

PACS numbers: 72.80.Ey, 85.30.De, 85.30.Tv, 85.40.Bh

Физика нанотранзисторов: 2D-электростатика MOS и модель виртуального истока

Ю. А. Кругляк

Одесский государственный экологический университет,
ул. Львовская, 15,
65016 Одесса, Украина

Подробно рассмотрена электростатика MOS. 1D-электростатика изгибает зоны, понижает барьер и позволяет потоку электронов двигаться от истока к стоку. 2D-электростатика деградирует транспорт электронов в полевых транзисторах, увеличивая подпороговый разброс и вызывая *DIBL*, который в свою очередь увеличивает выходную проводимость и уменьшает пороговое напряжение в короткоканальных транзисторах. Количественный учёт 2D-электростатики требует численного подхода; вместе с тем, все существенные эффекты физически понятны. 2D-электростатика разрушает функционирование транзисторов и приводит к: 1) подпороговому разбросу, большему, чем фундаментальный предел в 60 мВ/декада; 2) сдвигу проходных характеристик $\log_{10} I_{DS} \propto V_{GS}$ влево при увеличении напряжения на стоке (*DIBL*); 3) возникновению порога в зависимости от параметров затвора и напряжения на стоке; 4) низкому выходному сопротивлению. Когда эффекты 2D-электростатики сильны, затвор теряет контроль над током, и транзистор претерпевает пробой. Поскольку эти эффекты проявляются сильнее в короткоканальных транзисторах, их ещё называют эффектами короткого канала. По мере того как транзисторы становятся всё меньше и меньше, основной вызов, с которым встречаются схемотехники, сводится к контролю короткоканальных эффектов. Как правило, требуется численное моделирование. Возвращаясь к уравнению для тока, видим, что зависимость заряда от напряжений в локации виртуального истока имеет вполне приемлемое физическое объяснение. Что же касается скорости в этом уравнении, то далее построена адекватная физическая картина транспортных явлений в нанотранзисторах, исходя из обобщённой модели транспорта электронов ЛДЛ, и установлена связь её с традиционным подходом «сверху–вниз».

Докладно розглянуто електростатику MOS. 1D-електростатика вигинає зони, понижує бар'єр і уможливило потоку електронів рухатися від витоку до стоку. 2D-електростатика деградує транспорт електронів у польових транзисторах, збільшуючи підпороговий розкид і викликаючи *DIBL*,

який, в свою чергу, збільшує вихідну провідність і зменшує граничну напругу в короткоканальних транзисторах. Кількісне врахування 2D-електростатики вимагає чисельного підходу; разом з тим, всі істотні ефекти фізично зрозумілі. 2D-електростатика руйнує функціонування транзисторів і призводить до: 1) підпорогового розкиду, більшого, ніж фундаментальна границя у 60 мВ/декада; 2) зсуву прохідних характеристик $\log_{10} I_{DS} \propto V_{GS}$ вліво при збільшенні напруги на стоці (*DIBL*); 3) виникнення порогу, залежно від параметрів затвора та напруги на стоці; 4) низького вихідного опору. Коли ефекти 2D-електростатики є сильними, затвор втрачає контроль над струмом, і транзистор зазнає пробоя. Оскільки ці ефекти проявляються сильніше у короткоканальних транзисторах, їх ще називають ефектами короткого каналу. У міру того, як транзистори стають все менше і менше, основний виклик, з яким зустрічаються схемотехніки, зводиться до контролю короткоканальних ефектів. Як правило, потрібно чисельне моделювання. Повертаючись до рівняння для струму, бачимо, що залежність заряду від напруги в локації віртуального витоку має цілком прийнятне фізичне пояснення. Що ж стосується швидкості в цьому рівнянні, то далі побудовано адекватну фізичну картину транспортних явищ у нанотранзисторах на основі узагальненого моделю транспорту електронів ЛДЛ і встановлено зв'язок його з традиційним підходом «зверху–вниз».

The electrostatics of MOS is considered in detail. The 1D electrostatics bends zones, lowers the barrier and allows the flow of electrons to move from source to drain. The 2D electrostatics degrades electron transport in field-effect transistors by increasing the subthreshold swing and causing the *DIBL*, which, in turn, increases the output conductivity and reduces the threshold voltage in short-channel transistors. Quantitative accounting of the 2D electrostatics requires a numerical approach, but, at the same time, all the significant effects are physically understandable. The 2D electrostatics destroys the functioning of transistors and leads to: 1) a sub-threshold swing greater than the fundamental limit of 60 mV/decade; 2) the shift of the transfer characteristics, $\log_{10} I_{DS} \propto V_{GS}$, to the left with increasing drain voltage (*DIBL*); 3) the threshold depending on the gate parameters and voltage on the drain; 4) low output resistance. When 2D electrostatic effects are strong, the gate loses control of the current, and the transistor undergoes punch-through. Since these effects are more evident in short-channel transistors, they are also called as short-channel effects. As transistors get smaller and smaller, the main challenge with circuitry is to control the short-channel effects. As a rule, numerical modelling is required. Returning to the equation for the current, we see that the dependence of the charge on the voltages in the location of the virtual source has a completely acceptable physical explanation. As for the speed in this equation, it will constructed an adequate physical picture of the transport phenomena in nanotransistors based on the generalized model LDL of the electron transport, and its connection with the traditional top-down approach is established.

Ключевые слова: наноэлектроника, полевой транзистор, MOSFET, 2D-электростатика, метрика транзисторов, управление транзисторами, виртуальный источник.

Ключові слова: наноелектроніка, польовий транзистор, MOSFET, 2D-електростатика, метрика транзисторів, управління транзисторами, віртуальний витік.

Keywords: nanoelectronics, field effect transistor, MOSFET, 2D electrostatics, transistor metrics, transistor control, virtual source.

(Получено 29 ноября 2018 г.)

1. ВВЕДЕНИЕ

В продолжение научного обозрения физических принципов моделирования нанотранзисторов [1–3] рассмотрим 2D-электростатику MOS и обусловленные ею эффекты, а затем сформулируем ставшую уже классической модель виртуального истока без явного учёта возможности баллистического транспорта.

Мы до сих пор обсуждали, каким именно образом поверхностный потенциал полупроводниковой подложки изменяется с ростом напряжения на затворе, прибегая к рассмотрению лишь 1D-электростатики. В нанотранзисторах с короткими каналами проводимости потенциалы, подаваемые на исток и сток, создают сильное электрическое поле вдоль направления канала. Совершенно очевидно, что электростатический потенциал в короткоканальном MOSFET будет сильно изменяться в обоих x - и y -направлениях (рис. 1).

Учёт 2D-электростатики, особенно в нанотранзисторах, влечёт серьёзные последствия для режимов работы MOSFET. Как показано на рис. 2 слева, прикладывание большого напряжения на стоке сдвигает характеристику $\log_{10} I_{DS} \propto V_{GS}$ влево. Этот сдвиг на рис. 2 связан с $DIBL = -\Delta V_{GS} / \Delta V_{DS}$ (см. рис. 11 в [1]), где ΔV_{GS} есть такое изменение напряжения на затворе, которое необходимо, чтобы поддерживать ток стока неизменным при изменении напряжения на

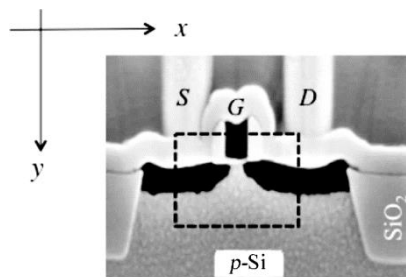


Рис. 1. Микрограмма сечения MOSFET, полученная на сканирующем электронном микроскопе (Texas Instruments, 1997), на которой хорошо видны все контакты и диэлектрическая прослойка, изолирующая затвор от канала проводимости.¹

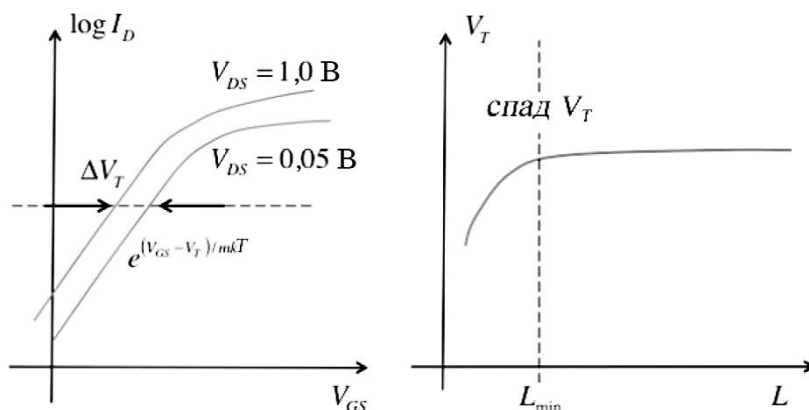


Рис. 2. Иллюстративные примеры того, как учёт $2D$ -электростатики сказывается на характеристиках короткоканальных транзисторов. Слева: $DIBL$, сдвигающее передаточные характеристики $\log_{10} I_{DS} \propto V_{GS}$; такое поведение может быть также интерпретировано как уменьшение порогового напряжения V_T с возрастанием напряжения на стоке. Справа: спад V_T как результат снижения порогового напряжения у нанотранзисторов.²

стоке на величину ΔV_{DS} . Если взять небольшой ток, соответствующий транзистору в режиме «ON» (пунктирная линия на рис. 2), то мы увидим, что большое напряжение на стоке уменьшает величину порогового напряжения на ΔV_T . Однако, полученное ранее выражение для V_T ((13) в [3]) не содержит зависимости от напряжения на стоке, и причина этого кроется в неучёте $2D$ -электростатики. Другое проявление $2D$ -электростатики состоит в зависимости V_T от длины канала проводимости (рис. 2, справа). Выходное сопротивление транзистора также связано с $2D$ -электростатикой. Эти и другие вопросы, связанные с $2D$ -электростатикой, обсудим ниже.

2. $2D$ -ЭЛЕКТРОСТАТИКА MOS

Учесть $2D$ -электростатику можно путём численного решения двумерного уравнения Пуассона, а если речь пойдёт о транзисторе с очень коротким каналом, то вынуждены будем решать и трёхмерное уравнение Пуассона. Численное моделирование неизбежно в проектировании промышленного производства современных нанотранзисторов. Наша же задача всё-таки состоит в том, чтобы построить качественно правильную физическую картину процессов в MOSFET.

$2D$ -уравнение Пуассона. Закон Гаусса для электрического смещения:

$$\nabla \cdot \mathbf{D}(x, y) = \rho(x, y), \quad (1)$$

где \mathbf{D} — вектор смещения, ρ — зарядовая плотность. Связь с электрическим полем даётся соотношением

$$\mathbf{D}(x, y) = \varepsilon_s \mathbf{E}(x, y), \quad (2)$$

где ε_s — диэлектрическая проницаемость полупроводника или оксидной плёнки, которые предполагаются пространственно однородными. Электростатический потенциал связан с электрическим полем:

$$\mathbf{E}(x, y) = -\nabla (x, y), \quad (3)$$

так что уравнение Пуассона в окончательном виде —

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = -\frac{\rho(x, y)}{\varepsilon_s}. \quad (4)$$

Нас особенно интересует решение уравнения Пуассона в подпороговой области и в самом начале режима инверсии, когда учёт 2D-электростатики позволяет понять природу эффекта *DIBL* и спада V_T . В подпороговом режиме

$$\rho(x, y) \approx q [N_D^+(x, y) - N_A^-(x, y)] \approx -qN_A, \quad (5)$$

где последнее соотношение следует из предположения, что учитываются в канале проводимости только p -допанты и что они полностью ионизированы и однородно распределены.

Слой диэлектрика и электрод затвора являются частью канала проводимости и должны быть включены в рассмотрение при вычислении $\psi(x, y)$. У слоя диэлектрика другая диэлектрическая проницаемость и электронным зарядом в этом слое обычно пренебрегают, так что уравнением Лапласа для этого слоя будет

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0. \quad (6)$$

В общем случае нужно численно решать систему уравнений (4) и (6) для нахождения электростатического потенциала. Мы же далее обсудим некоторые качественные подходы к пониманию решений этих уравнений.

Особое внимание далее мы уделим короткоканальным транзисторам, эффекты 2D-электростатики в которых существенны. В случае достаточно длинного канала проводимости потенциал меняется медленно вдоль направления канала, так что

$$\frac{\partial^2 \psi}{\partial x^2} \ll \frac{\partial^2 \psi}{\partial y^2} \quad (7)$$

и уравнение (4) сводится к одномерному уравнению Пуассона типа (57) в [2]. Значительная и существенная часть современной традиционной модели MOSFET построена в предположении (7). Этот подход известен как плавноканальное приближение (gradual channel approximation). Стандартный подход к моделированию короткоканальных MOSFET сводится к разработке модели для достаточно длинного канала проводимости с дальнейшим учётом эффектов, вызванных 2D-электростатикой [4–7].

Спад порогового напряжения и DIBL. Перепишем уравнение (4) с учётом (5) для режима истощения в виде

$$\frac{\partial^2 \psi}{\partial y^2} = \frac{qN_A}{\varepsilon_s} - \frac{\partial^2 \psi}{\partial x^2}. \quad (8)$$

Для n -MOSFET электростатический потенциал растёт от истока к стоку, так что $d\psi/dx > 0$. Электрическое поле, $-d\psi/dx$, также увеличивается от истока к стоку, так что кривизна потенциала $d^2\psi/dx^2$ положительная. Это также следует из результатов моделирования нанотранзисторов, представленных на рисунках 16 и 17 в [1]: как при низком, так и при высоком напряжении на стоке зависимость $E_c(x)$ имеет отрицательную кривизну, так что кривизна электростатического потенциала положительная.

Учитывая положительную кривизну $\psi(x)$, перепишем (8) в виде

$$\frac{\partial^2 \psi}{\partial y^2} = \frac{qN_A|_{eff}}{\varepsilon_s}, \quad (9)$$

где

$$N_A|_{eff} = \frac{qN_A}{\varepsilon_s} - \frac{\partial^2 \psi}{\partial x^2} < N_A. \quad (10)$$

Уравнение (9) есть одномерное уравнение Пуассона для электростатического потенциала вглубь полупроводниковой подложки с эффективной плотностью допирования, меньшей фактической плотности допирования. Согласно (13) из [3], пороговое напряжение связано с плотностью допирования следующим образом:

$$V_T = V_{FB} + \frac{\sqrt{2q\varepsilon_s N_A (2\psi_B)}}{C_{ox}} + 2\psi_B. \quad (11)$$

Поскольку учёт 2D-электростатики эффективно уменьшает N_A (10), можно ожидать понижения порогового напряжения. При укорочении длины канала кривизна потенциала $d^2\psi/dx^2$ увеличивается, что уменьшает эффективную плотность допирования и понижает пороговое напряжение. В этом состоит объяснение, почему V_T умень-

шается с укорочением канала проводимости (спад V_T).

Прибегая к тем же самым аргументам, можно понять явление *DIBL* и уменьшение V_T с увеличением напряжения на стоке при фиксированной длине канала: с ростом V_{DS} кривизна потенциала $d^2\psi/dx^2$ увеличивается, эффективная плотность допирования $N_A|_{eff}$ и V_T уменьшаются.

Учёт 2D-электростатики проявляется также в следующем. Барьер между истоком и каналом критически важен для функционирования транзистора. В идеале высота этого энергетического барьера полностью контролируется напряжением на затворе, а напряжение на стоке никак на неё не влияет (рис. 3, верх).

В реальном транзисторе потенциал стока достаёт-таки до вершины барьера и понижает его (рис. 3, низ). Через более низкий барьер

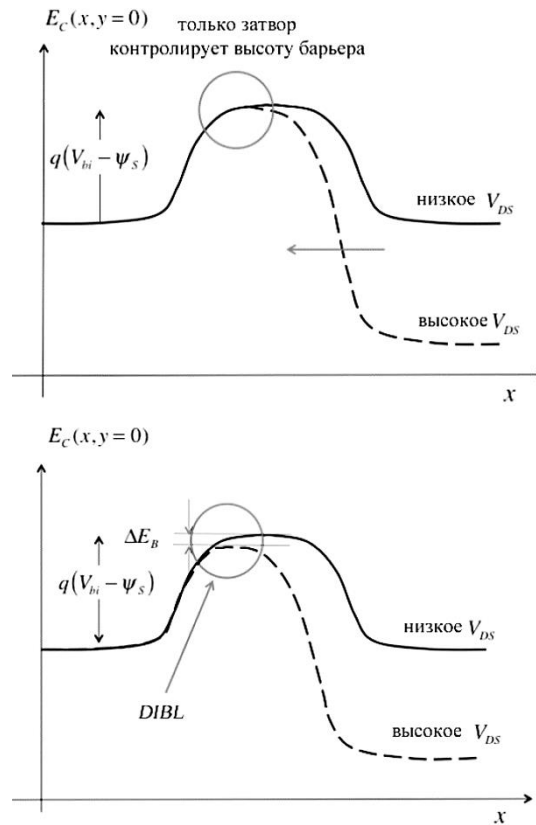


Рис. 3. Иллюстрация влияния напряжения на стоке V_{DS} на высоту барьера. Верх: напряжение V_{DS} не влияет на высоту барьера. Низ: напряжение V_{DS} немного понизило высоту барьера (эффект *DIBL*).³

проходит больше тока при том же напряжении на затворе. Можно сказать иначе: требуется меньшее напряжение на затворе для достижения нужного тока, поскольку барьер был понижен в результате влияния обоих потенциалов — и затвора и стока. Этот эффект *DIBL* сдвигает передаточную характеристику $\log_{10} I_{DS} \propto V_{GS}$ влево (рис. 2).

Этот подход с понижением барьера также позволяет объяснить, почему учёт *2D*-электростатики понижает эффективное допирование. Для сильно допированных подложек зоны труднее изогнуть, а вот при учёте *2D*-электростатики сток помогает затвору барьер понизить. Протяжённость области канала проводимости, находящейся под влиянием стока, зависит от *2D*-геометрии транзистора, к обсуждению которой мы переходим.

Геометрическое экранирование. Экранирование зарядов характерно для металлов и полупроводников. Если в какой-либо среде образовался «эффективный заряд», подвижные носители тока перестраиваются таким образом, чтобы нейтрализовать (экранировать) этот заряд. Характерное расстояние, на котором заряд экранируется, называют длиной экранирования, или иначе, дебаевской длиной L_D , а именно,

$$L_D = \sqrt{\frac{\varepsilon_s kT}{q^2 n_0}}, \quad (12)$$

где n_0 — электронная плотность.

В MOSFET реализуется иной вариант экранирования электрического поля. Рисунок 4 иллюстрирует так называемое «геометрическое экранирование».

Показаны три структуры MOSFET. Для DG ETSOI силовые линии электрического поля стока закорачиваются на верхнем и нижнем электродах затворов, проникая в канал лишь на расстояние Λ . Если величина Λ меньше длины канала проводимости L , $\Lambda < L$, электрическое поле стока не может «достать» начала канала проводимости и поэтому не может понизить барьер. В этом случае *DIBL*, если и есть, то очень небольшой. Точное значение длины геометрического экранирования определяется *2D*-геометрией транзистора, но интуитивно понятно, что чем плотнее окружён канал затворами, тем более эффективным будет геометрическое экранирование. На рисунке 4 наиболее сильное геометрическое экранирование реализуется в структуре DG ETSOI (наименьшее Λ), и поэтому ETSOI с двойным затвором меньше всего подвержено эффектам, связанным с *2D*-электростатикой.

Хотя расчёт величины Λ для произвольной геометрии затворов транзистора может оказаться весьма сложным [9, 10], эвристические соображения подсказывают, отчего же может зависеть вели-

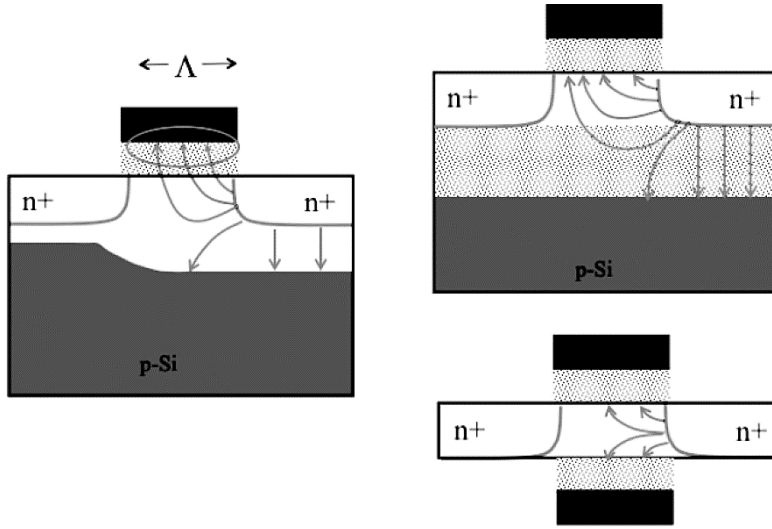


Рис. 4. Геометрическое экранирование в трёх типах MOSFET. Слева: массивная структура MOSFET. Справа наверху: однозатворный SG ETSOI. Справа внизу: DG ETSOI. В MOSFET высокого качества силовые линии электрического поля от стока проникают лишь на расстояние $\approx \Lambda$ в канал проводимости, поскольку большая часть линий закорачивается на верхнем и нижнем затворах [8].⁴

чина Λ . Прежде всего, обратимся к одномерному уравнению Пуассона вглубь кремниевой подложки:

$$\frac{d^2\psi}{dy^2} = \frac{qN_A}{\epsilon_S}. \quad (13)$$

Это уравнение можно переписать иначе, а именно,

$$\frac{d^2\psi}{dy^2} \approx \frac{V_G - \psi_S}{\Lambda^2}, \quad (14)$$

имея просто в виду, что если $V_G > \psi_S$, то $d^2\psi/dy^2$ будет положительным, а коэффициент пропорциональности выбран из соображений размерности. Приравнявая эти оба уравнения, находим, что

$$\frac{V_G - \psi_S}{\Lambda^2} = \frac{qN_A}{\epsilon_S}. \quad (15)$$

Воспользуемся известным решением в режиме истощения:

$$V_G = -\frac{Q(\psi_S)}{C_{ox}} + \psi_S = \frac{qN_A W_D}{C_{ox}} + \psi_S, \quad (16)$$

где W_D есть толщина слоя истощения. Из этих двух уравнений с использованием ((5) из [3]) находим

$$\Lambda = \sqrt{\frac{\varepsilon_s}{\varepsilon_{ox}} W_D t_{ox}}. \quad (17)$$

Из (8) и (4) получаем одномерное уравнение Пуассона вдоль канала проводимости:

$$\frac{d^2 \psi_S(x)}{dx^2} = \frac{\psi_S - V_G}{\Lambda^2} + \frac{qN_A}{\varepsilon_s}, \quad (18)$$

в котором мы сразу определились, что ищем поверхностный потенциал вдоль канала проводимости $\psi_S(x) = \underline{\psi_S(x)}|_{y=0}$. Заменой переменных

$$\phi = \psi_S - V_G + \frac{qN_A}{\varepsilon_s} \Lambda^2 \quad (19)$$

получаем простое дифференциальное уравнение

$$\frac{d^2 \phi}{dx^2} - \frac{\phi}{\Lambda^2} = 0, \quad (20)$$

в решения $\phi = \exp(\pm x/\Lambda)$ которого входит величина Λ , задаваемая по (17).

Итак, мы имеем характеристическую длину геометрического экранирования Λ , на протяжении которой возмущение, вызванное потенциалом стока, затухает. Если длина канала проводимости $L > \Lambda$, то такие эффекты как *DIBL* в короткоканальных транзисторах будут умеренными. Величины $L \approx 1,5-2\Lambda$ приемлемы для нынешних короткоканальных MOSFET. Как следует из (17), тонкие диэлектрические слои предпочтительны, так же как и тонкие области истощения. Как показано на рис. 4, в этих случаях силовые линии электрического поля стока, скорее всего, закоротятся либо на затворе, либо на незаряженной области подложки вместо того, чтобы достичь начала канала проводимости и понизить, пусть даже немного, барьер.

Итак, выше предложена эвристическая оценка характеристической длины геометрического экранирования Λ . Известны более строгие оценки Λ для различных структур MOS [9–11]. В целом, $\Lambda_{BULK} > \Lambda_{DG\ ETSOI}$. Транзистор тем лучше, чем меньше длина геометрического экранирования. Общая закономерность такова, что чем плотнее канал проводимости окружён проводящими электродами, в частности, затворными, тем меньше длина геометрического экранирования.

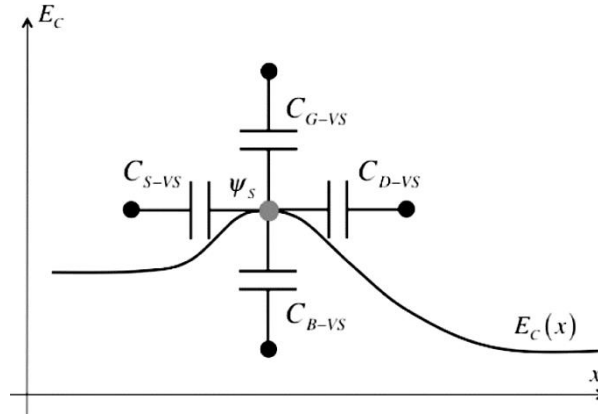


Рис. 5. Ёмкостная модель 2D-электростатики для массивной структуры MOS. Каждый конденсатор моделирует электростатическую связь электродов истока, стока, затвора и подложки с вершиной энергетического барьера.⁵

2D-электростатика: ёмкостная модель. На рисунке 5 показан полезный метод учёта 2D-электростатики. Каждый конденсатор на этом рисунке моделирует электростатическую связь всех четырёх терминалов MOSFET с вершиной барьера, виртуальным истоком VS . Сама вершина барьера располагается вблизи середины канала проводимости при низких значениях V_{DS} и сдвигается к истоку по мере увеличения напряжения на стоке, так что параметры конденсаторов зависят от напряжения на стоке [12]. Нужно решать 2D-уравнение Пуассона при заданной геометрии MOSFET для оценки ёмкости каждого конденсатора, однако, ёмкостной анализ сети конденсаторов полезен сам по себе и приводит к некоторым полезным общим результатам. Рисунок 5 предназначен для массивной структуры MOS. Аналогичные сети конденсаторов могут быть предложены для любых других структур MOS, таких как SG и DG SOI MOSFET [12].

Для анализа простой сети конденсаторов на этом рисунке мы воспользуемся принципом суперпозиции конденсаторов и вначале предположим, что напряжение на терминалы не подаётся, однако, заряд на вершине барьера ненулевой ($Q_s \neq 0$). Соответствующая этой ситуации сеть показана на рис. 6, слева.

Суммарная ёмкость на VS есть

$$C_{\Sigma} = C_{G-VS} + C_{S-VS} + C_{D-VS} + C_{B-VS}, \quad (21)$$

а соответствующий потенциал в локации VS есть

$$\psi_s = \frac{Q_s}{C_{\Sigma}}. \quad (22)$$

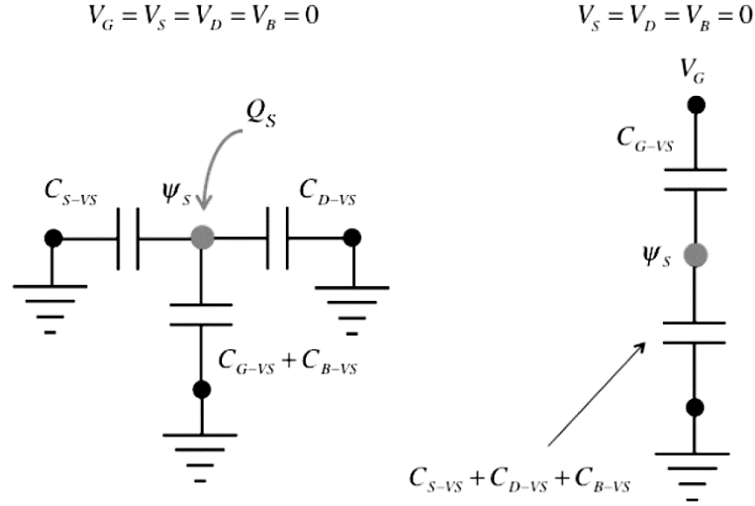


Рис. 6. Упрощённые сети конденсаторов без подачи напряжения. Слева: с зарядом Q_s на виртуальном источке (VS). Справа: с нулевым зарядом на VS и ненулевым напряжением на затворе.⁶

Теперь предположим, что напряжение на затвор подаётся, но остальные терминалы заземлены. Соответствующая сеть показана на рис. 6. Вклад в потенциал в локации VS теперь

$$\psi_s = \left(\frac{C_{G-VS}}{C_\Sigma} \right) V_G; \quad (23)$$

аналогичные вклады получаем от всех других терминалов: подайте на нужный терминал напряжение, а остальные терминалы заземлите. После суммирования всех вкладов от четырёх терминалов и потенциала от ненулевого заряда при нулевых напряжениях на всех терминалах имеем

$$\psi_s = \left(\frac{C_{G-VS}}{C_\Sigma} \right) V_G + \left(\frac{C_{S-VS}}{C_\Sigma} \right) V_S + \left(\frac{C_{D-VS}}{C_\Sigma} \right) V_D + \left(\frac{C_{B-VS}}{C_\Sigma} \right) V_B + \frac{Q_s}{C_\Sigma}. \quad (24)$$

Это уравнение сравним с аналогичным $1D$ -результатом ((11) из [3]), из которого следует, что

$$\psi_s = V_G + \frac{Q_s}{C_{ox}}. \quad (25)$$

Можно показать, что $2D$ -результат (24) сводится к $1D$ (25) в том случае, если ёмкость оксидного слоя затвора намного больше ёмко-

стей остальных терминалов. В этом случае потенциал на вершине барьера полностью контролируется напряжением на затворе, а остальные терминалы не оказывают влияния. Именно к такой ситуации стремятся все разработчики транзисторов: либо создают как можно бóльшую ёмкость затвора, утончая слой диэлектрика или подбирая диэлектрик с как можно более высокой диэлектрической проницаемостью, или максимально, насколько это возможно, уменьшают ёмкость остальных терминалов путём геометрического экранирования с целью электростатически изолировать остальные терминалы от влияния на виртуальный источник на вершине барьера.

Рассмотрим тот случай, когда прилагается напряжение только от затвора и стока, а остальные терминалы заземлены. В подпороговом режиме, когда заряд пренебрежимо мал, уравнение (24) упрощается до

$$\psi_s = \left(\frac{C_{G-VS}}{C_\Sigma} \right) V_G + \left(\frac{C_{D-VS}}{C_\Sigma} \right) V_D. \quad (26)$$

Только затвор и сток влияют на потенциал в локации VS :

$$\frac{\partial \psi_s}{\partial V_G} = \frac{C_{G-VS}}{C_\Sigma}, \quad \frac{\partial \psi_s}{\partial V_D} = \frac{C_{D-VS}}{C_\Sigma}. \quad (27)$$

Транзистор будет высокого качества, если контроль со стороны затвора намного эффективнее, чем со стороны стока: $\partial \psi_s / \partial V_G \gg \partial \psi_s / \partial V_D$. Также стремятся к тому, чтобы изменения потенциала соответствовали изменениям напряжения на затворе: $\partial \psi_s / \partial V_G \approx 1$.

Таким образом, критерии транзистора высокого качества сводятся к выполнению следующих условий:

$$C_{G-VS} \gg C_{D-VS}, \quad C_{G-VS} \approx C_\Sigma. \quad (28)$$

Утончение диэлектрического слоя увеличивает C_{G-VS} , а геометрическое экранирование уменьшает C_{D-VS} .

Конденсаторы в эквивалентной схеме можно напрямую связать с передаточными характеристиками MOSFET. Ток стока экспоненциально зависит от высоты барьера между истоком и каналом проводимости, например, по (7) из [1]:

$$I_{DS} \propto e^{-E_{SB}/kT} = e^{q\psi_s/kT}. \quad (29)$$

Перепишем уравнение (26) иначе:

$$\psi_s = \frac{V_G}{m} + \frac{DIBL}{m} V_D, \quad (30)$$

где

$$m \equiv \frac{C_{\Sigma}}{C_{G-VS}}, \quad DIBL \equiv \frac{C_{D-VS}}{C_{G-VS}}. \quad (31)$$

Теперь ток стока (29) можно с помощью (30) переписать иначе:

$$I_{DS} \propto e^{q\psi_s/kT} = e^{q(V_G + DIBL \cdot V_D)/mkT}. \quad (32)$$

Допороговый разброс при постоянном напряжении на стоке, согласно (4) из [1], определяется как

$$SS = \left[\frac{\partial(\log_{10} I_{DS})}{\partial V_{GS}} \right]^{-1} = 2,3 m kT \quad (33)$$

и даёт такое изменение напряжения на затворе, которое приводит к изменению тока в 10 раз. Допороговый разброс контролируется величиной $m \geq 1$, так что $SS \geq 60$ мВ/декада. Положив $C_{G-VS} = C_{ox}$ и $C_{B-VS} = C_D$, ёмкости полупроводниковой подложки в режиме истощения, из (31) получим

$$m = 1 + \frac{C_D}{C_{ox}} + \frac{C_{S-VS} + C_{D-VS}}{C_{ox}}. \quad (34)$$

Это выражение сравним с уравнением (31) из [3] для коэффициента эффекта подложки в режиме истощения, которое было получено для 1D-модели электростатики. Первое слагаемое, единица, соответствует идеальному допороговому разбросу. Второе слагаемое ответственно за 1D-эффект ((31) в [3]), который обязан падению напряжения на затворе и подложке в режиме истощения. Это слагаемое отсутствует для структуры ETSOI в режиме полного истощения в отличие от массивной структуры MOSFET. Третье слагаемое обязано учёту 2D-электростатики. Мы видим, что учёт 2D-электростатики увеличивает коэффициент эффекта подложки и тем самым увеличивается подпороговый наклон. Этот эффект, показанный на рис. 7, нежелателен, и его всегда стараются уменьшить.

Отметим также, что ёмкостная модель 2D-электростатики описывает эффект *DIBL*. Согласно (32), если увеличить V_D на ΔV_D с поддержанием тока стока постоянным, то необходимо уменьшить V_G . Требуемое уменьшение V_G будет

$$\Delta V_G = -DIBL \cdot \Delta V_D, \quad (35)$$

что фактически является определением эффекта *DIBL*.

Итак, ёмкостная модель фактически позволяет качественно понять эффекты, вызываемые строгим, путём численного решения

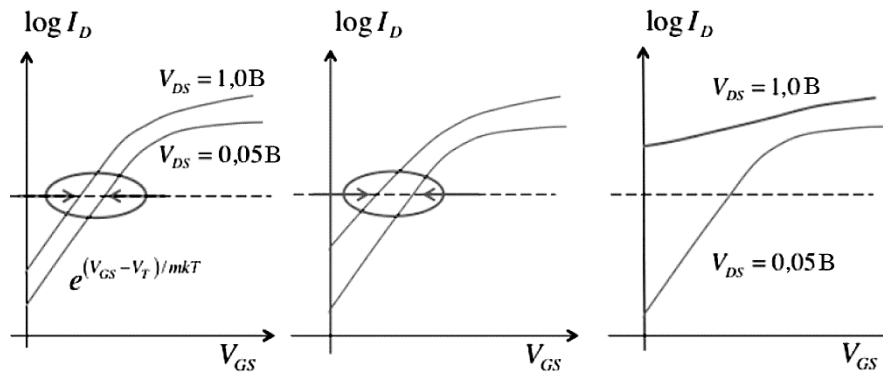


Рис. 7. Иллюстрация происхождения *DIBL* (слева) и увеличения допорогового разброса (по центру), а справа показан также эффект пробоя (punch through), который обсудим ниже.⁷

уравнения Пуассона, учёт $2D$ -электростатики. В качественных транзисторах роль ёмкости затвора должна доминировать во влиянии на виртуальный исток (вершину барьера) со стороны других терминалов, эквивалентные ёмкости которых увеличивают допороговый разброс и вызывают *DIBL*.

3. МАСШТАБИРОВАНИЕ ТРАНЗИСТОРОВ

Последние более чем 50 лет отмечены невиданным прогрессом в области миниатюризации интегральных схем, скейлинга транзисторов и самих электронных устройств. Если скейлинг транзисторов выполнен корректно, производительность интегральных схем улучшается. Выполняя скейлинг транзисторов, основную задачу видят в избавлении у короткоканальных нанотранзисторов от эффектов $2D$ -электростатики.

Глобальная цель скейлинга сформулирована на рис. 8: при уменьшении всех линейных размеров транзистора в k раз напряжения и токи всех его характеристик также должны уменьшиться в k раз.

Рисунок 2, справа иллюстрирует ожидаемое изменение порогового напряжения V_T с уменьшением длины канала проводимости MOSFET. Спад V_T в короткоканальных транзисторах вызван эффектами $2D$ -электростатики. Ниже некоторой минимальной длины канала L_{\min} напряжение V_T не только мало, но и слишком чувствительно к длине канала L . Ниже L_{\min} подпороговый разброс *SS* и *DIBL* становятся неприемлемыми, и транзистор может быть пробит. Эффект пробоя транзистора обсудим ниже. Задачу скейлинга в k раз видят также в том, чтобы уменьшить L_{\min} также в k раз таким образом, чтобы масштабированный транзистор с длиной $L = L_{\min}/k$

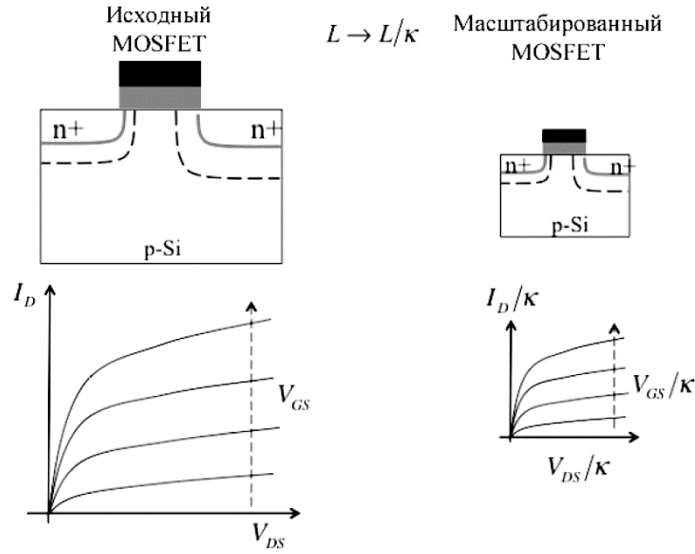


Рис. 8. Скейлинг транзистора. Слева показаны исходный транзистор и набор его характеристик. Справа показан масштабированный транзистор с коэффициентом скейлинга $k > 1$. Если масштабирование выполнено корректно, тогда в масштабированных характеристиках все токи и напряжения должны измениться с тем же коэффициентом скейлинга.⁸

не был подвержен короткоканальным $2D$ -эффектам.

Описанный выше подход к масштабированию транзисторов был предложен Робертом Деннардом более 40 лет назад [13] и многие десятилетия служит руководством по скейлингу не только транзисторов, но и интегральных схем и самих конечных электронных устройств. Основная идея скейлинга Деннарда заключается в том, чтобы все размеры уменьшить в k раз, увеличить допинг в это же число раз и уменьшить напряжение источника питания в k раз. Такая постановка задачи поддерживает электрическое поле в канале постоянным по ходу скейлинга. Более подробно скейлинг Деннарда заключается в следующем:

1) уменьшение всех размеров:

$$L, W \Rightarrow (L, W) / k, t_{ox} \Rightarrow t_{ox} / k, W_D \Rightarrow W_D / k, y_j \Rightarrow y_j / k; \quad (36)$$

2) увеличение допирования канала:

$$N_A \Rightarrow kN_A; \quad (37)$$

3) уменьшение напряжения источника питания:

$$V_{DD} \Rightarrow V_{DD} / k. \quad (38)$$

Величина W_D — это ширина области истощения, а y_j — это глубина p - n -перехода истока/стока.

Прикинем, как всё это работает, используя самые простые соображения. Прежде всего, электрическое поле $E \approx V_{DD}/L$, и поэтому V_{DD} и L уменьшаются одинаково с тем, чтобы электрическое поле в исходном и в масштабированном канале было одним и тем же.

Низкополевая скорость электронов есть подвижность, помноженная на электрическое поле. Предполагая, что подвижность не изменяется при скейлинге, скорость носителей тока также не изменится. Деннард предположил, что высокополевая скорость совпадает с v_{sat} , которая обычно служит материальным параметром, не изменяемым при масштабировании. В итоге, скорость носителей при скейлинге не меняется.

Важно масштабировать толщину области истощения

$$W_D = \sqrt{\frac{2\epsilon_s}{qN_A}(V_{bi} + V_{DD})},$$

где V_{bi} — встроенный потенциал p - n -переходов. Если $V_{DD} \gg V_{bi}$, то рост допинга в k раз и уменьшение V_{DD} в это же число раз приводит к уменьшению ширины области истощения в k раз. Если t_{ox} и y_j так же масштабировать, то эффекты 2D-электростатики в масштабированном транзисторе станут проявляться сильнее при длине канала проводимости в k раз короче, чем в исходном транзисторе. В результате величина L_{min} уменьшится приблизительно в k раз.

Ёмкость

$$C = \frac{\epsilon A}{t} [\Phi],$$

где t — толщина оксидного слоя или же области истощения. Поскольку все толщины уменьшаются в k раз, а площадь A — в k^2 раз, все ёмкости уменьшатся в k раз, однако C_{ox} , как ёмкость оксидного слоя на единицу поверхности, увеличится в k раз.

Рассмотрим влияние скейлинга Деннарда на некоторые важные величины. Заряд инверсионного слоя $Q = -C_{ox}(V_G - V_T)$. Поскольку C_{ox} растёт в k раз, а напряжения уменьшаются в это же число раз, то заряд инверсионного слоя на единицу поверхности не изменяется при масштабировании.

Теперь рассмотрим ток $I_{DS} = WQv$. Поскольку заряд и скорость не меняются при масштабировании, а ширина уменьшается в k раз, то и ток уменьшится в k раз.

Подводя итоги, приходим к следующему общему выводу из скейлинга Деннарда при сохранении постоянства электрического поля:

$$Q \Rightarrow Q, v \Rightarrow v, C \Rightarrow C/k, C_{ox} \Rightarrow kC_{ox}, I_{DS} \Rightarrow I_{DS}/k. \quad (39)$$

Пользуясь скейлингом Деннарда, проверим работоспособность масштабированных сетей. Задержка сети — это время, необходимое для удаления заряда CV_{DD} , накопленного сетевыми ёмкостями,

$$\tau = \frac{CV_{DD}}{I_{DS}}.$$

Как видим, задержка сети τ уменьшается в k раз. Мощность $P_D = V_{DD}I_{DS}$ уменьшается в k^2 раз. Плотность мощности в единицах Вт/м² не изменяется при скейлинге. Поскольку размер каждого транзистора уменьшается в k^2 раз, то плотность упаковки транзисторов D увеличивается в k^2 раз. Наконец, такая важная метрика как произведение мощности на задержку сети $P_D\tau$ уменьшается в k^3 раз.

Итак, суммируем результаты скейлинга Деннарда при сохранении постоянства электрического поля:

$$\begin{aligned} \tau &= CV_{DD}/I_{DS} \Rightarrow \tau/k, \quad P_D = V_{DD}I_{DS} \Rightarrow P_D/k^2, \\ P_D/A &\Rightarrow P_D/A, \quad D = \frac{N_{FET}}{A} \Rightarrow D \times k^2, \quad P_D\tau = CV_{DD}^2 \Rightarrow P_D\tau/k^3. \end{aligned} \quad (40)$$

Соблюсти скейлинг по Деннард не так легко, как кажется на первый взгляд, поскольку некоторые величины не масштабируются. Например, вспомните уравнение для порогового напряжения V_T (11). Напряжение плоских зон V_{FB} не масштабируется, а потенциал ψ_B относительно нечувствительный к скейлингу, так что в соответствии со сценарием Деннарда $V_T \Rightarrow V_T/k$. Учтём, однако, что нередко применяются нетрадиционные режимы допинга [7], что может сказаться на результатах скейлинга.

Стоковая область истощения зависит от $\sqrt{(V_{bi} + V_{DD})/N_A}$. Поскольку встроенный потенциал V_{bi} не масштабируется, то $W_D \Rightarrow W_D/k$ (36). Подпороговый разброс не чувствителен к скейлингу. Все эти факторы обнадёживают. Вместе с тем, разработчики транзисторов и чипов постоянно совершенствуют скейлинг с целью предложить устройства с ещё большей экономией мощности и меньшей задержкой.

В настоящее время масштабирование устройств испытывает серьёзные вызовы, а некоторые даже предрекают скорый конец эры скейлинга. Одна из проблем связана с оксидным слоем затвора, поскольку достигнут предел в утончении этого слоя вплоть до и даже меньше 1,5 нм, что соответствует нескольким молекулам SiO₂ [14]; дальнейшее утончение ведёт к чрезмерному протеканию тока. Это обстоятельство вынуждает переходить от плоских MOSFET к FinFET [15], которые обеспечивают лучший электростатический контроль при той же толщине оксидной плёнки [16]. Ещё один вызов дальнейшему масштабированию транзисторов связан с неудач-

ными попытками скейлинга подпорогового разброса SS . Задавшись максимально разрешённым током в режиме «OFF», величина SS должна быть лишь немного больше чем 60 мВ/декада и ток должен расти линейно выше порога, что требует $V_{DD} \approx 1$ В для достижения нужного значения тока I_{DS} в режиме «ON». В результате не удаётся масштабировать напряжение питания. Предложено несколько новых архитектур транзисторов для решения подобных проблем [17].

4. ПРОБОЙ ТРАНЗИСТОРОВ

Усиление 2D-эффектов может привести к пробую, т.е. к закорачиванию стока на исток. В этой ситуации ток идёт от стока к истоку в отсутствие или при минимальном влиянии напряжения на затворе (рис. 9).

На рисунке 9, слева влияние эффектов 2D-электростатики незначительное: подпороговый разброс лишь чуть больше 60 мВ/декада и $DIBL$ приемлемый. Когда 2D-эффекты становятся заметными (рис. 9, по центру), подпороговый разброс существенно возрастает, а $DIBL$ становится весьма большим, например, более 100 мВ/В. При доминировании 2D-эффектов (рис. 9, справа) наблюдается резкая деградация функциональности транзистора. Ток лишь слабо зависит от напряжения на затворе, а $DIBL$ вообще не поддаётся определению, поскольку подпороговые выходные характеристики при низких и высоких напряжениях на стоке даже не параллельны оси напряжений.

Рисунок 9 показывает также, как учёт 2D-электростатики влияет на выходные характеристики транзисторов. При достаточно

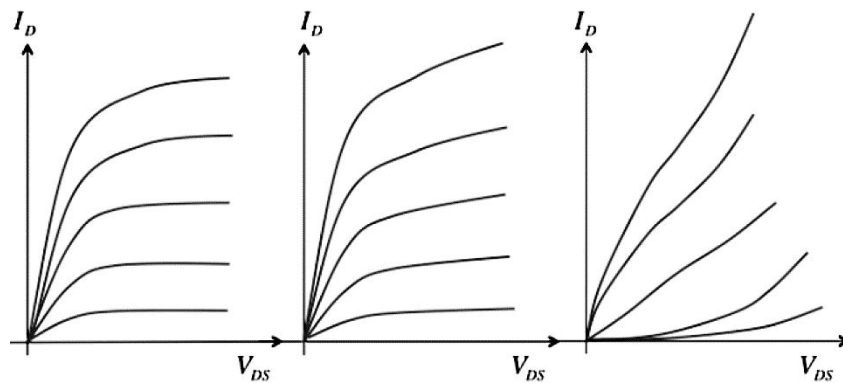


Рис. 9. Иллюстрация влияния 2D эффектов на выходные характеристики MOSFET. Слева: достаточно длинный канал проводимости с очень большим выходным сопротивлением. По центру: короткий канал с низким сопротивлением. Справа: транзистор повреждён пробоем.⁹

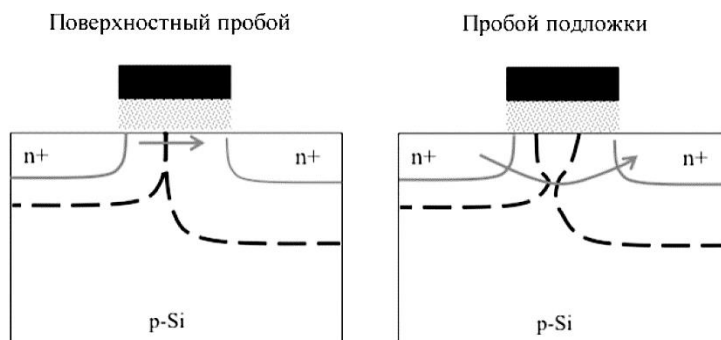


Рис. 10. Сечения MOSFET, иллюстрирующие границы истощения при поверхностном пробое (слева) и при пробое подложки (справа).¹⁰

длинном канале проводимости ток стока в области насыщения постоянный и выходное сопротивление приближается к бесконечности (рис. 7 из [1]). Для MOSFET с коротким каналом выходное сопротивление существенно уменьшается. Причину этого легко понять: ток пропорционален $(V_{GS} - V_T)$, а V_T уменьшается с ростом напряжения на стоке из-за эффекта *DIBL*. Не сразу очевидно, что роль *2D*-электростатики в подпороговой области, где формируется и измеряется *DIBL* и где заряд в канале пренебрежимо мал, так же важна, как и выше порога, где подвижный заряд в канале велик. Показано, однако, что в качественно собранных транзисторах одни и те же эффекты *DIBL* *2D*-электростатики наблюдаются как ниже, так и выше порогового напряжения [11, 18]. Наконец, справа на рис. 9 показаны выходные характеристики транзистора с пробоем: напряжение на стоке очень сильно влияет на ток даже в «области насыщения».

Пробой происходит тогда, когда электрическое поле стока покрывает весь канал проводимости вплоть до истока: приблизительно тогда, когда область истощения стока смыкается с областью истощения истока (рис. 10, слева).

Как показано справа на рис. 10, границы областей истощения могут иметь сложные профили как в результате режимов *2D*-допирования, так и вследствие эффектов *2D*-электростатики. Итак, перекрывание областей истощения может происходить как вблизи поверхности раздела Si/SiO₂, так и в подложке глубже, соответственно чему случается поверхностный пробой или же пробой подложки.

Критерий избежать пробоя $L > W_S + W_D$ служит лишь очень грубой оценкой. Лучшее пояснение даёт рис. 11.

Абсолютный пробой имеет место тогда, когда потенциал стока не просто покрывает весь канал проводимости и не просто понижает барьер, а ликвидирует барьер напрочь. В такой ситуации электроны

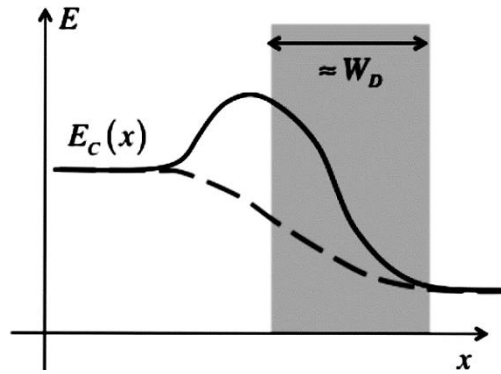


Рис. 11. Иллюстрация пробоя. Сплошная линия соответствует «хорошему» состоянию транзистора, пунктирная — пробоем. Серая зона показывает границы области истощения стока.¹¹

перетекают из истока в сток, не нуждаясь в действии потенциала затвора. Пробой можно определить иначе: как ситуацию, когда контроль над током со стороны стока так же эффективен, как и контроль со стороны затвора. Согласно ёмкостной модели (27), пробой происходит когда $C_{G-VS} = C_{D-VS}$. Реальное напряжение на стоке, вызывающее пробой, можно оценить численно путём решения 2D-уравнения Пуассона для транзистора конкретной структуры.

Итак, мы обсудили, как 2D-электростатика нарушает работу короткоканальных транзисторов. Как же функционирует электростатически хорошо сбалансированный MOSFET (рис. 12)? В таком транзисторе есть область вблизи начала канала проводимости, которая находится под сильным контролем со стороны затвора. В этой области поле dE_c/dx мало и она включает в себя вершину барьера, «экранированную» от влияния потенциала стока, что позволяет поддерживать незначительным *DIBL*.

Потенциал в районе вершины барьера контролирует высоту барьера, а тем самым, и ток стока в транзисторе. В идеале этот потенциал должен определяться только затвором, как в уравнении (25). На практике напряжение на стоке всегда оказывает некоторое влияние на потенциал в области вершины барьера, как например, согласно (26); особенно это касается короткоканальных MOSFET. Глобальная задача, стоящая перед дизайнерами транзисторов, заключается в том, чтобы гарантировать подчинение заряда инверсионного слоя на вершине барьера классическому 1D-результату, а именно:

$$Q|_{x=0} = -C_G (V_{GS} - V_T), \quad (41)$$

где $x = 0$ соответствует вершине барьера. Физически корректно предполагать, что именно 1D-электростатика MOS, учёт которой

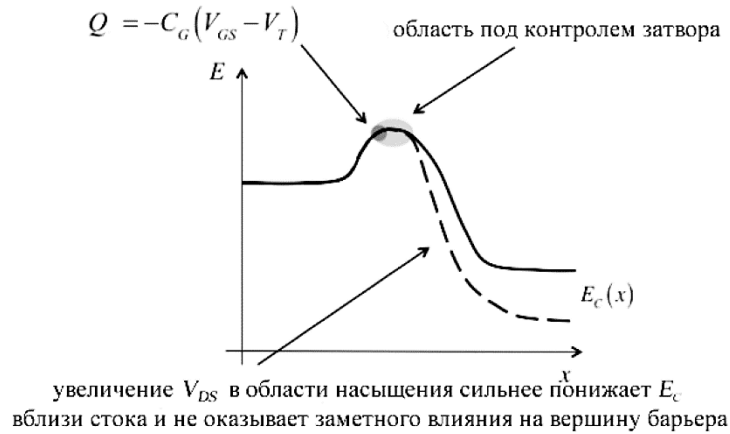


Рис. 12. Иллюстрация того, как работает электростатически хорошо сбалансированный MOSFET при большом напряжении как на стоке, так и на затворе. Как показывает пунктирная линия, рост напряжения на стоке увеличивает потенциал в области насыщения, не затрагивая барьер в начале канала, который находится под контролем затвора.¹²

приводит к (41), применима к вершине барьера, поскольку именно в этой локации кривизна потенциала $d^2\psi/dx^2 = 0$, и $2D$ -уравнение Пуассона сводится к $1D$ -уравнению. Последнее утверждение не совсем корректно, поскольку при учёте $2D$ -электростатики пороговое напряжение оказывается зависящим от напряжения на стоке:

$$V_T = V_{T0} - \delta V_{DS}, \quad (42)$$

где параметр δ учитывает *DIBL*.

Ток в MOSFET при высоком напряжении на стоке создаётся электронами, которые преодолевают барьер, диффундируют через небольшую область за вершиной барьера в начале канала, характеризующуюся низким полем, затем входят в область с высоким полем в стоковой части канала. Узким местом, которое ограничивает ток стока, является эта небольшая область с низким полем. Эта картина функционирования MOSFET фактически подобна тому, как работает биполярный транзистор, в котором роль истока играет эмиттер, низкополевая область в начале канала играет роль базы, а высокополевая область стока — это коллектор. Аналогия между MOSFET и биполярным транзистором очень тесная [19].

При низком напряжении на стоке ток пропорционален V_{DS} , однако, при высоком стоковом напряжении ток в хорошо сбалансированном MOSFET зависит от V_{DS} в меньшей степени. В транзисторе с длинным каналом ток действительно насыщается. Такое поведение имеет место потому, что сильный контроль затвора экранирует по-

тенциал в районе начала канала от влияния потенциала стока. Увеличение V_{DS} за пределы напряжения насыщения V_{DSAT} , прежде всего, увеличивает потенциал и электрическое поле в районе стокового конца канала проводимости. В хорошо сбалансированном MOSFET напряжение стока лишь несущественно влияет на потенциал в самом начале канала, что в итоге приводит к конечной выходной проводимости транзистора. Хотя и нет причин полагать, что роль 2D-электростатики ниже порога такая же, как и выше порога, однако, опыт применения модели виртуального истока для хорошо сбалансированных MOSFET свидетельствует, что это действительно так [18], и численное моделирование подтверждает этот вывод [11].

Изложенная картина влияния электростатики в качественно собранных транзисторах будет использована позже для более глубокого понимания физики нанотранзисторов.

5. МОДЕЛЬ ВИРТУАЛЬНОГО ИСТОКА

Упрощённая модель виртуального истока, модель VS нулевого уровня была изложена в [2]. Она построена с использованием простых и традиционных соображений в отличие от полноценной модели VS [18], которая была разработана целенаправленно с учётом физики нанотранзисторов. Сейчас мы уточним модель нулевого уровня и изложим исходную модель VS и физические принципы, лежащие в её основе.

Ток стока MOSFET в общем случае

$$I_{DS} = W \left| \underline{Q} \right|_{x=0} (V_{GS}, V_{DS}) \left| \left\langle v_x (V_{GS}, V_{DS}) \right\rangle \right|_{x=0}, \quad (43)$$

где $x = 0$ соответствует локации виртуального истока VS на вершине барьера. Ток непрерывен, так что мы вправе вычислять его там, где нам удобнее. В хорошо «темперированном» MOSFET на вершине барьера, как следует из анализа 1D-электростатики, $Q(V_{GS}, V_{DS}) \approx Q(V_{GS})$. Нужна ещё только небольшая поправка на $DIBL$. Далее мы уточним модель VS нулевого уровня, приняв во внимание эффекты электростатики MOS. Что же касается скорости электронов $\left\langle v_x (V_{GS}, V_{DS}) \right\rangle \big|_{x=0}$ в применении к нанотранзисторам,

нам придётся отказаться от концепции «сверху–вниз», которую мы до сих пор исповедовали, и воспользоваться обобщённой моделью транспорта электронов Ландауэра–Датта–Лундстрёма (ЛДЛ) [20–25], единственной полноценно пригодной для описания классического транспорта в нанoeлектронике.

Модель VS нулевого уровня исходит из двух отдельных выражений (28) из [2] для тока стока в линейной области и в области насы-

щения:

$$I_{DLIN} = \frac{W}{L} \mu |Q(V_{GS})| V_{DS}, \quad I_{DSAT} = W v_{sat} Q(V_{GS}), \quad (44)$$

которые показаны пунктирными прямыми на рис. 13. Реальная выходная характеристика, показанная на этом рисунке сплошной линией, получается плавным сочленением линейной области и области насыщения с помощью выражения для средней скорости, зависящей от напряжения на стоке, в виде

$$\langle v_x(V_{DS}) \rangle = F_{SAT}(V_{DS}) v_{sat}, \quad F_{SAT}(V_{DS}) = \frac{V_{DS}/V_{DSAT}}{[1 + (V_{DS}/V_{DSAT})^\beta]^{1/\beta}}, \quad (45)$$

где F_{SAT} — функция насыщения тока стока, а напряжение насыщения на стоке

$$V_{DSAT} = \frac{v_{sat} L}{\mu}. \quad (46)$$

Ранее заряд на вершине барьера был описан следующим образом:

$$Q(V_{GS}) = 0, \quad V_{GS} \leq V_T; \quad Q(V_{GS}) = -C_{ox}(V_{GS} - V_T), \quad V_{GS} > V_T; \quad (47)$$

$$V_T = V_{T0} - \delta V_{DS}.$$

Мы теперь понимаем, что C_{ox} в (47) нужно заменить ёмкостью C_G

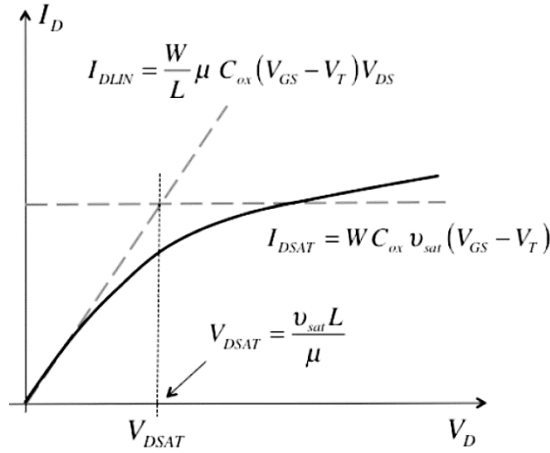


Рис. 13. Выходная характеристика n -MOSFET, включённого по схеме с общим истоком, при некотором фиксированном значении напряжения на затворе (сплошная линия). Пунктирные прямые дают ток в линейной области и в области насыщения согласно (44).¹³

(58) из [3], как последовательную комбинацию C_{ox} и ёмкости полупроводника в режиме инверсии C_S^{inv} , и что $C_G < C_{ox}$. Мы также понимаем теперь, как описать $Q(V_{GS}, V_{DS})$ ниже порога и таким образом можно уточнить модель VS , включив как подпороговую проводимость, так и надпороговую.

Подпороговая область. При напряжении на затворе ниже порогового напряжения MOSFET находится в подпороговом режиме. Рисунок 14 иллюстрирует ход зависимости $Q(V_{GS}) \propto V_{GS}$ в линейном масштабе и в логарифмическом. Ранее было показано ((45) в [3]), что поверхностная электронная плотность в массивном MOSFET в подпороговом режиме даётся выражением

$$Q(V_G) = -(m-1)C_{ox} \frac{kT}{q} e^{q(V_G - V_T)/mkT}. \quad (48)$$

Аналогичный результат ((96) в [3]) был получен для ETSOI только с $m = 1$. Ключевым для обоих случаев является экспоненциальная зависимость $Q(V_G) \propto \exp[q(V_{GS} - V_T)/mkT]$ в подпороговом режиме.

Из (43) и (48) получаем уравнение для тока в массивном MOSFET в подпороговом режиме, а именно:

$$I_{DS} = W(m-1)C_{ox} \frac{kT}{q} e^{q(V_G - V_T)/mkT} \langle v_x \rangle \Big|_{x=0}. \quad (49)$$

Вспомним, что, согласно ёмкостной модели (34),

$$m = 1 + \frac{C_\Sigma}{C_{ox}}, \quad (50)$$

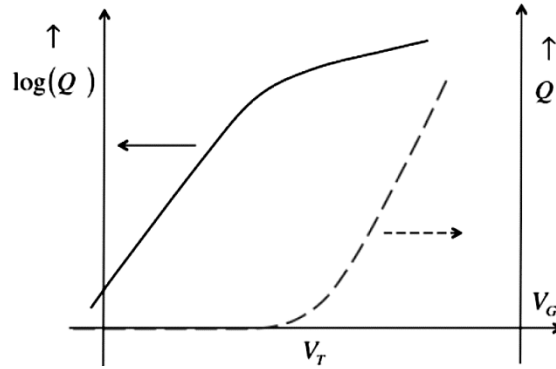


Рис. 14. Зарядовая плотность в инверсионном слое в зависимости от напряжения на затворе (пунктирная кривая справа), а также в логарифмическом масштабе (сплошная кривая слева).¹⁴

где C_{Σ} есть суммарная ёмкость, влияющая на виртуальный исток VS . В случае массивного MOSFET это — сумма ёмкостей всех четырёх терминалов (21).

Для тока стока в подпороговом режиме знаем также подпороговый разброс ((4) из [1])

$$SS = \left[\frac{\partial(\log_{10} I_{DS})}{\partial V_{GS}} \right]^{-1} = 2,3 m \frac{kT}{q} \text{ [В/декада]}. \quad (51)$$

Подпороговый разброс обычно приводится в мВ/декада; $SS < 100$ мВ/декада считается приемлемым подпороговым разбросом.

Рисунок 15 показывает, почему подпороговый разброс является такой важной метрикой транзисторов.

Низкие значения I_{OFF} , когда электронная схема устройства не потребляет избыточную мощность в режиме ожидания, предпочтительнее в любых устройствах. Для заданного значения I_{OFF} метрика SS определяет то напряжение источника питания V_{DD} , которое обеспечит нужный ток I_{ON} . Высокие значения последнего благоприятны для ускоренного функционирования электронной схемы, поскольку все ёмкости схемы могут заряжаться и разряжаться быстрее. У транзистора с меньшим SS нужное значение тока I_{ON} достигается при меньшем напряжении источника питания V_{DD} . Электронная схема будет работать с той же скоростью, но поскольку мощность пропорциональна V_{DD}^2 , то схема будет диссипировать меньшую мощность. Размещая в настоящее время несколько миллиардов транзисторов на одной интегральной схеме, потребляемая

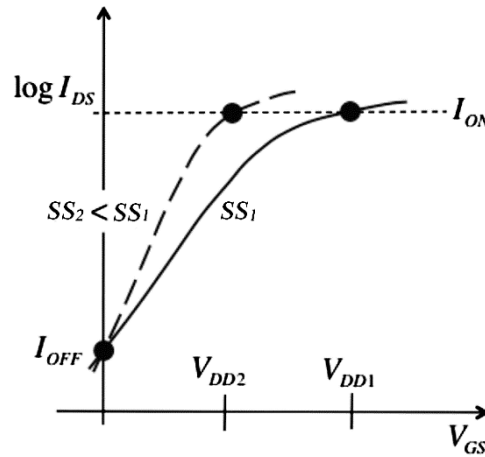


Рис. 15. Иллюстрация того, как подпороговый разброс SS связан с напряжением источника питания V_{DD} . Ток в режиме «ON» соответствует максимальному напряжению источника питания, поданному на затвор.¹⁵

мощность приобретает критическое значение — как активная мощность в рабочем режиме пропорциональная V_{DD}^2 , так и мощность в режиме ожидания, определяемая значением I_{OFF} .

Согласно (51), наименьший подпороговый разброс SS при комнатной температуре равен 60 мВ/декада. Полностью истощённые транзисторы такие, как ETSOI MOSFET, характеризуются значением $m = 1$ и наиболее благоприятны для достижения наименьших значений SS . Выше порога ток растёт приблизительно линейно с увеличением напряжения на затворе, так что минимальный предел для SS , равный 60 мВ/декада, определяет и нижний предел напряжения источника питания V_{DD} . На практике сейчас этот нижний предел составляет 1 В. Напомним, что скейлинг Деннарда транзисторов при сохранении электрического поля постоянным требует масштабирования V_{DD} (38) с каждой новой технологической генерацией. Поскольку у подпорогового разброса SS есть нижний предел (при комнатной температуре!), то мощность источника питания масштабированию уже не поддаётся. Это обстоятельство и составляет суть критической ныне проблемы диссипации мощности интегральных схем [26].

Итак, у подпорогового разброса есть нижний предел с учётом того, что $m \geq 1$. Однако, какова его природа, откуда он берётся? Это поясняет рис. 16.

Ток стока образуют электроны, эмитированные истоком и преодолевшими барьер, которые далее скатываются к стоку. Вероятность этой термоионной эмиссии экспоненциально зависит от высоты барьера E_B , а именно:

$$P_{S \rightarrow D} = e^{-E_B/kT}. \quad (52)$$

Эта экспоненциальная зависимость влечёт за собой экспоненциаль-

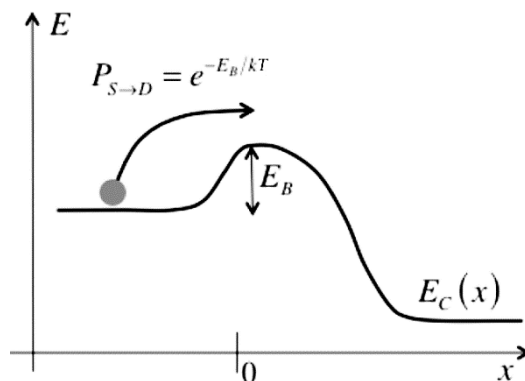


Рис. 16. Термоионный ток в MOSFET.¹⁶

ную зависимость $Q \propto V_{GS}$ и ограничивает подпороговый разброс $SS \geq 60$ мВ/декада при комнатной температуре. Предложены физические подходы, позволяющие преодолеть это ограничение [27, 28].

Остаётся обсудить среднюю скорость электронов в канале проводимости $\langle v_x \rangle|_{x=0}$ в подпороговом режиме; приведённое выше выражение (15) применимо лишь выше порога. В подпороговом режиме средняя скорость в локации виртуального истока (вершина барьера) определяется выражением

$$\langle v_x \rangle|_{x=0} = \frac{D}{L} \left(\frac{n_s(0) - n_s(L)}{n_s(0)} \right), \quad (53)$$

где D есть коэффициент диффузии электронов. Из простой модели термоионной эмиссии следует, что

$$n_s(L) / n_s(0) = e^{-qV_{DS}/kT},$$

так что

$$\langle v_x \rangle|_{x=0} = \frac{D}{L} (1 - e^{-qV_{DS}/kT}) = \frac{kT}{q} \frac{\mu}{L} (1 - e^{-qV_{DS}/kT}), \quad (54)$$

где было использовано соотношение Эйнштейна между коэффициентом диффузии электронов и их подвижностью. Подставляя (54) в (49), получаем стандартное выражение для тока в подпороговом режиме [7]:

$$I_{DS} = \mu C_{ox} \frac{W}{L} (m-1) \left(\frac{kT}{q} \right)^2 e^{q(V_{GS}-V_T)/mkT} (1 - e^{-qV_{DS}/kT}). \quad (55)$$

Отметим ещё один момент. Из рис. 16 совсем не очевидно, что электроны должны диффундировать через весь канал проводимости. Представляется физически более корректным, что электроны диффундируют через низкополевую область канала, а затем более сильное электрическое поле быстро проносит их через оставшуюся часть канала. В соответствии с такой моделью длину канала проводимости L надо бы заменить меньшим значением $l < L$. Исходя даже из этого обстоятельства, на практике затруднительно определить предэкспоненциальный фактор, так что уравнение (55) обеспечивает удовлетворительное описание реальных транзисторов в целом [7].

Из-под порога в область выше порога. Уравнение (48) даёт зависимость $Q(V_{GS})$ ниже порога, а в режиме сильной инверсии, согласно (58) из [3], $Q(V_{GS}) = -C_G^{inv} (V_{GS} - V_T)$, однако, переход из подпорого-

вого режима в режим сильной инверсии является постепенным, что необходимо учитывать. Это особенно важно при моделировании и при численном решении уравнений Пуассона–Больцмана для $Q(\psi_s)$, лежащем в основе так называемых моделей поверхностного потенциала [29].

Поведение $Q \propto V_{GS}$ можно описать эмпирически. Одно из выражений было предложено таким [30]:

$$Q(V_{GS}) = -m C_G^{inv} \left(\frac{kT}{q} \right) \ln \left(1 + e^{q(V_{GS} - V_T)/mkT} \right). \quad (56)$$

Для $V_{GS} \ll V_T$ можно воспользоваться приближением $\ln(x) \approx 1 + x$ и переписать (56) в виде

$$Q(V_{GS}) = -m C_G^{inv} \left(\frac{kT}{q} \right) e^{q(V_{GS} - V_T)/mkT}, \quad (57)$$

которое оказывается близким к уравнению (48). В практических расчётах предэкспоненциальный множитель не является существенным, так что эмпирическое выражение (57) вполне приемлемо.

При $V_{GS} \gg V_T$ уравнение (56) даёт уже известный правильный результат:

$$Q(V_{GS}) = -C_G^{inv} (V_{GS} - V_T). \quad (58)$$

Таким образом, эмпирическое уравнение (56) пригодно для описания поведения $Q \propto V_{GS}$ от подпороговой области до области с сильной инверсией. Модель виртуального истока VS использует несколько улучшенную версию уравнения (46), обеспечивая тем самым лучшее сочетание областей со слабой и сильной инверсией [18].

В заключение отметим тесную связь между токами I_{OFF} и I_{ON} . Мы видели, что $I_{OFF} \propto \exp[(V_{GS} - V_T)/mkT]$, а ток $I_{ON} \propto V_{GS} - V_T$, так что в итоге

$$\ln I_{OFF} \propto I_{ON}. \quad (59)$$

Системные дизайнеры могли бы уменьшить пороговое напряжение V_T с тем, чтобы увеличить I_{ON} , что улучшит быстродействие, однако это влечёт за собой экспоненциальный рост I_{OFF} и, как следствие, увеличение диссипации мощности в режиме ожидания. Рисунок 17 может служить примером такого компромисса применительно к технологии 65 нм NMOS. Этот фундаментальный компромисс является непосредственным следствием физических принципов функционирования MOSFET.

Уравнение (46) даёт значение напряжения на стоке V_{DSAT} в режиме сильной инверсии. Модель виртуального истока VS сводит

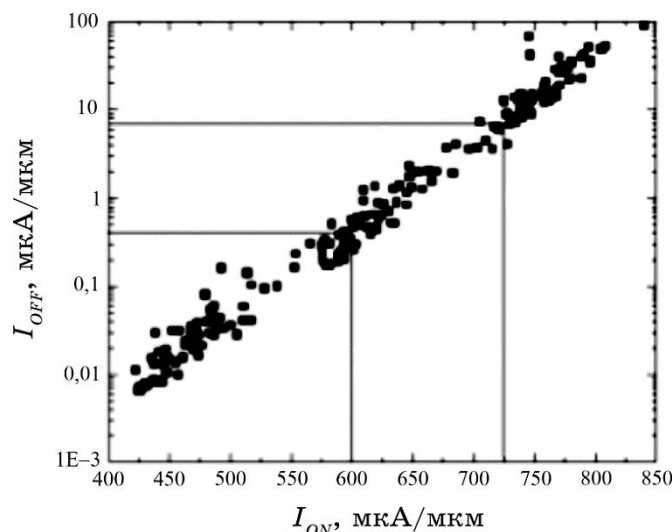


Рис. 17. График зависимости $\log_{10} I_{OFF} \propto I_{ON}$ для технологии 65 нм NMOS [31].¹⁷

V_{DSAT} в режиме сильной инверсии к kT/q в подпороговом режиме эмпирически. Характерная ошибка такой эвристической процедуры меньше 10% [18].

Как свидетельствует рис. 7 из [2], модель VS с высокой точностью вписывается в экспериментальные данные для нанотранзисторов. Это удивительно, поскольку такие параметры, как μ и v_{sat} , имеют очевидный и ясный физический смысл в том случае, когда длина канала проводимости во много раз превосходит среднюю длину свободного пробега, но это не та ситуация, которая имеет место в нанотранзисторах. Остаётся пока рассматривать эти величины как эмпирические параметры. Далее будет развита теория транспорта электронов в нанотранзисторах, и мы увидим, что этим параметрам может быть придан ясный физический смысл.

Подведём итоги. Подробно рассмотрена электростатика MOS. $1D$ -электростатика изгибает зоны, понижает барьер и позволяет потоку электронов двигаться от истока к стоку. $2D$ -электростатика деградирует транспорт электронов в полевых транзисторах, увеличивая подпороговый разброс и вызывая $DIBL$, который в свою очередь увеличивает выходную проводимость и уменьшает пороговое напряжение в короткоканальных транзисторах. Количественный учёт $2D$ -электростатики требует численного подхода; вместе с тем все существенные эффекты физически понятны.

$2D$ -электростатика разрушает функционирование транзисторов и приводит к: 1) подпороговому разбросу большему, чем фундаментальный предел в 60 мВ/декада; 2) сдвигу проходных характеристик

$\log_{10} I_{DS} \propto V_{GS}$ влево при увеличении напряжения на стоке (*DIBL*); 3) возникновению порога в зависимости от параметров затвора и напряжения на стоке; 4) низкому выходному сопротивлению. Когда эффекты 2D-электростатики сильны, затвор теряет контроль над током и транзистор претерпевает пробой. Поскольку эти эффекты проявляются сильнее в короткоканальных транзисторах, их ещё называют эффектами короткого канала. По мере того, как транзисторы становятся всё меньше и меньше, основной вызов, с которым встречаются схемотехники, сводится к контролю короткоканальных эффектов. Как правило, требуется численное моделирование.

Возвращаясь к уравнению (43), видим, что зависимость $Q|_{x=0}(V_{GS}, V_{DS})$ имеет вполне приемлемое физическое объяснение.

Что же касается $\langle v_x(V_{GS}, V_{DS}) \rangle|_{x=0}$ в этом уравнении, то далее будет построена адекватная физическая картина транспортных явлений в нанотранзисторах, исходя из обобщённой модели транспорта электронов ЛДЛ, и установлена связь её с традиционным подходом «сверху–вниз».

В основу настоящего обзора положены лекции Марка Лундстрёма «Fundamentals of Nanotransistors» [25], прочитанные им в 2016 году в рамках инициативы [Purdue University/nanohub.org/u](http://PurdueUniversity/nanohub.org/u).

ЦИТИРОВАННАЯ ЛИТЕРАТУРА—REFERENCES

1. Yu. A. Kruglyak, *Nanosistemi, Nanomateriali, Nanotehnologii*, **16**, No. 2: 201 (2018) (in Russian); Ю. А. Кругляк, *Наносистемы, наноматериалы, нанотехнологии*, **16**, вып. 2: 201 (2018).
2. Yu. A. Kruglyak, *Nanosistemi, Nanomateriali, Nanotehnologii*, **16**, No. 2: 233 (2018) (in Russian); Ю. А. Кругляк, *Наносистемы, наноматериалы, нанотехнологии*, **16**, вып. 2: 233 (2018).
3. Yu. A. Kruglyak, *Nanosistemi, Nanomateriali, Nanotehnologii*, **16**, No. 3: 465 (2018) (in Russian); Ю. А. Кругляк, *Наносистемы, наноматериалы, нанотехнологии*, **16**, вып. 3: 465 (2018).
4. R. F. Pierret, *Semiconductor Device Fundamentals* (New York: Addison Wesley: 1996).
5. B. Streetman and S. Banerjee, *Solid State Electronic Devices* (New York: Prentice Hall: 2005).
6. Y. Tsiividis and C. McAndrew, *Operation and Modeling of the MOS Transistor* (New York: Oxford Univ. Press: 2011).
7. Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices* (New York: Oxford Univ. Press: 2013).
8. D. Frank, Yuan Taur, and H.-S. Philip Wong, *IEEE Device Research Conf. Technical Digest*, 18–21 (1999).
9. D. Frank, Yuan Taur, and H.-S. Philip Wong, *IEEE Electron Device Lett.*, **19**: 385 (1998).
10. J. Wang, P. Solomon, and M. Lundstrom, *IEEE Trans. Electron Dev.*, **51**:

- 1361 (2004).
11. Q. Xie, J. Xu, and Yu. Taur, *IEEE Trans. Electron Dev.*, **59**: 1569 (2012).
12. R. Koh, H. Kato, and H. Matsumoto, *Jap. J. Appl. Phys.*, **35**, Pt. 1, No 2B: 996 (1996); <https://doi.org/10.1143/JJAP.35.996>.
13. R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Ridout, E. Bassous, and A. R. LeBlanc, *IEEE J. Solid-State Circuits*, **51**: 256 (1974).
14. Ch. Hu, *Modern Semiconductor Devices for Integrated Circuits* (London: Pearson India: 2009).
15. Ch. Hu, *What Else Besides FinFET?*, www.synopsys.com/community/resources/events/keynote-finfet.html.
16. X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and Ch. Hu, *International Electron Devices Meeting. Technical Digest*, 67 (1999).
17. M. Jeong, B. Doris, J. Kedzierski, K. Rim, and M. Yang, *Science*, **306**: 2057 (2004).
18. A. Khakifirooz, O. M. Nayfeh, and D. A. Antoniadis, *IEEE Trans. Electron Dev.*, **56**: 1674 (2009).
19. E. O. Johnson, *RCA Review*, **34**: 80 (1973).
20. S. Datta, *Lessons from Nanoelectronics: A New Perspective on Transport* (Singapore: World Scientific: 2012).
21. M. Lundstrom and C. Jeong, *Near-Equilibrium Transport. Fundamentals and Applications* (Singapore: World Scientific: 2013).
22. Yu. O. Kruhliak and M. V. Strikha, *Ukr. Fiz. Zhurn. Ohlyady*, **10**: 3 (2015) (in Ukrainian); Ю. О. Кругляк, М. В. Стриха, *Укр. фіз. журн. Огляди*, **10**: 3 (2015).
23. Yu. A. Kruglyak, *Nanoehlektronika 'Snizu-Vverkh'* (Odessa: TES: 2015) (in Russian); Ю. А. Кругляк, *Нанoeлектроника «снизу-вверх»* (Одесса: ТЭС: 2015).
24. S. Datta, *Lessons from Nanoelectronics. Part A: Basic Concepts* (Singapore: World Scientific: 2017).
25. M. Lundstrom, *Fundamentals of Nanotransistors* (Singapore: World Scientific: 2018); www.nanohub.org/courses/NT.
26. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Yuan Taur, and H.-S. Philip Wong, *Proc. IEEE*, **89**: 259 (2001).
27. J. Appenzeller, Y.-M. Lin, J. Knoch, and Ph. Avouris, *Phys. Rev. Lett.*, **93**: 196805-1 (2004); <https://doi.org/10.1103/PhysRevLett.93.196805>.
28. S. Salahuddin and S. Datta, *Nano Lett.*, **8**: 405 (2008).
29. G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klassen, *IEEE Trans. Electron Dev.*, **53**: 1979 (2006).
30. G. T. Wright, *Electron Lett.*, **21**: 221 (1985).
31. A. L. Steegen, R. Mo, R. Mann et al., *International Electron Devices Meeting. Technical Digest*, 64 (2005).

Odesa State Environment University,
15 Lvivs'ka Str.,
65016 Odesa, Ukraine

¹ Fig. 1. A micrograph of the MOSFET cross section obtained with a scanning electron microscope (Texas Instruments, 1997), on which all contacts and a dielectric layer isolating the gate from the

conduction channel are clearly visible.

² **Fig. 2.** Illustrative examples, how the 2D electrostatics takes into account the characteristics of short-channel transistors. Left: *DIBL*, shifting the transfer characteristics $\log_{10} I_{DS} \propto V_{GS}$; this behaviour can also be interpreted as a decrease in the threshold voltage V_T with an increase in the drain voltage. Right: decrease V_T because of lowering the threshold voltage of nanotransistors.

³ **Fig. 3.** Illustration of the effect of voltage on the drain V_{DS} on the barrier height. Upper: the voltage V_{DS} does not affect the barrier height. Bottom: the voltage V_{DS} slightly lowered the barrier height (*DIBL* effect).

⁴ **Fig. 4.** Geometric screening in three types of MOSFET. Left: a massive MOSFET structure. To the right is the SG SG ETSOI single-gate. On the bottom right: DG ETSOI. In a high-quality MOSFET, the electric field lines from the drain penetrate only a distance $\approx \Lambda$ into the conduction channel, since most of the lines are short-circuited on the upper and lower gates [8].

⁵ **Fig. 5.** Capacitive model of the 2D electrostatics for the massive structure of MOS. Each capacitor is modelling the electrostatic coupling of the source, drain, gate and substrate electrodes to the top of the energy barrier.

⁶ **Fig. 6.** Simplified capacitor networks without voltage supply. Left: with charge Q_s on virtual source (*VS*). Right: with zero charge on *VS* and non-zero voltage on the gate.

⁷ **Fig. 7.** An illustration of the origin of the *DIBL* (left) and the increase in the subthreshold swing (centre), and the right side also shows the punch through effect, which we discuss below.

⁸ **Fig. 8.** Scaling of the transistor. The source transistor and its characteristics are shown on the left. On the right is a scaled transistor with a scaling factor $k > 1$. If the scaling is correct, then in scaled characteristics, all currents and voltages should change with the same scaling factor.

⁹ **Fig. 9.** Illustration of influence of the 2D effects on the output characteristics of MOSFET. On the left: a sufficiently long conduction channel with very high output impedance. Centre: short channel with low resistance. Right: the transistor is damaged by a punch-through.

¹⁰ **Fig. 10.** MOSFET cross-sections illustrating the depletion boundaries for surface punch-through (on the left) and in the bulk punch-through (right).

¹¹ **Fig. 11.** Illustration of a punch-through. The solid line corresponds to the ‘good’ state of the transistor; the dotted line corresponds to the punch-through. The grey zone shows the boundaries of the drainage depletion region.

¹² **Fig. 12.** An illustration, how an electrostatically well-balanced MOSFET works at high voltage, both on the drain and on the gate. As the dashed line shows, the increase in voltage on the drain increases the potential in the saturation region, without affecting the barrier at the beginning of the channel, which is under the control of the gate.

¹³ **Fig. 13.** The output characteristic of the *n*-MOSFET, connected in the common-source circuit, at some fixed value of the gate voltage (solid line). The dashed lines give a current in the linear region and in the saturation region according to (44).

¹⁴ **Fig. 14.** The charge density in the inversion layer as a function of the gate voltage (dashed curve on the right), and in the logarithmic scale (solid curve on the left).

¹⁵ **Fig. 15.** An illustration, how the subthreshold swing *SS* is related to the voltage of the power source V_{DD} . The current in the ‘ON’ mode corresponds to the maximum voltage of the power supply applied to the gate.

¹⁶ **Fig. 16.** Thermionic current in MOSFET.

¹⁷ **Fig. 17.** Dependence $\log_{10} I_{OFF} \propto I_{ON}$ plot for the 65 nm NMOS technology [31].