическое энергопотребление. Поэтому, чем меньше подпороговый размах, тем лучше. Минимально возможное значение фактора неидеальности $m \sim 1$, что соответствует минимально возможному (идеальному) значению размаха

$$S_{\min} = \ln 10 \varphi_T \cong 60 \text{ мВ/декаду}. \quad (4.3.5)$$

Отметим, что это значение определяется фундаментальным механизмом активационного переноса носителей из истока в сток через барьер канала. Для того чтобы уменьшить S-фактор, необходимы приборы на других физических принципах, например, туннелирования через барьер переменной высоты.

В коммерческих МОПТ обычно выполняется условие $C_{it} \ll C_D$, и фактор неидеальности находится в диапазоне $m \sim 1.1 \dots 1.6$, что соответствует $S \sim 70 \dots 100$ мВ/декаду. При воздействии внешних и внутренних неравновесных факторов (ионизирующее облучение и горячие носители канала) фактор неидеальности может существенно вырасти – $m \sim 1.5 \dots 2.5$ (90..150 мВ/декаду) – за счет накопления радиационно-индуцированных поверхностных состояний ($C_{it} \geq C_D$).

4.4. Статические подпороговые токи утечки

Нежелательный ток между стоком и истоком в закрытом МОПТ I_{OFF} (подпороговый ток утечки) является одной из главных проблем наноэлектронных приборов. Используя результаты пунктов 4.1 и 4.3, получаем для подпороговых токов n -МОПТ выражение

$$I_D(V_G) = \text{const} \times \exp\left(\frac{\varphi_S - 2\varphi_F}{\varphi_T}\right) = I_T \exp\left(\frac{V_G - V_T}{m\varphi_T}\right), \quad (4.4.1)$$

где I_T – ток стока при напряжении, равном пороговому напряжению.

Подпороговые токи утечки – это токи, протекающие между стоком и истоком при нулевом смещении на затворе

$$I_{OFF} = I_T \exp\left(-\frac{V_T}{m\varphi_T}\right) = I_T 10^{-\frac{V_T}{S}}. \quad (4.4.2)$$

В современных интегральных транзисторах токи утечки лежат в пределах $10^{-11} \dots 10^{-8}$ А/мкм.

С точки зрения принципов скейлинга, пороговое напряжение должно уменьшаться в той же мере, что и напряжение питания. Для достижения хороших функциональных характеристик по быстродействию напряжение питания V_{DD} должно в 3-4 раза быть больше, чем пороговое напряжение V_T .

Здесь мы сталкиваемся с двумя противоречивыми следствиями снижения порогового напряжения. С одной стороны, низкое значение порога способствует увеличению рабочего тока $I_{DSAT} \sim \sim (V_G - V_T)^a$ ($a \sim 1.0 \dots 1.5$) и быстродействия. С другой, уменьшение V_T приводит к увеличению статических токов утечки. По этой причине пороговое напряжение нельзя делать очень маленьким. Для обеспечения минимально необходимой величины отношения токов транзистора в открытом и закрытом транзисторе (динамического диапазона) в 3...4 декады

$$\frac{I_{ON}}{I_{OFF}} \geq 10^3 \quad (4.4.3)$$

требуется пороговое напряжение, по меньшей мере, 0.2...0.3В.

4.5. Влияние обратного смещения на подложке

В реальных схемах часто только исток является заземленным контактом, в то время как подложка остается незаземленной. При этом транзистор может работать в режиме, когда переход исток-подложка оказывается смещенным, чаще всего, в обратном направлении. Для подложки p -типа это соответствует $V_{SB} = V_S - V_B > 0$ (т.е. на p -подложку подается отрицательное относительно истока напряжение). Если между стоком и истоком нет приложенного напряжения ($V_D = V_S$), то потенциал истока равен потенциалу канала ($V_C = V_S$).

Введем электронный химический потенциал $\zeta_n = E_C - E_F$, определяющий локальную концентрацию электронов $n = N_C e^{-\zeta_n / \varphi_T}$. Если $V_{SB} = 0$, то уровень Ферми одинаков по всей толщине структуры, и электрохимический потенциал не зависит от поперечной координаты (рис. 4.3):

$$\begin{aligned} \varphi_S(V_{SB} = 0) + \zeta_n(V_{SB} = 0) = \\ = E_G/2q + \varphi_F = \text{const}(x). \end{aligned} \quad (4.5.1)$$

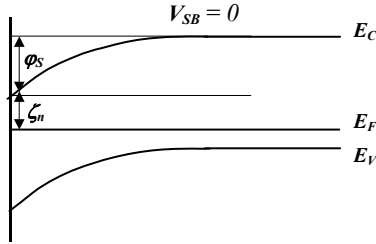


Рис. 4.3. Зонная диаграмма при нулевом смещении ($V_{SB} = 0$) на подложку относительно истока

После приложения обратного смещения к подложке $V_{SB} = V_S - V_B > 0$ происходит расщепление уровня Ферми на два квазиуровня – электронный в канале E_{Fn} и дырочный E_{Fp} в подложке (рис. 4.4), причем в канале имеем $np = n_i^2 e^{-V_{SB}/\varphi_T} \ll n_i^2$ и $E_{Fp} - E_{Fn} = qV_{SB}$.

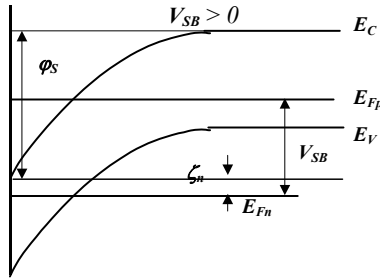


Рис. 4.4. Зонная диаграмма при обратном смещении ($V_{SB} > 0$) на подложку относительно истока

Все это похоже на слабелегированную сторону обратносмещенного pn -перехода. При таком смещении из истока в подложку течет небольшой ток (для объемных МОПТ он не существен, но важен для КНИ технологий, когда подложка изолирована)

$$\varphi_S(V_{SB}) + \zeta_n(V_{SB}) = E_G/2q + \varphi_F + V_{SB}. \quad (4.5.2)$$

4.6. Пороговое напряжение по отношению напряжения между затвором и подложкой

Химический потенциал в надпороговом режиме определяется плотностью носителей в канале и не зависит от обратного смещения на подложке

$$\zeta_n(V_{SB}) \cong \zeta_n(V_{SB} = 0) \cong E_G/2q - \varphi_F. \quad (4.6.1)$$

Используя (4.5.2), получаем, что соответствующий поверхностный потенциал, напротив, непосредственно зависит от V_{SB} :

$$\varphi_{th}(V_{SB}) = \varphi_S(V_{SB} = 0) + V_{SB} \cong 2\varphi_F + V_{SB}. \quad (4.6.2)$$

Пороговое напряжение на затворе зависит, естественно, от того, относительно чего мы прикладываем напряжение на затвор. Для того чтобы определить пороговое напряжение относительно *истока*, необходимо сначала определить пороговое напряжение относительно *подложки* (как если бы она была заземлена)

$$V_{T,GB} = V_G - V_B = \varphi_{MS} + 2\varphi_F + V_{SB} + \frac{qN_{A}x_d(2\varphi_F + V_{SB})}{C_O}. \quad (4.6.3)$$

Учитывая связь между электрическими смещениями на разных электродах

$$V_{GS} = V_G - V_B - (V_S - V_B) = V_{GB} - V_{SB}, \quad (4.6.4)$$

получаем, что пороговое напряжение относительно *истока* имеет вид

$$V_T \equiv V_{T,GS} = \varphi_{MS} + 2\varphi_F + \frac{qN_{A}x_d(2\varphi_F + V_{SB})}{C_O}. \quad (4.6.5)$$

В дальнейшем мы будем подразумевать под пороговым напряжением МОПТ только пороговое напряжение, определенное относительно истока.

4.7. Зависимость порогового напряжения от обратного смещения на подложке

Как видно из формулы (4.6.5), обратное смещение на подложке увеличивает эффективный заряд обедненной области и соответственно увеличивает по абсолютной величине пороговое напряжение. Используя соотношения типа (4.6.5), получаем, что зависимо-

сти порогового напряжения для n -канального (знак «+») и p -канального (знак «-») МОПТ имеют вид

$$V_T(V_{SB}) = V_{T0} \pm \frac{\sqrt{2qN_B\epsilon_S}}{C_O} \left(\sqrt{2\varphi_F + V_{SB}} - \sqrt{2\varphi_F} \right) \equiv \\ \equiv V_{T0} \pm \gamma \left(\sqrt{2\varphi_F + V_{SB}} - \sqrt{2\varphi_F} \right), \quad (4.7.1)$$

где $V_{T0} = V_T(V_{SB} = 0)$, $\gamma \equiv \sqrt{2qN_B\epsilon_S}/C_O = \text{ГАММА}$ – параметр, традиционно используемый для характеристики степени легирования подложки.

Как мы видим (рис. 4.5), при приложении обратного смещения к переходу исток-подложка пороговое напряжение n -МОПТ возрастает, а в p -канальном МОПТ пороговое напряжение уходит в отрицательную сторону.

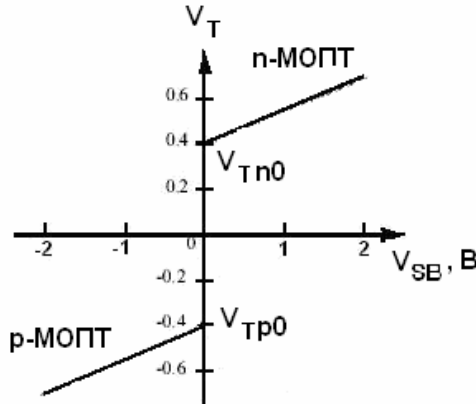


Рис. 4.5. Зависимости порогового напряжения n - и p -МОП транзисторов от обратного смещения на подложке относительно истока

Наклон зависимости порогового напряжения от обратного смещения определяется формулой

$$\frac{\partial V_T}{\partial V_{SB}} = \frac{\gamma}{2\sqrt{2\varphi_F + V_{SB}}} = \frac{C_D(V_{SB})}{C_O} \cong \frac{\epsilon_S}{\epsilon_i} \frac{d_{ox}}{x_d} \cong 3 \frac{d_{ox}}{x_d}. \quad (4.7.3)$$

Как видно из формулы (4.7.3), влияние подложки возрастает с увеличением толщины подзатворного окисла и увеличением степени легирования подложки (рис. 4.6)

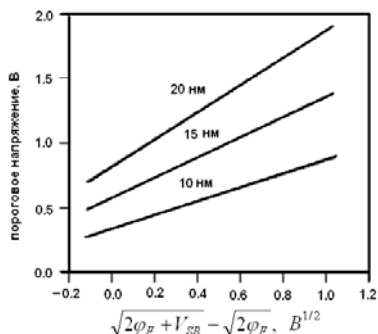


Рис. 4.6. Влияние напряжения на подложке на пороговое напряжение МОПТ с разной толщиной подзатворного окисла

4.8. Важность эффекта подложки в реальных схемах

В идеале, у каждого транзистора на чипе исток должен быть закорочен с подложкой. На практике это условие не всегда реализуется, поскольку оно ведет к усложнению конструкции и удорожанию схемы.

Например, рассмотрим ячейку И-НЕ с двумя типами заземления истоков (рис. 4.7). При локальном заземлении истоков каждого из *n*-канальных транзисторов обратное смещение истокового перехода всегда равно нулю и пороговое напряжение одинаково для всех *n*-МОПТ. При заземлении истока только одного нижнего *n*-МОП транзистора потенциал истока верхнего *n*-МОПТ больше нуля, и подложка этого транзистора с нулевым потенциалом оказывается обратносмещенной относительно истока. Соответственно, эффективное пороговое напряжение верхнего транзистора становится больше эквивалентного ему порогового напряжения нижнего транзистора.

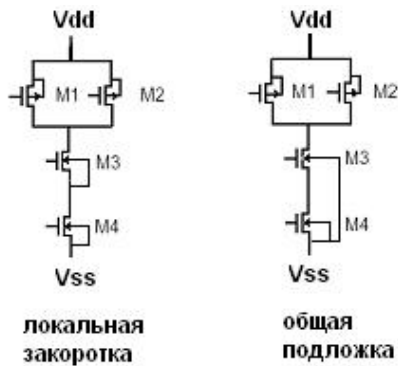


Рис. 4.7. Два типа заземления подложки в логическом венти́ле И-НЕ

Рассогласование пороговых напряжений разных транзисторов приводит к уменьшению рабочего тока и замедлению быстродействия и другим нежелательным эффектам. По этой причине практически никогда не изготавливают логические вентиля с тремя входами.

4.9. Напряжение между стоком и истоком

Ранее был рассмотрен случай, когда исток и сток находятся под одинаковым потенциалом, и электрохимический потенциал не меняется вдоль канала ($V_S = V_D$). Теперь рассмотрим случай, когда исток и подложка заземлены ($V_{SB} = 0$), а на сток подается напряжение V_{DS} (для определенности будем всегда рассматривать n -МОПТ с p -подложкой с длиной канала L). Зонная диаграмма, показывающая энергетические поверхности краев зон в двух направлениях, представлена на рис. 4.8.

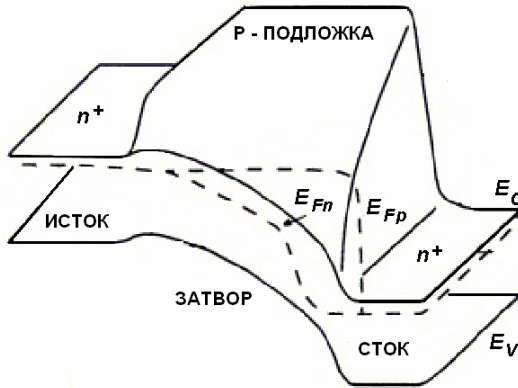


Рис. 4.8. Зонные диаграммы, показывающие поверхности потенциальной энергии МОП транзистора при напряжении $V_{DS} > 0$

В этом случае переход стока оказывается обратносмещенным относительно подложки:

$$V_{DS} = V_D - V_S = V_{DB} - V_{SB} = V_{DB}(y = L). \quad (4.9.1)$$

С другой стороны, это же значение является разностью электрохимических потенциалов между стоком и истоком. Соответственно, значение электрохимического потенциала канала в данной точке представляет собой локальное значение обратного смещения канала относительно подложки в этой точке

$$V_C(y) = V_C(y) - V_S = V_{DB} - V_{SB} = V_{DB}(y), \quad (4.9.2)$$

которое меняется от нуля в истоке до V_{DS} в стоке. Иногда распределение электрохимического потенциала в канале называют просто распределением потенциала. При этом недопустимо отождествлять

этот потенциал с электростатическим потенциалом, поскольку это может привести к ошибкам и непониманию.

4.10. Приближение плавного канала

На рис. 4.9 показано схематическое сечение МОП транзистора, в котором ток течет между истоком (S) и стоком (D) в канале, направленном по оси y .

Для анализа работы МОП транзистора необходимо использовать существенные приближения, главным из которых со времен ранних работ Шокли является приближение плавного канала (gradual channel approximation).

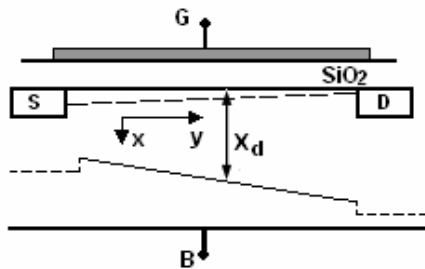


Рис. 4.9. Сечение МОП транзистора

Приближение плавного канала состоит в том, что в каждой точке канала можно записать «плоское» уравнение электронейтральности для локальных значений поверхностных концентраций (ср. 3.5.2)

$$N_G(y) = n_S(y) + N_A x_d(y). \quad (4.10.1)$$

Это приближение справедливо при выполнении формального неравенства

$$\frac{\partial E_x}{\partial x} \gg \frac{\partial E_y}{\partial y}, \quad (4.10.2)$$

которое, в свою очередь, выполняется, когда электрические смещения и поля по направлению оси y существенно меньше, чем в направлении оси x (см. рис. 4.9).

Отметим, что потенциал затвора не зависит от V_{DS} и не меняется по всей его длине (в отличие от поверхностной плотности заряда на затворе)

$$V_{GB} = \varphi_{MS} + 2\varphi_F + V_C(y) + \frac{q}{C_O} (n_S(y) + N_A x_d (2\varphi_F + V_C(y))) = \text{const}(y). \quad (4.10.3)$$

Анализ (4.10.3) показывает, что по мере увеличения y возрастает локальный потенциал канала V_C и соответственно поверхностная плотность заряда обедненного слоя. Это означает, что локальная плотность носителей в канале $n_S(y)$ уменьшается по мере приближения к стоку.

4.11. Плотность электронов вдоль канала при $V_{DS} > 0$

Приближение плавного канала позволяет записать плотность электронов в канале в каждой его точке в виде

$$\begin{aligned} qn_S(y) &= C_O \left[V_{GS} - \underbrace{\left(V_T(V_C(y)) + V_C(y) \right)}_{V_{T,GB}} \right] \cong \\ &\cong C_O \left[V_{GS} - \left(V_{T0} + \left\langle \frac{\partial V_T}{\partial V_{SB}} \right\rangle V_C(y) \right) - V_C(y) \right] = \\ &= C_O \left[V_{GS} - V_{T0} - \overbrace{\left(1 + \left\langle \frac{\partial V_T}{\partial V_{SB}} \right\rangle \right)}^n V_C(y) \right], \quad (4.11.1) \end{aligned}$$

где введен n – безразмерный коэффициент влияния подложки.

Он определяется тем, насколько пороговое напряжение чувствительно относительно обратного смещения на подложку:

$$\left\langle \frac{\partial V_T}{\partial V_{SB}} \right\rangle = \left\langle \frac{C_D}{C_O} \right\rangle = \frac{\varepsilon_S}{\varepsilon_i} \frac{d_{ox}}{x_d (2\varphi_F + \langle V_{SB} \rangle)} \cong \frac{\gamma}{2\sqrt{2\varphi_F + 0.5V_{DS}}} < 1. \quad (4.11.2)$$

Отметим, что коэффициент влияния подложки n по смыслу и значению очень близок введенному в п. 4.1 безразмерному коэффициенту m .

Считая коэффициент влияния подложки постоянной величиной ($1 < n < 2$), можно приближенно рассчитать распределения плотности носителей вдоль канала

$$qn_S(y) = C_O(V_{GS} - V_{T0} - nV_C(y)). \quad (4.11.3)$$

Плотность электронов у истока в этом приближении постоянна: $qn_S(0) = C_O(V_{GS} - V_{T0})$, а вблизи стока

$$qn_S(L) = C_O(V_{GS} - V_{T0} - nV_{DS}). \quad (4.11.4)$$

При некотором значении $V_{DS} = V_{DSAT} = (V_{GS} - V_{T0})/n$, которое называется напряжением насыщения, плотность электронов в канале вблизи стока оказывается формально равной нулю. Дополнительное напряжение на стоке $V_{DS} - V_{DSAT} > 0$ падает на pn -переходе стока и практически перестает влиять на распределение потенциала в канале. Это означает, что зависимость тока в канале от напряжения V_{DS} «насыщается», т. е. практически перестает расти.

4.12. Простейшая модель ВАХ МОПТ

Пренебрегая диффузионной составляющей тока, полный ток в канале шириной Z можно записать в виде

$$I_D = Zq n_S(y) \mu_n E(y) = Z \mu_n C_O (V_{GS} - V_{T0} - nV_C(y)) \frac{dV_C(y)}{dy}, \quad (4.12.1)$$

где μ_n – подвижность носителей (электронов) в канале. Последнюю формулу можно переписать в форме

$$I_D dy = Z \mu_n q n_S(V_C(y)) dV_C(y). \quad (4.12.2)$$

С учетом (4.11.3), обе стороны уравнения (4.12.2) можно проинтегрировать с граничными условиями

$$V_C(0) = 0; \quad V_C(L) = V_{DS}. \quad (4.12.3)$$

Для наглядности это можно сделать графически, заметив, что ток через канал пропорционален площади трапеции в зависимости $n_S(u_C(y))$ (рис. 4.10).

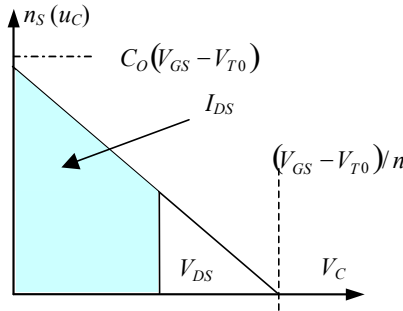


Рис. 4.10. Зависимость плотности электронов в канале как функция электрохимического потенциала (см. 4.11.4). Ток канала пропорционален заштрихованной области

Из соотношения (4.11.4) и рис. 4.10 очевидно, что вычисление тока стока таким способом невозможно при больших V_{DS} . Действительно, формальный интеграл от (4.12.2) дает выражение

$$I_D = (Z/L)\mu_n C_o \left(V_{GS} - V_{T0} - \frac{n}{2} V_{DS} \right) V_{DS}, \quad (4.12.4)$$

в котором ток $V_{DS} > (V_G - V_{T0})/n$ начинает уменьшаться с ростом V_{DS} (рис. 4.11).

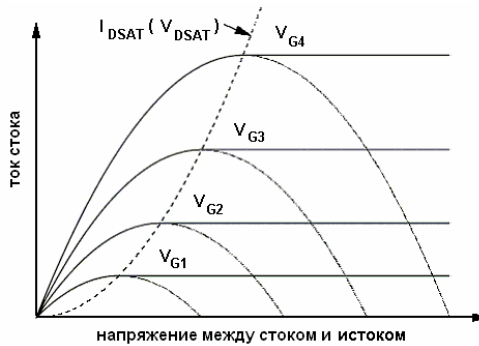


Рис. 4.11. Схематические выходные ВАХ МОПТ для разных напряжений на затворе

В качестве тока насыщения волевым образом используется максимальное значение (4.12.4). То есть ток насыщения транзистора I_{DSAT} вводится с помощью искусственной процедуры

$$\frac{dI_D}{dV_{DS}} = 0 \Rightarrow I_{D_{\max}} = I_{D_{SAT}} = (V_{GS} - V_{T0})/n. \quad (4.12.5)$$

Таким образом, в этом простейшем приближении вольт-амперная характеристика (ВАХ) МОПТ выражается кусочно-непрерывной функцией

$$I_D = \begin{cases} \beta \left(V_{GS} - V_{T0} - \frac{n}{2} V_{DS} \right) V_{DS}, & V_{DS} < V_{D_{SAT}} \\ \beta \frac{(V_{GS} - V_{T0})^2}{2n}, & V_{DS} > V_{D_{SAT}}. \end{cases} \quad (4.12.6)$$

где $\beta \equiv (Z/L)\mu_n C_O$ – удельная крутизна МОПТ.

Это базисное уравнение, полученное еще в начале 60-х гг. XX в., до сих пор остается основой практически для всех моделей, используемых в модифицированном виде, в том числе и для проектирования наноэлектронных МОПТ с длиной канала < 100 нм (BSIM4). Причина этого состоит в том, что ВАХ транзисторов, в том числе самых современных, имеют довольно простой вид, и для их описания достаточно (особенно при использовании множества подгоночных параметров) использования простых компактных моделей, не требующих громоздких расчетов.

Тем не менее, с физической точки зрения, описанная простейшая модель является совершенно неудовлетворительной. Причина нефизического поведения модели в некоторых режимах работы кроется в самом виде изначального уравнения (4.12.1), в котором игнорируется диффузионный ток. Отметим, что диффузионный ток присутствует в канале уже в рамках самой же простейшей модели хотя бы из-за наличия ненулевого градиента концентрации вдоль канала в соотношении (4.11.3). Другим важнейшим недостатком описанного подхода является отсутствие явного решения уравнения непрерывности в канале. Последовательной процедурой является решение уравнения непрерывности для плотности тока в канале транзистора с учетом диффузионно-дрейфового характера транспорта носителей (глава 7).

NB. Интересно, что в рамках простейшей модели процедуру интегрирования уравнения (4.12.1) можно провести по-другому. Например, переписать соотношение (4.12.1) в виде

$$I_D dR(y) = I_D \frac{dy}{Z \mu_n q n_s (V_C(y))} = dV_C(y), \quad (4.12.7)$$

и затем формально проинтегрировать его, считая I_D константой.

С физической точки зрения, это была бы более последовательная процедура, по сравнению с интегрированием (4.12.2), так как она соответствовала бы естественному суммированию последовательных сопротивлений элементов канала dR с шириной Z и длиной dy . Но такой подход не используется, поскольку не приводит к простым и замкнутым аналитическим выражениям для ВАХ.

4.13. Насыщение скорости носителей в канале

При выводе выражения для ВАХ МОПТ мы считали, что подвижность носителей в канале является константой. На деле подвижность определяется многими параметрами и, в частности, зависит от величины (по y) электрического поля вдоль канала E .

С ростом тянущего электрического поля рассеивание носителей в канале усиливается, и дрейфовая скорости насыщается на своем максимальном уровне порядка тепловой скорости носителей в канале (рис. 4.12)

$$v_{th} \sim 10^7 \text{ см/с.} \quad (4.13.1)$$

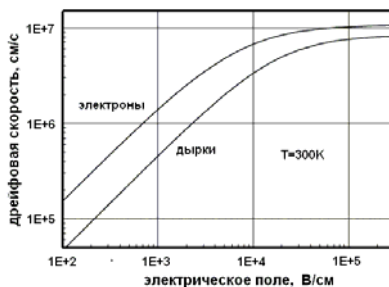


Рис. 4.12. Зависимости дрейфовой скорости электронов и дырок от электрического поля в кремнии при комнатной температуре

Основным механизмом насыщения дрейфовой скорости при этом, как и в объемном случае, является интенсивное рассеяние на оптических фононах. Тем не менее, в КНИ транзисторах с ультратонким телом (глава 8) существенную роль начинает играть рассеяние на границах раздела (см. главу 1. и рис. 4.13).

Хорошая количественная физическая модель зависимости подвижности от тянущего поля в каналах отсутствует, и поэтому широко используются эмпирические модели.

Например, зависимость дрейфовой скорости носителей от тянущего поля $v_{dr}(E)$ приближенно аппроксимируется формулой

$$v_{dr}(E) = \frac{\mu_0 E}{1 + E/E_{SAT}}, \quad (4.13.2)$$

где μ_0 (SPICE параметр UO) – подвижность носителей в канале транзистора в слабых тянущих полях; v_{SAT} – скорость насыщения носителей в канале транзистора (SPICE параметр VSAT).

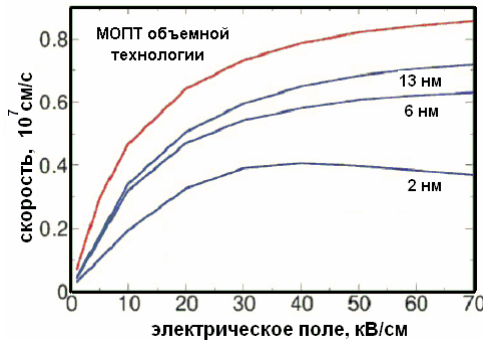


Рис. 4.13. Зависимости дрейфовой скорости носителей в каналах МОПТ объемной технологии (верхняя кривая) и МОПТ по технологии «кремний-на-изоляторе» с разной толщиной кремниевого тела

Типичные значения скоростей насыщения в каналах МОПТ

$$v_{SAT} = \begin{cases} 8 \times 10^6 \text{ см/с, (электроны),} \\ 6 \times 10^6 \text{ см/с, (дырки).} \end{cases} \quad (4.13.3)$$

Вместо скорости насыщения часто используют значение электрического поля, соответствующего насыщению E_{SAT} , иногда определяемого как

$$E_{SAT} = \frac{v_{SAT}}{\mu_0}. \quad (4.13.4)$$

Практика показывает, что использование соотношения (4.3.12) совместно с формулой (4.3.14) (условно говоря, модель 1) соответствует относительно медленному нарастанию дрейфовой скорости и приводит к ее недооценке при промежуточных значениях тянущих электрических полей (рис. 4.14).

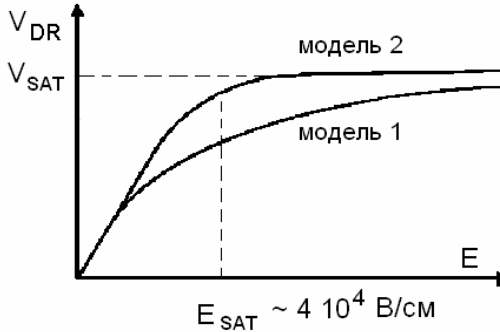


Рис. 4.14. Два варианта модели дрейфовой скорости носителей в канале как функции тянущего поля вдоль канала

Поэтому на практике чаще всего используется, условно говоря, модель 2, в которой параметр электрического поля насыщения определяется как

$$E_{SAT} = \frac{2v_{SAT}}{\mu_0} \cong (4 \dots 8) \times 10^4 \text{ В/см}, \quad (4.13.5)$$

а зависимость дрейфовой скорости от тянущего поля представляется кусочным выражением

$$v_{dr}(E) = \begin{cases} \frac{\mu_0 E}{1 + E/E_{SAT}}, & E < E_{SAT}, \\ v_{SAT}, & E \geq E_{SAT}. \end{cases} \quad (4.13.6)$$

В отличие от модели 1, где $v(E_{SAT}) = v_{SAT}/2$, модель 2 дает $v(E_{SAT}) = v_{SAT}$, и соответствует более быстрому насыщению скорости, что лучше согласуется с экспериментальными данными.

4.14. Механизмы насыщения тока канала

Таким образом, насыщение тока в канале МОПТ может происходить за счет двух физически различных механизмов.

А. Электростатическое запираение канала

Насыщение тока в данной модели достигается за счет электростатического запираения канала, которое начинается, когда плотность носителей в канале равна нулю (см. 4.11.4):

$$n_s(y=L) = 0 \Leftrightarrow V_{DS} = (V_{GS} - V_T)/n. \quad (4.14.1)$$

При больших напряжениях V_{DS} электрическое поле у истока перестает расти из-за эффекта экранирования приложенного внешнего напряжения полем объемного заряда обедненного слоя и поверхностного заряда на затворе.

Ток насыщения в канале (4.12.6) в этом случае можно представить в виде произведения средней плотности заряда в канале на максимальную дрейфовую скорость, зависящую в этом случае только от затворного напряжения,

$$I_{DSAT} = \frac{Z}{L} \mu_n C_O \frac{(V_{GS} - V_T)^2}{2n} = Z \underbrace{\frac{C_O (V_{GS} - V_T)}{2}}_{q \langle n_s \rangle} \underbrace{\frac{\mu_n (V_{GS} - V_T)}{nL}}_{v_{dr, max}}. \quad (4.14.2)$$

Б. Насыщение дрейфовой скорости

Согласно экспериментальным данным, параметризованным в эмпирической формуле (4.13.2), при достаточно сильных полях в канале дрейфовая скорость носителей насыщается вне зависимости от затворного напряжения. Это означает, что при выполнении неравенства

$$\frac{\mu_n (V_{GS} - V_T)}{2nL} > v_{th}$$

ток насыщения в канале можно представить в виде

$$I_{DSAT} = Z C_O (V_{GS} - V_T) v_{th}. \quad (4.14.3)$$

С физической точки зрения это означает, что все носители в канале движутся с максимальной скоростью, близкой к их тепловой скорости.

Сделаем оценки для современных МОП транзисторов. Считая, что $L = 0.1$ мкм; $V_{DD} - V_T = 1.5$ В; $\mu_n = 500$ см²/(В×с), $n = 1.2$, получа-

ем, что формальная максимальная дрейфовая скорость $v_{dr,max} \approx 3 \times 10^7$ см/с оказывается больше тепловой скорости. Таким образом, современные цифровые ИС работают с максимальной дрейфовой скоростью, и для оценки их максимального тока более подходит формула (4.14.3).

Характер насыщения легко определить из вида ВАХ. При электростатическом запираении, характерном для длинноканальных транзисторов, имеет место традиционная квадратичная зависимость

$$I_{DSAT} \sim (V_{GS} - V_T)^2.$$

Для короткоканальных транзисторов с насыщением дрейфовой скорости имеем

$$I_{DSAT} \sim (V_{GS} - V_T).$$

В реальности в современных приборах экспериментально наблюдается, как правило, промежуточная ситуация, когда зависимости тока насыщения от затворного напряжения имеют вид

$$I_{DSAT} \sim (V_{GS} - V_T)^n, \quad (4.14.4)$$

где $1 \leq n \leq 2$.

4.15. Формула для ВАХ МОП-транзистора с учетом насыщения дрейфовой скорости (BSIM3-4)

Итак, уменьшение подвижности и насыщение дрейфовой скорости при возрастании тянущего электрического поля в канале приводит к появлению дополнительного механизма насыщения тока в канале МОПТ.

Рассчитаем ВАХ МОП транзистора с учетом возможного насыщения дрейфовой скорости. Как и в случае длинноканальных транзисторов с электростатическим запираением, распределение плотности электронов вдоль канала при малом смещении между стоком и истоком $V_{DS} < V_G - V_T$ записывается в виде (см. п. 4.11)

$$qn_s(y) = C_o (V_G - V_T - n V(y)),$$

$$n \cong 1 + \frac{C_D}{C_O} = 1 + \frac{\varepsilon_S}{\varepsilon_i} \frac{d_{ox}}{x_d}, \quad (4.15.1)$$

где n – коэффициент влияния подложки (в руководстве BSIM эта величина обозначается как A_{bulk}).

Начальное уравнение для дрейфового тока записывается в том же виде, как и в (4.12.1), но только с учетом зависимости подвижности от тянущего поля (4.13.2)

$$I_D = Z C_O [V_G - V_T - nV(y)] \mu_0 \frac{dV/dy}{1 + (dV/dy)/E_{SAT}}. \quad (4.15.2)$$

Как и ранее, эту формулу можно переписать в форме

$$I_D \left(dy + \frac{dV}{E_{SAT}} \right) = Z C_O (V_G - V_T - nV) \mu_0 dV, \quad (4.15.3)$$

и проинтегрировать обе стороны полученного уравнения от истока ($y = 0$) до стока ($y = L$):

$$I_D = \mu_0 \frac{Z}{L} \frac{C_o}{1 + \frac{V_{DS}}{E_{SAT}L}} \left(V_G - V_T - n \frac{V_{DS}}{2} \right) V_{DS}. \quad (4.15.4)$$

Сравнивая (4.15.4) и (4.12.4), видим, что для напряжений малых смещений (в крутой области характеристики $V_{DS} < V_{DSAT}$) можно грубо считать, что

$$I_D = \frac{\text{длинноканальный } I_D}{1 + V_{DS} / E_{SAT}L}. \quad (4.15.5)$$

4.16. Ток насыщения МОПТ

Ток насыщения определяется как обычно (см. п. 4.12), приравняв формально производную нулю $dI_D/dV_{DS} = 0$, что дает напряжение насыщения

$$V_{DSAT} = E_{SAT} L \left(\sqrt{1 + \frac{2(V_G - V_T)}{n E_{SAT}L}} - 1 \right). \quad (4.16.1)$$

Подставляя (4.16.1) в выражение для тока (4.15.4), можно получить формулу для тока насыщения МОПТ. Полученная при этом формула имеет громоздкий и непрозрачный вид. Кроме того, как мы уже отмечали, процедура определения тока насыщения по условию $dI_D/dV_{DS} = 0$ не является физически корректной. Поэтому часто используют более простой и, как оказалось, более «точный» способ определения напряжения насыщения

$$\frac{1}{V_{DSAT}} \cong \frac{n}{V_{GS} - V_T} + \frac{1}{E_{SAT}L}. \quad (4.16.2)$$

Если $E_{SAT} L \gg V_{GS} - V_T$, то вторым слагаемым в (4.16.2) и эффектом насыщения дрейфовой скорости можно пренебречь, и насыщение тока происходит за счет электростатического запираания

$$V_{DSAT} = (V_G - V_T) / n. \quad (4.16.3)$$

Такая ситуация имеет место для длинноканальных МОПТ (когда L велико) и/или для случая слабой инверсии (когда V_{GS} чуть больше V_T).

В противном случае, когда $E_{SAT} L \ll V_{GS} - V_T$, мы имеем насыщение тока МОПТ за счет насыщения скорости носителей в канале МОПТ.

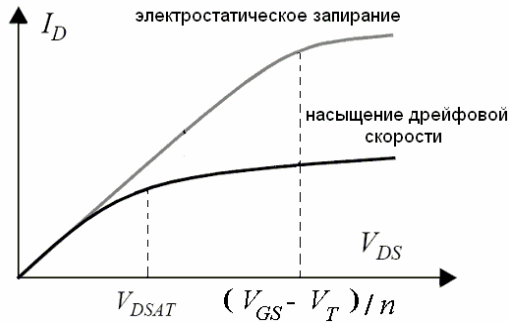


Рис. 4.15. Выходные ВАХ транзистора, рассчитанного по формуле для электростатического запираания (4.14.2) и с использованием модифицированной формулы (4.16.6) для случая $E_{SAT} L < V_{GS} - V_T$

Для справки, выражение для напряжения насыщения, используемое в системе моделей BSIM3-5, имеет вид

$$V_{DSAT} = \frac{E_{SAT} L (V_{GSTeff} + 2 \varphi_T)}{n E_{SAT} L + V_{GSTeff} + 2 \varphi_T}, \quad (4.16.4)$$

где V_{GSTeff} — стандартное обозначение эффективного затворного напряжения, используемое в руководстве BSIM3-4,

$$V_{GSTeff} \equiv q n_s (V_{GS}) / C_O. \quad (4.16.5)$$

Отметим, что в надпороговой области $V_{GSTeff} \cong V_{GS} - V_T$.

Соответственно, ток насыщения МОПТ представляется в виде формулы

$$I_{DSAT} = \frac{\frac{Z}{2L} \mu_0 C_O (V_G - V_T)}{\frac{n}{V_G - V_T} + \frac{1}{E_{SAT} L}} = \left(\frac{Z}{2nL} \right) \frac{\mu_0 C_O (V_G - V_T)^2}{1 + (V_G - V_T) / n E_{SAT} L}. \quad (4.16.6)$$

В зависимости от соотношения параметров, получаются предельные случаи (рис. 4.15) насыщение дрейфовой скорости ($E_{SAT} L \ll V_{GS} - V_T$)

$$I_{DSAT} = (Z/L) C_O (V_G - V_T) v_{SAT} ;$$

и электростатическое запираение канала ($E_{SAT} L \gg V_{GS} - V_T$)

$$I_{DSAT} = \frac{Z}{2nL} \mu_0 C_O (V_G - V_T)^2 .$$

Используя для оценок $E_{SAT} \cong 8 \times 10^4$ В/см; $n \cong 1.2$; $V_{DD} = 1.8$ В; $V_T = 0.25$ В, по формуле (4.16.4) можно легко получить $V_{DSAT} \cong 0.5$ В ($L = 100$ нм) и $V_{DSAT} \cong 0.3$ В ($L = 50$ нм).

5. Физические процессы в каналах МОПТ

5.1. Механизмы рассеяния носителей в канале

Подвижность носителей в канале определяется, главным образом, тремя механизмами рассеяния.

А. Рассеяние на заряженных центрах

На границе раздела Si-SiO₂, либо в пределах 1-2 нм в окисле расположено большое количество заряженных дефектов. Как правило, их концентрация находится в диапазоне 10¹⁰...10¹² см⁻². Сечение кулоновского рассеяния на таких дефектах можно грубо оценить следующим образом. «Радиус» дефекта определяется из равенства потенциальной энергии кулоновского центра и средней энергии свободных носителей ($\sim kT$)

$$kT \sim \frac{q^2}{\epsilon_0 r_0}.$$

Тогда сечение рассеяния на кулоновских центрах имеет следующую температурную зависимость:

$$\sigma \propto r_0^2 \propto \frac{q^4}{T^2}.$$

Вспоминая, что тепловая скорость носителей $\propto \sqrt{T}$, получаем оценку для среднего времени столкновений с заряженными дефектами $\tau \sim (\sigma N_t v_{th})^{-1} \sim T^{3/2}$.

Соответственно, подвижность, связанная с рассеянием только на ионизированных примесях,

$$\mu_l = (q/m)\tau \sim T^{3/2}. \quad (5.1.1)$$

Полученная формула основана на очень грубых приближениях; достаточно сказать, что рассеяние на кулоновских центрах в инверсионном слое не носит, как мы предполагали, трехмерного характера. Тем не менее, несмотря на это, температурная зависимость подвижности (точнее крутизны) $T^{3/2}$ (5.1.1) часто (хотя и не всегда) и с хорошей точностью выполняется при измерениях.

Важно иметь в виду, что механизм рассеяния на кулоновских центрах существенен только в том случае, если плотность электронов в канале мала. В надпороговом режиме потенциал заряженных

примесей эффективно экранируется свободными носителями в канале, и роль рассеяния на примесях существенно уменьшается.

В. Рассеяние на фононах

Рассеяние на колебаниях решетки (фононах) зависит от чисел заполнения (количества) фононов ($N_{ph} \sim T$) и усиливается с ростом температуры

$$\mu_{phon} \propto \tau \propto \frac{1}{N_{ph} v_{th}} \propto \frac{1}{T} \frac{1}{T^{1/2}} \propto T^{-3/2}. \quad (5.1.2)$$

В отличие от кулоновского механизма, рассеяние на фононах слабо зависит от величины прижимающего электрического поля и плотности носителей в канале и одинаково проявляет себя как в надпороговом, так и в подпороговом режиме работы транзистора.

С. Рассеяние на шероховатостях границы раздела

В надпороговом режиме ($V_{GS} > V_T$) доминирующим становится механизм рассеяния на шероховатостях поверхности. Идеальная поверхность (или граница раздела) отражает электроны зеркальным образом и не вносит вклад в рассеяние. Реальная поверхность раздела кремния с окислом имеет несовершенства («шероховатости», surface roughness), которые обуславливают случайный («диффузный») характер рассеивания носителей, что приводит к «трению» системы носителей в канале о границу раздела. Подвижность, обусловленная рассеянием на шероховатости поверхности, пропорциональна времени столкновения с поверхностью (см. также (3.10.3))

$$\mu_{SR} \sim \frac{q}{m} \tau \sim \frac{q}{m} \frac{x_{inv}}{v_T} \sim \frac{q}{m} \frac{kT}{E_{eff} v_T} \propto \frac{1}{E_{eff}}. \quad (5.1.3)$$

Как мы видим, рассеивание на шероховатостях возрастает с увеличением прижимающего электрического поля в канале.

5.2. Универсальная подвижность в надпороговом режиме

Предполагается, что все механизмы рассеяния независимы и их вероятности складываются по правилу Матиссена

$$\frac{1}{\tau} = \frac{1}{\tau_{phon}} + \frac{1}{\tau_{SR}} + \frac{1}{\tau_I} \quad (5.2.2)$$

В подпороговой области проводимость канала определяется в основном концентрацией носителей, которая может изменяться на

много порядков. Изменение подвижности играет существенную роль, главным образом, в надпороговом режиме $V_{GS} > V_T$, когда концентрация носителей в канале велика и изменяется не очень сильно. Как мы уже указывали, в этом режиме рассеяние на заряженных центрах подавляется, и подвижность начинает определяться только рассеянием на фононах и на шероховатостях границы раздела

$$\mu_{univer} \cong \left(\frac{1}{\mu_{phon}} + \frac{1}{\mu_{SR}} \right)^{-1} \cong \frac{\mu_{phon}}{1 + \text{const} \frac{\mu_{phon} E_{eff}}{v_T}}. \quad (5.2.2)$$

В литературе часто пишут об «универсальной» зависимости подвижности. Под этим понимается, что зависимости от прижимающего электрического поля E_{eff} для широкого класса приборов ложатся на одну кривую вне зависимости от толщины окисла, длины и ширины канала, легирования подложки и т.д.

На практике зависимость подвижности от прижимающего поля часто аппроксимируют эмпирическим соотношением

$$\mu(E_{eff}) = \frac{\mu_0}{1 + (E_{eff}/E_0)^a}. \quad (5.2.3)$$

Типичные значения эмпирических параметров приведены в табл. 5.1.

Таблица 5.1

Параметры модели подвижности электронов и дырок от прижимающего электрического поля в канале МОП транзисторов

	Электроны	Дырки
μ_0 (см ² /В с)	540	185
a	1.8	1
E_0 (МВ/см)	0.7	0.45

Величина μ_0 зависит от температуры, прижимающего электрического поля, ориентации границы раздела (например, в поверхностях разной ориентации границы раздела Si-SiO₂ имеем μ_0 ($\langle 110 \rangle$) $\cong 0.8 \mu_0$ ($\langle 100 \rangle$)), безразмерная константа в знаменателе зависит от размеров шероховатостей границы раздела.

Для лучшего описания экспериментальных данных приходится брать разные выражения для эффективного прижимающего поля в канале для электронов и дырок

$$E_{eff} = (q/\epsilon_S)(N_A x_d + n_S/2) \quad (\text{электроны}), \quad (5.2.4)$$

$$E_{eff} = (q/\epsilon_S)(N_D x_d + n_S/3) \quad (\text{дырки}), \quad (5.2.5)$$

Как мы видим, для дырок приходится брать «неуниверсальную» оценку для прижимающего электрического поля (ср. 3.3.10).

5.3. Зависимость подвижности от прижимающего поля и температуры

В режиме обеднения и слабой инверсии подвижность растет за счет увеличения экранирования примесей. В промежуточной области подвижность относительно слабо зависит от поля (для оценок $\mu \sim 32500 E_{eff}^{-1/3}$). В надпороговом режиме (когда $E_{eff} > 1$ МВ/см) эффективная подвижность начинает падать за счет рассеяния на шероховатостях.

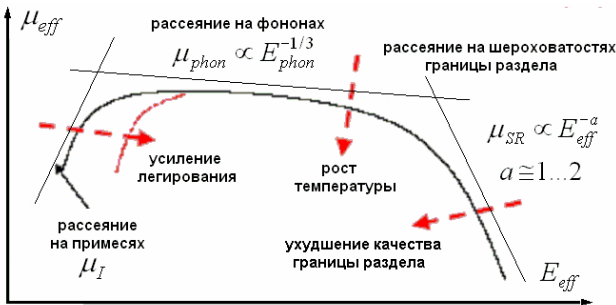


Рис. 5.1. Схематическое представление зависимости подвижности в канале от прижимающего эффективного поля для разных механизмов рассеяния

Физическая причина уменьшения подвижности состоит в усилении рассеяния на шероховатостях границы раздела, что можно описать формулой (5.2.2). Эффективное прижимающее поле не является непосредственно измеряемой величиной, и поэтому на прак-

тике очень часто подвижность моделируют с помощью простой аппроксимационной функции от затворного напряжения

$$\mu = \frac{\mu_0}{1 + \theta(V_{GS} - V_T)}, \quad (5.3.1)$$

где эмпирические параметры μ_0 и θ определяются согласованно для подгонки экспериментальных данных, θ находится в диапазоне значений от 0.08 до 0.4 В⁻¹. В целом, можно отметить, что с уменьшением проектной нормы типичные значения прижимающего электрического поля возрастают, и подвижность носителей в канале падает.

Характерное увеличение подвижности с ростом температуры, обусловленное рассеянием на заряженных примесях (см. 5.1.1), становится существенным только для сильнолегированной подложки (рис. 5.2).

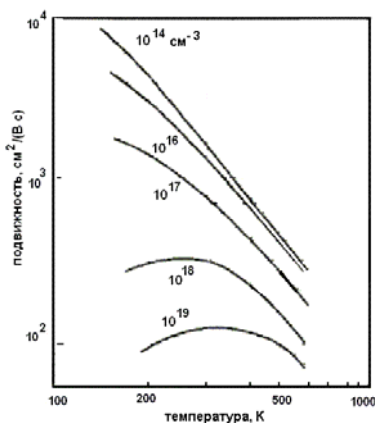


Рис. 5.2. Зависимости подвижности носителей в канале от температуры при разных значениях уровней легирования

Одновременное уменьшение подвижности носителей в надпороговом режиме и порогового напряжения с ростом температуры приводит к появлению характерной точки пересечения ВАХ МОПТ для разных температур (рис. 5.3).

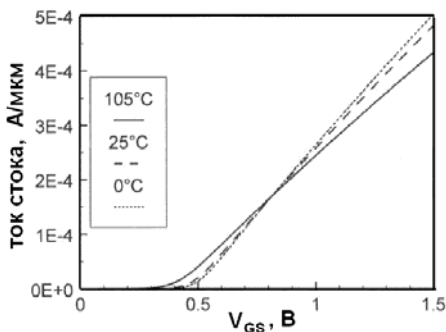


Рис. 5.3. Типичные передаточные вольт-амперные характеристики *n*-МОПТ для разных температур измерения

Такое поведение ВАХ при разных температурах свойственно практически всем типам транзисторов и имеет воспроизводимый характер, и этот факт иногда используется в аналоговой схемотехнике для разработки схем с высокой температурной стабильностью. Точка пересечения соответствует концентрации носителей в канале $\sim 10^{11} \text{ см}^{-2}$. С физической точки зрения интересно то, что справа от этой точки канал ведет себя как металл (рост сопротивления с ростом температуры), а слева – как изолятор (уменьшение сопротивления с ростом температуры).

5.4. Повышение подвижности с использованием технологии напряженного кремния

Подвижность является важнейшей характеристикой МОПТ, непосредственно определяющей быстродействие всей системы. Величина подвижности в полупроводнике зависит от величины эффективной массы носителей в нем, которая, в свою очередь, определяется фундаментальными свойствами кристаллической решетки. Эти свойства могут изменяться, подвергая решетку воздействию механического напряжения (это величина с размерностью «сила на единицу площади»). В частности, таким образом можно изменять эффективную массу носителей тока. Самый распространенный вариант приложения механических напряжений состоит в использовании кремний – германиевых сплавов (см. также п. 3.13).

Кремний и германий являются химически совместимыми элементами и могут образовывать сплав. При этом постоянная решет-

ки германия (~ 0.56 нм) приблизительно на 4 % больше постоянной решетки кремния (~ 0.54 нм). Поэтому, если слой кремния поместить на подложку из кремний-германиевого сплава, то кремний оказывается под действием растягивающего механического напряжения, действующего по всем направлениям (т.н. биаксиального или двухосного растяжения). Механическое напряжение может уменьшать эффективное значение эффективной массы носителей, что приводит к увеличению их подвижности. Впервые эта технология была реализована для 90 нм проектной нормы. Сечение n -канального транзистора, изготовленного по технологии напряженного кремния, представлено на рис. 5.4. Толщина слоя напряженного кремния составляет ~ 10 нм, а толщина слоя $\text{Si}_{1-x}\text{Ge}_x$ – около 30 нм. Существуют другие варианты такой технологии, включающие в себя сжимающие и растягивающие механические напряжения (одноосные и двухосные). Выигрыш в подвижности в подобных технологиях довольно значителен и составляет несколько десятков процентов. Этот выигрыш может достигать максимальных значений ~ 80 % для электронов (приблизительно от 400 до $700 \text{ см}^2/\text{В}\cdot\text{с}$) и ~ 70 % для дырок (100 до $170 \text{ см}^2/\text{В}\cdot\text{с}$) для биаксиально-(двухосно) растянутого слоя кремния, что достигается при доле германия, равной приблизительно 30 %.

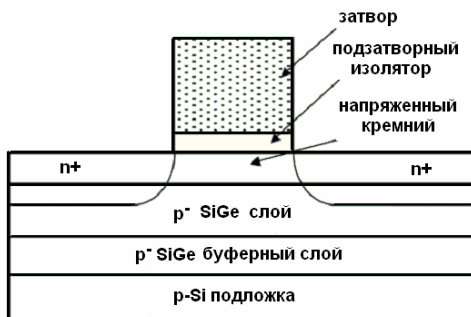


Рис. 5.4. Сечение n -МОП транзистора со слоем напряженного кремния

Существуют другие варианты создания механических напряжений в области канала. В частности, для этого можно использовать боковую STI – изоляцию мелкими канавками (STI, см. п. 5.9).

Проблемы подходов, основанных на напряженном кремнии, состоят в высокой дефектности слоев такого кремния; плохой со-

вместимости с технологиями «кремний-на-изоляторе»; низкой технологичности и низком проценте выхода (особенно для больших концентраций германия), а также высокой стоимости. Модели механического напряжения включены в систему параметров BSIM4.

5.5. Зависимость подвижности эффекта поля от спектра поверхностных состояний

Экспериментальное определение подвижности носителей в канале не является однозначной и простой задачей. Сложность состоит в том, что для однозначного определения подвижности необходимо независимое определение плотности носителей в канале, что можно сделать, как правило, только для тестовых структур с большой площадью (например, используя эффект Холла).

Обычно различают подвижность проводимости и подвижность эффекта поля. Подвижности проводимости определяются по наклону крутой области выходной ВАХ $I_D(V_{DS})$

$$\mu_0 = \frac{L/Z}{C_O(V_{GS} - V_T)} \left. \frac{\partial I_D}{\partial V_{DS}} \right|_{V_{GS}=\text{const}} \quad (5.5.1)$$

Подвижность проводимости формально дает значение микроскопической подвижности, но следует помнить, что в этом случае плотность носителей в канале оценивается через эмпирическое значение порогового напряжения, которое, вообще говоря, не определяется строго и однозначно.

Подвижность эффекта поля определяется по крутизне передаточной ВАХ $I_D(V_{GS})$

$$\mu_{FE} = \frac{L/Z}{C_O V_{DS}} \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{V_{DS} \rightarrow 0} \quad (5.5.2)$$

Если считать, что ток в надпороговой области точно выражается через пороговое напряжение

$$I_D = \frac{Z}{L} \mu_0 q n_S V_{DS} = \frac{Z}{L} \mu_0 C_O (V_G - V_T) V_{DS}, \quad (5.5.3)$$

то оба определения подвижности совпадают: $\mu_0 = \mu_{FE}$.

В реальности зависимость плотности носителей в канале от затворного напряжения не является в точности линейной, поскольку поверхностный потенциал продолжает слабо расти даже в надпо-

роговой области ($\varphi > 2\varphi_F$), что приводит к перезарядке поверхностных состояний, соответствующей деформации зависимости $V_{GS}(\varphi_S)$ и появлению зависимости μ_{FE} от плотности поверхностных состояний D_{it} . Действительно, используя известное соотношение (см. п. 3.5)

$$qn_S = C_O(V_G - \varphi_{MS} - \varphi_S) - q(N_A x_d(\varphi_S) - N_{ot}(\varphi_S)), \quad (5.5.4)$$

можно получить

$$\begin{aligned} \mu_{FE}(D_{it}) &= \frac{L/Z}{C_O V_{DS}} \left(\frac{\partial I_D}{\partial n_S} \right) \left(\frac{\partial n_S}{\partial \varphi_S} \right) \left(\frac{\partial \varphi_S}{\partial V_{GS}} \right) \Bigg|_{V_{DS} \rightarrow 0} = \\ &= \frac{\mu_0}{1 + \frac{C_O + C_D + q^2 D_{it}}{C_{inv}}}. \end{aligned} \quad (5.5.5)$$

Таким образом, подвижность эффекта поля определяется не только процессами рассеяния носителей в канале, но и процессами перезарядки поверхностных состояний.

При сильной инверсии (когда C_{inv} очень велико) отличие подвижности проводимости от подвижности эффекта поля незначительно. В то же время, в области вблизи порога и/или при высокой плотности поверхностных состояний величина μ_{FE} может отличаться от подвижности проводимости μ_0 на десятки процентов.

Плотность поверхностных состояний является величиной, чувствительной к внешним и внутренним воздействиям, например радиации и/или горячим носителям в канале. Крутизна при таких воздействиях может деградировать как за счет уменьшения подвижности (усиления рассеяния носителей на заряженных дефектах), так и за счет усиленной перезарядки ПС.

Изменение подвижности эффекта поля при нестабильной величине плотности ПС выражается соотношением

$$\begin{aligned} \mu_{FE}(D_{it0} + \Delta D_{it}) &= \frac{\mu_0}{1 + \frac{C_O + C_D + q^2(D_{it0} + \Delta D_{it})}{C_{inv}}} = \\ &= \frac{\mu_{FE}(D_{it0})}{1 + \frac{\Delta D_{it}}{C_{inv} + C_O + C_D + q^2 D_{it0}}}. \end{aligned} \quad (5.5.6)$$

Для описания деградации крутизны в надпороговой области ВАХ от поглощенной дозы ионизирующего излучения D часто используется эмпирическая формула

$$\beta(D) = \frac{\beta(D=0)}{1 + \alpha_{it} \Delta N_{it}(D)}, \quad (5.5.7)$$

где дозовое изменение крутизны выражается через изменение интегральной плотности поверхностных состояний ΔN_{it} и эмпирическую константу $\alpha \sim 10^{-12} \text{ см}^2$. Методики определения ΔN_{it} фактически сводятся к определению неких эффективных значений энергетической плотности ΔD_{it} , что означает в практическом плане эквивалентность (5.5.6) и (5.5.7).

В силу указанных выше причин, очень трудно разделить экспериментально деградацию крутизны за счет увеличения степени рассеяния и перезарядки поверхностных состояний. Экспериментально этот вопрос не исследован детально, но есть основания полагать, что при облучении крутизна деградирует, главным образом, за счет перезарядки радиационно-индуцированных ПС.

5.6. Короткоканальные эффекты в МОП транзисторах и электростатическое качество

В идеале конфигурация МОПТ должна обеспечивать полный контроль заряда в канале зарядом на затворе. В значительной степени это определяется геометрической конфигурацией транзистора. Для того чтобы охарактеризовать степень идеальности транзистора, часто используют понятие электростатического качества (electrostatic integrity).

Электростатическое качество выражается в степени выполнения двух следующих требований:

- 1) каждый заряд на затворе индуцирует заряд в канале;

2) заряд в стоке не должен влиять на заряд в канале.

В реальных транзисторах с любой технологической нормой эти требования не выполняются в полной мере. Этому мешает следующие факторы:

1) Заряд в обедненной области и на поверхностных состояниях. В контексте управляемости тока заряды легирующих примесей в обедненной области являются паразитными. Тем не менее, для транзисторов, изготовленных по объемной технологии, заряд обедненного слоя устранить невозможно, поскольку он является фактором, обеспечивающим необходимую электростатику;

2) Геометрический фактор. Часть силовых линий от затвора уходит в сторону и индуцирует дополнительный паразитный заряд на краях. Возникает также паразитная электростатическая связь (взаимная емкость) между затвором и стоком, каналом и стоком. Особенно существенны такие эффекты в современных короткоканальных транзисторах, в которых все более значительной становится роль краевых эффектов.

Математическое выражение электростатического качества в планарной структуре выражается условием

$$\frac{\partial E_x}{\partial x} \gg \frac{\partial E_y}{\partial y}, \frac{\partial E_z}{\partial z}. \quad (5.6.1)$$

Одним из путей борьбы с геометрическими эффектами является переход от неидеальных планарных структур к объемным. Речь идет о необходимости максимально окружать активную область кремния затворами непланарной конфигурации структуры (triple-gate, П-gate, Ω-gate) (см. также п. 8.12).

5.7. Геометрические эффекты порогового напряжения

В идеальных (длинноканальных) МОПТ пороговое напряжение V_T не зависит от длины канала и является локальной характеристикой, т.е. определяется локальными значениями параметров, например, $V_T = f(d_{ox}, N_A)$. Для коротких и узких каналов существенную роль начинают играть краевые эффекты. Часть силовых линий от затвора уходит на края и/или на другие электроды. В этом случае количество заряда в канале и, соответственно, пороговое напряжение становится глобальной характеристикой всего транзи-

сторы $V_T = f(d_{ox}, N_A, L, x_j)$ и зависит, в частности, от длины канала и конфигурации стока.

Приближение плавного канала в этом случае работает плохо, и нужно вносить поправки. Одна из главных поправок состоит в том, что вместо условия локальной электрической нейтральности нужно записывать условие глобальной (интегральной) электронейтральности.

Напомним, что условие локальной электронейтральности выражается в равенстве локальных значений поверхностных плотностей [Кл/см²] в длинноканальных приборах (см. 3.5.2)

$$+qN_G - qn_{inv} - qN_A x_d = 0.$$

Интегральное условие электронейтральности для короткоканальных приборов выражается в равенстве полных зарядов [Кл]

$$+Q_G - Q_{inv} - Q_{depl} + Q_{drain} = 0. \quad (5.7.1)$$

Последнее слагаемое в (5.7.1), обозначающее заряд на электроде стока, ответственно за паразитный эффект понижения барьера между стоком и истоком полем стока (т.н. DIBL, см. п. 5. 10).

5.8. Эффект спада порогового напряжения («roll-off») для коротких каналов

В соответствии с идеей глобальной электронейтральности, пороговое напряжение зависит от *полного* заряда обедненной области Q_{depl} , контролируемого затвором,

$$V_T = V_{FB} + 2\phi_F + \frac{Q_{depl}}{C_O Z L}. \quad (5.8.1)$$

При коротком канале часть отрицательного заряда акцепторов компенсируется не только зарядом на затворе, но и положительным зарядом доноров в стоке и истоке. Часть обедненной области стока и истока “заползает” под затвор, уменьшая Q_{depl} и, соответственно, пороговое напряжение.

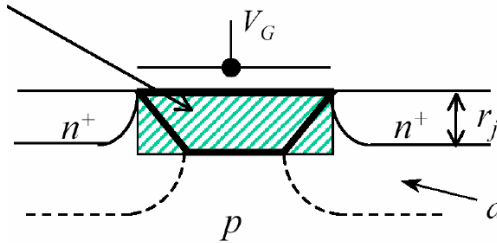


Рис. 5.5. Схематическое представление разделения активной области обеднения (заштриховано) на части, контролируемые затвором, а также стоком и истоком

Таким образом, пороговое напряжение уменьшается с уменьшением длины канала. Строго говоря, в этом случае, пороговое напряжение необходимо определять, решая трехмерное уравнение Пуассона. Именно так поступают, используя громоздкие пакеты технологического проектирования TCAD. Для схмотехнических пакетов программ необходимы простые компактные модели.

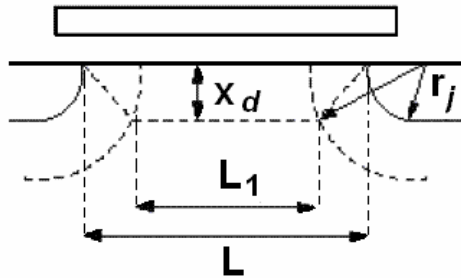


Рис. 5.6. Схема, поясняющая процедуру приближенного решения уравнений электрической нейтральности для всего транзистора

Грубое моделирование можно провести на основе соображений полной электронейтральности (5.7.1), основываясь на схеме, приведенной на рис. 5.6.

Полный заряд обедненной области можно оценить из грубой геометрической оценки полного заряда в обедненной области

$$Q_{depl} = qN_A x_d \frac{L + L_1}{2}. \quad (5.8.2)$$

Отсюда приведенная длина канала становится равной

$$L_1 \cong L - 2 \left(\sqrt{(r_j + x_d)^2 - x_d^2} - r_j \right) = L \left(1 - 2 \frac{r_j}{L} \left(\sqrt{1 + \frac{2x_d}{r_j}} - 1 \right) \right), \quad (5.8.3)$$

и в соответствии с (5.8.1), пороговое напряжения выражается формулой

$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{d0}}{C_O} \left(1 - \frac{r_j}{L} \left(\sqrt{1 + \frac{2x_d}{r_j}} - 1 \right) \right). \quad (5.8.4)$$

Из (5.8.4) следует, что пороговое напряжение уменьшается с уменьшением длины канала. Это действительно имеет место для короткоканальных приборов (рис. 5.7).

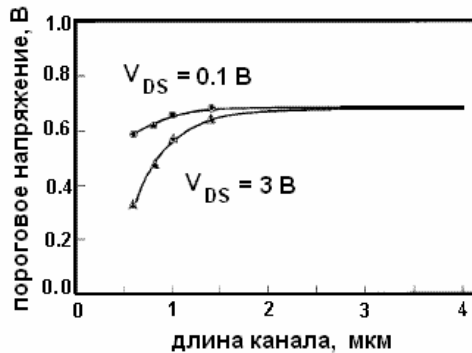


Рис. 5.7. Зависимость порогового напряжения от длины канала для транзисторов при разных напряжениях на стоке

Как видно из рис. 5.7, этот эффект сильно зависит от напряжения на стоке, что является выражением DIBL-эффекта (см. п. 5.10). Способы борьбы с эффектом спада порога состоят, во-первых, в формировании более тонких слоев *pn*-перехода (уменьшения r_j); во-вторых, в уменьшении толщины обедненной области x_d . И то, и другое достигается за счет увеличения концентрации легирующей примеси в подложке в районе канала или у стока (истока).

Практическое следствие геометрического эффекта состоит в том, что пороговое напряжение короткоканальных транзисторов гораздо сложнее контролировать, чем при изготовлении в длинноканальных МОПТ.

5.9. Эффекты узкого канала и общая характеристика геометрических эффектов порога

Эффекты узкого канала определяются способами боковой изоляции транзисторов. Основными технологиями боковой изоляции являются локальное окисление кремния (LOCOS) и изоляция мелкими канавками (shallow trench isolation, STI) (рис. 5.8). Технология STI является более современной (одновременно и более сложной, и дорогой) и практически доминирует при нормах проектирования менее 250 нм.

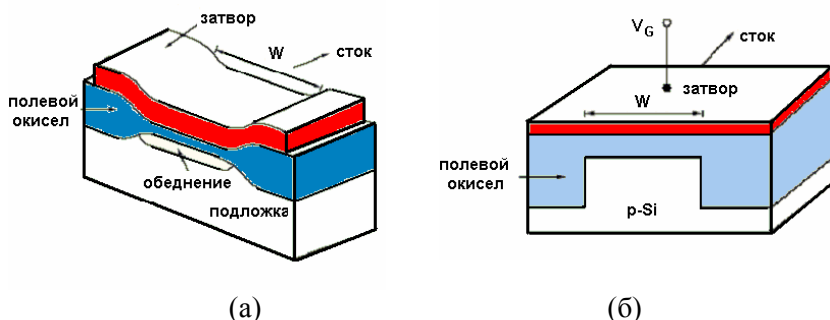


Рис. 5.8. Основные виды боковой изоляции современных транзисторов: (а) LOCOS, (б) STI

Эти технологии могут иметь противоположные зависимости величины порогового напряжения от ширины канала. Для технологии LOCOS характерен т.н. «птичий клюв» (рис. 5.9). Поэтому на краях полевого окисла образуется паразитные области обеднения. Это приводит к увеличению порогового напряжения с уменьшением ширины канала W . В этом случае говорят о «нормальном» эффекте узкого канала.

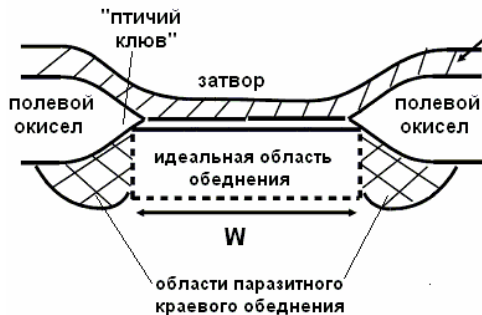


Рис. 5.9. Идеальная и реальная область контроля зарядом затвора в боковой изоляции типа LOCOS

Для боковой изоляции мелкими канавками скорее характерен «обратный» эффект узкого канала – уменьшение порогового напряжения с уменьшением ширины канала (рис. 5.10).

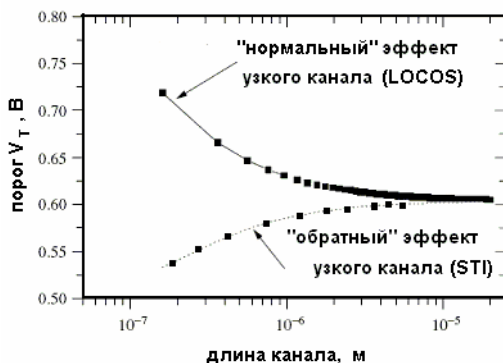


Рис. 5.10. Типичные зависимости порогового напряжения от ширины канала для технологий боковой изоляции LOCOS и STI

Следует «обратный» эффект узкого канала, который определяется особенностями профиля толщины изоляции на границе тонкого подзатворного и толстого окисла в STI канавке. Поэтому он не является универсальным, и для некоторых STI технологий имеет место «нормальный» эффект узкого канала.

5.10. Индуцированное стоком понижение барьера (DIBL)

При приложении напряжения на сток часть силовых линий от заряда в стоке оканчиваются в области канала и индуцируют дополнительный заряд в канале. Это влияние напряжения на стоке V_{DS} на эффективное значение порогового напряжения называется DIBL (Drain Induced Barrier Lowering). Этот эффект эквивалентен тому, что потенциальный барьер между стоком и истоком контролируется не только затвором, но и стоком (рис. 5.11).

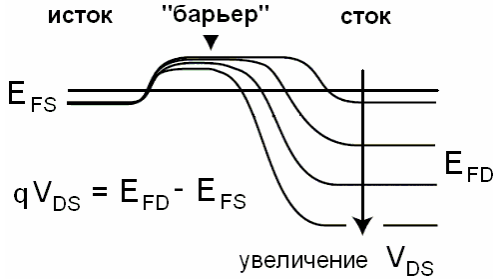


Рис. 5.11. Понижение барьера между стоком и истоком, индуцированное напряжением на стоке (DIBL)

Чем больше V_{DS} , тем больше положительный заряд в стоке индуцирует отрицательный заряд в канале, тем меньше пороговое напряжение. Поле стока начинает понижать барьер между стоком и истоком (DIBL). Сток начинает играть роль паразитного затвора. Эффект DIBL становится заметным при длине канала $L < 0.5$ мкм.

Зависимость порогового напряжения от напряжения на стоке

$$V_T(V_{DS}) \cong V_T(V_{DS} = 0) - \Delta V_T^{DIBL}(L, V_{DS}) \quad (5.10.1)$$

определяется еще и обратным смещением на подложке, и изменение порогового напряжения из-за DIBL-эффекта можно оценить по эмпирической формуле

$$\Delta V_T^{DIBL}(L, V_{DS}) \cong a_{DIBL} V_{DS} \cong \frac{\text{const}}{L^2} (2\varphi_F + V_{SB})^{1/2} V_{DS}, \quad (5.10.2)$$

где эмпирическая константа зависит от технологии.

Главным следствием эффекта DIBL является то, что токи в подпороговой области также начинают сильно зависеть от напряжения

на стоке. Соответственно, статические токи утечки через закрытый транзистор I_{off} становятся нестабильными (возрастают).

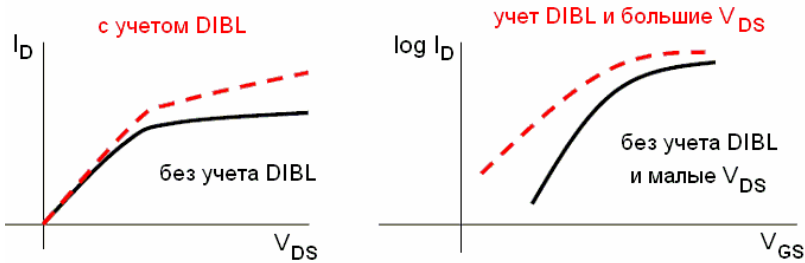


Рис. 5.12. Влияние DIBL для выходных (слева) и передаточных (справа) характеристик МОПТ

Технологические способы подавления DIBL реализуются через:
 уменьшение толщины r_j -переходов стока и истока;
 уменьшение толщины подзатворного окисла d_{ox} , что улучшает электростатический контроль затвором;
 увеличение уровня легирования подложки.

Таким образом, DIBL подавляется, а электростатическое качество МОПТ улучшается, если уменьшается отношение толщины подзатворного окисла к длине канала d_{ox}/L и глубина залегания pn -перехода стока и истока r_j , а также увеличивается уровень легирования подложки N_A .

5.11. Паразитные токовые эффекты короткого канала

Эффекты прокола: при недостаточном легировании кремневой подложки и малой длине канала обедненные области стока и истока при случайной локальной флуктуации количества легирующей примеси могут смыкаться, сильно понижая тем самым потенциальный барьер между стоком и истоком (рис. 5.13).

Через участки смыкания между стоком и истоком начинает течь большой ток, и этот эффект носит название прокола (punchthrough).

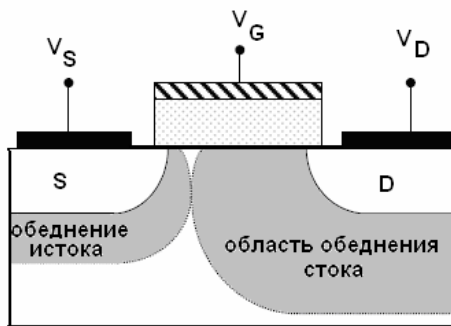


Рис. 5.13. Эффект смыкания областей пространственного заряда истока и стока

Естественный метод борьбы с проколом состоит в увеличении уровня легирования подложки, которое уменьшает толщину обедненных областей.

Эффект паразитного биполярного транзистора: структура МОП транзистора (исток – подложка (барьер) – сток) имеет большое сходство со структурой биполярного транзистора (эмиттер – база (барьер) – коллектор). Для получения биполярного эффекта в МОП транзисторе не хватает только внешнего дырочного тока в p -подложку, играющего роль тока базы.

Одним из механизмов, способных обеспечить внешний «базовый» ток и вызвать биполярный эффект в МОПТ, является ударная ионизация. В районе стока в канале образуется сильное электрическое поле, которое приводит к процессу ударной ионизации, выражающейся в генерации электронно-дырочных пар. Электроны уходят в сток, а дырки текут в контакт подложки, и за счет падения напряжения на подложке стоковый переход оказывается прямосмещенным. Потенциал подложки за счет избыточных дырок оказывается более положительным, что вызывает инжекцию электронов из истока, которые попадают в сток не через канал, а через объем подложки. Более подробно этот вопрос рассмотрен в главе 6.

5.12. Оптимизация структуры истоков и стоков

Минимизация геометрических эффектов предполагает уменьшение глубины залегания pn -переходов. С другой стороны, малая глубина pn -перехода стока r_j приводит к нежелательному увеличению сопротивления стока $R_{SD} \sim 1/r_j$.

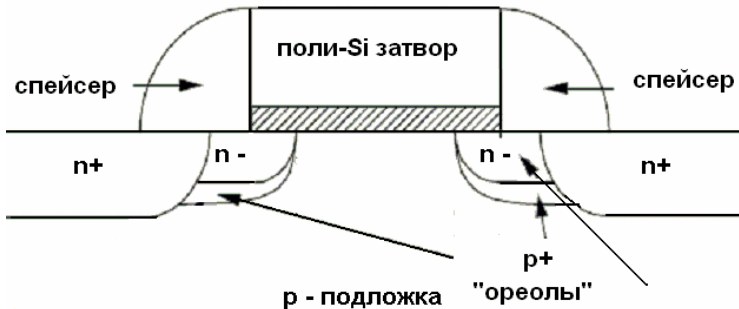


Рис. 5.14. Структура современного транзистора

Оптимизирующее техническое решение состоит в использовании тонких и коротких слаболегированных пристроек (extensions) к массивным истоку и стоку, удлиняющих сток и исток в сторону канала (рис. 5.14). Такие области иногда называют слаболегированными стоками (Lightly Doped Drain, LDD). Эти области уменьшают геометрические эффекты короткого канала, и в то же время не приводят к существенному увеличению сопротивлений стока-истока. Необходимо подчеркнуть, что эти области «слабо» легированы ($10^{18} \dots 10^{19} \text{ см}^{-3}$) только по сравнению с n^+ - областями стоков и истоков ($5 \times 10^{19} \dots 1 \times 10^{20} \text{ см}^{-3}$). Толщина этих областей составляет в современных транзисторах несколько десятков нанометров и имеет ограничение снизу. Например, уменьшение глубины pn -переходов до 10 нм приводит к увеличению удельного поверхностного сопротивления стоков и истоков до 10 кОм/квadrat. К сожалению, слаболегированные n -области вблизи канала уменьшают пороговое напряжение транзистора, что особенно заметно для очень малых длин канала. Для компенсации этого эффекта используется допол-

нительное легирование, с помощью которого создается тонкий p^+ -слой в виде «ореола» (“halo”), окружающего LDD области. Ореольное легирование снижает DIBL эффекты, уменьшает подпороговые утечки и вероятность смыкания обедненных областей истока и стока.

Технологическим вариантом компенсации уменьшения порогового напряжения является легирование в «кармашек» (pocket), которое отличается от ореола только тем, что охватывает не всю LDD-область, а только ее часть, примыкающую к (и)стоку. На рис. 5.14 изображен промежуточный случай между ореолом и «кармашком».

5.13. Моделирование выходного сопротивления МОПТ

Одна из проблем в короткоканальных приборах состоит в том, что ток в режиме насыщения (на пологом участке) продолжает заметно расти при увеличении V_{DS} . Для цифровой схмотехники, работающей на большом сигнале, это не так существенно, но в аналоговых схемах это является большой проблемой, поскольку приводит к изменению выходного сопротивления транзистора. Приращение тока в режиме насыщения тока МОПТ, с точки зрения его малосигнальной эквивалентной схемы, равносильно добавлению параллельно идеальному источнику тока некоторого, обычно довольно большого внутреннего сопротивления:

$$R_{OUT} = \left(\frac{dI_{SAT}}{dV_{DS}} \right)^{-1}. \quad (5.13.1)$$

Считается, что увеличения тока при $V_{DS} > V_{DSAT}$ происходят за счет трех эффектов:

- а) модуляции длины канала (channel length modulation, CLM);
- б) эффекта понижения барьера стоком (DIBL);
- в) паразитного тока ударной ионизации (SCBE), который становится заметен только при больших V_{DS} .

Классификацию этих эффектов проводят на основе анализа зависимости тока и выходного сопротивления МОПТ от V_{DS} (рис. 5.15).

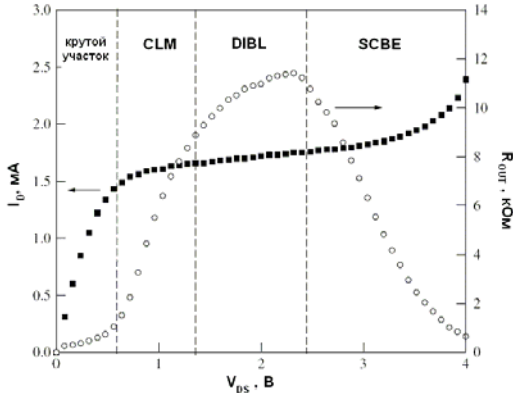


Рис. 5.15. Выходная ВАХ и выходное сопротивление короткоканального МОПТ

С некоторой степенью произвола считается, что участок сильного роста R_{out} обусловлен модуляцией длины канала (CLM); участок ВАХ, где R_{out} максимально приписывается DIBL эффекту, а участок уменьшения R_{out} – эффекту ударной ионизации (SCBE).

Ток канала в режиме насыщения является гладкой функцией V_{DS} , и поэтому для моделирования этого участка используют линейную аппроксимацию

$$\begin{aligned}
 I_{DS}(V_{GS}, V_{DS} > V_{DSAT}) &\cong \\
 &\cong I_{DSAT}(V_{GS}, V_{DSAT}) \left(1 + \frac{V_{DS} - V_{DSAT}}{V_A} \right), \quad (5.13.2)
 \end{aligned}$$

где параметр

$$V_A \equiv I_{DSAT} \left[\frac{\partial I_{DS}(V_{GS}, V_{DS})}{\partial V_{DS}} \right]^{-1} = \frac{I_{DSAT}}{R_{OUT}} \quad (5.13.3)$$

определяется как аналог напряжения Эрли в биполярном транзисторе.

Далее предполагается, что все механизмы, дающие вклад в напряжение Эрли (V_A), являются независимыми и могут вычисляться отдельно друг от друга, что эквивалентно суммированию обратных значений параллельных выходных сопротивлений, соответствующих трем механизмам:

$$\frac{1}{V_A} = \frac{1}{V_{ACLM}} + \frac{1}{V_{ADIBL}},$$

$$R_{OUT} = R_{OUT}(DIBL) + R_{OUT}(ACLM). \quad (5.13.4)$$

Сопротивление из-за DIBL эффекта обусловлено уменьшением порога V_T с ростом V_{DS} :

$$V_T(V_{DS}) \cong V_{T0} - \text{const} \times V_{DS}. \quad (5.13.5)$$

Как легко видеть, это сопротивление не зависит от V_{DS}

$$R_{out}(DIBL) = \left(\frac{dI_{DSAT}}{dV_{DS}} \right)^{-1} = \left(\frac{dI_{DSAT}}{dV_T} \frac{dV_T}{dV_{DS}} \right)^{-1} \cong \text{const}(V_{DS}). \quad (5.13.6)$$

5.14. Эффект модуляции длины канала

При увеличении V_{DS} плотность электронов вблизи стока уменьшается, пока не станет близкой к нулю (при $V_{DS} = V_{DSAT}$). При дальнейшем увеличении смещения на стоке происходит т.н. отсечка (pinch-off), и все добавочное напряжение $V_{DS} - V_{DSAT}$ начинает падать на обедненной области pn -перехода стока.

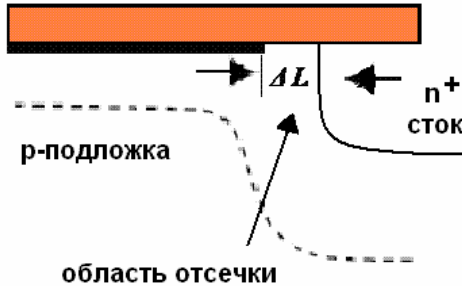


Рис. 5.16. Область отсечки и эффективное сокращение длины канала

Обедненная область стока увеличивается, а эффективная длина канала становится меньше на величину ΔL . Этот эффект называется эффектом модуляции длины канала (Channel Length Modulation, CLM).

За счет этого эффекта ток МОПТ в режиме насыщения продолжает расти приблизительно по линейному закону (рис. 5.17):

$$I_{DSAT} \propto 1/(L - \Delta L) \sim 1 + \Delta L / L \sim 1 + \lambda V_{DS}. \quad (5.14.1)$$

Эмпирический параметр λ (параметр λ в ранних моделях) имеет порядок $\sim(0.08\dots 0.10) 1/V$.

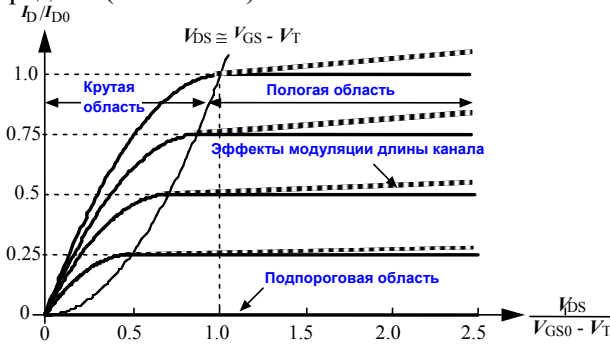


Рис. 5.17. Выходные ВАХ МОПТ, нормированные на средний ток насыщения I_{D0}

В новых системах моделей в BSIM3-4 моделирование модуляции длины канала производится с помощью соотношения $\Delta L = (V_{DS} - V_{DSAT}) / E_{pinch}$, где E_{pinch} – электрическое поле в области отсечки:

$$\begin{aligned}
 I_D &= I_{DSAT} \left(1 + \frac{\Delta L}{L} \right) = \\
 &= I_{DSAT} \left(1 + \frac{V_{DS} - V_{DSAT}}{E_{pinch} L} \right) \equiv I_{DSAT} \left(1 + \frac{V_{DS} - V_{DSAT}}{V_{ACLM}} \right). \quad (5.14.2)
 \end{aligned}$$

В главе 6 мы более подробно рассмотрим процессы в районе отсечки и получим приближенное аналитическое выражение, описывающее зависимости длины области отсечки от напряжения между стоком и истоком.

5.15. Паразитные сопротивления стока и истока

В короткоканальных МОПТ возрастает роль паразитных элементов. Металлические контакты, а также сток и исток, особенно слаболегированные, являются проводниками с конечным сопротивлением (рис. 5.18), что ухудшает характеристики МОПТ.

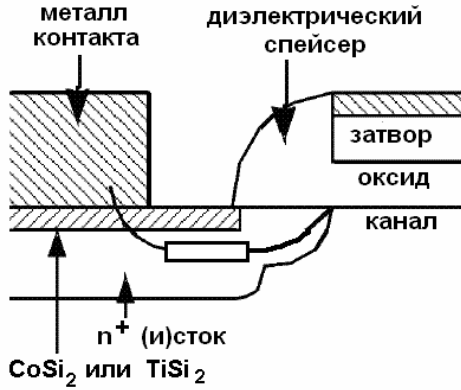


Рис. 5.18. Паразитное сопротивление от контакта до начала канала

При наличии заметного сопротивления стока и истока внутренние потенциалы не совпадают с внешними (приложенными) напряжениями (см. эквивалентную схему на рис. 5.19).

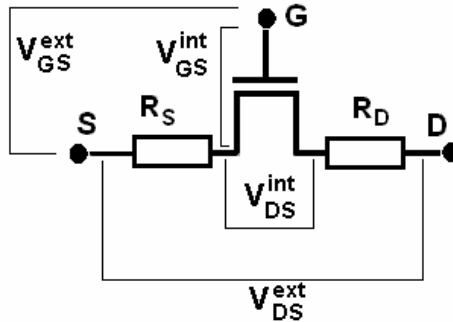


Рис. 5.19. Эквивалентная схема транзистора с ненулевыми сопротивлениями стока и истока

Используя эквивалентную схему на рис. 5.19, записываем связь внешних и внутренних (обозначаемых индексом int) напряжений:

$$\begin{aligned}
 V_{GS}^{int} &= V_{GS} - I_D R_S, \\
 V_{DS}^{int} &= V_{DS} - I_D (R_S + R_D).
 \end{aligned}
 \tag{5.15.1}$$

Из этих выражений легко получить выражение эффективной крутизны МОПТ для линейного режима работы транзистора

$$g_m \equiv \frac{\partial I_D}{\partial V_{GS}^{ext}} = \frac{g_{m0}}{1 + g_{D0}(R_D + R_S) + g_{m0}R_S}, \quad (5.15.2)$$

$$\text{где } g_{m0} = \frac{\partial I_D}{\partial V_{GS}^{int}}; \quad g_{D0} = \frac{\partial I_D}{\partial V_{DS}^{int}}. \quad (5.15.3)$$

Из формулы видно, что сопротивление истока сильнее влияет на уменьшение крутизны. Из-за этого часто слабо легирует только сток, что усложняет проектирование разводки схемы.

При длине канала 0.1 мкм сопротивление стока-истока уменьшает ток в канале I_{DSAT} на $\sim 10-20\%$. Считается, что допустимые значения паразитных сопротивлений стокового и истокового перехода определяются условием $R_S + R_D < 0.2R_C$, где R_C – сопротивление канала открытого транзистора. Сопротивление контактов растет с уменьшением глубины p-n переходов стока и истока; поэтому эта проблема усугубляется с миниатюризацией приборов. В более продвинутых технологиях в качестве материала контактов используются силициды металлов с более высокой проводимостью по сравнению с сильнолегированным кремнием. При этом исток и сток являются контактами Шотки.

NB. Так как проводимость транзистора в линейной области выражается формулой (β – удельная крутизна)

$$g_{D0} = \beta (V_{GS} - V_T),$$

то формула (5.15.2) дает такой же качественный вид зависимости крутизны от затворного напряжения, как и в формуле (5.3.1). Соотношение (5.3.1) часто связывают с уменьшением подвижности за счет увеличения прижимающего тока. Действительно, эти эффекты невозможно отличить, основываясь только на виде единственной ВАХ. Для однозначного разделения этих эффектов требуются, как правило, громоздкие исследования, например, получение ВАХ нескольких транзисторов, изготовленных по одной технологии, но имеющих разную ширину каналов.

Подобное смешение физических и схемотехнических или чисто приборных эффектов является типичным в микроэлектронике, что существенно затрудняет анализ и адекватное моделирование работы приборов.

5.16. Паразитные емкости стока и истока

В идеале заряды на затворе и на стоке не должны зависеть друг от друга. В реальных транзисторах они связаны через взаимные паразитные емкости (рис. 5.20).

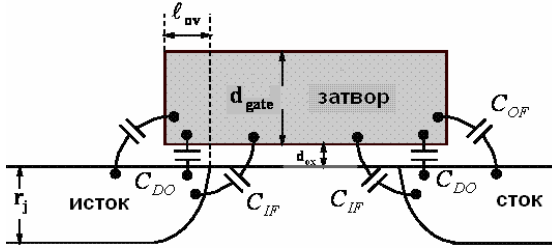


Рис. 5.20. Схема паразитных емкостей перекрытия МОПТ

Емкость прямого перекрытия (direct overlap)

$$C_{DO} = W l_{OV} C_O = \frac{\varepsilon_i W l_{OV}}{d_{ox}}, \quad (5.16.1)$$

где l_{OV} – длина перекрытия между (и)стоком и затвором, W – ширина канала.

Дополнительные два слагаемых обусловлены наличием краев и соответствующими поправками к формуле для плоского конденсатора.

Емкость внешних краев (outer fringe)

$$C_{OF} = \frac{2\varepsilon_{spacer}\varepsilon_0 W}{\pi} \ln\left(1 + \frac{d_{gate}}{d_{ox}}\right). \quad (5.16.2)$$

Емкость внутренних краев (inner fringe)

$$C_{IF} = \frac{2\varepsilon_{Si}\varepsilon_0 W}{\pi} \ln\left(1 + \frac{x_J}{2d_{ox}}\right). \quad (5.16.3)$$

Обратите внимание, что силовые линии взаимной электростатической связи в одном случае идут через спейсер (краевую изоляцию затвора), а в другом – через кремний, что выражается в разных значениях диэлектрической проницаемости.

Для типичных значений $d_{gate}/d_{ox} \approx 40$ и $x_J/d_{ox} \approx 20$, $l_{OV} \sim (2...3)d_{ox}$, получаем оценку $C_{OF}/W \approx 0.08$ фФ/мкм и $C_{IF}/W \approx 0.16$ фФ/мкм (в закрытом режиме). Полная паразитная емкость перекрытия

$$C_{OV}(V_G = 0) = C_{DO} + C_{OF} + C_{IF} \sim 10 \varepsilon_i \varepsilon_0 W \cong 0.3 \text{ (фФ/мкм)} \times W. \quad (5.16.4)$$

6. Эффекты сильных электрических полей

6.1. Квазидвумерная модель распределения сильных электрических полей в районе стока

При $V_{DS} > V_{DSAT}$ неоднородность поперечного поля становится сравнимой с неоднородностью продольного; в области стока нарушается приближение плавного канала (gradual channel approximation, GCA). Часть силовых линий из стока замыкается не на исток, а на затвор (рис. 6.1).

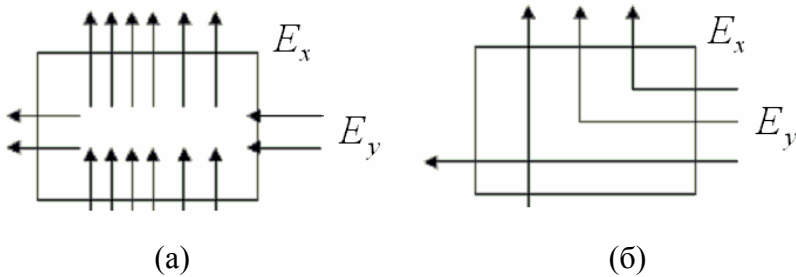


Рис. 6.1. Распределение силовых линий электрического поля в области вблизи стока: (а) в приближении плавного канала; (б) схематичное представление реального распределения

В области отсечки нужно решать 2D уравнение Пуассона (хотя бы приближенно). Характерные длины задачи ΔL – по горизонтали, x_j (глубина залегания pn -перехода стока) – по вертикали. Силовые линии веером расходятся из области канала, замыкаясь на положительном заряде в стоке. Поэтому электроны из канала распространяются по глубине залегания pn -перехода, так что их объемная концентрация становится больше концентрации акцепторов в низколегированной области стока.

Приближенное решение двумерного уравнения Пуассона будем искать в области (боксе) ($\Delta L \times x_j$) (рис. 6.2):

$$\frac{\partial E_y}{\partial y} + \frac{\partial E_x}{\partial x} \cong -\frac{q n_s}{x_j \epsilon_s} - \underbrace{\frac{q N_A}{\epsilon_s}}_{N_A \ll n_s / x_j} . \quad (6.1.1)$$

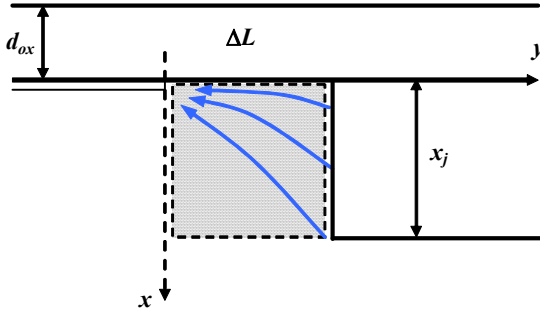


Рис. 6.2. Область отсечки в районе стока, в которой приближенно решается двумерное уравнение Пуассона

В качестве граничных условий примем $E_x(x = -x_j) = 0$ на нижней границе области и $\varepsilon_i E_{ox} = \varepsilon_S E_S$ на границе раздела Si-SiO₂. Для того чтобы двумерную задачу приближенно свести к одномерной, оценим одно из слагаемых уравнения Пуассона следующим образом:

$$\frac{\partial E_x}{\partial x} \cong - \frac{\overbrace{(\varepsilon_i / \varepsilon_S) F_{ox}}^{E_S}}{x_j} = - \frac{V_g - \varphi_{MS} - 2\varphi_F - \varphi}{\underbrace{(\varepsilon_S / \varepsilon_i) d_{ox} x_j}_{=l^2}}, \quad (6.1.2)$$

где

$$l \equiv [(\varepsilon_S / \varepsilon_i) x_j d_{ox}]^{1/2} \quad (6.1.3)$$

характерная длина задачи (SPICE параметр lint).

Плотность электронов в районе отсечки оценивается как

$$\frac{q n_S}{x_j \varepsilon_S} = \frac{C_o (V_G - \varphi_{ms} - 2\varphi_F - V_{DSAT})}{x_j \varepsilon_S} = \frac{V_G - \varphi_{ms} - 2\varphi_F - V_{DSAT}}{l^2}. \quad (6.1.4)$$

Таким образом, двумерная задача свелась к обыкновенному дифференциальному уравнению

$$\frac{d^2 \varphi}{dy^2} = \frac{\varphi - V_{DSAT}}{l^2}, \quad (6.1.5)$$

имеющему общее решение

$$\varphi(y) - V_{DSAT} = C_1 \sinh\left(\frac{y}{l}\right) + C_2 \cosh\left(\frac{y}{l}\right). \quad (6.1.6)$$

Используя граничные условия в начале области отсечки ($y = 0$)

$$\left. \frac{d\varphi}{dy} \right|_{y=0} = E_{SAT}; \quad \varphi(y=0) = V_{DSAT}, \quad (6.1.7)$$

получаем распределение потенциала в области отсечки

$$\varphi(y) = V_{DSAT} + E_{SAT}l \sinh\left(\frac{y}{l}\right). \quad (6.1.8)$$

Записывая отсюда падение потенциала в области отсечки

$$V_{DS} - V_{DSAT} = E_{SAT}l \sinh\left(\frac{\Delta L}{l}\right), \quad (6.1.9)$$

получаем длину области отсечки как функцию смещения V_{DS}

$$\Delta L = l \sinh^{-1}\left(\frac{V_{DS} - V_{DSAT}}{E_{SAT}l}\right). \quad (6.1.10)$$

6.2. Моделирование максимальных электрических полей в канале МОПТ

Максимальное поле в районе стока приводит к нежелательным эффектам ударной ионизации, зарядке окисла и образованию горячих носителей.

Используя (6.1.8) и (6.1.10), получаем выражение для электрического поля в области отсечки

$$E(y) = -\frac{d\varphi(y)}{dy} = E_{SAT} \cosh\left(\frac{y}{l}\right) = E_{SAT} \sqrt{\left(\sinh\left(\frac{y}{l}\right)\right)^2 + 1}, \quad (6.2.1)$$

$$\sinh \frac{y}{l} = \frac{\varphi(y) - V_{DSAT}}{E_{SAT}l}. \quad (6.2.2)$$

Отсюда можно получить формулу для максимального электрического поля в конце канала E_m

$$E_m = E_{SAT} \sqrt{\left(\frac{V_{DS} - V_{DSAT}}{E_{SAT}l}\right)^2 + 1}. \quad (6.2.3)$$

Если выполняется условие $(V_{DS} - V_{DSAT})/E_{SAT}l \gg 1$, то соотношение (6.2.3) записывается в совсем простой форме

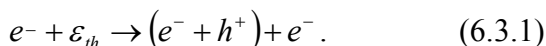
$$E_m \cong \frac{V_{DS} - V_{DSAT}}{l}, \quad (6.2.4)$$

где характерная длина экранирования в районе отсечки l выражается формулой (6.1.3).

6.3. Горячие носители

В сильных электрических полях каналов n -МОПТ небольшая часть электронов может набрать достаточно большую энергию ($\sim 1..3$ эВ). Неравновесные электроны с большой энергией (т.н. «горячие носители») могут быть источником дефектообразования на границе раздела Si-SiO₂ (при энергиях 2.0...2.5 эВ). То же самое, но в значительно меньшей степени (из-за меньшей подвижности дырок), имеет место в дырочных каналах p -МОПТ. Заряженные дефекты на границе раздела, образованные горячими носителями, так же, как и радиационные дефекты, приводят к деградации параметров транзисторов, а именно: (а) сдвигам порогового напряжения из-за зарядки окисла отрицательным зарядом (в этом главное отличие от случая радиации); (б) уменьшению подвижности эффекта поля и соответственно быстродействия за счет образования поверхностных состояний.

Другими эффектами горячих носителей являются процессы ударной ионизации и размножения носителей в районе стока по схеме



Пороговая энергия ударной ионизации ($\sim 1.5 E_G$) находится в диапазоне $\varepsilon_{th} \sim 1.1...1.8$ эВ.

В конечном итоге эффект ударной ионизации приводит (а) к появлению заметного тока в подложку (рис. 6.3); (б) инжекции электронов в подзатворный окисел и увеличению тока затвора.

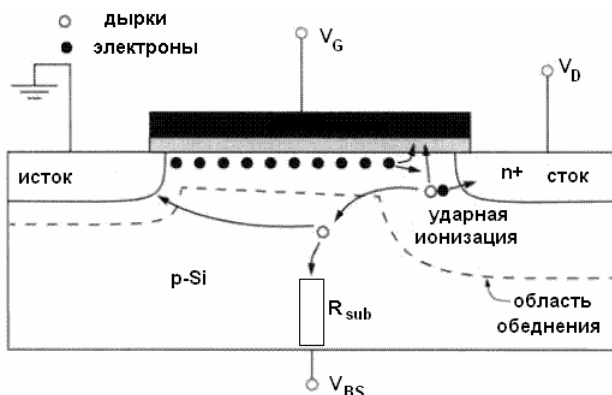


Рис. 6.3. Схема образования паразитного тока подложки при ударной ионизации в районе стока n -МОПТ для КМОП объемной технологии

Кроме того, паразитный дырочный ток в подложку может приводить к нежелательному включению паразитной биполярной структуры (см. пп. 5.11 и 8.6).

Необходимо упомянуть полезное использование горячих носителей. Как мы уже отмечали, электроны с энергией больше разрыва краев зон проводимости окисла SiO_2 и кремния (~ 3.1 эВ) могут инжектироваться в окисел. Этот эффект используется для перезаписи информации в плавающих затворах EEPROM (флэш-памяти).

6.4. Методы борьбы с горячими носителями

Деградация прибора, вызванная горячими носителями, приводит к уменьшению срока службы транзистора. Уменьшение электрического поля в канале в районе стока увеличивает надежность прибора. Для этой цели используются слаболегированные стоки (lightly doped drain, LDD). Целью использования слабого легирования стока является уменьшение пика электрического поля в районе стока.

Сильное электрическое поле канала практически не проникает в сильнолегированную область стока ($> 10^{20} \text{ см}^{-3}$) и резко спадает на очень малых длинах экранирования (~ 1 нм). Если в районе стока внедрить небольшую слаболегированную область ($\sim 2 \times 10^{15} \text{ см}^{-3}$), то поле начнет спадать не так резко, увеличивая эффективную длину канала и, соответственно, уменьшая свои пиковые значения в конце канала (рис. 6.4).

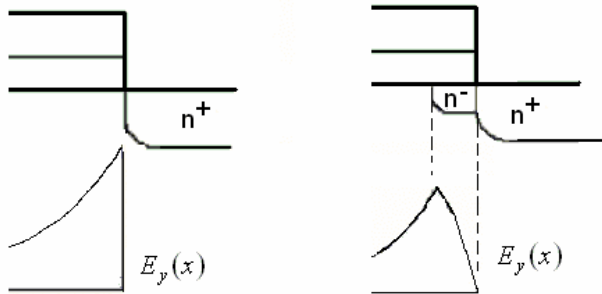


Рис. 6.4. Схема распределения электрического поля вдоль канала, иллюстрирующая уменьшение пикового значения электрического поля в слаболегированном стоке

Эмпирически эффект снижения максимального электрического поля в конце канала при использовании слаболегированного участка стока часто описывают формулой

$$E_m \cong \frac{V_{DS} - V_{DSAT}}{l + l_{LDD}}. \quad (6.4.1)$$

Как упоминалось выше, тонкие расширения стока улучшают также электростатические характеристики транзистора (см. п. 5.12). Расчеты показывают, что оптимальной структурой, с точки зрения минимизации электрических полей, является расположение LDD областей под затвором. Если LDD область вылезает из-под затвора, то структура плохая. Недостатком использования LDD является некоторое увеличение последовательного сопротивления стока и истока.

Другим эффективным методом борьбы с горячими носителями является понижение напряжения питания в духе общих идей скейлинга (см. главу 2). Нужно отметить, что негативные эффекты от горячих носителей существенно снизились (но не исчезли), когда напряжение питания стало меньше барьера между краями зоны проводимости в окисле и кремнии (~ 3.1 эВ).

Наконец, еще одним методом повышения устойчивости приборов к воздействию горячих носителей является использование в качестве подзатворных изоляторов более

устойчивых к дефектообразованию материалов (например, т.н. оксинитридов кремния $\text{SiO}_x\text{N}_{2-x}$).

6.5. Разогрев носителей и «удачливые» (lucky) электроны

В электрическом поле вдоль канала транзистора электроны набирают энергию и передают ее фононам, что соответствует переходу электрической энергии в тепло. В сильном электрическом поле повышается не только средняя кинетическая энергия (т.е. температура) колебаний атомов решетки T_L , но еще в большей степени и температура электронов T_e .

Тем не менее, проблему представляет не большинство электронов, обладающее средней энергией (пускай и повышенной по сравнению с равновесной температурой), а та небольшая доля электронов, сумевших получить аномально большую (по сравнению со средней) кинетическую энергию. Именно такие электроны могут вызывать процессы ударной ионизации, инжекции в окисел и дефектообразования.

Доля носителей с энергией превышающей $\varepsilon_{th} \sim 1.5$ эВ в случае равновесного максвелловского распределения ничтожна. В сильном поле распределение электронов по энергии становится сильно анизотропным, т.е. в распределении появляется выраженный «хвост» высокоэнергетических электронов (рис. 6.5).

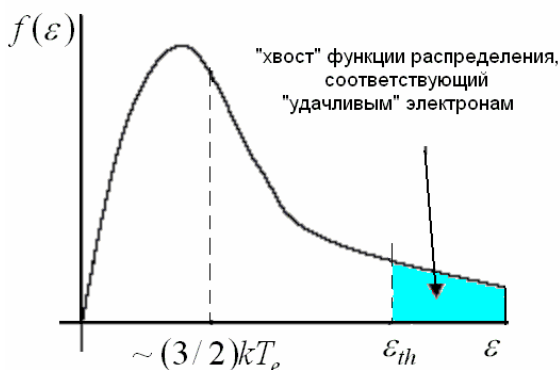


Рис. 6.5. Энергетическое распределение электронов в сильных электрических полях

Вклад в ударную ионизацию вносят только электроны из высокоэнергетического «хвоста» распределения (заштрихованная область на рис. 6.5) Формирование этого «хвоста» носит стохастический (т.е. случайный) характер. Дело в том, что скорость процесса потери электроном энергии лимитируется стохастическими процессами рассеяния на т.н. оптических фононах. Среднее количество актов рассеяния на длине канала L современного транзистора (< 100 нм) оценивается по формуле L/l_ε (l_ε – длина пробега электрона по энергии, см. главу 1) и измеряется единицами. Для пуассоновского процесса со средним числом событий ~ 1 характерны большие относительные флуктуации. Это означает, что вероятность электрона пройти весь канал без потери энергии достаточно велика. Если напряжение между стоком и истоком превосходит некоторую пороговую энергию ε_{th} ($qV_{DS} > \varepsilon_{th}$), то горячие (высокоэнергетические) носители могут вызывать разнообразные процессы деградации.

Для приближенного количественного описания роли горячих носителей до сих пор важную роль играет введенная Шокли концепция «удачливых» (lucky) носителей, т.е. носителей, избежавших столкновений с потерей энергии до того, как они успели набрать в электрическом поле заданную энергию. Если средняя длина пробега по энергии l_ε , то пуассоновская вероятность отсутствия столкновения до набора энергии ε в электрическом поле E равна

$$p = \exp\left(-\frac{\varepsilon}{qEl_\varepsilon}\right). \quad (6.5.1)$$

Этим же выражением можно оценить долю носителей с энергией, превышающей ε .

6.6. Моделирование ударной ионизации в канале

При ударной ионизации в канале n -МОПТ электроны уходят в сток, а дырки инжектируются в подложку, формируя дырочный ток I_{sub} .

Уравнение непрерывности для тока электронов в канале с учетом рождения электронов в канале за счет ударной ионизации имеет вид:

$$\frac{dI_n}{dy} = \alpha(E(y))I_n, \quad (6.6.1)$$

где $\alpha(E)$ ([1/см]) – зависящая от электрического поля вероятность процесса ударной ионизации на единицу длины канала.

Зависимость вероятности ударной ионизации от электрического поля можно записать, основываясь на модели удачливых носителей:

$$\alpha(E) = A_i \exp\left(-\frac{B_i}{E}\right), \quad (6.6.2)$$

где эмпирические константы для электронов $A_i \cong 2 \times 10^6$ см⁻¹, $B_i \cong 1.7 \times 10^6$ В/см. Для дырок B_i существенно больше (поскольку подвижность дырок меньше); поэтому эффект горячих носителей для дырок гораздо слабее (рис. 6.6). Решение уравнения (6.6.1) имеет вид

$$I_n(L) = I_n(y=0) \exp\left(\int_0^L \alpha(y') dy'\right). \quad (6.6.3)$$

Как видно из (6.6.3), ток в конце канала больше тока в его начале. Здесь нас будет интересовать только случай, когда показатель экспоненты в (6.6.3) много меньше единицы. Этот случай слабого размножения очень далек, например, от пробоя, вызванного ударной ионизацией.

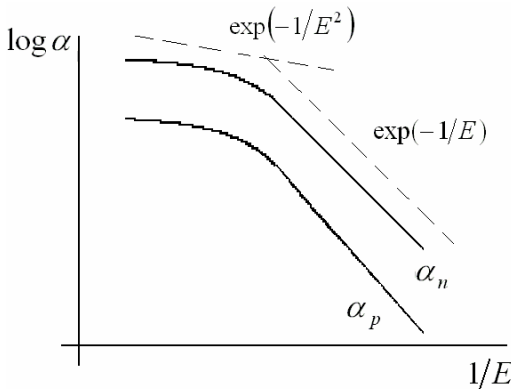


Рис. 6.6. Зависимость вероятности коэффициента ударной ионизации для электронов (α_n) и дырок (α_p)

В указанном случае соотношение (6.6.3) можно приближенно представить в форме суммы тока канала и тока подложки

$$I_n(L) \cong I_d(0) \left(1 + \int_0^L \alpha(y') dy' \right) \cong I_d(0)(1 + M) \cong I_n + I_{sub}, \quad (6.6.4)$$

где введена величина безразмерного фактора размножения

$$M \equiv \int_0^L \alpha(y') dy'. \quad (6.6.5)$$

Действительно, вторая составляющая электронного тока в канале в (6.6.4) в точности равна дырочному току, уходящему в контакт подложки,

$$I_{sub} = I_{D0} \int_0^L \alpha(y') dy' = M I_{D0}. \quad (6.6.6)$$

Используя результаты для распределения полей в районе стока (см. п. 6.1), получаем

$$E(y) = E_{SAT} \cosh\left(\frac{y}{l}\right) \propto e^{y/l}, \quad \frac{dE(y)}{dy} \cong \frac{E(y)}{l}. \quad (6.6.7)$$

Эти формулы дают возможность рассчитать коэффициент умножения тока

$$\begin{aligned} M &= \int_0^L \alpha\left(\frac{B_i}{E}\right) dy = \int \alpha\left(\frac{B_i}{E}\right) \frac{d(1/E)}{\left(-\frac{1}{E^2}\right) \frac{dE}{dy}} = \\ &= \int \alpha\left(\frac{B_i}{E}\right) l \frac{d(1/E)}{(-1/E)} \cong \frac{\overbrace{E_m}^{\max E} l}{B_i} \int_{E_m} A_i \exp\left(-\frac{B_i}{E}\right) d(B_i/E) \cong \\ &\cong \frac{E_m l}{B_i} A_i \exp\left(-\frac{B_i}{E_m}\right). \end{aligned} \quad (6.6.8)$$

Наконец, используя упрощенное выражение для максимального электрического поля в конце канала (6.2.4), получаем конечную формулу для тока ударной ионизации в подложку

$$I_{SUB} = I_D \underbrace{\left(\frac{A_i}{B_i}\right)}_{\sim 1.2 B^{-1}} (V_D - V_{DSAT}) \exp\left(-\frac{B_i l}{V_D - V_{DSAT}}\right). \quad (6.6.9)$$

Эта формула хорошо описывает экспериментальную зависимость тока подложки от электрических смещений на затворе и стоке (рис. 6.7).

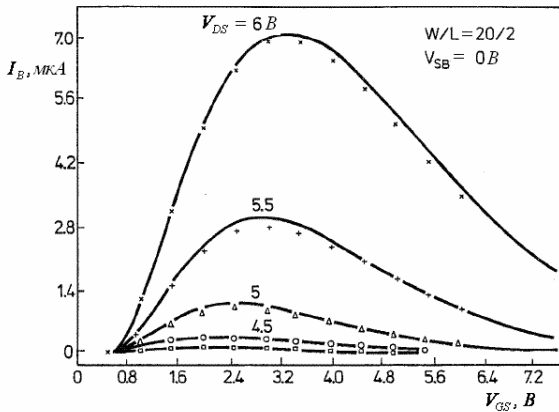


Рис. 6.7. Зависимости тока подложки от затворного напряжения при разных смещениях на стоке

Эта зависимость имеет ярко выраженный максимум. При малых V_{GS} ток подложки мал, поскольку мал ток канала, а при увеличении V_{GS} при заданном V_{DS} ток подложки снижается из-за уменьшения электрического поля в районе стока.

6.7. Влияние тока подложки на работу МОПТ

Ток, текущий в подложку, может существенно повлиять на работу транзистора. Это связано с тем, что существует несколько механизмов уменьшения эффективного порогового напряжения за счет дырок, инжектированных в подложку.

В КМОП схемах объемных технологий (когда дырочный ток уходит на контакт подложки) реализуется резистивный механизм (рис. 6.8).

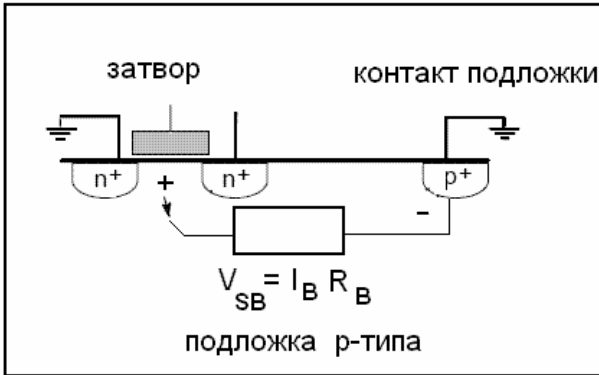


Рис. 6.8. Схема резистивного механизма понижения эффективного порогового напряжения МОПТ

Этот механизм обусловлен тем, что резистивное падение напряжения между активной областью канала и контактом подложки, с электрической точки зрения, прикладывает прямое смещение между истоком и подложкой. Это, с одной стороны, может приводить к открыванию перехода исток-подложка; с другой стороны, это уменьшает пороговое напряжения транзистора.

В схемах КНИ («кремний-на-изоляторе») технологий, в которых активная область (т.н. «тело») КНИ транзистора изолирована от подложки слоем скрытого окисла, реализуется емкостной эффект. Положительный заряд дырок накапливается в теле КНИ транзистора, повышая его потенциал, что эквивалентно приложению положительного смещения на тело относительно истока.

Это приводит к уменьшению порогового напряжения и появлению на выходных ВАХ характерных изломов, т.н. «кинков» (рис. 6.9).

В любом случае инжекция дырок в подложку может приводить к запуску механизма паразитного биполярного транзистора. Если прямое смещение между подложкой и истоком оказывается $V_{BS} \geq 0.6$ В, то электроны начинают инжектироваться из истока и через подложку течь в сток, увеличивая ток стока и уменьшая выходное сопротивление R_{OUT} .

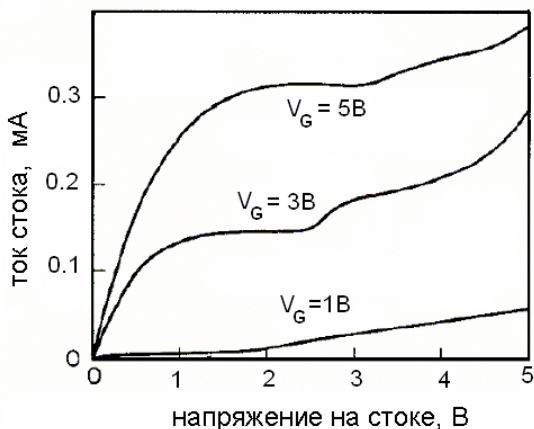


Рис. 6.9. Кинк-эффект на выходных ВАХ КНИ МОПТ для разных значений напряжения на затворе

Следует отметить, что МОПТ можно рассматривать как плохой (низкий коэффициент усиления) биполярный транзистор с длиной базы, приблизительно соответствующей длине канала $L_B \cong L_{eff}$ в режиме отсечки.

6.8. Влияние горячих носителей на срок службы МОПТ

Проблема предсказания процессов деградации от горячих носителей состоит в том, что эти процессы очень медленные. Для ускорения эксперимента необходимо максимизировать скорость деградации. Проблема в том, что измерять скорость деградации сложно, но можно легко измерять ток подложки. Хотя ток подложки не влияет напрямую на деградацию, но он является ее индикатором.

Энергия активации ударной ионизации составляет $\varepsilon_i \sim 1.4$ эВ, поэтому ток подложки из-за ударной ионизации можно оценить по формуле для «удачливых» носителей

$$I_{SUB} = C I_D \exp\left(-\frac{B_i}{E_m}\right) = C I_D \exp\left(-\frac{\overbrace{\varepsilon_i}^{\sim 1.4 \text{ эВ}}}{\underbrace{\lambda_e}_{8 \text{ нм}} E_m}\right), \quad (6.8.1)$$

где λ_e – длина пробега по энергии в кремнии ($\sim 7-8$ нм), E_m – максимальное значение электрического поля в канале, C – эмпирическая константа.

Энергия активации для заброса электронов в зону проводимости окисла $\varepsilon_B \sim 3.1$ эВ, поэтому ток в затвор можно оценить с помощью такого же подхода, как и в (6.8.1):

$$I_{gate} = C' I_D \exp\left(-\frac{\varepsilon_B}{\lambda_e E_m}\right). \quad (6.8.2)$$

Конечно же, из-за большего значения энергетического барьера активационный ток затвора на несколько порядков ниже, чем ток подложки. Обычно затворный ток оказывается меньше, чем $10^{-11} \dots 10^{-12}$ А, так что его трудно измерить.

Скорость деградации транзистора в сильных полях – это скорость генерации дефектов на границе раздела Si-SiO₂ («поверхностных состояний»). Энергия активации для процесса генерации дефектов на границе раздела окисла $\varepsilon_{IT} \sim 4 \dots 5$ эВ. Тогда скорость генерации поверхностных состояний G_{it} будет степенным образом зависеть от тока подложки

$$G_{IT} \propto \exp\left(-\frac{\varepsilon_{IT}}{\lambda_e E_m}\right) \propto \exp\left(-\frac{\varepsilon_i}{\lambda_e E_m}\right)^{\frac{\varepsilon_{IT}}{\varepsilon_i}} \propto (I_{SUB})^{\varepsilon_{IT}/\varepsilon_B} \approx (I_{SUB})^3. \quad (6.8.3)$$

Отказ (обычно параметрический, а не функциональный) наступает, когда количество поверхностных состояний достигает некоторого критического значения Q_{ITcrit} . Тогда срок службы МОПТ за счет деградации горячими носителями τ_{HCE} можно оценить с помощью полуэмпирического выражения с некоторой эмпирической константой C

$$\tau_{HCE} = \frac{Q_{ITcrit}}{G_{IT}} = C I_{SUB}^{-3}. \quad (6.8.4)$$

В статическом режиме (DC) ток в КМОП схемах не течет, поэтому нет никаких эффектов горячих носителей. Деградация имеет место только в момент переключения, когда через структуру течет ток и в каналах обоих транзисторов большие электрические поля.

Срок службы в режиме высокочастотных переключений $\tau_{HCE}(AC)$ можно оценить, используя величину доли времен переключения $duty (< 1)$ в полном цикле функционирования,

$$\tau_{HCE}(AC) = \frac{\tau_{HCE}(DC)}{duty}. \quad (6.8.5)$$

Эффекты горячих носителей по-разному проявляют себя в каналах разного типа.

(1) Эти эффекты заметно сильнее выражены в n -канальных МОПТ по сравнению с p -канальными МОПТ, что обусловлено разной величиной подвижности в канале.

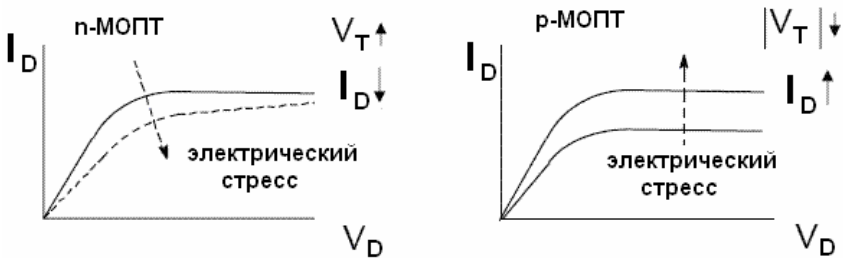


Рис. 6.10. Качественный характер деградации от горячих носителей в n -канальных (слева) и p -канальных (справа) транзисторах

(2) Для транзисторов обоих типов эффекты горячих носителей чаще всего приводят к отрицательной зарядке окисла. Поэтому пороговое напряжение смещается в сторону более положительных значений. Для n -канальных приборов это уменьшает ток, а для p -канальных – увеличивает (рис. 6.10).

(3) Для очень коротких приборов (< 0.18 мкм) начинается также инжекция дырок в окисел, что может привести к сдвигу порога в положительную сторону и уменьшению тока стока.

6.9. Методика прогнозирования срока службы транзистора по отношению к воздействию горячих носителей

Выберем критерий параметрического отказа в виде изменения тока насыщения на 10 % ($|\Delta I_{DSAT}|/I_{DSAT} > 0.1$) и измерим срок

службы в наиболее жестком электрическом режиме. Наиболее жесткий электрический режим можно определить по пику тока подложки (рис. 6.7), причем для уменьшения времени эксперимента напряжения могут быть выбраны выше номинального напряжения питания V_{DD} . Ток подложки играет роль индикатора эффективности генерации горячих носителей, хотя энергия активации ударной ионизации в несколько раз меньше энергии дефектообразования.

Строим двойной логарифмический график времени параметрического отказа от значения тока подложки (рис. 6.11). Эта экспериментальная зависимость имеет линейный характер (см. также 6.8.4), что позволяет провести экстраполяцию на номинальное значение напряжения питания, соответствующее более мягким условиям.

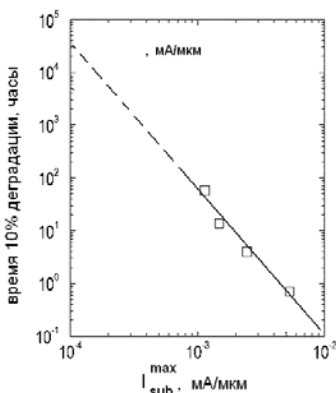


Рис. 6.11. Двойная логарифмическая зависимость времени 10 %-ной деградации крутизны n -МОПТ от максимального тока подложки

Технологическими методами увеличения срока службы транзистора являются:

использование подзатворных изоляторов, более устойчивых к возникновению дефектов ($\epsilon_{IT}/\epsilon_i > 3$) – нитридизация окисла кремния, оксинитриды SiO_xN_{1-x} ;

использование низколегированных стоков (LDD) для снижения максимальных значений электрических полей канале транзистора (см. п. 6.6).

Эффекты горячих носителей в очень малых МОПТ выражены слабо по следующим причинам:

1) снижение напряжения питания приводит к тому, что скорость инъекции носителей в окисел и создания дефектов также уменьшается из-за недостаточности энергии носителей;

2) очень малая толщина подзатворного изолятора приводит к тому, что вероятность дефектообразования уменьшается;

3) с другой стороны, в наноразмерных транзисторах даже несколько дефектов могут являться причиной заметной деградации прибора.

7. Диффузионно-дрейфовая модель тока в МОПТ

7.1. Введение

Существующие простейшие модели ВАХ МОПТ основаны идеологически на формальном интегрировании выражения для омического тока, игнорируя последовательный учет диффузионного тока, что приводит к необходимости кусочного вида описания вольт-амперной характеристики отдельно для надпорогового и подпорогового участка, а также для линейного режима и режима насыщения тока транзистора (см. главу 4). Последовательное описание ВАХ с необходимостью требует решения уравнения непрерывности для тока в канале. Действительно, уравнение непрерывности для тока является неотъемлемой частью т.н. фундаментальной системы уравнений, включающей в себя также уравнение Пуассона, граничные условия и уравнения переноса носителей тока. Например, теория биполярного транзистора изначально базируется на решении уравнений непрерывности для тока неосновных носителей в базе. В этой главе описан подход, основанный на явном аналитическом решении для уравнения непрерывности тока в канале, что позволяет получить в компактном виде замкнутые выражения для тока, а также для распределения электрических полей, потенциалов и плотности носителей в канале при любом напряжении на затворе. Для этого необходимо более детально рассмотреть задачу о распределении электрохимического потенциала в канале транзистора.

7.2. Электрохимический потенциал в канале МОПТ

Хорошо известно, что разность напряжений между контактами представляется разностью электрохимических потенциалов. Электрохимический потенциал является суммой электрического и химического потенциалов и его распределение отслеживает распределение локального уровня Ферми. В частности, распределение электрохимического потенциала вдоль канала можно записать в виде (здесь используется обозначение $\zeta(y) = |\zeta(y)| > 0$)

$$\mu(y) = q\phi_s(y) + q\zeta(y). \quad (7.2.1)$$

При $V_{DS} = 0$ электрохимический потенциал в кремнии постоянен $\mu(x, y) = \text{const}$. При ненулевом смещении между стоком и истоком $V_{DS} > 0$, изменение электрохимического потенциала вдоль канала, как уже обсуждалось в главе 4, есть локальное значение обратного смещения между каналом и подложкой (рис. 7.1)

$$\mu(y)/q = V_{CB}(y) = \varphi_S(y) + \zeta(y).$$

Отметим, что в физической литературе под химическим потенциалом чаще всего подразумевают величину со знаком, противоположным знаку в выражении (7.2.1). В этом случае химический потенциал оказывается положительной величиной для вырожденных систем и отрицательной для невырожденных.

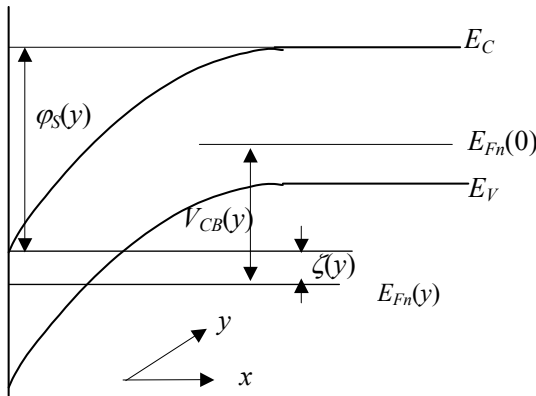


Рис. 7.1. Зонная диаграмма n -МОПТ, показывающая локальное значение электрохимического потенциала в точке y по длине канала

В этой главе мы будем рассматривать канал МОПТ в приближении невырожденного электронного газа, при котором удобнее пользоваться соглашением о знаке химического потенциала, определенного как в (7.2.1). Кроме того, следует иметь в виду, что в литературе очень часто электрохимический потенциал μ (стандартное обозначение) называют просто химическим потенциалом.

Изменение поверхностного электрического потенциала характеризуется изменением положения краев зон

$$\varphi_S(y) = -E_C(y)/q, \quad (7.2.2)$$

а изменение химического потенциала вдоль канала характеризуется изменением объемной концентрации электронов вдоль канала

$$n = N_C \exp(-(E_C(x) - E_{Fn})/kT) \equiv N_C \exp(-\zeta(y)/kT). \quad (7.2.3)$$

Приложенное внешнее напряжение между стоком и истоком приводит, в общем случае, как к изменению электрического потенциала (электрических полей), так и к изменению химического потенциала (т.е. к пространственному перераспределению плотности носителей), но с выполнением условия

$$V_{DS} = \varphi_S(L) - \varphi_S(0) + \zeta(L) - \zeta(0). \quad (7.2.4)$$

Хорошо известно, что величина тока определяется градиентом электрохимического потенциала μ . Поэтому в отсутствие тока оба слагаемых не являются независимыми величинами и их пространственные изменения в точности компенсируют друг друга, как это имеет место в равновесном p -переходе. В неравновесных условиях пространственные изменения химического и электрического потенциала в МОП транзисторах связаны между собой только условием электронейтральности и могут управляться за счет изменения напряжения на внешних электродах.

Отметим принципиальную разницу в распределении электрохимического потенциала в однородном проводнике и канале МОП транзистора, а также в большинстве других нанoeлектронных структур. В однородном полупроводнике (либо в металле) концентрация носителей вдоль длины образца и, соответственно, химический потенциал остаются постоянными, поскольку приложенное напряжение никак не влияет на объемную локальную электронейтральность (рис. 7.2, а).

В приборных структурах электронейтральность обеспечивается за счет перераспределения поверхностной плотности заряда разного знака в планарных слоях. Поэтому, как мы уже отмечали выше, в канале n -МОПТ плотность электронов уменьшается по мере приближения к стоку и химический потенциал изменяется (рис. 7.2, б).

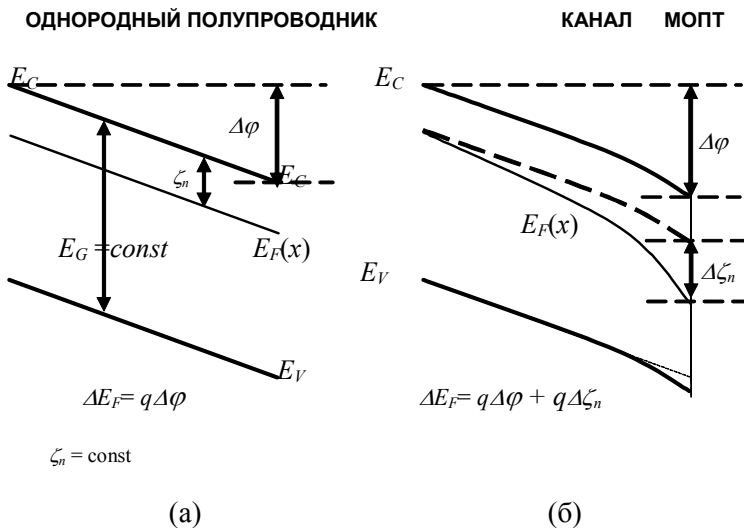


Рис. 7.2. Распределение электрических и химических потенциалов в однородном полупроводнике (а) и в канале n -МОПТ (б)

Вообще говоря, абсолютное изменение химического потенциала от стока до истока при любом режиме работы транзистора невелико и не превышает нескольких φ_T . Именно по этой причине в традиционных моделях не делается различия между полным электрохимическим и электрическим потенциалом. Однако игнорирование изменений химического потенциала вдоль канала МОПТ приводит к фактическому игнорированию процессов диффузионного тока, что делает невозможным получить единое выражение для тока, справедливое во всех режимах работы транзистора.

7.3. Полная плотность тока в канале МОПТ

Как известно, полная плотность тока, в том числе и в канале МОПТ J_S ($[J_S] = A/cm$), является суммой дрейфовой и диффузионной составляющих

$$J_S = J_{DIFF} + J_{DR}. \quad (7.3.1)$$

Эту сумму можно записать в виде градиента от электрохимического потенциала и переписать в несколько иной форме:

$$\begin{aligned}
 J_S &= q\mu_n n_S \frac{dE_{Fn}}{dy} = q\mu_n n_S \left(\frac{d\varphi}{dy} + \frac{d\zeta_n}{dy} \right) = \\
 &= q\mu_n n_S \frac{d\varphi}{dy} \left(1 + \frac{d\zeta_n}{d\varphi} \right) \equiv \sigma_S E (1 + \kappa), \quad (7.3.2)
 \end{aligned}$$

где введена σ_S – двумерная (2D) проводимость канала, выражаемая формулой Друде

$$\sigma_S = 1 / \rho_S = q\mu_n n_S = q^2 \tau n_S / m. \quad (7.3.3)$$

Кроме того, в выражении (7.3.3) введен безразмерный параметр κ , характеризующий отношение диффузионной и дрейфовой составляющих тока канала:

$$\kappa \equiv \frac{J_{DIFF}}{J_{DR}}, \quad (7.3.4)$$

который предполагается приближенно постоянным вдоль канала при заданных электрических смещениях на электродах.

7.4. Отношение диффузионной и дрейфовой компонент тока как управляющий параметр

Как видно из (7.3.1) и (7.3.2), отношение диффузионной и дрейфовой составляющих тока (7.3.4) может быть определено через отношение градиентов изменения химического и электрического потенциала вдоль канала

$$\kappa \equiv \frac{J_{DIF}}{J_{DR}} = \left| \frac{d\zeta}{d\varphi_S} \right|. \quad (7.4.1)$$

Этот безразмерный параметр является функцией напряжения на затворе и является ключевым управляющим параметром диффузионно-дрейфовой модели. Для вычисления этого параметра напомним, что в приближении плавного канала вдоль всей его длины имеется

$$V_{GS} = \varphi_{MS} + \varphi_S(y) + \frac{q}{C_O} (n_S(\zeta(y)) + N_A x_d(\varphi_S(y))). \quad (7.4.2)$$

Плотность электронов в канале определяется только локальным значением химического потенциала ζ , все остальные величины зависят только от электрического потенциала φ_S .

При приложении смещения между стоком и истоком распределения электрохимического и электрического потенциалов вдоль канала будут, вообще говоря, отличаться друг от друга, и их можно рассматривать как независимые переменные. Например, локальное значение плотности электронов в канале является функцией только химического потенциала $n_S(\zeta)$ и, вообще говоря, не зависит от электрического потенциала φ_S . Считая химический и электрический потенциалы независимыми переменными, дифференцируем обе стороны (7.4.2) по φ_S и, с учетом того, что левая часть этого соотношения не зависит от координаты вдоль канала ($V_{GS} = \text{const}$), получаем

$$\kappa = \frac{C_O + C_D}{C_{inv}}. \quad (7.4.3)$$

Здесь введена емкость инверсионного слоя (см. п. 3.19)

$$C_{inv} \equiv \left(q \frac{dn_S}{d\zeta} \right). \quad (7.4.4)$$

В случае невырожденного канала мы получили в п. 3.19 для этой величины выражение

$$C_{inv} = \frac{q n_S}{2\varphi_T} \left(1 + \frac{N_A x_d}{n_S + N_A x_d} \right) \cong \begin{cases} \frac{q n_S}{2\varphi_T} = \frac{C_O (V_G - V_T)}{2\varphi_T}, & V_G > V_T, \\ \frac{q n_S}{\varphi_T}, & V_G < V_T. \end{cases} \quad (7.4.5)$$

Тогда для надпорогового режима имеем

$$\kappa \cong 2\varphi_T \frac{1 + C_D/C_O}{V_{GS} - V_T} \cong 2\varphi_T \frac{n}{V_{GS} - V_T} \ll 1, \quad (7.4.6)$$

где n – фактор влияния подложки, определенный в п. 4.11. Иными словами, это означает, что диффузионный ток канала в надпороговом режиме много меньше дрейфового тока ($j_{DR} \gg j_{DIFF}$).

Напротив, в подпороговом режиме ($V_G < V_T$), когда концентрация носителей в канале и C_{inv} очень мала, то из (7.4.3) и (7.4.5) име-

ем $\kappa \gg 1$ и, следовательно, доминирует диффузионный ток ($j_{DR} \ll j_{DIFF}$). Именно такое положение вещей имеет место в идеальных МОПТ. В частности, подпороговый ток почти не зависит от напряжения V_{DS} , поскольку почти целиком определяется диффузионной компонентой.

7.5. Уравнение непрерывности

Ключевым пунктом рассматриваемого подхода является аналитическое решение уравнения непрерывности для плотности тока в канале.

Для стационарного случая уравнение непрерывности имеет вид

$$\frac{dJ_S}{dy} = 0. \quad (7.5.1)$$

В этой главе мы будем считать электроны в канале невырожденными. В этом случае, используя (7.3.2) и считая κ постоянным вдоль канала, из (7.5.1) получаем уравнение для распределения электрического тока вдоль канала

$$\frac{dE}{dy} \cong \frac{\kappa}{\varphi_T} E^2. \quad (7.5.2)$$

Решение этого уравнения с граничным условием $E(y=0) = E(0)$ дает

$$E(y) = \frac{E(0)}{1 - \frac{E(0)}{\varphi_T} \kappa y}, \quad (7.5.3)$$

где электрическое поле в районе истока $E(0)$ подлежит определению из граничного условия.

7.6. Интегральное граничное условие

Одна из стратегий получения вольт-амперной характеристики транзистора состоит в определении зависимости электрического поля в районе истока как функции напряжения между стоком и истоком $E(0) \equiv E(y=0, V_{DS})$.

Зная эту зависимость и зависимость плотности носителей в канале в районе истока как функцию V_{GS} (или φ_S), можно воспользоваться формулой (7.3.2) для вычисления тока в районе истока,

который в силу постоянства тока вдоль канала и будет током всего МОПТ.

Для определения электрического поля у истока используем интегральное граничное условие, которое будет выражать условие (7.2.4),

$$(1 + \kappa) \int_0^L E(y) dy = V_{DS}. \quad (7.6.1)$$

Используя условие (7.6.1) и выражение (7.5.3), можно получить соотношение для электрического поля в районе истока

$$E(0) = \frac{\varphi_T}{\kappa L} \left[1 - \exp\left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T}\right) \right]. \quad (7.6.2)$$

Тогда с учетом (7.5.3) получаем распределение электрического поля вдоль канала, справедливое для всех режимов работы МОПТ,

$$E(y) = \frac{\varphi_T}{\kappa L} \frac{1 - \exp\left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T}\right)}{1 - \frac{y}{L} \left[1 - \exp\left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T}\right) \right]}. \quad (7.6.3)$$

Как видно из формулы (7.6.3), распределение электрического поля вдоль канала сильно зависит от режима работы прибора.

В надпороговом режиме ($\kappa \ll 1$) поле почти постоянно $E(y) \cong E(0) = V_{DS}/L$, в то время как в подпороговом режиме ($\kappa \gg 1$) электрическое поле везде очень мало $\sim \varphi_T/L$ и резко возрастает только ближе к стоку (рис. 7.3).

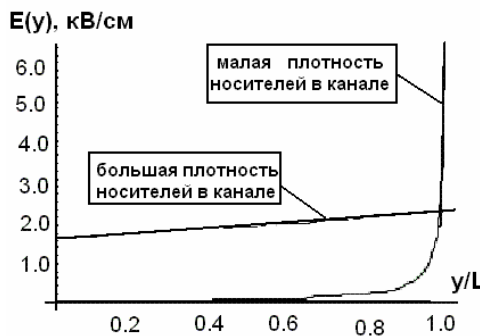


Рис. 7.3. Распределения электрического поля вдоль канала при одинаковой V_{DS} , но для разных значений n_S

Заметное возрастание электрического поля в районе стока имеет место также и в случае насыщения (когда показатель экспоненты в 7.6.3 становится больше единицы).

7.7. Распределение электрического и химического потенциалов вдоль канала

Зная распределение электрического поля, легко получить распределение потенциалов. Интегрируя (7.6.3) по y , получаем распределение электростатического потенциала вдоль канала длиной L

$$\varphi_S(y) - \varphi_S(0) = -\frac{\varphi_T}{\kappa} \ln \left[1 - \frac{y}{L} \left[1 - \exp \left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T} \right) \right] \right] \quad (7.7.1)$$

и распределения химического потенциала вдоль канала

$$\zeta(y) - \zeta(0) = -\varphi_T \ln \left[1 - \frac{y}{L} \left[1 - \exp \left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T} \right) \right] \right]. \quad (7.7.2)$$

Эти соотношения автоматически удовлетворяют условию (7.2.4), которое означает, что полное падение электрохимического потенциала равно напряжению, приложенному между истоком и стоком.

При этом для любого напряжения на затворе V_G (и соответствующего ему κ (см. п. 7.4)) полное падение напряжения между стоком и истоком разделяется на части, обусловленные независимым изменением электрического и химического потенциалов, следующим образом:

$$\varphi(L) - \varphi(0) = \frac{V_{DS}}{1 + \kappa}; \quad \zeta(L) - \zeta(0) = \frac{\kappa V_{DS}}{1 + \kappa}. \quad (7.7.3)$$

Как следует из (7.7.3), полное приложенное напряжение V_{DS} падает в надпороговой области за счет разности электрического потенциала, а в подпороговой – за счет разности химического потенциала (разности плотностей носителей в канале).

7.8. Общее выражение для тока в диффузионно-дрейфовой модели

Как отмечалось выше, если известны плотность электронов у истока $n_S(0)$ и зависимость электрического поля у истока от напряжения V_{DS} , то вольт-амперную характеристику МОПТ можно представить формулой

$$I_{DS} = Z(1 + \kappa)\mu_0 n_S(0)E(y = 0, V_{DS}). \quad (7.8.1)$$

Тогда, используя (7.6.2), общее выражение для тока может быть записано в компактной форме

$$I_D = q \frac{Z}{L} D n_S \frac{1 + \kappa}{\kappa} \left[1 - \exp\left(-\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\phi_T}\right) \right]. \quad (7.8.2)$$

Формула (7.8.2) является общим выражением, справедливым во всех режимах работы МОПТ, непрерывным образом описывающим переход между ними, включая переход от подпороговой области к надпороговой и от линейного тока режима к режиму насыщения.

Все величины, входящие в (7.8.2), можно представить как явные функции поверхностного потенциала ϕ_S . Действительно, задавая значение ϕ_S , можно последовательно рассчитать значения $n_S(\phi_S)$ (3.9.2), $V_G(\phi_S)$ (3.5.6), $C_{inv}(\phi_S)$ (7.4.5) и $\kappa(\phi_S)$ (7.4.3) и, наконец, полного тока (7.8.2). Таким образом, описанная модель принадлежит к классу моделей, основанных на поверхностном потенциале (potential surface PSP), набирающих популярность в последнее время. К этому же классу относится известная модель EKV, рассматриваемая как конкурент моделей BSIM.

Главным достоинством моделей такого типа является возможность непрерывного описания ВАХ от подпороговой к надпороговой области, главным недостатком – отсутствие в этих моделях такого важного параметра как пороговое напряжение. Значение порогового напряжения V_T легко (хотя и не вполне однозначно) определяется экспериментально, что кардинальным образом упрощает идентификацию всех других параметров транзистора.

Описанная в этом разделе диффузионно-дрейфовая модель не имеет указанного недостатка, поскольку все величины в ней рассчитываются как явные функции даже не поверхностного потенциала, а плотности электронов в канале. Это дает возможность ис-

пользовать вместо зависимости от поверхностного потенциала $n_S(\varphi_S)$ (3.9.2) более удобную на практике интерполяционную зависимость от затворного напряжения $n_S(V_G)$ (4.2.2). Последнее означает, что формулу (7.8.2) можно представить как компактную замкнутую и явную зависимость от затворного напряжения.

7.9. ВАХ в надпороговой области

Покажем, что диффузионно-дрейфовая модель воспроизводит результаты стандартных моделей для разных режимов работы транзистора, отличаясь от них единством описания. Напомним, что стандартные модели основаны на кусочном описании разных режимов работы транзистора.

А. Крутая область ВАХ (триодный режим)

В случае сильной инверсии ($n_S > N_A \chi_d$, $V_{GS} > V_T$) имеем хорошие приближения для (7.4.6)

$$qn_S = C_O(V_G - V_T) \text{ и } \kappa \cong 2\varphi_T \frac{n}{V_{GS} - V_T} \ll 1.$$

Тогда, если показатель экспоненты в (7.9.2) меньше единицы:

$$\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{\varphi_T} < 1, \quad (7.9.1)$$

то экспоненту можно разложить до линейного слагаемого, выразив ток транзистора через удельную проводимость однородного двумерного сопротивления (7.3.3):

$$I_D = \frac{Z}{L} q \mu_n n_S V_{DS} \cong \frac{Z}{L} \sigma_S V_{DS}. \quad (7.9.2)$$

В. Режим насыщения

Напряжение насыщения за счет электростатического запираения естественно определить следующим образом:

$$\frac{\kappa}{1 + \kappa} \frac{V_{DSAT}}{\varphi_T} = 2. \quad (7.9.3)$$

Отсюда получаем

$$V_{DSAT} \cong 2\varphi_T \frac{1 + \kappa}{\kappa} \cong \frac{V_{GS} - V_T}{1 + C_D/C_O} = \frac{V_{GS} - V_T}{n} = \frac{V_{GS} - V_T}{1 + \gamma/\sqrt{2}\varphi_F}, \quad (7.9.4)$$

(ср. (4.16.3)).

Если $V_{DS} > V_{DSAT}$, то ток в длинноканальном транзисторе насыщается на уровне

$$I_{SAT} = q \frac{Z}{L} D_n n_s \frac{1 + \kappa}{\kappa} \cong \frac{Z}{L} \frac{\mu_n C_o (V_{GS} - V_T)^2}{2n}, \quad (7.9.5)$$

(ср. (4.14.2)).

Таким образом, общую формулу для тока (7.8.2) в надпороговом режиме можно записать в приближенном виде

$$I_D(V_{GS}, V_{DS}) = \frac{Z}{L} \mu_0 C_o \frac{(V_G - V_T)^2}{2n} \left\{ 1 - \exp \left[-2n \frac{V_{DS}}{V_{GS} - V_T} \right] \right\}. \quad (7.9.6)$$

С учетом определения напряжения и тока насыщения (7.9.4) и (7.9.5) можно представить (7.9.6) в виде еще более простой формулы

$$I_D = I_{SAT} \left(1 - \exp \left[-2 \frac{V_{DS}}{V_{DSAT}} \right] \right), \quad (7.9.7)$$

которая дает правильные значения тока насыщения и проводимости длинноканального транзистора.

Последовательное решение уравнения непрерывности для плотности тока в канале с учетом зависимости подвижности от тянущего электрического поля приводит к неявным формам вольт-амперной характеристики МОП транзистора [7.4]. Физический механизм насыщения тока стока транзистора определяется параметром a , характеризующим степень насыщения скорости носителей

$$a = \frac{\mu_0 (V_G - V_T)}{v_{SAT} L}. \quad (7.9.8)$$

При выполнении условия $a \ll 1$, эффекты насыщения дрейфовой скорости не играют существенной роли и приводят только лишь к перенормировке подвижности

$$\mu(E) \cong \frac{\mu_0}{1 + V_{DS}/E_{SAT} L}. \quad (7.9.9)$$

В этом случае, электрическое поле в районе истока насыщается за счет перераспределения заряда на затворе и насыщается раньше, чем становится заметной нелинейность дрейфовой скорости от поля у истока. В противоположном случае, когда $a \gg 1$, критическое значение продольного поля в окрестности истока достигается в

пределах линейности зависимости $E(0)$ от V_{DS} , и насыщение дрейфовой скорости достигается раньше электростатического за-пираания канала. В промежуточном случае $a \sim 1$ требуется числен-ный расчет.

Тем не менее, существует хорошее приближение, позволяющее записать единое выражение для ВАХ транзистора в замкнутой форме в духе компактного моделирования (ср. 7.9.6)

$$I_D = \frac{W}{L} \mu_0 \frac{C_{ox} (V_G - V_T)^2}{2n} \times \left(1 - \exp \left(- \frac{2n V_{DS}}{V_G - V_T} \frac{1}{(1 + V_{DS}/E_{SAT}L)} \right) \right). \quad (7.9.10)$$

Линейный участок ВАХ для малых значений V_{DS} дается разло-жением экспоненты в ряд Тейлора

$$I_D \cong \frac{W}{L} \mu_0 \frac{C_{ox} (V_G - V_T) V_{DS}}{1 + V_{DS}/E_{SAT}L}, \quad (7.9.11)$$

а тип насыщения определяется, главным образом, длиной канала транзистора. На рис. 7.4. показаны трехмерные зависимости тока стока МОП транзистора, рассчитанные по формуле (7.9.10) как функции напряжений на затворе и стоке для разных длин канала, но при прочих равных параметрах.

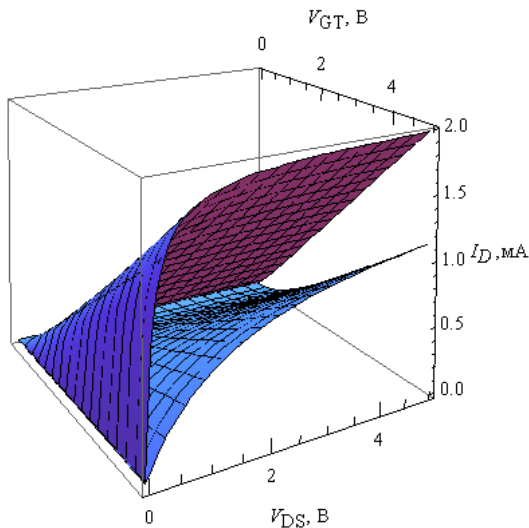


Рис. 7.4. Трехмерные зависимости тока стока I_D как функции затворного напряжения $V_{GT} = V_G - V_T$ и напряжения на стоке V_{DS} , рассчитанные по формуле (7.9.10) для разных длин канала $L = 0.06$ мкм (верхняя поверхность) и 1 мкм (нижняя поверхность) ($d_{ox} = 5$ нм, $W = 1$ мкм, $E_{SAT} = 2 \times 10^4$ В/см, $\mu_0 = 300$ см²/(В с))

Видно, что максимальный ток короткоканального транзистора превышает ток длинноканального транзистора менее чем в два раза при отношении длин каналов $1/0.06 \cong 16$, что является следствием разных механизмов насыщения тока для разных значений L .

7.10. Подпороговый режим

Подпороговый режим, или режим обеднения, соответствует смещениям на затворе ниже порогового значения V_T , где $n_s \ll N_A x_d$ и $\kappa \gg 1$. В этих условиях общая формула (7.8.2) дает

$$I_D \cong q \frac{Z}{L} D n_s \left[1 - \exp\left(-\frac{V_{DS}}{\phi_T}\right) \right]. \quad (7.10.1)$$

Плотность носителей в инверсионном слое в подпороговой области сводится к выражению (см. 4.1.5)

$$qn_S = C_D \varphi_T \exp\left(\frac{\varphi - 2\varphi_F}{\varphi_T}\right) \cong C_D \varphi_T \exp\left(\frac{V_G - V_T(V_{DS})}{m \varphi_T}\right). \quad (7.10.2)$$

Отметим, что в современных транзисторах в формуле (7.10.2) должна быть учтена зависимость порогового напряжения $V_T(V_{DS}) \cong V_T(V_{DS} = 0) - \Delta V_T^{DIBL}(L, V_{DS})$, связанная с эффектом снижения барьера между стоком и истоком (DIBL, см. п. 5.10).

Формула (7.10.1) соответствует чисто диффузионному току, что выражается формальными пределами $\kappa \rightarrow \infty$ и $\kappa/(1+\kappa) \rightarrow 0$. На практике чисто диффузионного тока не бывает даже в подпороговой области, что выражается в конечных значениях κ . Особенно это заметно для транзисторов с большой плотностью поверхностных состояний. В подпороговой области емкость поверхностных состояний, как правило, существенно превосходит емкость инверсионного слоя $q^2 D_{it} \gg C_{inv}$. Заполнение поверхностных состояний вдоль канала определяется здесь распределением химического потенциала. Тогда формула (7.4.3) принимает более точный вид

$$\kappa = \frac{C_O + C_D}{C_{inv} + q^2 D_{it}}, \quad (7.10.3)$$

который отражает общий химический потенциал для подвижных зарядов в канале и локализованных зарядов на дефектах (поверхностных состояниях).

Используя (7.10.3) и определение (4.1.3), можно выразить комбинацию $(1+\kappa)/\kappa$

$$\frac{1+\kappa}{\kappa} = \frac{m}{1 + C_D/C_O} \equiv \frac{m}{1 + \eta_D} \quad (7.10.4)$$

через безразмерную величину m , непосредственно связанную с измеряемой величиной логарифмического размаха напряжений подпорогового ВАХ (см. п. 4.3).

С учетом (7.10.4) общая формула для ВАХ (7.8.2) приобретает в подпороговой области вид

$$I_D^{sub} = q \frac{Z}{L} D_n n_S(V_{GS}, V_{DS}) \frac{m}{1 + \eta_D} \left[1 - \exp\left[-(1 + \eta_D) \frac{V_{DS}}{m \varphi_T}\right] \right]. \quad (7.10.5)$$

Ток в подпороговом режиме насыщается при $V_{DS} > 2m\varphi_T/(1 + \eta_D)$ на уровне

$$\begin{aligned}
 I_{DSAT}^{sub} &= \frac{Z}{L} D_n n_S (V_{GS}, V_{DS}, y=0) \frac{1+\kappa}{\kappa} = \\
 &= \frac{Z}{L} D_n \frac{C_D m \varphi_T}{1+\eta_D} \exp\left(\frac{V_G - V_T(V_{DS})}{m \varphi_T}\right). \quad (7.10.6)
 \end{aligned}$$

При малой плотности поверхностных состояний $q^2 D_{it} \ll C_D$ соотношения (7.10.5) и (7.10.6) совпадают с классической формулой (7.10.1). В приборах с большой плотностью поверхностных состояний (например, подвергнутых ионизирующей радиации или с high-K подзатворным окислом) более точной является формула (7.10.5). Экспериментальное различие будет иметь место в величине напряжения насыщения в подпороговой области. Обычный подход предсказывает напряжения насыщения $\approx (2-3) \varphi_T \cong 50-70$ мВ при комнатных температурах, в то время как формула (7.10.5) предсказывает напряжение насыщения $\sim (2-3) m \varphi_T$. Действительно, область линейности при облучении (и соответствующем увеличении m -фактора) может расширяться до значений ~ 100 мВ.

Наличие широкого диапазона линейной проводимости в подпороговой области говорит о том, что дрейфовая компонента тока даже в глубоком обеднении не бывает, как правило, пренебрежимо мала по сравнению с диффузионной. Согласно (7.7.3) доля дрейфовой компоненты в полном токе равна $(1+\kappa)^{-1}$. При этом κ не бывает, как правило, слишком велика в силу наличия поверхностных состояний, энергетическая плотность которых обычно находится в пределах $10^{10} \dots 10^{12}$ эВ⁻¹ см⁻². Механизм влияния плотности поверхностных состояний на величину доли дрейфового тока качественно можно объяснить следующим образом.

Физической причиной появления диффузионной компоненты тока в канале МОПТ является наличие градиента плотности подвижных носителей в инверсионном слое, навязанное условием локальной электронейтральности вдоль канала. Наличие поверхностных состояний, способных легко обмениваться носителями с каналом, дает возможность выполнить условие электронейтральности за счет экранирования изменяющегося по продольной координате у затворного заряда N_G зарядом на поверхностных ловушках. Это приводит к тому, что градиент плотности подвижных носителей

может быть существенно меньшим, что эквивалентно уменьшению доли диффузионного тока.

7.11. Время пролета электрона через длину канала

Важным параметром, характеризующим максимальное для данного типа МОПТ быстродействие, является время пролета носителя через канал, точнее время, необходимое носителю, для преодоления расстояния от истока до стока. Очевидно, в линейном режиме это время определяется величиной дрейфовой скорости носителя в тянущем электрическом поле, а в режиме насыщения – процессом диффузии носителя через канал. В описанном в этой главе подходе это легко выразить в виде единого выражения. Для этого необходимо записать интеграл для времени пролета в эффективном электрическом поле $(1 + \kappa)E(y)$

$$\tau_{TT} = \int_0^L \frac{dy}{\mu_n (1 + \kappa) E(y)}. \quad (7.11.1)$$

Выполняя интегрирование с использованием распределения электрического поля вдоль канала (7.6.3), с учетом соотношения Эйнштейна

$$D_n = \mu_n \varphi_T, \quad (7.11.3)$$

получаем

$$\tau_{TT} = \frac{1}{1 + \kappa} \left(\frac{L}{\mu E(0)} - \frac{\kappa L^2}{2D_n} \right). \quad (7.11.4)$$

С учетом выражения для $E(0)$ (7.6.2), получаем компактную формулу для времени пролета, справедливую для всех режимов работы МОПТ:

$$\tau_{TT} = \frac{L^2}{2D_n} \frac{\kappa}{1 + \kappa} \coth \left(\frac{\kappa}{1 + \kappa} \frac{V_{DS}}{2\varphi_T} \right). \quad (7.11.5)$$

Нетрудно видеть, что в линейном надпороговом режиме (когда $\kappa \ll 1$ и аргумент гиперболического котангенса меньше единицы) эта формула дает время дрейфового пролета как в обычном резисторе:

$$\tau_{TT} \cong \frac{L^2}{\mu_n V_{DS}}. \quad (7.11.6)$$

Для режима насыщения в надпороговой области (когда $V_{DS} > V_{DSAT}$, но $\kappa \ll 1$) с учетом (7.4.6) получаем

$$\tau_{TT} \cong \frac{L^2}{2D_n} \frac{\kappa}{1+\kappa} \cong \frac{L^2}{\mu_n} \frac{1+\eta_D}{V_{GS}-V_T}. \quad (7.11.7)$$

Наконец, в подпороговом режиме работы транзистора, когда $\kappa \gg 1$, выражение для времени пролета с учетом принимает вид

$$\tau_{TT} \cong \frac{L^2}{2D_n}, \quad (7.11.8)$$

что соответствует времени диффузии через длину канала, аналогичной времени диффузии через базу биполярного транзистора.

Вообще говоря, полученное общее выражение для времени пролета, как и весь подход в целом, применим не только для МОП транзисторов, но и для других структур, где соотношение диффузионного и дрейфового токов может заметно меняться при изменении внешних параметров, в частности, биполярного транзистора.

7.12. Транспортное уравнение Больцмана в канале

Уравнение Больцмана в одномерном приближении можно записать в виде

$$v_y \frac{\partial f}{\partial y} - qE(y) \frac{\partial f}{\partial p_y} = -\frac{\delta f}{\tau}, \quad (7.12.1)$$

где δf – неравновесная часть функции распределения, τ – время релаксации по импульсу.

Вдоль канала имеется неоднородное распределение потенциальной энергии $U(y) = -q\varphi_S(y)$, химического $\zeta(y)$ и электрохимического $\mu(y)$ потенциалов.

Записывая симметричную часть функции распределения в локальном виде

$$\begin{aligned} f_0(\varepsilon) &= \left(1 + \exp\left(\frac{\varepsilon(\mathbf{p}) + U(y) - \mu(y)}{k_B T}\right) \right)^{-1} = \\ &= \left(1 + \exp\left(\frac{\varepsilon(\mathbf{p}) - q\zeta(y)}{k_B T}\right) \right)^{-1}, \end{aligned} \quad (7.12.2)$$

представляем левую часть уравнения Больцмана в виде

$$v_y \frac{\partial f}{\partial y} - qE \frac{\partial f}{\partial p_y} = \left(-\frac{\partial f_0}{\partial \varepsilon} \right) v_y \frac{\partial \mu(y)}{\partial y}. \quad (7.12.3)$$

Отсюда неравновесная часть функции распределения представляется в виде

$$\delta f = \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \tau v_y \frac{\partial \mu(y)}{\partial y} = (1 + \kappa) \left(-\frac{\partial f_0}{\partial \varepsilon} \right) q \tau v_y E(y). \quad (7.12.4)$$

Плотность тока в канале

$$\begin{aligned} J_S &= (1 + \kappa) E(y) q^2 \int \frac{v_s v_v d^2 p}{(2\pi \hbar)^2} \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \tau v_y^2 \equiv \\ &\equiv (1 + \kappa) \sigma_S E(y), \end{aligned} \quad (7.12.5)$$

где была введена двумерная проводимость

$$\sigma_S = q^2 \int \frac{v_s v_v d^2 p}{(2\pi \hbar)^2} \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \tau v_y^2 = \frac{q^2 \tau n_S}{m}. \quad (7.12.6)$$

Соотношение (7.12.6) справедливо как для вырожденного, так и для невырожденного случая и носит название формулы Друде:

$$\int \frac{v_s v_v d^2 p}{(2\pi \hbar)^2} \left(-\frac{\partial f_0}{\partial \varepsilon} \right) v_y^2 = \frac{1}{m} \int g_{2D}(\varepsilon) \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \varepsilon d\varepsilon = \frac{n_S}{m}. \quad (7.12.7)$$

Соотношение (7.12.7) дает возможным образом ввести понятие квантовой емкости C_Q и энергии диффузии ε_D :

$$\int g_{2D}(\varepsilon) \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \varepsilon d\varepsilon \equiv C_Q \varepsilon_D / q^2, \quad (7.12.8)$$

где введена энергия диффузии

$$\varepsilon_D = \frac{\int g_{2D} \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \varepsilon d\varepsilon}{\int g_{2D}(\varepsilon) \left(-\frac{\partial f_0}{\partial \varepsilon} \right) d\varepsilon} \quad (7.12.9)$$

и квантовая емкость

$$C_Q = \int g_{2D}(\varepsilon) \left(-\frac{\partial f_0}{\partial \varepsilon} \right) d\varepsilon \cong q^2 \frac{dn_S}{d\zeta}. \quad (7.12.10)$$

Квантовая емкость не является геометрической характеристикой и определяется сжимаемостью электронного газа в канале. В МОП транзисторах квантовая емкость имеет смысл удельной емкости инверсионного слоя C_{inv} (глава 3), последовательно присоединенной к геометрической емкости затвора.

8. Транзисторы технологии «кремний-на-изоляторе»

8.1. Мотивация КНИ

Несмотря на то, что экспериментально продемонстрировано, как обычная планарная объемная КМОП технология может быть использована вплоть до длины канала ~ 15 нм, совсем не очевидно, что такое сокращение размеров будет сопровождаться соответствующим улучшением характеристик приборов. Проблема состоит в ухудшающемся электростатическом контроле заряда в кремнии затвором и, соответственно, плохой воспроизводимости характеристик транзисторов ультрамалых размеров.

Технология «кремний-на-изоляторе» (КНИ) рассматривается как естественная преемница объемной технологии. Считается, что последовательное развитие технологии КНИ должно привести к достижению предельных характеристик кремниевых КМОП схем.

Исторически МОП транзисторы, изготовленные по технологии «кремний-на-изоляторе» (КНИ МОПТ, международный термин «Silicon-on-Insulator», SOI MOSFET), появились достаточно давно как элементы схем специального применения.

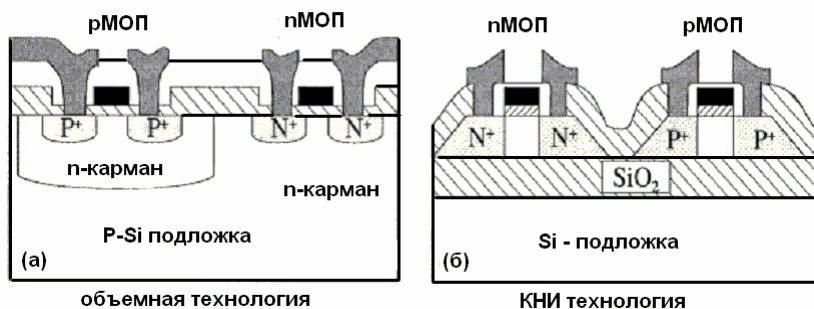


Рис. 8.1. Сравнение (а) объемной и (б) КНИ технологий

Первые КНИ транзисторы имели структуру «кремний-на-сапфире», и их главным достоинством считалось отсутствие радиационно-индуцированных токов в *pn*-переходах стока и истока. До начала 90-х гг. XX в. КНИ технологии не рассматривались как серьезная альтернатива коммерческой объемной технологии, глав-

ным образом, из-за своей дороговизны. Со временем выяснилось, что КНИ МОПТ имеют существенные функциональные преимущества над транзисторами обычных объемных технологий с неизолированной подложкой.

В настоящее время КНИ структуры активно проникают в коммерческие технологии (рис. 8.2). Предполагается, что этот процесс будет идти еще быстрее по мере повышения степени интеграции и, в конце концов, КНИ станет доминирующей коммерческой технологией.

Активная область КНИ МОП транзистора, именуемая базой или «телом» (международный термин *body*), представляет собой тонкую кремниевую пленку между контактами стока и истока, изолированную со всех сторон слоями окислов. Скрытый окисел, или захороненный окисел (международный термин *buried oxide, BOX*), в КНИ структурах изолирует активную область прибора от подложки (см. рис. 8.1), а вертикальная изоляция (как правило, STI) делает невозможными токи утечки и тиристорный эффект между двумя соседними приборами.

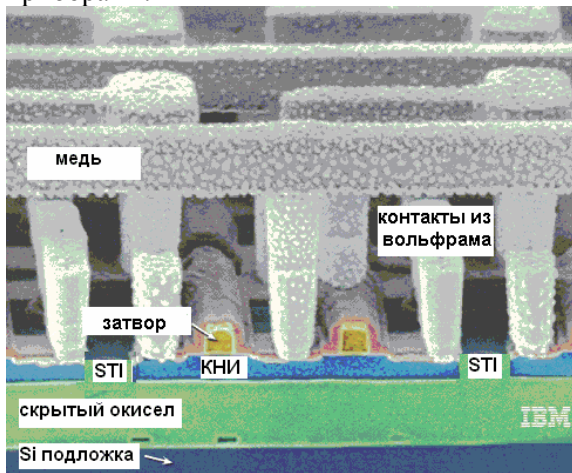


Рис. 8.2. Фотография сечения частично обедненной КНИ ИС

8.1. Преимущества КНИ МОПТ

Транзисторы и изготовленные по технологии КНИ обладают рядом важных преимуществ над транзисторами объемных технологий с аналогичной проектной нормой.

1. Область стока/истока расположена прямо над скрытым окислом, поэтому изоляция скрытым окислом имеет существенно меньшую емкость, чем изоляция *pn*-переходом в объемных транзисторах, что объясняется большей толщиной скрытого окисла и меньшей диэлектрической проницаемостью окисла по сравнению с кремнием. Это дает уменьшение паразитных емкостей (на 30...50 %) и соответствующее увеличение быстродействия из-за уменьшения времени задержки RC.

2. По той же причине, из-за уменьшения паразитной емкости, уменьшается динамическое энергопотребление (приблизительно на 30 % при той же тактовой частоте и напряжении питания).

3. Улучшается электростатическое качество транзисторов за счет подавления геометрических короткоканальных эффектов. Повышается (на ~ 15 %) плотность интеграции за счет уменьшения расстояния между транзисторами.

4. В КНИ схемах отсутствуют эффекты влияния общей подложки на пороговые напряжения, как это имеет место в схемах объемной КМОП технологии.

5. КНИ транзисторы могут иметь очень малое (близкое к минимальному) значение подпорогового размаха (~60 мВ/декаду при комнатных температурах), что позволяет снизить пороговое напряжение до 0.3 В, не увеличивая статические токи утечки. Соответственно, это позволяет уменьшать напряжения питания и динамическое энергопотребление.

6. Уменьшаются перекрестные помехи между линиями в смешанных схемах.

7. Из-за высокой степени изоляции перекрываются пути для развития паразитного тиристорного эффекта, часто имеющего место в *pnpr*-структурах объемных КМОП технологий. Отсутствуют ионизационные токи в *pn*-переходах при внешних импульсных ионизационных воздействиях. Поэтому КНИ МОПТ чрезвычайно устойчивы к воздействию импульсной радиации, однако могут возникать проблемы с толстым скрытым окислом, где может накапли-

ваться радиационно-индуцированный заряд и появляться дополнительный канал утечек.

8. Боковая межприборная изоляция (например, изоляция типа «птичий клюв» (LOCOS), мелкими канавками (STI)) дает возможность более компактного расположения элементов и более простой технологии изготовления, поскольку нет необходимости в карманах и глубоких канавках, как это имеет место в объемной технологии.

8.2. Различные конфигурации КНИ МОПТ

Толщина кремниевой базы КНИ МОПТ может быть различной. Различают КНИ структуры с толстым слоем кремния ($d_S \geq 1$ мкм), которые используются в мощных приборах, и тонкопленочные структуры ($d_S \leq 0.2 \dots 0.3$ мкм), используемые в интегральных схемах высокой степени интеграции.

Среди тонкопленочных МОПТ выделяют приборы двух типов: полностью обедненные (ПО) и частично обедненные (ЧО). Подобная классификация основана на сравнении толщины кремниевой базы d_S и толщины слоя обеднения x_d . Толщина слоя кремния полностью обедненной КНИ МОПТ (международный термин fully depleted SOI, FD SOI MOSFET) структуры обычно не превосходит 50 нм с типичным значением ~ 30 нм. В частично обедненных (ЧО) КНИ структурах (partially depleted, PD SOI) $d_S > 50$ нм с типичными значениями $\sim 150 \dots 200$ нм. Естественным условием полного обеднения является то, что толщина обедненной области больше толщины базы $x_d > d_S$. Ясно, что выполнение этого условия зависит от уровня легирования базы. В короткоканальных транзисторах условие полного обеднения зависит еще и от длины канала. Это связано с тем, что электрод стока влияет на условие обеднения в коротком канале, существенно облегчая этот процесс. Расчетная фазовая диаграмма условий полного обеднения показана на рис. 8.3.

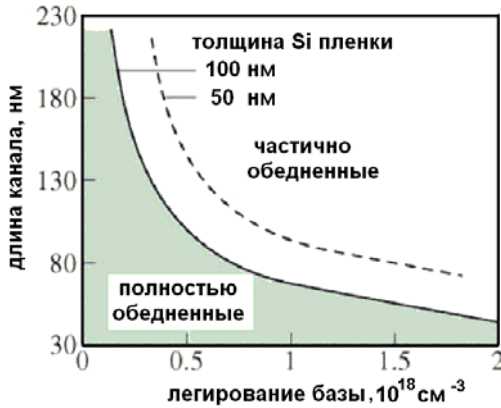


Рис. 8.3. Области полного и частичного обеднения для разных длин канала и уровней легирования при разных значениях толщины кремниевой базы

Как можно видеть на рис. 8.3, для обеспечения условия частичного обеднения в теле с толщиной 100 нм и длиной канала 50 нм требуется на порядок более сильное легирование, чем в транзисторе с той же толщиной тела, но с длиной канала 150 нм.

Критерием, по которому можно различать полностью и частично обедненные КНИ транзисторы, является сравнение максимальной толщины обедненной области в районе истока ($x_d(0) \cong x_d(y=0, \varphi_S = 2\varphi_F)$) и в районе стока ($x_d(L) \cong x_d(y=L, \varphi_S = 2\varphi_F + V_{DS})$) с толщиной слоя кремния внутренней подложки d_S .

При выполнении условий $x_d(L) < d_S$ и $x_d(L) > x_d(0) > d_S$ имеем соответственно частично (PD) и полностью обедненные (FD) случаи. В промежуточном случае, когда $x_d(L) > d_S > x_d(0)$, говорят о т.н. динамическом обеднении.

Существуют различные варианты конфигураций КНИ МОП транзисторов (рис. 8.4), включающие в себя возможность управления напряжением как с верхнего затвора (front gate), так и со стороны подложки (нижнего затвора, back gate).

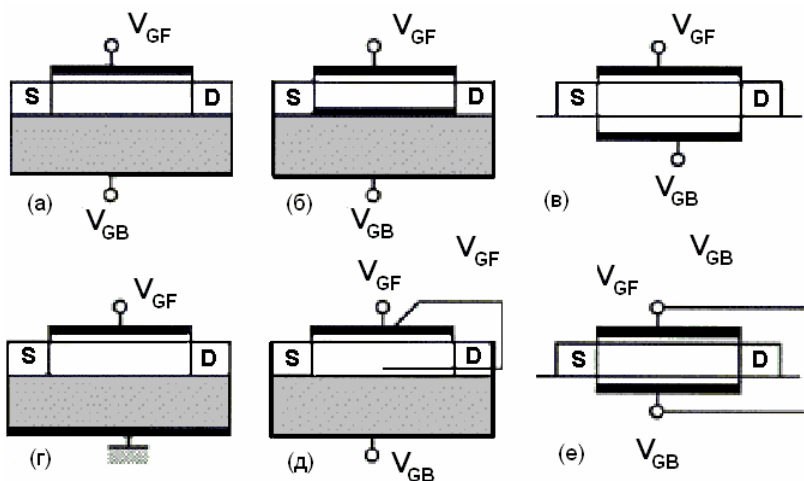


Рис. 8.4. Схематические структуры мыслимых КНИ МОПТ: (а) полностью обедненные, (б) транзистор с сильнолегированным дельта-слоем (pulsed doping) в кремниевой пленке, (в) с полевой обкладкой (field plate), (г) с заземленной основой (ground plane, GP), (д) динамический порог, (е) транзистор с двойным затвором

Для улучшения электростатики и повышения электростатической целостности прибора в структуру прибора можно внедрять сильнолегированные высокопроводящие слои. Они могут быть расположены непосредственно над скрытым окислом либо под ним и играть роль дополнительного управляющего затвора.

8.3. Частично обедненные КНИ МОПТ

В настоящее время наиболее широко распространены КНИ транзисторы с частичным обеднением (ЧО КНИ МОПТ) (рис. 8.5). Это означает, что толщина обедненного слоя $x_d < d_s$ оказывается меньше толщины кремниевой базы (тела). В этом смысле частично обедненный КНИ МОПТ очень похож на транзистор, изготовленный по объемной технологии, что обеспечивает преемственность и эволюционность технологии.

Главной особенностью, обуславливающей качественное отличие КНИ технологий, является изолированность кремниевой базы от

контакта подложки, характерная для большинства коммерческих ЧО КНИ технологий.

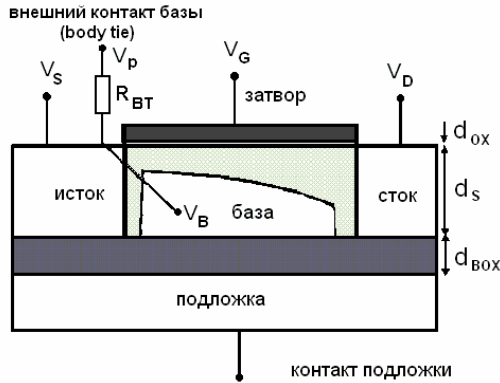


Рис. 8.5. Сечение частично обедненного КНИ МОПТ

Эта особенность обуславливает то, что потенциал внутренней подложки в КНИ ЧО МОПТ не фиксирован и является плавающим, что может приводить к многочисленным, часто нежелательным эффектам плавающей базы (floating body effects). Эти эффекты обусловлены зарядом, приходящим и уходящим из базы за счет различных, часто случайных и неконтролируемых механизмов, таких как ионизация пролетающей одиночной частицы, ударная ионизация в сильном электрическом поле вблизи стока, эффекты GIDL, туннелирование через тонкий подзатворный окисел.

База частично обедненного КНИ МОПТ емкостным образом связана со всеми выводами транзистора, что означает увеличение (уменьшение) потенциала базы при увеличении (уменьшении) потенциалов каждого из электродов. Чаще всего база оказывается в прямосмещенном относительно истока режиме, что обуславливает уменьшение эффективного порогового напряжения (рис. 8.6).

Текущее значение потенциала базы становится функцией предыдущего электрического состояния прибора. На схемном уровне это приводит к переходным эффектам различного рода и нежелательному разбросу времен распространения сигнала в цепях в зависимости от электрической предыстории конкретной цепи.

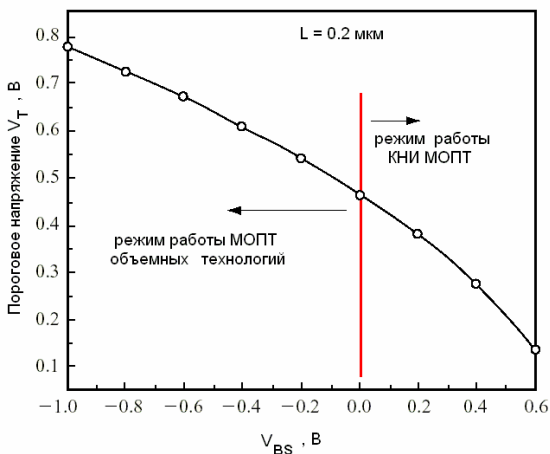


Рис. 8.6. Зависимость порогового напряжения n -МОПТ как функция смещения на подложку (базу) относительно истока

Эффекты предистории зависят от многих параметров, таких как времена нарастания и спада входного сигнала, нагрузочной емкости, деталей конфигурации и т.п. Разброс времен задержки в одном транзисторе из-за эффектов электрической предистории может составлять до 8 %.

Эффекты плавающего тела приводят также к нестационарным нестабильностям и гистерезисам выходных и передаточных ВАХ частично обедненных КНИ МОПТ. Для борьбы с нежелательными эффектами плавающей базы иногда используют дополнительный контакт базы, который, как правило, соединяет базу с контактом истока (body tie) и фиксирует ее потенциал. Такая процедура усложняет процесс изготовления и увеличивает эффективные геометрические размеры транзистора.

8.4. Кинк-эффект в частично обедненных КНИ МОПТ

Для определенности будем рассматривать n -канальный частично обедненный МОПТ с p -базой, на выход которого подается напряжение V_{DS} . При достаточно большом напряжении V_{DS} в сильном электрическом поле в окрестности стока происходят процессы ударной ионизации и образования электронно-дырочных пар (см. п. 6.3). Электроны уходят в сток, а дырки выбрасываются

электрическим полем обедненной области в нейтральную область кремния, накапливаясь в изолированной подложке.

Избыточный заряд дырок (основных носителей) за очень малые времена диэлектрической релаксации τ_{diel} достигает своего стационарного значения $Q_h^{(kink)}$, распределяясь на площади базы и уменьшая толщину обедненного слоя x_d и плотность заряда в нем:

$$V_G - \varphi_{MS} = \varphi_S + \frac{1}{C_O} \left(qn_S + qN_A x_D - \frac{Q_h^{(kink)}}{AC_O} \right). \quad (8.5.1)$$

Условие электрической нейтральности и фиксированность смещения на затворе приводят к тому, что инжекция дырок в подложку сопровождается увеличением заряда инверсионного слоя и порогового напряжения:

$$\Delta V_T^{(kink)} \cong - \frac{Q_h^{(kink)}}{AC_O}. \quad (8.5.2)$$

На выходной ВАХ транзистора образуется скачок тока (т.н. «кинк») (рис. 8.7).

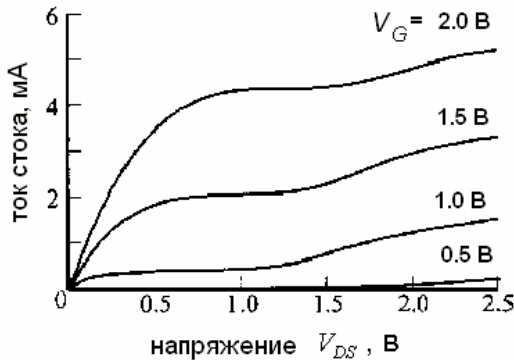


Рис. 8.7. Кинк-эффект на выходных характеристиках ЧО КНИ МОПТ

Амплитуда этого скачка и потенциал изолированной базы определяются балансом процесса рекомбинации в теле и ударной ионизации. В цифровой технике кинк-эффекты могут приводить даже к некоторому увеличению быстродействия за счет повышения тока насыщения транзистора, хотя и вызывают разбросы времен задержки для разных элементов.

С другой стороны, нестабильность значений выходного напряжения и выходного сопротивления в насыщении (см. п. 5.13) делает частично обедненные КНИ МОПТ неудобным для аналоговых приложений.

8.5. Паразитный биполярный эффект

При дальнейшем увеличении смещения на стоке инжекция дырок в базу начинает преобладать над рекомбинацией, что приводит к повышению потенциала базы:

$$V_B = \frac{Q_{body}}{C_{body}}, \quad (8.7.1)$$

где емкость базы (тела) с частичным обеднением складывается из полной емкости верхнего и нижнего затворов и емкостей истока и стока (рис. 8.8).

$$C_{body} = C_{BG} + C_{BOX} + C_{BS} + C_{BD}. \quad (8.7.2)$$

Когда потенциал базы (т.е. прямое смещение базы относительно заземленного истока) достигает значения ~ 0.7 В, то *pn*-переход истока открывается, и реализуется режим паразитного биполярного транзистора.

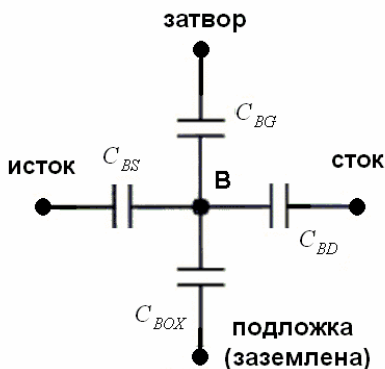


Рис. 8.8. Эквивалентная емкостная схема изолированной базы КНИ транзистора

Электроны из истока устремляются в тело КНИ структуры, играющей роль базы паразитного биполярного транзистора. Часть электронов рекомбинирует с избыточными дырками, другая часть

электронов достигает стока КНИ МОПТ, играющего здесь роль коллектора.

Долю электронов, достигающую коллектор α_T , имеющую смысл коэффициента переноса для паразитной биполярной структуры, можно оценить через пуассоновскую вероятность диффузии из истока в сток без рекомбинации

$$\alpha_T = \exp\left(-\frac{t_D}{t_T}\right) \cong \exp\left(-\frac{L_C^2}{2D_n t_T}\right) = \exp\left(-\frac{L_C^2}{2L_D^2}\right), \quad (8.7.3)$$

где t_D и t_L – времена диффузии и жизни электрона в кремниевой базе, L_G и L_D – длина канала и диффузионная длина соответственно. Коэффициент усиления паразитной биполярной структуры оценивается по стандартной формуле

$$\beta = \frac{\alpha_T}{1 - \alpha_T}. \quad (8.7.4)$$

В отличие от настоящего биполярного транзистора, где α_T очень близко к единице и $\beta \sim 100$, коэффициент усиления паразитной биполярной структуры имеет порядок $\beta \sim 10$.

8.6. Полностью обедненные КНИ МОПТ

В КНИ МОПТ полностью обедненного типа обедненная область занимает всю толщину кремниевого тела и имеет фиксированный заряд. Электростатика таких транзисторов такова, что основные носители (дырки), инжектированные в базу, сразу прижимаются электрическим полем к *pn*-переходу истока, локально открывают его и быстро рекомбинируют с электронами, поступающими из истока. Поэтому в полностью обедненных приборах практически не происходит накопление основных носителей, и эффекты плавающего потенциала по сравнению с частично обедненными транзисторами практически отсутствуют.

Контакт подложки в полностью обедненном КНИ транзисторе можно рассматривать как своеобразный нижний затвор (back gate, BG), который можно использовать для юстировки порогового напряжения основного канала от верхнего затвора (front gate, FG) (рис. 8.9). Это объясняется тем, что полное обеднение означает отсутствие квазинейтральной области в базе КНИ транзистора. По-

этому в полностью обедненном КНИ транзисторе силовые электрические линии от верхнего затвора могут достигать нижнего затвора. Таким образом, нижний и верхний затвор оказываются электрически связанными через уравнение планарной электростатической эквивалентности многослойной структуры.

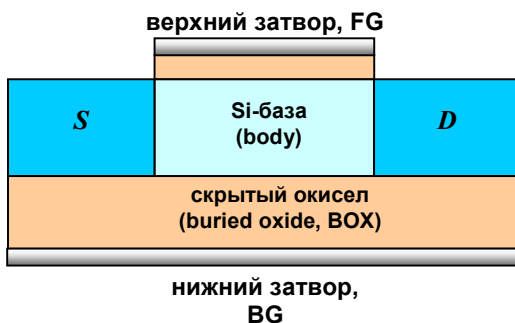


Рис. 8.9. Схематичная структура МОП транзистора КНИ структуры с заземленной подложкой. Рабочий канал формируется у границы с подзатворным окислом

Поверхностные потенциалы на двух границах раздела оказываются, таким образом, электрически связанными друг с другом. Это приводит к тому, что результаты измерения при изменении напряжения на верхнем затворе зависят от напряжения на нижнем затворе. Модели транзисторов с полным обеднением должны учитывать электростатическую связь двух затворов.

Прикладывая напряжение на верхний затвор (V_{GF}) либо на нижний затвор (V_{GB}), инверсионный слой в КНИ транзисторе может быть получен как на границе раздела кремния с тонким подзатворным окислом, так и на границе раздела с толстым скрытым окислом. В полностью обедненных КНИ МОПТ нет возможности управлять пороговым напряжением с помощью легирования подложки, и поэтому в них желательно иметь металлические затворы с возможностью контролируемого изменения работы выхода (см. п. 3.13).

8.7. Эффекты саморазогрева

Скрытый окисел сильно затрудняет отвод тепла в подложку. В результате КНИ структуры обладают большими тепловыми постоянными времени, и для них типичны эффекты т.н. саморазогрева. Повышение температуры функционирования из-за омического тепловыделения при больших токах приводит к снижению подвижности. Таким образом, в КНИ транзисторах часто наблюдаются эффекты отрицательного дифференциального сопротивления (рис. 8.10).

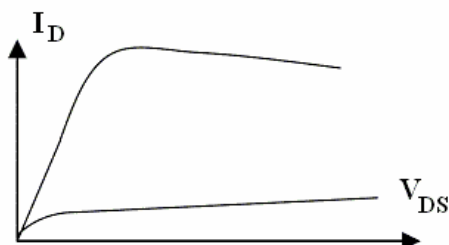


Рис. 8.10. Эффекты саморазогрева в КНИ МОПТ при больших токах

Саморазогрев характерен для обоих типов КНИ МОПТ, но более ярко выражен в КНИ МОПТ с полным обеднением.

8.8. Влияние обратного напряжения на подложке на пороговое напряжение

Так же, как и в МОПТ объемной технологии, пороговое напряжение в полностью обедненных КНИ МОПТ может управляться напряжением на нижнем затворе (подложке). Напряжение на нижний затвор подается относительно заземленного истока, причем, в отличие от объемного МОПТ, диэлектрическая изоляция позволяет любую полярность напряжения. Например, нижний затвор, представляющий собой сильнолегированную подложку кремния, можно поддерживать в состоянии аккумуляции основных носителей (D), инверсии (I) и обеднения (D).

Схематический вид передаточных вольт-амперных характеристик изображен на рис. 8.11, где V_{BG} является параметром. Каждая кривая может быть объяснена изменением этого управляющего па-

раметра. Например, прикладывая положительное смещение на нижний затвор V_{GB} и переводя таким образом нижний канал в режим инверсии, мы уменьшаем, что легко понять из соображений электронейтральности, пороговое напряжение для верхнего транзистора V_{TF} . Напротив, аккумуляция основных носителей в базе на нижней границе раздела приводит к увеличению порогового напряжения V_{TF} .

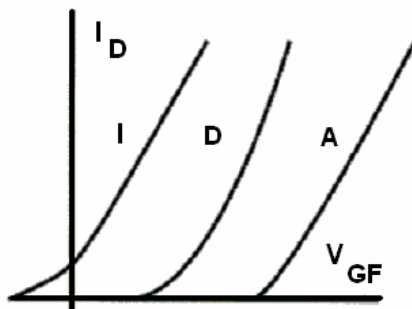


Рис. 8.11. Типичные характеристики полностью обедненного n -канального КНИ МОПТ для случая аккумуляции (А), обеднения (D) или инверсии (I) на нижней границе раздела

Пороговое напряжение верхнего транзистора линейно уменьшается по мере увеличения V_{GB} от режима аккумуляции до инверсии вблизи нижней границы раздела (глава 9).

8.9. Ультратонкие КНИ МОПТ

Уменьшение толщины кремниевого тела КНИ МОПТ (конечно, речь здесь идет только о полностью обедненных приборах) приводит к уменьшению подпорогового размаха (S -фактора) и соответственно, к уменьшению подпороговых токов утечки (рис. 8.12).

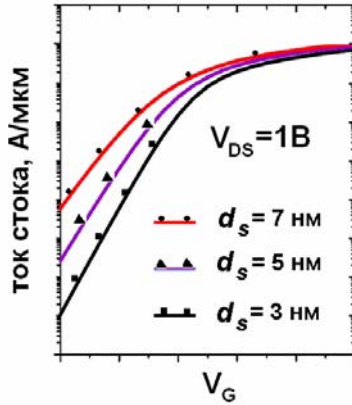


Рис. 8.12. Подпороговые токи утечки для ультратонких КНИ МОПТ с разной толщиной базы

С другой стороны, подвижность носителей в канале с уменьшением толщины кремневой пленки заметно падает (рис. 8.13).

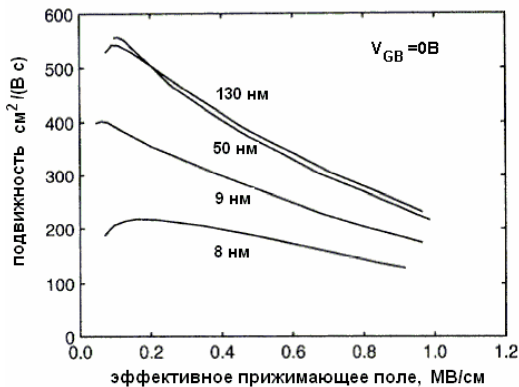


Рис. 8.13. Зависимости подвижности от прижимающего электрического поля для КНИ транзисторов с разной толщиной кремневой базы

Это объясняется тем, что в очень тонких базах толщина кремневой пленки становится сопоставимой с толщиной инверсионного слоя (порядка нескольких нанометров). Это означает, что транспорт носителей в канале носит по существу объемный характер, и появляется дополнительный механизм рассеяния на шероховато-

стях границы раздела со скрытым окислом. Другим источником деградации подвижности в канале КНИ транзисторов с ультратонкими базами является технологическое загрязнение кремниевой пленки неконтролируемыми примесями при изготовлении. Более того, для транзисторов с малым объемом активной области даже случайные флуктуации количества атомов легирующей примеси в базе могут приводить к недопустимому разбросу индивидуальных значений порогового напряжения и проводимости, а также уменьшению подвижности отдельного транзистора. По этим причинам база полностью обедненного КНИ транзистора в идеале должна быть вообще не легирована примесями, что технологически является трудновыполнимой задачей.

8.10. Сравнение полностью и частично обедненных КНИ МОПТ

Традиционно считается, что наиболее перспективными являются КНИ структуры с полным обеднением. Такая технология обеспечивает наилучшие характеристики приборов (отсутствие эффектов плавающего тела, более крутой подпороговый наклон, минимизация короткоканальных эффектов и паразитных емкостей) и считается наиболее приемлемой в диапазоне длин канала менее 100 нм. Тем не менее, полностью обедненные КНИ МОПТ существенно менее технологичны и более трудны для изготовления. Например, пороговое напряжение ПО КНИ МОПТ очень чувствительно к толщине кремниевой базы, а поддержание постоянства толщины базы является трудной технологической задачей.

Преимущество полностью обедненных приборов над частично обедненными в части электростатического качества и отсутствия короткоканальных эффектов также не столь бесспорно. Например, из-за отсутствия квазинейтральной области в полностью обедненных КНИ МОПТ существует электростатическая связь стока с затвором через скрытый окисел (рис. 8.14).

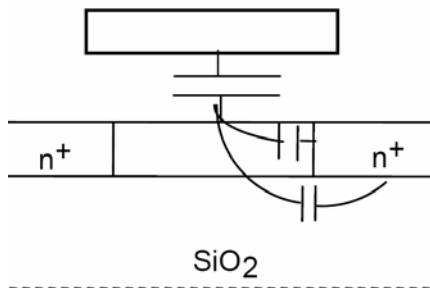


Рис. 8.14. Иллюстрация дополнительной электрической связи стока ПО КНИ МОПТ с затвором через силовые линии электрического поля, проходящие через скрытый окисел

Это обстоятельство ухудшает электростатическое качество МОПТ полностью обедненного типа и усиливает короткоканальные эффекты. В частности, это касается относительного изменения порогового напряжения от длины канала (рис. 8.15).

Уменьшение абсолютной величины спада порогового напряжения как функции длины канала для полностью обедненных приборов на рис. 8.15 обусловлено только уменьшением глубины залегания переходов стока и истока.

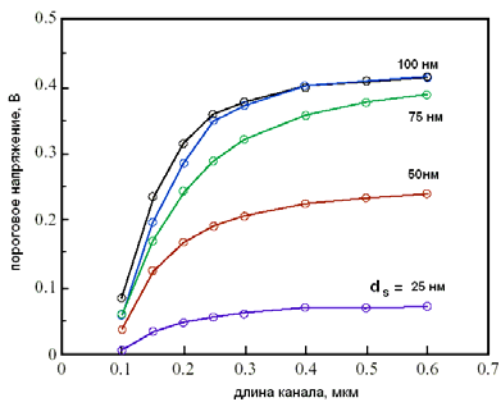


Рис. 8.15. Типичный вид зависимостей порогового напряжения от длины затвора для полностью обедненных КНИ МОПТ и толщины базы к толщине скрытого окисла (в нм) по сравнению с частично обедненными КНИ МОПТ (верхняя кривая)

