

СХЕМОТЕХНИКА И ПРОЕКТИРОВАНИЕ CIRCUIT ENGINEERING AND DESIGN

УДК 621.3

Применение табличных моделей туннельных эффектов для ускорения SPICE-моделирования нанометровых МОП-транзисторов

В.А. Бачманов¹, И.В. Заболотнов¹, А.В. Лапин^{1,2}

¹АО «Ангстрем-Т» (г. Москва)

²Национальный исследовательский университет «МИЭТ»

Implementation of Table Models of Tunneling Effects for Acceleration SPICE-Simulation of Nanometer MOSFET Transistors

V.A. Bachmanov¹, I.V. Zabolotnov¹, A.V. Lapin²

¹SC “Angstrom-T”, Moscow

²National Research University of Electronic Technology

Приведены результаты реализации методики ускорения SPICE-моделирования применительно к учету туннельных эффектов в нанометровых МОП-транзисторах. Показано, что замена сложных моделей этих эффектов табличными представлениями существенно ускоряет моделирование схем, содержащих указанные транзисторы.

Ключевые слова: SPICE-моделирование; BSIM3; эффекты туннелирования.

The results of the implementation of the accelerated SPICE-simulation method applied to the account of the tunneling effects in nanometer MOSFET have been presented. It has been shown that the replacement of complex models of these effects by tabular representation significantly accelerates simulation of the circuits containing these transistors.

Keywords: SPICE-simulation; BSIM3; tunneling affects.

Введение. Снижение проектно-технологических норм и рабочих напряжений КМОП СБИС порождает появление в их конструкции новых физических эффектов. Так, за последние 15–20 лет только в МОП-транзисторах обнаружено более 30 таких эффектов [1]. Из них особое внимание заслуживают явления туннелирования, поскольку заметно снижают такие преимущества КМОП-схем, как почти нулевое потребление в статическом режиме и отсутствие гальванической связи между входами и выходами элементов. Поэтому указанные эффекты подлежат полному учету в транзисторных моделях.

© В.А. Бачманов, И.В. Заболотнов, А.В. Лапин, 2015

Уточнение транзисторной модели эффектами туннелирования токов через подзатворный окисел и затворного индуцирования токов в областях стока, истока приводит к многократному замедлению SPICE-счета схем либо к его прерыванию. Степень замедления SPICE-счета определяется многими факторами и поэтому колеблется от 3 до 100 раз и более, так как зависит от характера функций для описания утечек генераторами тока, способов отсечения в этих функциях лишних ветвей и приемов блокировки сингулярностей. Степень замедления зависит и от класса моделируемых схем, а также от того, какой из модельных уровней, нижний или верхний, дополняется новыми эффектами.

Нижний уровень соответствует встроенным в симулятор моделям схемных компонентов и доступен только разработчикам модели и симулятора. Верхний уровень соответствует подсхемам и потому доступен пользователям симулятора. На нижнем уровне модифицируются устаревшие модели и внедряются новые, на верхнем выполняется, как правило, только модификация, т.е. составляется подсхема из устаревшей встроенной модели компонента и встроенных моделей других компонентов, необходимых для ее уточнения. Такой подход к оперативному уточнению встроенных моделей для схемных компонентов всех типов практикуется в кремниевом производстве уже более 10 лет. В случае транзисторов данный факт свидетельствует не столько о хроническом запаздывании работ на уровне встроенных моделей, сколько о чрезмерных трудностях экстракции параметров для таких новейших моделей, как BSIM4, BSIM5, PSP и HiSIM [2, 3].

Таким образом, выявление и устранение причин замедления SPICE-процесса моделями уровня подсхем приобретает повышенную практическую значимость, а решение этой задачи применительно к туннельным эффектам в нанометровых МОП-транзисторах – характер первоочередной необходимости.

Построение и исследование транзисторной модели подсхемного уровня. Наиболее применяемая встроенная модель МОП-транзистора, «свободная» от эффектов туннелирования и затворного индуцирования, – BSIM3v3. Дополним ее на уровне подсхемы (рис.1,а) генераторами тока, имитирующими эти эффекты. Рассмотрим подробнее эффекты туннелирования и затворного индуцирования [1, 2–4].

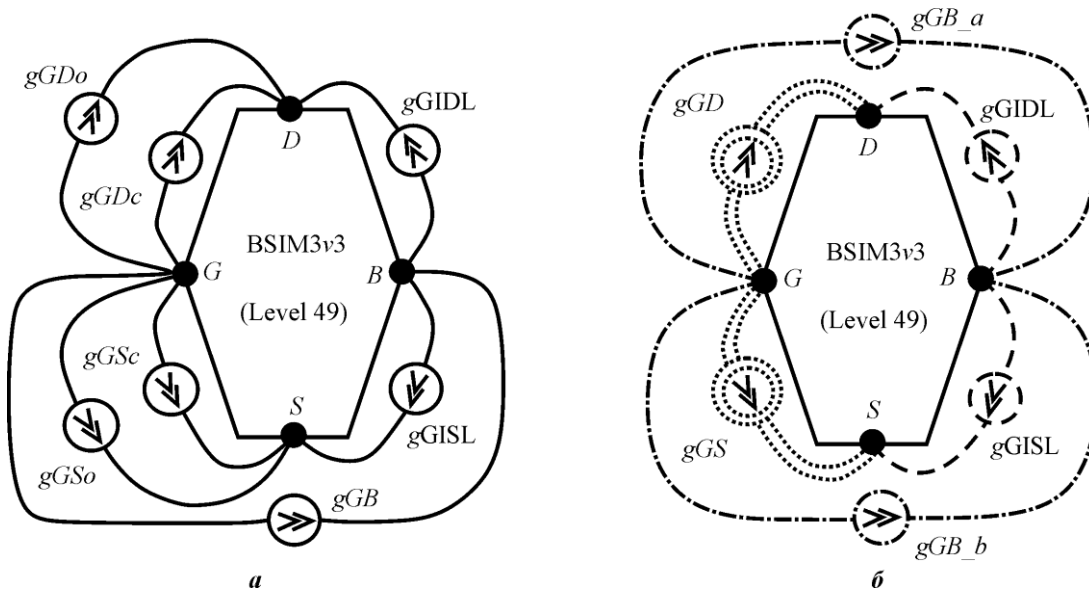


Рис.1. Преобразование аналитической модели (а) в таблично-аналитическую (б): — аналитический вид; только композиция табличных данных; - - - - декомпозиция и сжатие; --- только сжатие

Эффект туннелирования есть следствие снижения толщины подзатворного окисла до 5 нм. Ниже этого значения вероятность туннелирования зарядов через диэлектрическую пленку резко возрастает [2]. Реализуется так называемое прямое туннелирование в транзисторе, когда между затвором G и стоком D , истоком S и подложкой B текут токи. Их величина зависит от переменных: TOX , W_e , L_e , M , V_{G-S} , V_{G-D} и V_{G-B} , где TOX – толщина окисла под затвором; W_e , L_e – эффективные ширина и длина транзисторного канала соответственно; M – число параллельно включенных транзисторов; $V_{G-S} = V_G - V_S$, $V_{G-D} = V_G - V_D$ и $V_{G-B} = V_G - V_B$ – перепады напряжения между электродами. Соответствующие генераторы тока будем обозначать gGS , gGD и gGB .

Затворное индуцирование возможно по двум причинам: повышение уровня легирования областей стока, истока и снижение толщины подзатворного окисла. В сущности, здесь активизируется иной тип туннелирования. Так, при электрическом поле, обедняющем сильно легированный сток закрытого n -транзистора, область обеднения оказывается настолько узкой, что становится возможным переход электронов из валентной зоны в зону проводимости стока (Band-To-Band Tunneling, BTBT). Эта генерация электронно-дырочных пар и питает ток утечки между стоком и p -подложкой. Аналогичный процесс происходит в области истока. Они конкурируют, ослабляя напряженность обедняющего поля в направлениях $G-D$ и $G-S$ перепадами напряжения V_{G-S} и V_{G-D} соответственно. Усиливаются указанные компоненты поля соответственно разностями V_{D-S} и V_{S-D} . Процессы называются GIDL (Gate Induced Drain Leakage) и GISL (Gate Induced Source Leakage), а генераторы тока для имитации этих процессов – $gGIDL$ и $gGISL$.

Данный подход, во-первых, позволяет явления туннелирования и индуцирования кратко именовать эффектами туннелирования в МОП-транзисторах, а соответствующие токи утечки – туннельными, во-вторых, локализует узлы подключения одноименных генераторов тока в подсхеме (см. рис.1,а), в-третьих, определяет управляющие перепады напряжения для генераторов $gGIDL$ и $gGISL$.

Каждый из генераторов gGS и gGD обычно представляется двумя компонентами: gGS_c , gGS_o и gGD_c , gGD_o (рис.1,а). Первая из них (с индексом «с») зависит от площади канала, а вторая – только от его ширины. Для n -транзистора эти компоненты задаются формулами:

$$gGS_c = 0,5W_eL_eM[0,5\{\text{sign}(V_{G-S}) + 1\}A1C \cdot V_{G-S} \times \exp\{-B1C|V_{G-S}|(1 + K1C|V_{G-S}| + P1C(V_{G-S})^2)\} - \quad (1)$$

$$- 0,5\{\text{sign}(V_{G-S}) - 1\}A2C \cdot V_{G-S} \cdot \exp\{-B2C|V_{G-S}|(1 + K2C|V_{G-S}| + P2C(V_{G-S})^2)\},$$

$$gGS_o = 0,5W_eM[0,5\{\text{sign}(V_{G-S}) + 1\}A1O \cdot V_{G-S} \times \exp\{-B1O|V_{G-S}|(1 + K1O|V_{G-S}| + P1O(V_{G-S})^2)\} - \quad (2)$$

$$- 0,5\{\text{sign}(V_{G-S}) - 1\}A2O \cdot V_{G-S} \exp\{-B2O|V_{G-S}|(1 + K2O|V_{G-S}| + P2O(V_{G-S})^2)\}].$$

Здесь значение TOX учтено в коэффициентах $A1C$, $A2C$ и $A1O$, $A2O$, которые, как и коэффициенты $B1C$, $B2C$, $B1O$, $B2O$, $K1C$, $K2C$, $K1O$, $K2O$, $P1C$, $P2C$, $P1O$, $P2O$, получены путем экстракции. Коэффициенты зависят от вариаций технологического процесса (Fast, Typical, Slow). В случаях симметричной конструкции транзистора выражения для gGD_c и gGD_o получают соответственно из выражений (1) и (2) заменой в них V_{G-S} на V_{G-D} . Для транзистора p -типа перед правыми частями соотношений (1), (2) стоит знак минус, а размеры канала, множитель M и коэффициенты имеют другие значения.

В отличие от выражений (1), (2) ток генератора gGB есть функция двух переменных V_{G-B} и V_{D-S} , которые в данном случае можно представить в виде пары сложных аргументов – кластеров напряжений (КН):

$$cla = V_{G-B} + \frac{V_{D-S}}{D1B}, \quad (3)$$

$$clb = V_{G-B} + \frac{V_{D-S}}{D2B}, \quad (4)$$

где $D1B, D2B$ – коэффициенты.

С учетом введенных обозначений формула для gGB имеет вид

$$gGB = 0,5W_e L_e M [0,5\{\text{sign}(V_{G-B}) + 1\}A1B \cdot cla \times \\ \times \exp\{-B1B|cla|(1 + K1B|cla| + P1B \cdot cla^2)\} - \\ - 0,5\{\text{sign}(V_{G-B}) - 1\}A2B \cdot clb \exp\{-B2B|clb|(1 + K2B|clb| + P2B \cdot clb^2)\}]. \quad (5)$$

Здесь толщина окисла учтена в значениях коэффициентов $A1B, A2B$, а переменная V_{G-B} в операторе «sign» имеет вспомогательное назначение, определяя нужный знак для переключателей типа

$$0,5\{\text{sign}(X) \pm 1\}, \quad (6)$$

отсекающих лишние ветви функции $gGB(cla, clb)$ (см. выражения (1), (2)). В случае p -транзистора формулы (3)–(5) имеют такой же вид, отличаясь лишь значениями W_e, L_e, M и коэффициентов.

Ток генератора $gGISL$ для n -транзистора также зависит от двух переменных V_{S-D} и V_{G-D} , но они охватываются одним КН вида

$$cls = V_{S-D} - GGI \cdot V_{G-D} - EGI, \quad (7)$$

где GGI, EGI – коэффициенты.

При этом сам ток описывается выражением

$$gGISL = 0,5\{\text{sign}(cls) + 1\}AGI \cdot W_e M \left[\frac{cls}{3TOX} \right] \exp\left\{-3TOX \frac{BGI}{cls}\right\}, \quad (8)$$

где AGI и BGI – коэффициенты. Толщина окисла задается в явном виде, и применяется один из переключателей типа (6) при $X = cls$.

В случаях симметричной конструкции транзистора выражение для $gGIDL$ получается из формулы (8) только заменой cls на кластер

$$cld = V_{D-S} - GGI \cdot V_{G-S} - EGI. \quad (9)$$

Для p -транзистора перед правой частью выражения (8) ставится минус, меняются значения коэффициентов и W_e, M , а КН в (7), (9) заменяются на кластеры:

$$cls = -V_{S-D} + GGI \cdot V_{G-D} - EGI, \quad (10)$$

$$cld = -V_{D-S} + GGI \cdot V_{G-D} - EGI. \quad (11)$$

Непосредственное применение формул (1)–(11) в соответствии с под схемой на рис.1,а приводит к многократному замедлению SPICE-моделирования схем. Например, при норме технологического процесса 120 нм, напряжении питания 1,2 В, $T = 25^\circ\text{C}$ и

коэффициентах в выражениях (1), (2), (5), (8), экстрагированных из данных измерения технологического процесса Turical ($TOX = 2,4$ нм), расчет цифровых схем замедляется в 12 раз, схем памяти в 19 раз, а генераторных кольцевых схем – в 108 раз (симулятор Hspice ver. A-2008.03-SP1). Причина такого значительного замедления не столько в возросшем за счет формул (1)–(11) количестве операций, сколько в их способности кратковременно прерывать сходимость вычислений к верному решению. Это проявляется на кривой динамического тока потребления генераторных схем (рис.2,а, интервал T_g –12), где в отдельные моменты модельного времени наблюдаются нереальные по амплитуде всплески. Для сравнения на рис.2,б показана реальная кривая потребления.



Рис.2. Графики тока потребления: а – неверный результат; б – верный (T_g – момент включения генерации)

В SPICE-симуляторе подобные ситуации распознаются по комплексу косвенных признаков, после чего модельное время прерывается для активизации специальных процедур по восстановлению сходимости. Их выполнение и занимает основное время счета.

Функции (1)–(11) способны оказывать наибольшее влияние на сходимость вычислений через элементы матрицы Якоби, содержащие члены вида

$$\frac{\partial g(V)}{\partial V}, \quad (12)$$

где $g(V)$ – генератор тока, управляемый напряжением.

Следовательно, источниками прерывания сходимости являются те функции в выражениях (1)–(11), значение которых меняется наиболее резко с изменением V . К таким относятся в первую очередь переключатели (6), входящие в состав всех выражений (1), (2), (5), (8), а также показатель экспоненты в формуле (8), содержащий сингулярность.

Логично предположить, что источник нереальных функциональных всплесков может находиться и в самом симуляторе. Например, им может быть не вполне корректный алгоритм дифференцирования выражений, задаваемых аналитически. Поэтому задача ускорения SPICE-счета в рассматриваемом здесь аспекте получает, по меньшей мере, два способа решения:

- 1) поиск аналитических форм, тождественных функциям (1), (2), (5), (8) и находящихся в области корректных вычислений указанного алгоритма дифференцирования;
- 2) табулирование функций (1), (2), (5), (8) с целью использования в симуляторе другого метода дифференцирования, предназначенного для работы с аппроксимируемыми выражениями.

В первом способе применяются известные тождества

$$0,5\{\text{sign}(X) \pm 1\}X \equiv 0,5\{X | \pm X\} \quad (13)$$

для генераторов (1), (2), (8), где место аргумента X занимают соответственно переменные V_{G-S} и cls (для генератора (5) найти тождественную замену не удалось). Время счета сократилось вдвое, а максимальная амплитуда всплесков тока потребления снизилась на два порядка. Затем генератор (8) был дополнен ограничением на величину показателя экспоненты. Этим принцип тождественности нарушается лишь формально, так как назначение экспоненциального множителя – обнулять (8) при активизации сингулярности в его показателе – реализуется раньше действия ограничения. В итоге время счета сократилось в 2,5 раза, а максимальная амплитуда токовых всплесков уменьшилась почти на три порядка. Большого эффекта в рамках первого способа добиться не удалось. Подтвердилось предположение о некорректности SPICE-дифференцирования функций, задаваемых аналитически.

Реализация табличного способа. Полученный результат одновременно является обоснованием табличного способа ускорения SPICE-счета. Кроме того, он объясняет известный факт, в соответствии с которым активизация туннельных эффектов на нижнем модельном уровне (например, в последних версиях модели BSIM4) замедляет SPICE-счет всего в 3 раза. Действительно, на нижнем уровне имеется возможность отказаться от рассмотренного алгоритма дифференцирования, используя вместо него предварительно полученные соотношения вида (12). В этом контексте табличный способ ускорения SPICE-счета приближается по эффективности к нижнему модельному уровню, поскольку соотношения для производных от аппроксимирующих выражений тоже применяются в симуляторе в виде готовых формул. При этом функции (1), (2), (5), (8) используются в табличном виде, а это значит, что содержащиеся в них операции заменяются на относительно нетрудоемкие операции: считывание табличных данных – аппроксимация – интерполяция.

К другим особенностям табличного способа относятся:

- отсутствие проблем с отсечением лишних ветвей в генераторах (1), (2), (5), (8) и блокировкой сингулярности в генераторе (8);
- композиция и декомпозиция табличных данных на уровне ветвей и генераторов;
- сжатие табличных данных заменой двумерного табулирования генераторов (5), (8) одномерным;
- управление уровнем детализации эффектов (1), (2), (5), (8) посредством задания постоянных или переменных шагов (табулирования) подходящей величины;
- параметризация табличных данных значениями W_e , L_e , M и температуры T (SPICE-опции *scale*, *m* и *tc1*, *tc2* соответственно);
- управление нелинейностью аппроксимации (SPICE-опция *delta*);
- сглаживание табличных данных (SPICE-опция *smooth*).

Реализация табличного способа ускорения SPICE-счета снизила показатели замедления с 12 до 2, с 19 до 2,3 и со 108 до 6, т.е. по эффективности данный способ действительно близок к нижнему модельному уровню.

Ветви генераторов, лишние в каком-либо поддиапазоне изменения аргумента, отсекаются просто тем, что табулируются только нужные рабочие ветви. Результаты заносятся в итоговую таблицу. Например, первая ветвь (первое слагаемое) в (1) является лишней при $V_{G-S} < 0$, а вторая – при $V_{G-S} > 0$ (рис.3,а). В точке $V_{G-S} = 0$ обе ветви проходят через 0 из-за множителя V_{G-S} в их выражениях. Поэтому и график результирующей зависимости проходит через 0 (см. рис.3,а). Аналогично отсекаются лишние ветви и в генераторе (2) (рис.3,б). В обоих примерах подразумевается, что переключатели типа (6) предварительно удалены из (1), (2). В этом случае табулирование выполняет еще и композицию рабочих ветвей, т.е. заменяет эти переключатели.

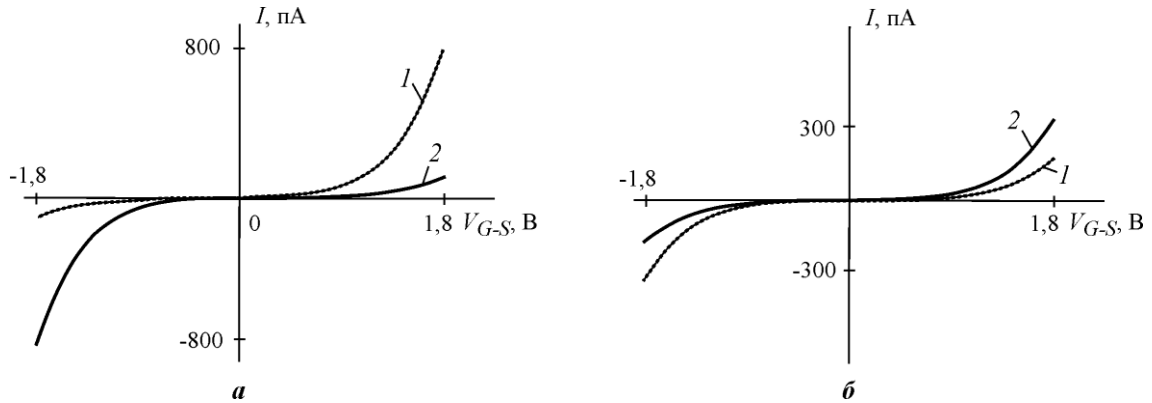


Рис.3. Табулирование и композиция рабочих ветвей для генераторов gGS_c (а) и gGS_o (б):
1 – рабочие ветви (точки – табличные значения); 2 – лишние ветви

Генераторы gGS_c и gGS_o имеют общие узлы подключения и управления (см. (1), (2)) и нижнюю часть рис.1,а. Данное обстоятельство позволяет объединить их в один генератор gGS для снижения накладных затрат при построении схемной модели в симуляторе. На уровне таблиц это можно сделать по формуле

$$gGS = M \cdot WLREL \cdot [gGS_c] + M \cdot WREL \cdot [gGS_o] = M \cdot WREL \cdot (LREL \cdot [gGS_c] + [gGS_o]), \quad (14)$$

где в квадратных скобках – таблицы, свободные от масштабирующих множителей M , $WLREL$ и $WREL$.

Преобразование (14) корректно, когда gGS_c и gGS_o имеют одинаковые значения параметров $tc1$, $tc2$, $delta$, $smooth$.

Генератор gGD можно получить из gGS заменой узла S на D в цепях подключения и управления. Аналогичные действия выполняются и для p -транзистора. Такая композиция табличных данных (рис.1,б) снижает время SPICE-счета на 10–20 %.

Вольт-амперной характеристикой генератора gGB , как следует из (3)–(5), является семейство зависимостей типа $f(V_{G-B}, V_{D-S})$ при параметре V_{D-S} либо V_{G-B} . График этого семейства для транзистора n -типа, где роль параметра отведена, например, переменной V_{D-S} , показан на рис.4.

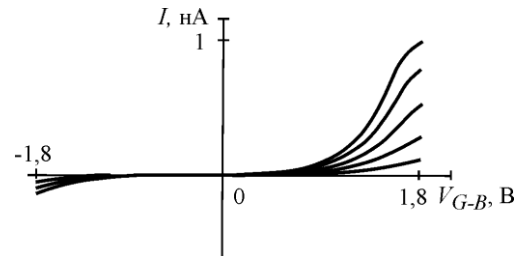


Рис.4. Вольт-амперная характеристика генератора gGB

Табулирование генератора gGB невозможно без декомпозиции его на две компоненты gGB_a и gGB_b в соответствии с кластерами (3), (4), поскольку в SPICE-симуляторах не предусмотрено задание двумерных таблиц. При этом каждая из компонент

$$gGB_a = 0,5W_e L_e \cdot M \cdot A1B \cdot cla \times \exp \left\{ -B2B|cla|(1 + K1B|cla| + P1B \cdot cla^2) \right\}, \quad V_{G-B} > 0, \quad (15)$$

$$gGB_b = 0,5W_e L_e \cdot M \cdot A2B \cdot clb \times \exp \left\{ -B2B|clb|(1 + K2B|clb| + P2B \cdot clb^2) \right\}, \quad V_{G-B} < 0, \quad (16)$$

будет функцией, производящей семейство типа $f(V_{G-B}, V_{D-S})$ при параметре V_{D-S} либо V_{G-B} .

Вариация в (3), (4) каждой из переменных V_{D-S} , V_{G-B} в пределах от $-1,8$ до $+1,8$ В дает диапазоны $-2,6 < cla < 2,6$ В и $-2,1 < clb < 2,1$ В. Развертка в них значений cla , clb с каким-нибудь шагом (постоянным или переменным) переводит производящие функции (15), (16) в одномерные таблицы $gGB_a(cla)$ и $gGB_b(clb)$. Пример таких таблиц в графической форме представлен на рис.5,а,б (значения cla , clb задаются непосредственно). Если для них организовать развертку переменной V_{G-B} посредством внутреннего цикла, а переменной V_{D-S} посредством внешнего цикла и на каждом шаге получать значения cla , clb по формулам (3), (4), то в системе координат V_{G-B} , I на выходе 1-мерных таблиц будут формироваться семейства, указанные на рис.5,в,г (значения cla , clb вычисляются по формулам (3), (4)). Другими словами, таблицы рис.5,а,б содержат в сжатом виде информацию, представленную на рис.5,в,г.

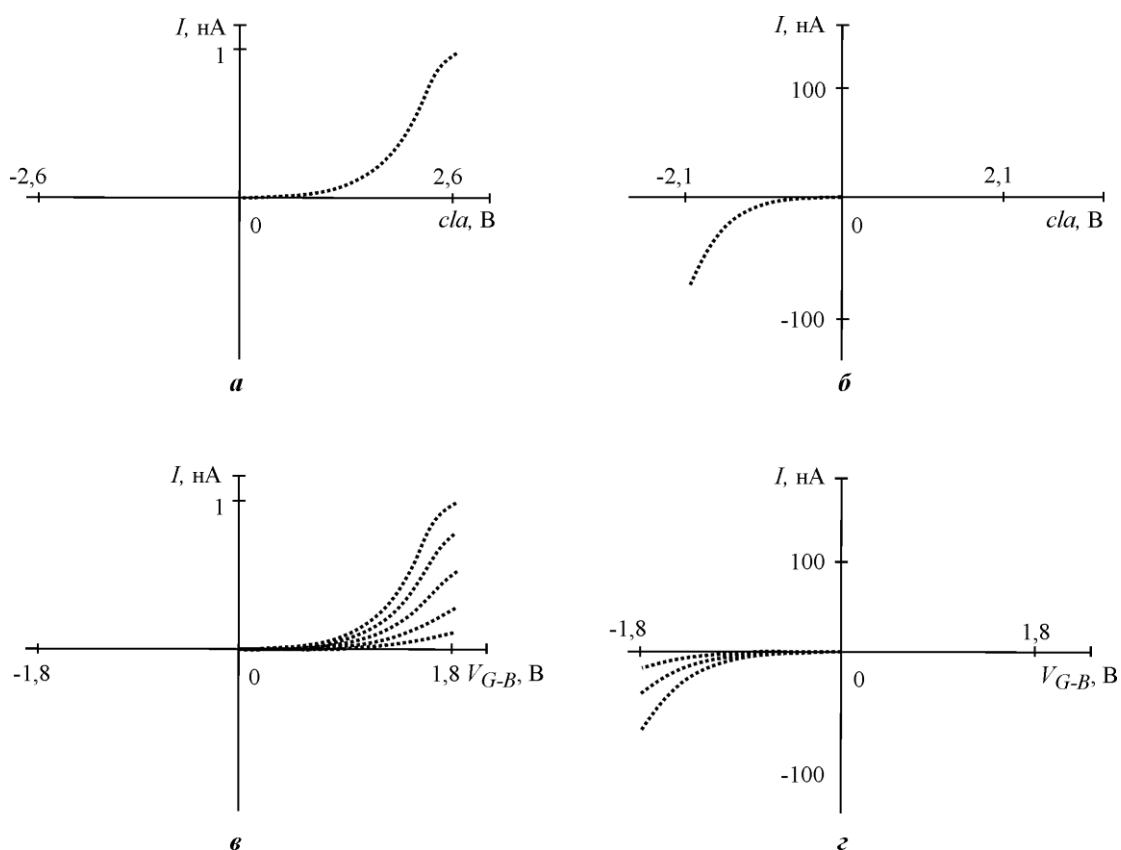


Рис.5. Получение (а, б) и применение (в, г) 1-мерных таблиц $gGB_a(cla)$ (а, в) и $gGB_b(clb)$ (б, г)

Композиция семейств на рис.5,в,г выполняется во входном файле симулятора параллельным включением табличных генераторов $gGB_a(cla)$, $gGB_b(clb)$ и введением пары источников типа E , управляемых по формулам (3), (4). Получаемые в результате значения соответствуют кривым рис. 4 с высокой точностью. Для транзистора p -типа рассмотренные преобразования генератора gGB выполняются аналогично.

Ток генератора $gGISL$ тоже зависит от двух переменных, но они охватываются одним КН (см. (8), (7)). Поэтому в данном случае выполняется только операция сжатия информации, т.е. получение одномерной таблицы $gGISL(cis)$ с последующим введением ее во входной файл симулятора вместе с источником типа E , управляемым по фор-

муле (7). Такие же действия выполняются и для генератора *gGIDL*, а все вместе – и для *p*-транзистора. На этом завершаются все преобразования транзисторной модели, заявленные на рис. 1.

Заключение. Аналитическое задание на входе SPICE-симулятора сложных функций вида (1), (2), (5), (8) нарушает в нем сходимость вычислений, из-за чего время счета схем возрастает до неприемлемых значений. От этого недостатка свободны табличные эквиваленты тех же функций. Данное свойство позволяет оперативно дополнять новыми эффектами транзисторные модели третьего поколения путем табличного задания функций на верхнем модельном уровне, т.е. на уровне подсхем.

Литература

1. **Денисенко В.В.** Новые физические эффекты в нанометровых МОП-транзисторах // Компоненты и технологии. – 2009. – №12. – С. 157–162.
2. **Prabhakar V.S.V., Lal Kishore K.** Device Modeling Solutions to Reduce GIDL Current in Low Power VLSI Circuits // International J. of Electrical, Electronics and Data Communication. – 2013. – Vol. 1. – Iss. 9. – P. 21.
3. **Денисенко В.В.** Компактные модели МОП-транзисторов для SPICE в микро- и наноэлектронике. – М.: Физматлит, 2010. – 408 с.
4. Model definition of MOS model 11, level 1102. NXP Semiconductors, July 2011.

Статья поступила
после доработки 19 августа 2015 г.

Бачманов Владимир Александрович – руководитель дизайн-группы АО «Ангстрем-Т» (г. Москва). *Область научных интересов:* машинный анализ электронных схем.

Заболотнов Игорь Вячеславович – заместитель генерального директора АО «Ангстрем-Т» (г. Москва). *Область научных интересов:* технологическая среда проектирования систем-на-кристалле.

Лапин Александр Владимирович – аспирант кафедры проектирования и конструирования интегральных микросхем МИЭТ, инженер-программист дизайн-группы АО «Ангстрем-Т» (г. Москва). *Область научных интересов:* программирование, алгоритмы в САПР на этапе логического проектирования. **E-mail:** lapin@angstrem-t.com

Вниманию читателей журнала «Известия высших учебных заведений. Электроника»

Оформить годовую подписку на электронную
копию журнала можно на сайтах

- Научной электронной библиотеки: www.elibrary.ru
- Национального цифрового ресурса «Рукопт»: www.rucont.ru
- Урал-Пресс: www.ural-press.ru