Accurate Estimation of Total Leakage in Nanometer-Scale Bulk CMOS Circuits Based on Device Geometry and Doping Profile

Saibal Mukhopadhyay, *Student Member, IEEE*, Arijit Raychowdhury, *Student Member, IEEE*, and Kaushik Roy, *Fellow, IEEE*

***Abstract*—Dramatic increase of subthreshold, gate and reverse biased junction band-to-band-tunneling (BTBT) leakage in scaled devices results in the drastic increase of total leakage power in a logic circuit. In this paper, a methodology for accurate estimation of the total leakage in a logic circuit based on the compact modeling of the different leakage current in nanoscaled bulk CMOS devices has been developed. Current models have been developed based on the device geometry, two-dimensional doping profile, and operating temperature. A circuit-level model of junction BTBT leakage has been developed. Simple models of the subthreshold current and the gate current have been presented. Also, the impact of quantum mechanical behavior of substrate electrons, on the circuit leakage has been analyzed. Using the compact current model, a transistor has been modeled as a sum of current sources (SCS). The SCS transistor model has been used to estimate the total leakage in simple logic gates and complex logic circuits (designed with transistors of 25-nm effective length) at room and elevated temperatures.**

***Index Terms*—Band-to-band-tunneling (BTBT), doping profile, estimation, gate direct tunneling, halo doping, subthreshold leakage, technology scaling.**

I. Introduction

AGGRESSIVE scaling of CMOS devices in each technology generation has resulted in higher integration density and performance. Simultaneously, supply-voltage scaling has reduced the switching energy per device. However, the leakage current (i.e., the current flowing through the device in its "off" state) has increased significantly with technology scaling [1], [2]. Hence, the estimation of the total leakage is absolutely necessary for designing low-power logic circuits.

Among different leakage mechanisms in scaled devices [1], three major ones can be identified as: subthreshold leakage, gate leakage, and reverse biased drain-substrate and source-substrate junction band-to-band-tunneling (BTBT) leakage [1]. With technology scaling, each of the different leakage components increases, which has two major implications in the leakage estimation and low-power logic design. First, this results in a large increase of the total leakage. Moreover, each of the leakage components becomes equally important in nanoscaled devices. Hence, the relative magnitudes of the leakage components play a major role in low-leakage logic design. For example, it has been shown in [3] that the effectiveness of a standard leakage-reduction technique, known as "transistor stacking," strongly depends on the contribution of the gate and the subthreshold leakage to the total leakage. Fig. 1 shows the different leakage components of NMOS devices of $L_{\rm eff} = 25$ nm, $L_{\rm eff} = 50$ nm, and $L_{\rm eff} = 90$ nm taken from [4]. It can be observed that for the 90-nm device, the major leakage component is the subthreshold leakage, but in the scaled devices, contributions of the junction leakage and the gate leakage have significantly increased. Moreover, the magnitudes of each of these components strongly depend on the device geometry (namely, channel length, oxide thickness, and transistor width), the doping profiles and temperature [1], [5]. Hence, an accurate estimation of leakage in nanoscaled logic circuits should be able to: 1) estimate the gate, the subthreshold, and the junction tunneling leakage separately along with the total leakage; 2) estimate the effect of variation in doping profile, transistor geometry, and temperature on individual leakage component; and 3) account for all the different physical mechanisms that can modify the leakage components to prevent both underestimation and overestimation. In this paper we have developed a methodology for estimation of the gate, the subthreshold, the junction BTBT and the total leakage of a logic circuit for different primary input vectors, based on the knowledge of: 1) the device geometry; 2) the two-dimensional (2-D) doping profile of the device; and 3) the operating temperature. Although, a number of previous work are reported on the estimation of leakage in logic circuits [6], [7] they have

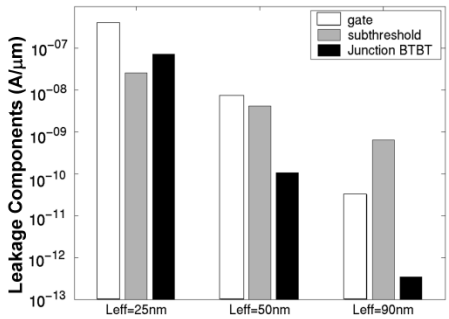


Fig. 1. Contribution of different leakage components in NMOS devices [4] at different technology generation. The leakage values are extracted using device simulation in MEDICI. $V_{\rm DD}$ values are chosen following ITRS guideline (0.7 V at 25 nm, 0.9 V at 50 nm, and 1.2 V at 90 nm).

Manuscript received October 9, 2003; revised April 7, 2004. This work was supported in part by the Semiconductor Research Corporation, in part by the Gigascale Silicon Research Center, in part by the Intel Corporation, and in part by the IBM Corporation. An earlier version of this paper was published in the Proceedings of the Design Automation Conference, 2003, pp. 169–174, Anaheim, CA. This paper was recommended by Associate Editor S. Saxena.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: sm@ecn.purdue.edu; araycho@ecn.purdue.edu; kaushik@ecn.purdue.edu)

Digital Object Identifier 10.1109/TCAD.2004.842810

only considered the subthreshold leakage. However, as shown in Fig. 1, the gate and the junction BTBT leakage are also becoming extremely important and thus cannot be neglected for estimation of total leakage (in some recent work [8] logic-level estimation technique for gate leakage is reported). Moreover, all of the previous estimation methodologies start with either the precharacterization of the basic logic gates or use device models which are extremely empirical (although accurate) in nature. Hence, these methods do not allow us to estimate the effect of variations in device level parameters, such as transistor geometry or doping profile, on the total leakage in a logic circuit. Any change of these parameters require either a new set of precharacterization or a reevaluation of a large number of empirical parameters and thus increasing the total computation time by a significant amount. In nanoscaled devices such a study is extremely useful for designing low-leakage logic circuits. In particular, through this work we tried to make the following contributions to the existing art.

1) We have developed a physics-based, compact circuit level model of junction BTBT leakage in a MOSFET with "halo" [5] and "retrograde" doping [5]. We have used a Gaussian approximation of the doping profile [4], which more accurately describes the actual 2-D doping profile. To the best of our knowledge, there is no existing circuit-level model for the junction BTBT leakage for bulk-MOSFET considering the 2-D doping profile. This model can be effectively used to study the impact of the junction BTBT leakage in CMOS circuits.
2) A simple and reasonably accurate model of the subthreshold current has been presented based on the Gaussian approximation of the 2-D doping profile. The developed subthreshold current model accounts for the quantization of the electron energy in the substrate [5] on the threshold voltage and hence its impacts on the leakage in logic circuits. We have shown that the quantum mechanical effects results in a substantial reduction in the subthreshold leakage. We have used the gate leakage model presented in [9] and [10].
3) The compact models of the leakage components have been used to model a transistor as a sum of current sources (SCS) for accurate leakage estimation. A numerical solver has been developed to evaluate leakage in simple logic gates by solving the Kirchoff's Current Law (KCL) at intermediate nodes, using SCS transistor model. We have used an existing algorithm given in [7] for the calculation of the total leakage of a logic circuit by adding the individual leakage contribution of its constituent gates. We have verified the leakage estimation technique on simple logic gates, such as INVERTER, NAND, and NOR gate, and on complex logic circuits, such as an adder and multiplier.

The methodology of the proposed estimation can take input from the device level (geometry, doping profile, and temperature), circuit level (transistor level description of different logic gates), and logic level (logic-level description of the circuit, algorithm by which leakage of logic circuit is calculated from the leakage value of logic cells). This allows a device-circuit-logic coestimation of leakage which can lead to a device-circuit-logic co-optimization.

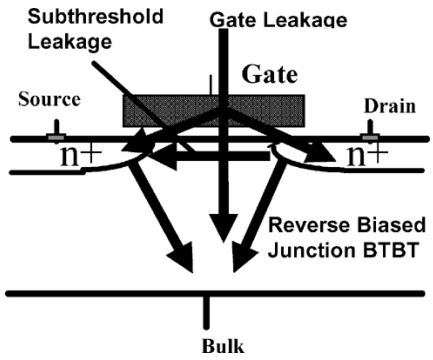


Fig. 2. Major leakage components in a transistor.

II. Leakage Current Components and Their Impact in the Logic Circuit

In nanometer-scaled devices, the major leakage components are (see Fig. 2): 1) subthreshold leakage; 2) gate-oxide tunneling leakage; and 3) reverse-bias drain-substrate and source-substrate junction BTBT leakage. There are other leakage components, like gate-induced-drain-leakage (GIDL), punchthrough current, etc., but those are not very serious at normal mode of operations. GIDL will be of concern in cases where $V_{\rm GD} < 0$ and pass-gate logic is definitely a part of it. However, for the range of $V_{\rm DD}$ suggested by ITRS, we observe that $V_{\rm GD} = -V_{\rm DD}$ does not result in any significant GIDL. Each of these three leakage components is increasing significantly with technology scaling. In this section, we will briefly discuss the effect of technology scaling on each of these leakage components; their strong dependence on device geometry, doping profile, and temperature; the correlation of these components in the device and circuit level. With technology scaling, the supply voltage needs to be scaled down to reduce the dynamic power and maintain reliability. However, this requires the scaling of the device threshold voltage ($V_{\rm th}$) to maintain a reasonable gate over drive [1]. The threshold voltage ($V_{\rm th}$) scaling and the $V_{\rm th}$ reduction due to short channel effects (SCEs) [like drain-induced-barrier-lowering (DIBL), $V_{\rm th}$ roll off] [1], [2], [5], result in an exponential increase in the subthreshold current. To control the SCE and to increase the transistor drive strength, oxide thickness needs to be scaled down in each technology generations. The aggressive scaling of oxide thickness results in a high direct tunneling current through the gate insulator of the transistor [1], [5]. On the other hand, scaled devices require the use of the higher substrate doping density and the application of the "halo" profiles (implant of the high doping region near the source and drain junctions of the channel) to reduce the depletion region width of the source-substrate and drain-substrate junctions [1], [5]. A lower depletion-region width helps to control the SCE. The high-doping density near the source-substrate and drain-substrate junctions cause significantly large BTBT current through these junctions under high reversed bias [1], [5]. From the above discussion we can conclude that each of the leakage components is increasing with technology scaling.

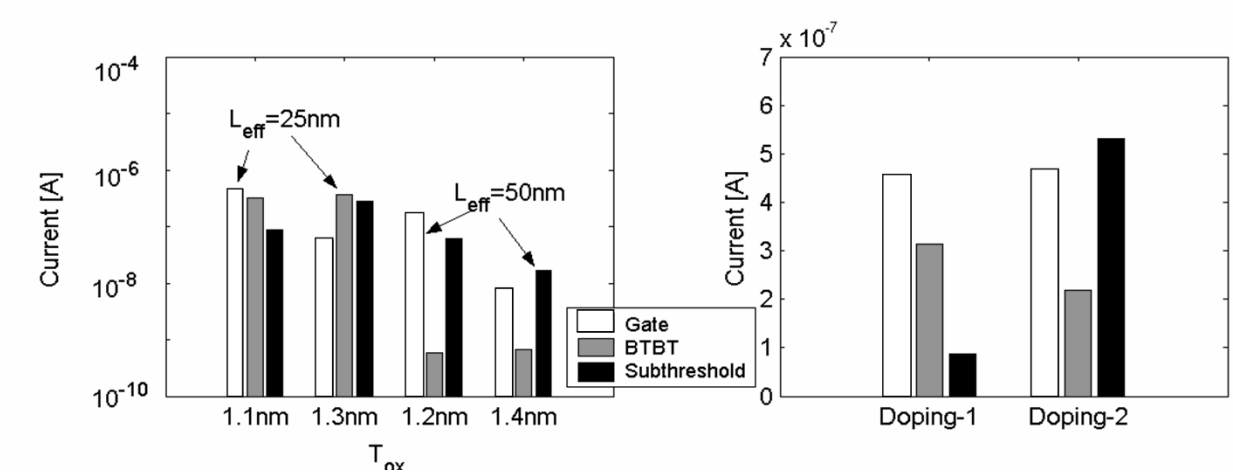


Fig. 3. Variation of different leakage components with (a) technology generation and oxide thickness; and (b) doping profile. "Doping-1" has a stronger halo profile than "Doping-2". The leakage values are extracted using device simulation in MEDICI. ($V_{\rm DD}$ is chosen as: 0.7 V at 25 nm, 0.9 V at 50 nm).

Fig. 3 shows the different leakage components of NMOS devices of 25 and 50-nm physical gate length [4] at different oxide thickness based on the result of MEDICI [11] device simulation. $V_{\rm DD} = 0.7$ V and $V_{\rm DD} = 0.9$ V were used in the simulations for 25 and 50 nm devices, respectively. Also, only the oxide thickness was varied in the simulations for a particular technology node (keeping doping constant). The gate leakage and the subthreshold leakage are strongly correlated through oxide thickness. A high oxide thickness results in low gate leakage. Although according to the long-channel MOSFET theory, higher oxide thickness helps to increase the threshold voltage, it will worsen the SCE [5]. If the SCE is not very high [e.g., in the 50-nm device in Fig. 3(a)] increasing $T_{\rm ox}$ may reduce the subthreshold leakage. However, in a nanoscale device where SCE is extremely severe (e.g., in the 25-nm device in the present case), an increase in the oxide thickness will increase the subthreshold leakage [Fig. 3(a)]. Similarly, the subthreshold leakage and the junction BTBT are strongly coupled through the doping profile. Fig. 3(b) shows the different leakage components of a 25-nm device at different doping profile (oxide thickness and $V_{\rm DD}$ were kept constant). A strong "halo" doping reduces the subthreshold current but results in a high BTBT. Reduction of the halo-strength lowers the BTBT, but increases subthreshold current considerably [Fig. 3(b)]. From the above discussion, we can conclude that magnitude of the leakage components and their relative dominance on each other depends strongly on device geometry and doping profile.

The basic physical mechanisms governing the different leakage-current components have different temperature dependence. Subthreshold current is governed by the carrier diffusion that increases with an increase of temperature. Since tunneling probability of an electron through a potential barrier does not depend directly on temperature, the gate and the junction BTBT is expected to be less sensitive to temperature variations. However, increase of temperature reduces the band-gap of silicon [12], which is the barrier height for tunneling in BTBT. Hence, the junction BTBT is expected to increase with an increase in temperature. Fig. 4 shows the effect of temperature variation on individual leakage component of the previously mentioned 25-nm NMOS device based on the device simulation. From

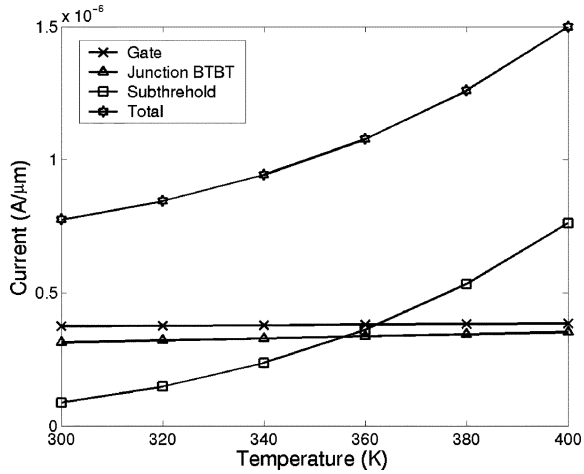


Fig. 4. Simulation result for variation of different leakage components with temperature (without considering the impact of quantum correction) for NMOS device of $L_{\rm eff} = 25$ nm [4]. The leakage values are extracted using device simulation in MEDICI. $V_{\rm DD}$ of 0.7 V is used at 25 nm.

Fig. 4, it is observed that the subthreshold leakage increases exponentially with temperature, the junction BTBT increases slowly with temperature and the gate leakage is almost independent of temperature variation. Fig. 4 shows that for that particular NMOS device, at $T = 300$ K (a possible temperature in the stand-by mode) the gate leakage is the dominant leakage component. However, the subthreshold and the BTBT become dominant at $T = 400$ K (a possible temperature in the active mode). Hence, we can conclude that the individual leakage component and the total leakage depends strongly on temperature (or mode of operation).

It is evident from previous discussions that in nanoscaled devices, all of the different leakage components become important and their magnitude depends strongly on the device structure, the doping profile, and the temperature. However, along with the absolute magnitude, the relative contributions of the individual leakage component to the total leakage also have a strong impact on the leakage of a logic gate and logic circuit. Fig. 5 shows the leakage of a two-input stack (designed with devices of $L_{\rm eff} = 25$ nm) for different input vectors "00," "01," and

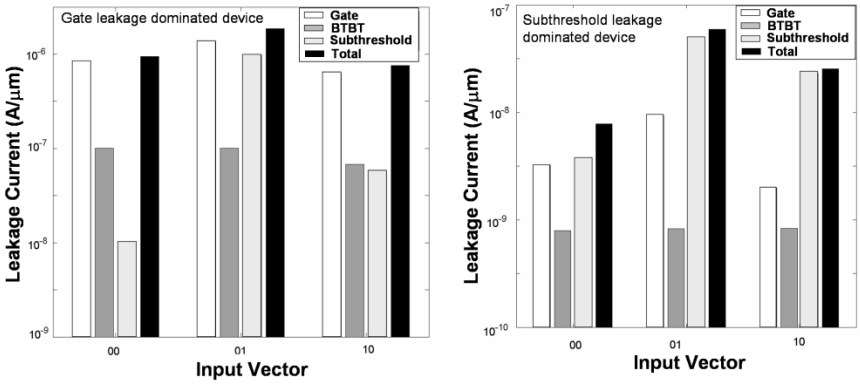


Fig. 5. Leakage of 2-T stack for different input vector with NMOS devices of $L_{\rm eff} = 25$ nm. (a) With the gate leakage dominant devices and (b) with the subthreshold leakage dominated devices. When the gate leakage is the dominant leakage component "10" is the minimum leakage vector, whereas when the subthreshold leakage is the dominant leakage component "00" is the minimum leakage vector [3]. The leakage values are extracted using device simulation in MEDICI. $V_{\rm DD}$ of 0.7 V is used at 25 nm.

"10." The stack in Fig. 5(a) is designed with an NMOS transistor which has higher gate leakage, whereas the stack in Fig. 5(b) is designed with transistors with higher subthreshold leakage. The total leakage currents in both types of devices were kept the same. It can be observed that in Fig. 5(a), the minimum leakage vector is "10," whereas in Fig. 5(b) the minimum leakage vector is "00." A more detailed discussion of the effect of gate leakage to subthreshold leakage ratio in the leakage of a logic gate can be found in [3]. Hence, we can conclude that the total leakage of a logic gate and logic circuits strongly depends on the relative magnitude of the different leakage components and hence, an accurate estimation of leakage should consider each of these components and their interaction.

III. Leakage Estimation Steps

It is evident from the discussions in the previous sections that in scaled devices, each of the leakage components is extremely important and the leakage strongly depends on the transistor geometry, the doping profile, and the temperature. Hence, the proposed estimation method of total leakage of a logic circuit starts with the description (device geometry, doping profile) of the transistor and the operating temperature. The different steps used to estimate the total leakage are shown in Fig. 6. First, the leakages for a device are modeled. Using the models of individual leakage components a transistor is described as an SCS. Based on the SCS model, the leakage current of basic logic gates are calculated. This calculation is done using a numerical solver written in MATLAB. In the next step, an algorithm is used to calculate the leakage of a given logic circuit (gate level description) based on the leakage of the logic gates. In the present work we have used the algorithm given in [7]. In this algorithm the circuit leakage for an applied input vector is obtained by adding the leakage of the individual logic gates (input vector dependent estimation of leakage). However, the SCS model can also be used to describe a transistor in SPICE. A transistor level description of the logic circuit in SPICE can then be used to calculate the total circuit leakage. The outputs of the estimation tool are the subthreshold, gate and BTBT leakage components along with the total leakage of the circuit. The following sections elaborate each of the steps shown in the Fig. 6.

IV. Modeling Leakage Components

This section describes the general approach used to formulate the model for the junction BTBT, the subthreshold, and the gate leakage in a MOSFET. The formulation, developed for NMOS transistors, can be easily extended to PMOS transistors. Device structures with 2-D nonuniform channel ("super halo" channel doping) and source/drain (S/D) doping profiles have been considered while deriving these models. A schematic of the device structure (symmetric about the middle of the channel) is shown in Fig. 7 [13]. The 2-D doping profile in the channel and the source/drain region are approximated as Gaussian functions [4], [13]. The Gaussian doping profile in the channel ($N_a(x,y)$) and S/D ($N_{\rm sd}(x,y)$) can be represented as [4], [13]

$$x > 0$$
$$N_a(x,y) = A_p \Gamma_{xa}(x) K_{ya}(y) + N_{\rm SUB}$$

where

$$K_{ya}(y) = \exp\left(\frac{-(y-\alpha_a)^2}{\sigma_{ya}^2}\right)$$

and

$$\left[\begin{array}{ll}\Gamma_{xa}(x) = \exp\left(\frac{-(x-\beta_a)^2}{\sigma_{xa}^2}\right), & 0 \le x \le \beta_a \\ = 1, & x > \beta_a\end{array}\right]. \tag{1}$$

Similarly, the S/D doping [$N_{\rm sd}(x,y)$] can be represented as

$$x > 0$$
$$N_{\rm sd}(x,y) = A_{\rm sd} \Gamma_{\rm xsd}(x) K_{\rm ysd}(y)$$

where

$$K_{\rm ysd}(y) = \exp\left(\frac{-(y)^2}{\sigma_{\rm ysd}^2}\right)$$

and

$$\left[\begin{array}{ll}\Gamma_{\rm xsd}(x) = \exp\left(\frac{-(x-\beta_{\rm sd})^2}{\sigma_{\rm xsd}^2}\right), & 0 \le x \le \beta_{\rm sd} \\ = 1, & x > \beta_{\rm sd}\end{array}\right] \tag{2}$$

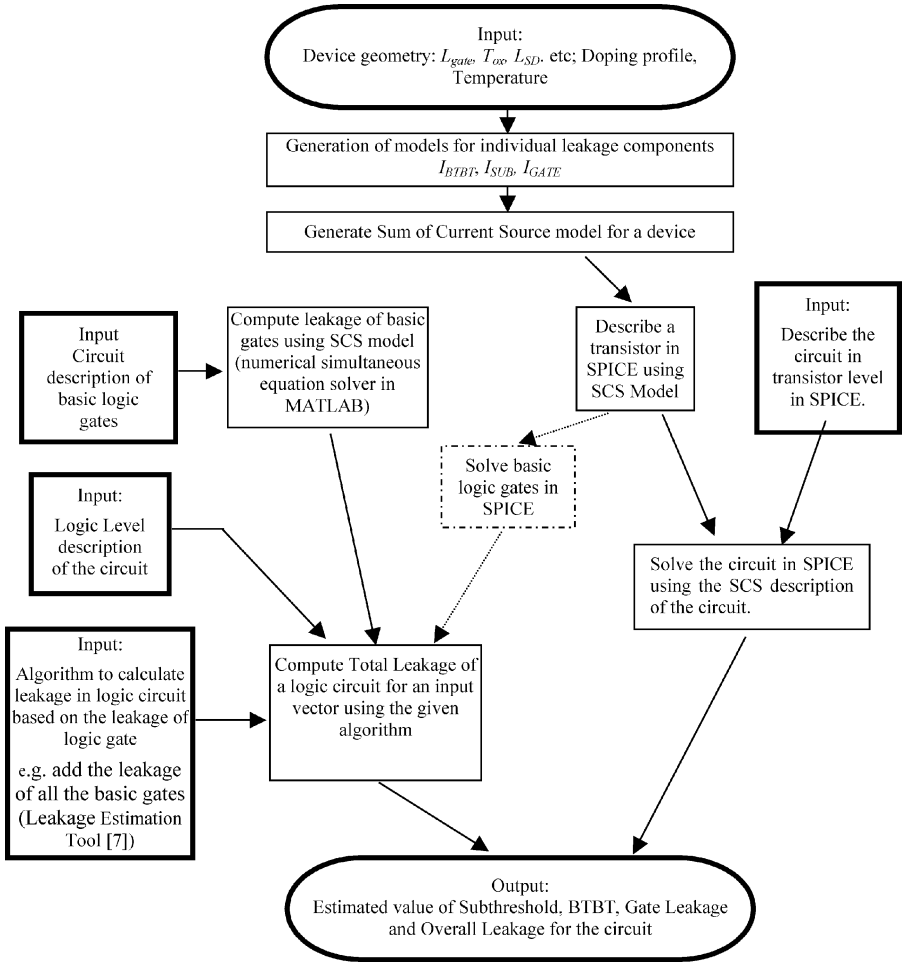


Fig. 6. Leakage estimation steps.

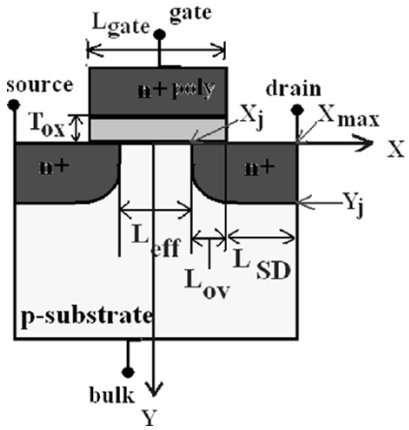


Fig. 7. Device structure.

where A_p and $A_{\rm sd}$ represent the peak "halo" and S/D doping, respectively. $N_{\rm SUB}$ is the constant uniform doping in the bulk and is much less compared to contributions from Gaussian profiles at and near the channel and S/D regions. Parameters α_a, $\alpha_{\rm sd}$ $(=0)$, β_a, and $\beta_{\rm sd}$ control the positions and σ_{ya}, σ_{xa}, and $\sigma_{\rm ysd}$, $\sigma_{\rm xsd}$ control the variances of the Gaussian profiles in channel and S/D regions [4], [13]. The detailed process for the extraction of these parameters from device measurement results is given in [13]. The described method extracts the Gaussian parameters from the measurement of $I_{\rm DS}$ at different values of $V_{\rm GS}$ and V_{BS} data in the subthreshold region [13]. The method uses a global optimization procedure to obtain the doping profile parameters shown in (1) and (2) by fitting the MEDICI simulation result to the experimental measurement results [13]. Unless otherwise specified in this paper, we have used NMOS ($N_{\rm ref}$) and PMOS ($P_{\rm ref}$) transistors with $L_{\rm eff} = 25$ nm, $W_{\rm eff} = 1$ μm, and channel doping profile $\alpha_a = 0.018$ μm, $\sigma_{ya} = 0.018$ μm $\beta_a = 0.016$ μm, $\sigma_{xa} = 0.016$ μm, and S/D profile from [4]. Fig. 8 shows the nature of the doping profiles for the device $N_{\rm ref}$.

A. Modeling Junction Band-to-Band Leakage Current ($I_{\rm BTBT}$)

A high electric field across a reverse biased p-n junction causes significant current to flow through the junction due to tunneling of electrons from the valence band of the p-region to the conduction band of the n-region (causing the generation of holes in the p-region) as shown in Fig. 9 [5]. From Fig. 9, it is evident that for such tunneling to occur the total voltage drop across the junction (applied reverse bias ($V_{\rm app}$) + built-in voltage($\psi_{\rm bi}$) must be more than the band-gap. Since silicon is an indirect band-gap semiconductor, the BTBT current in

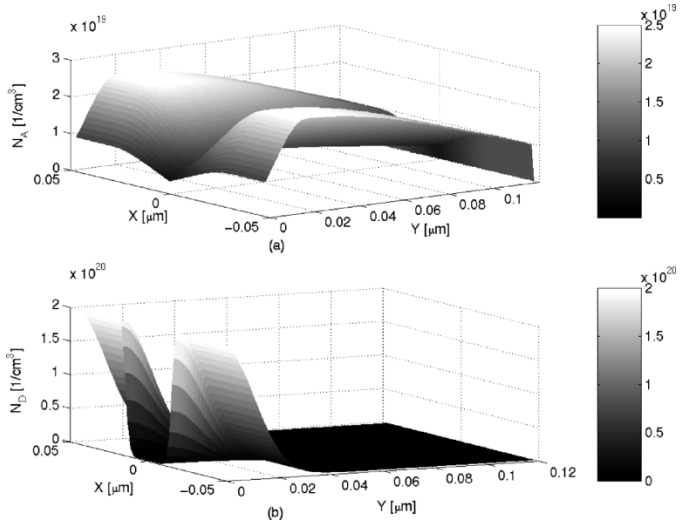


Fig. 8. 2-D Gaussian profile for (a) channel and (b) source-drain region.

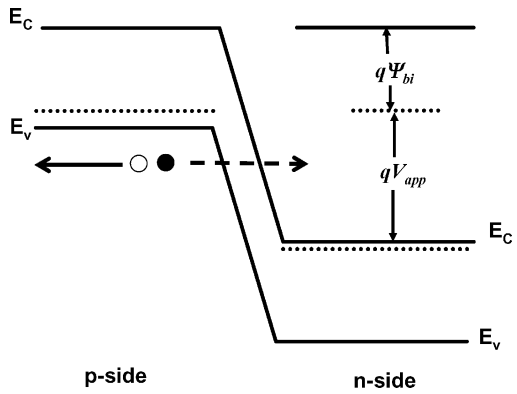


Fig. 9. Physical picture of valence band electron tunneling in a reversed bias p-n junction. The tunneling current density through a silicon p-n junction is given in [5], [14].

silicon involves the emission or absorption of phonon(s) [5]. The current density of junction BTBT is given by [5]

$$J_{b-b} = A\frac{\xi V_{\rm app}}{E_g^{1/2}}\exp\left(-B\frac{E_g^{3/2}}{\xi}\right)$$
$$A = \frac{\sqrt{2m^*}q^3}{4\pi^3\hbar^2} \qquad B = \frac{4\sqrt{2m^*}}{3q\hbar} \tag{3}$$

where, m^* is effective mass of electron, E_g is energy band-gap, ξ is the electric field at the junction, q is electronic charge, and $\hbar$ is the reduced Plank's constant.

In a NMOSFET, when the drain or the source is biased at a potential higher than that of the substrate, a significant BTBT current flows through the drain-substrate and the source-substrate junctions. The total junction BTBT current in the MOSFET is the sum of the currents flowing through the drain-substrate and source-substrate junctions and is given by

$$I_{\rm BTBT} = w_{\rm eff}\int_l J_{b-b}(x,y)dl\Big|_{\rm drain} + w_{\rm eff}\int_l J_{b-b}(x,y)dl\Big|_{\rm source}$$
$$l := \text{Junction line} \equiv \text{solution of the equation}$$
$$N_a(x,y) = N_d(x,y) \tag{4}$$

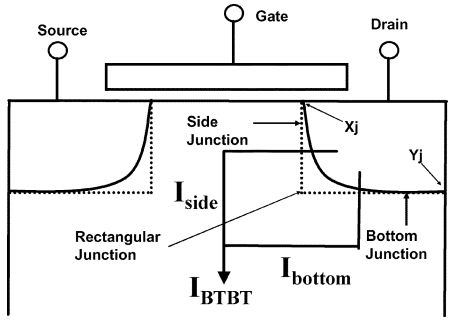


Fig. 10. "Rectangular junction" approximation.

where $w_{\rm eff}$ is the effective width and $J_{b-b}(x,y)$ is the current density at a point (x,y) at the junction. The integration in (4) has to be done along the junction line "l" (Fig. 10) within the tunneling region, i.e., for all values (x,y) for which $(V_{\rm app} + \psi_{\rm bi}(x,y)) > E_g/q$. Based on (3) and (4), the numerical method for evaluating the junction BTBT current is given below.

Numerical Approach for the Calculation of the Junction BTBT Leakage:

For each of the drain-substrate and source-substrate junction do the following

Step 1) Obtain the definition of the junction line "l" given by

$$l = \{(x,y) : N_a(x,y) = N_d(x,y)\}. \tag{5}$$

Step 2) At each point (x,y) of the junction line "l," obtain the electric field $\xi(x,y)$. Let us assume that $xp(y)$ and $xn(y)$ is the width of the depletion region on the substrate side and source/drain side respectively. Using Poisson's equation and the continuity of the electric field at the junction, we obtain the following equation connecting $xp(y)$ and $xn(y)$:

$$\xi(x,y) = \int_{-xp(y)}^{x}\frac{q}{\varepsilon_{si}}\rho(x_1,y)dx_1 = \int_{-xp(y)}^{x}\frac{q}{\varepsilon_{si}}\left[N_{\rm sd}(x_1,y) - N_a(x_1,y)\right]dx_1$$
$$\text{for} -xp(y) \le x_1 \le x$$
$$\xi(x,y) = \int_{x}^{xn(y)}\frac{q}{\varepsilon_{si}}\rho(x_1,y)dx_1 = \int_{x}^{xn(y)}\frac{q}{\varepsilon_{si}}\left[N_{\rm sd}(x_1,y) - N_a(x_1,y)\right]dx_1$$
$$\text{for } x \le x_1 \le xn(y).$$

Using the continuity of the electric field at the junction

$$\int_{-xp(y)}^{x}\frac{q}{\varepsilon_{si}}\left[N_{\rm sd}(x_1,y) - N_a(x_1,y)\right]dx_1 = \int_{x}^{xn(y)}\frac{q}{\varepsilon_{si}}\left[N_{\rm sd}(x_1,y) - N_a(x_1,y)\right]dx_1 \tag{6}$$

is obtained, where ε_{si} is the permittivity of silicon. Since $\xi(x) = -dv/dx$, the solution for the electrostatic potential gives another equation connecting $xp(y)$ and $xn(y)$. Using the boundary condition $V = 0$ at $x_1 = -xp(y)$ and $V = V_{\rm bi} + V_{\rm app}$ at $x_1 = xn(y)$, and the continuity of the potential at

the junction ($x_1 = x$) we obtain the following relation [14]:

$$\int_{-xp(y)}^{x} \xi(x_1, y)dx_1 = (V_{\rm bi} + V_{\rm app}) - \int_{x}^{xn(y)} \xi(x_1, y)dx_1$$

or

$$\int_{-xp(y)}^{x}\left[\int_{-xp(y)}^{x1} \frac{q}{\varepsilon_{\rm si}}\left[N_{\rm sd}(x_1, y) - N_a(x_1, y)\right]dx_1\right]dx_1 = V_{\rm bi}(x, y) - \int_{x}^{xn(y)}\left[\int_{x1}^{xn(y)} \frac{q}{\varepsilon_{\rm si}}[N_{\rm sd}(x_1, y) - N_a(x_1, y)]\,dx_1\right]dx_1 \quad (7)$$

where $V_{\rm bi}(x, y)$ is the built-in potential at the point (x, y) in the junction. The built-in potential is given by [14]

$$V_{\rm bi}(x, y) = \left(\frac{1}{q}\right)\Big[E_g - \big(E_c(x, y) - E_F(x, y)\big)_{\rm n-side} - \big(E_F(x, y) - E_V(x, y)\big)_{\rm p-side}\Big] \quad (8)$$

where E_g is the band-gap, $E_C(x, y)$ is the bottom of the conduction band in the n-side, $E_F(x, y)$ is the Fermi-level, and $E_V(x, y)$ is the top of the valence band in the p-side (all values are evaluated at the point (x, y)). For nondegenerate values of $N_a(x, y)$ and $N_{\rm sd}(x, y)$ the built-in potential is given by [14]

$$V_{\rm bi}(x, y) = \left(\frac{kT}{q}\right)\ln\left(\frac{N_a(x, y)N_{\rm sd}(x, y)}{n_i^2}\right). \quad (9)$$

However, the S/D and halo doping can easily go beyond the nondegenerate limit. In that case, $V_{\rm bi}(x, y)$ can be obtained using the following relations [12]:

$$N_a(x, y) = N_V F_{1/2}\left(E_V(x, y) - \frac{E_F(x, y)}{kT}\right)$$
$$N_{\rm sd}(x, y) = N_C F_{1/2}\left(E_F(x, y) - \frac{E_C(x, y)}{kT}\right) \quad (10)$$

where N_C and N_V are effective density of states for conduction band and valence band, respectively. The modified Fermi-Dirac integral $F_{1/2}$ can be evaluated numerically or using the approximate analytical expressions given in [12]. The impact of band-gap narrowing due to heavy doping can also be included to make the calculations more accurate [11]. However, we have not included it in this paper.

Step 3) Obtain the current density at the point (x, y) using (3) and (5) as given below

$$J_{b-b}(x, y) = A\frac{\xi(x, y)V_{\rm app}}{E_g^{1/2}}\exp\left(-B\frac{E_g^{3/2}}{\xi(x, y)}\right) \quad \text{if } V_{\rm bi}(x, y) + V_{\rm app} \geq \frac{E_g}{q}$$
$$= 0, \quad \text{if } V_{\rm bi}(x, y) + V_{\rm app} < \frac{E_g}{q} \quad (11)$$

where $V_{\rm app} = V(\text{drain, substrate}) = V_{\rm db}$ for the drain-substrate junction and $V_{\rm app} = V(\text{source, substrate}) = V_{\rm sb}$.

Step 4) Using (4) and (11), we can calculate the current through each of the junctions

$$I_{\rm BTBT}|_{\rm drain/source} = w_{\rm eff}\int_{l} J_{b-b}(x, y)dl\Big|_{\rm drain/source} = w_{\rm eff}\sum_{(x,y)\in l} J_{b-b}(x, y)\Big|_{\rm drain/source}. \quad (12)$$

Step 5) Add the currents due to the source and the drain junction to get the total current.

The above method accurately accounts for the complex 2-D doping profile. However, it requires numerical solution of (5), numerical solution of nonlinear simultaneous differential (6)–(7), and numerical integration of (12). The computational complexity of the model makes it unsuitable for a circuit simulation. Hence, based on this numerical procedure, we have developed a reasonably accurate analytical model for the junction BTBT current which captures the physical effects of the 2-D doping profile and can be efficiently used for circuit simulation.

Development of the Analytical Model for the Calculation of the Junction BTBT Leakage: For a symmetrical MOSFET, the drain and the source junction will have the same expression for the junction tunneling current. Hence, here we have described the model for the drain junction only. To achieve an analytical model we first approximated the complex junction *'l'* as a rectangular one (Fig. 10). Using the "rectangular junction" approximation the total current through a junction is given by:

$$I_{\rm BTBTdrain} = I_{\rm side} + I_{\rm bottom} = w_{\rm eff}\int_{\rm side} J_{b-b}(X_j, y)dy + w_{\rm eff}\int_{\rm bottom} J_{b-b}(x, Y_j)dx \quad (13)$$

where X_j is the position of the side junction and Y_j is the position of the bottom junction (Fig. 10). The rectangular junction approximation eliminates the need for a numerical solution of (5). The lateral junction depth X_j and vertical junction depth Y_j are found by analytically solving the following equations (see the Appendix):

$$N_{\rm sd}(X_j, y = 0) = N_a(X_j, y = 0)$$
$$N_{\rm sd}(x = x_{\max}, Y_j) = N_a(x_{\max}, Y_j). \quad (14)$$

In this paper, we have presented the derivation of the current due to side junction. The current due to the bottom junction can be derived following a similar procedure. The current due to the side junction is given by

$$I_{\rm side} = w_{\rm eff}\int_{y1}^{y2} J_{b-b}(X_j, y)dy = \int_{y1}^{y2} A\frac{\xi(X_j, y)V_{\rm app}}{E_g^{1/2}}\exp\left(-\frac{BE_g^{3/2}}{\xi(X_j, y)}\right)dy \quad (15)$$

where y_1 to y_2 is the tunneling region. However, due to the nonuniform doping in the substrate and the drain region, this integration cannot be solved analytically. Hence, we approximate the integral using an average tunneling current density ($J_{\mathrm{b-bside}}$), which is determined by the average electric field (ξ_{side}) across the junction. This is given by the following equation:

$$\begin{aligned} I_{\mathrm{side}} &= w_{\mathrm{eff}} \left| y_2 - y_1 \right| J_{\mathrm{b-bside}} \\ &= \int_{y1}^{y2} A \frac{\xi_{\mathrm{side}} V_{\mathrm{app}}}{E_g^{1/2}} \exp\left(-\frac{\mathrm{BE}_g^{3/2}}{\xi_{\mathrm{side}}} \right) dy \end{aligned} \tag{16}$$

where ξ_{side} is given by

$$\begin{aligned} \xi_{\mathrm{side}} &= \frac{1}{|y2 - y1|} \int_{y1}^{y2} \xi(X_j, y) dy \\ &= \frac{1}{|y2 - y1|} \int_{y1}^{y2} \int_{xp}^{X_j} \frac{q}{\varepsilon_{si}} \left[N_{\mathrm{sd}}(x,y) - N_a(x,y) \right] dx \end{aligned} \tag{17}$$

where $\xi(X_j, y)$ is the electric field at the junction of the differential diode of length dy. However, for a nonuniform 2-D profile, the expression become too complicated to be solved analytically. To simplify the derivation, keeping the essential information of the electric field, we defined the average field as

$$\xi_{\mathrm{side}} = \sqrt{\frac{2qN_{\mathrm{aside}}N_{\mathrm{dside}}(V_{\mathrm{app}} + V_{\mathrm{biside}})}{\varepsilon_{si}(N_{\mathrm{aside}} + N_{\mathrm{dside}})}}. \tag{18}$$

This is the field at the junction of the p–n junction with p-side and n-side doping equal to N_{aside} and N_{dside}, respectively. The validity of the above "step junction" simplification can be justified by analyzing the doping profile of differential diode "dy" at depth y (Fig. 11). It can be observed from Fig. 11, that for practical values of the doping profiles, the junction can be assumed as a step junction with doping at the p side ($N_{\mathrm{aside}}(X_j, y)$) and n side ($N_{\mathrm{dside}}(X_j, y)$) given by

$$\begin{aligned} &N_{\mathrm{aside}}(X_j, y) \\ &\quad = N_a(X_j, y) \\ &\quad = A_p \Gamma_{xa}(X_j) K_{ya}(y) \\ &N_{\mathrm{dside}}(X_j, y) \\ &\quad = N_{\mathrm{sd}}(x = \beta_{\mathrm{sd}}, y) - N_a(x = \beta_a, y) \\ &\quad = A_{\mathrm{sd}} \Gamma_{\mathrm{xsd}}(x = \beta_{\mathrm{sd}}) K_{\mathrm{ysd}}(y) - A_p \Gamma_{xa}(x = \beta_a) K_{ya}(y) \\ &\quad = A_{\mathrm{sd}} K_{\mathrm{ysd}}(y) - A_p K_{ya}(y). \end{aligned} \tag{19}$$

Using the observations from Fig. 11 and (19), N_{aside} is given by

$$N_{\mathrm{aside}} = \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} N_{\mathrm{aside}}(X_j, y) dy. \tag{20}$$

The above integration is transformed into the form of error function and evaluated (see the Appendix for details). N_{dside} can also be obtained similarly and given by

$$N_{\mathrm{dside}} = \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} N_{\mathrm{dside}}(X_j, y) dy. \tag{21}$$

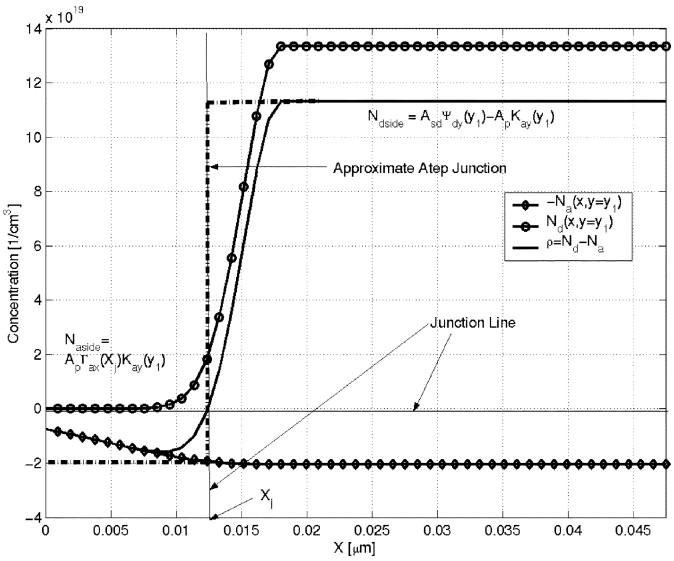


Fig. 11. "Step junction" approximation.

Using (8)–(10) built-in potential V_{biside} is obtained as [14]

$$\begin{aligned} V_{\mathrm{biside}} &= \frac{kT}{q} \ln\left(\frac{N_{\mathrm{aside}} N_{\mathrm{dside}}}{n_i^2} \right) \\ &\quad \text{for nondegenerate doping} \\ &= \frac{E_g}{q} + \left(\frac{kT}{q} \right) \left[F_{1/2}^{-1}\left(\frac{N_{\mathrm{aside}}}{N_V} \right) + F_{1/2}^{-1}\left(\frac{N_{\mathrm{dside}}}{N_C} \right) \right] \\ &\quad \text{for degenerate doping.} \end{aligned} \tag{22}$$

For simplicity, the whole side junction is assumed to be tunneling (i.e., $y_1 = 0$ and $y_2 = Y_j$). For bottom junction $x_1 = X_j$ and $x_2 = x_{\max}$. Using expressions from (20)–(22), into (18) ξ_{side} (and similarly ξ_{bottom}) can be obtained. ξ_{side} (and ξ_{bottom}) can be used to obtain J_{side} (similarly J_{bottom}). If $(V_{\mathrm{app}} + V_{\mathrm{biside}}) < E_g/q$, then no tunneling occurs and J_{side} is zero (similar argument holds for J_{bottom}). Hence, the total junction BTBT current in the drain junction is given by

$$\begin{aligned} I_{\mathrm{BTBTdrain}} &= w_{\mathrm{eff}} \left| Y_j \right| J_{\mathrm{side}} + w_{\mathrm{eff}} \left| x_{\max} - X_j \right| J_{\mathrm{bottom}} \\ J_{\mathrm{side}} &= A \frac{\xi_{\mathrm{side}} V_{\mathrm{app}}}{E_g^{1/2}} \exp\left(-\frac{\mathrm{BE}_g^{3/2}}{\xi_{\mathrm{side}}} \right) \\ &\quad \text{when } (V_{\mathrm{app}} + V_{\mathrm{biside}}) \geq \frac{E_g}{q} \\ &= 0, \quad \text{otherwise} \\ J_{\mathrm{bottom}} &= A \frac{\xi_{\mathrm{bottom}} V_{\mathrm{app}}}{E_g^{1/2}} \exp\left(-\frac{\mathrm{BE}_g^{3/2}}{\xi_{\mathrm{bottom}}} \right) \\ &\quad \text{when } (V_{\mathrm{app}} + V_{\mathrm{bibottom}}) \geq \frac{E_g}{q} \\ &= 0, \quad \text{otherwise.} \end{aligned} \tag{23}$$

The "step junction" approximation eliminates the numerical complexity of (6), (7) and (12) by the analytical evaluation of (16), (18), and (23).

For a 25-nm transistor, the comparison of the analytical model given in (23) and the simulated data from MEDICI [11] shows close match for small reverse and forward substrate bias (Fig. 12). However, deviations are observed at high forward

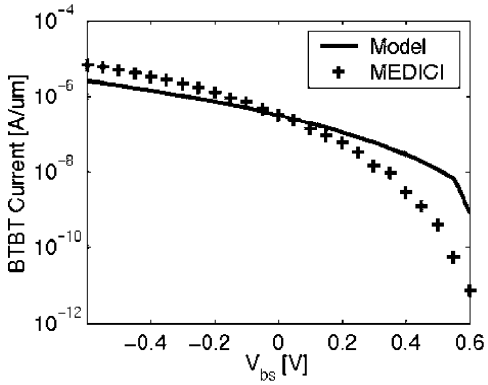


Fig. 12. Variation of BTBT current with substrate bias. Comparison of analytical result with simulated data from MEDICI for NMOS transistor with $L_{\rm eff} = 25$ nm and doping profile: $\alpha_a = 0.018\ \mu$m, $\sigma_{ay} = 0.018\ \mu$m $\beta_a = 0.016\ \mu$m, $\sigma_{\rm ax} = 0.016\ \mu$m. $V_{\rm dd}$ of 0.7 V is used at 25 nm.

(i.e., low $V_{\rm app}$) and reverse (i.e., very high $V_{\rm app}$) substrate bias. At high $V_{\rm app}$, the average electric field (calculated using the average doping density) used in the model is considerably less than the peak field (at the peak doping region). Since the tunneling is dominated by the peak field, the analytical current is less than the simulated one at high $V_{\rm app}$. In the low bias region, reduction of $V_{\rm app}$ considerably reduces the tunneling volume. The model does not consider the reduction of the tunneling volume. Moreover, the derived field is based on the abrupt junction approximation which also predicts a higher field. Hence, the evaluated current is higher than the simulated current at low $V_{\rm app}$ (i.e., high forward substrate bias). Also, at a high gate voltage: 1) small increase in the potential near the substrate side of the side junction and 2) nonnegligible voltage drop at the S/D series resistance caused by the high "on" current flowing through the transistor reduce the effective applied reverse bias across the junction. Hence, the BTBT current reduces by a small amount. Exact modeling of these effects requires calculation of the tunneling rate at each point, which makes formulation of a compact circuit model of the currents extremely difficult. To take care of these effects an empirical function ($\lambda(V_{\rm app})$) and an empirical gate correction factor (δ_g) have been introduced in the model. With these corrections the current due to the drain junction (or source) is given by

$$I_{\rm BTBTcorrected} = a_0 I_{\rm BTBTdrain}\left(\lambda(V_{\rm app})\right)\left(1 - \delta_g V_G\right) \quad (24)$$

where $\lambda(V_{\rm app})$ is an empirical function (for drain-substrate junction $V_{\rm app} = V_{\rm db}$ and for source-substrate junction $V_{\rm app} = V_{\rm sb}$). From previous discussion and from Fig. 12 we can conclude that the empirical function should have the following nature:

$$\lambda(V_{\rm app}) \begin{cases} > 1 \text{ and increasing}, & \text{for } V_{\rm app} \to V_{\rm dd} \\ = C = \frac{I_{\rm BTBTmeasured}}{I_{\rm BTBTcomputed}}, & \text{for } V_{\rm app} = V_{\rm dd} \\ < 1, \text{ decreasing and} \to 0, & \text{for } V_{\rm app} \to 0 \end{cases} . \quad (25)$$

Based on (25), we can design $\lambda(V_{\rm app})$ as a polynomial function. We found that a cubic function gives a reasonably good fit and the accuracy increases with the use of higher degree polynomial. However, a polynomial of degree "n" requires "$n+1$" fitting parameters. Hence, a cubic polynomial will require four fitting parameters and the number will be higher for higher degree polynomials. Moreover, it has been observed that a change in the doping profile changes the fitting parameters by a considerable amount and hence makes it difficult to study the effect of doping profile variation on the current. From the previous discussion about the origin of the error in the analytical model, we can observe that the source of error is the underestimation of the electric field (high $V_{\rm app}$) and overestimation of the tunneling volume (low $V_{\rm app}$). An increase in the electric field exponentially increases the current and the tunneling volume tends to reduce at a linear rate with the reduction of the applied voltage. Hence, the auxiliary function should exponentially depend on the applied voltage ($e^x \to 1 + x$ for small x). Based on this observation, we have designed the auxiliary function $\lambda(V_{\rm app})$ as follows:

$$\lambda(V_{\rm app}) = \exp\left(\frac{\alpha_{\rm btbt} V_{\rm app}^{1+m}}{\beta_{\rm btbt} + \gamma_{\rm btbt} V_{\rm app}^{m}}\right) - 1 \quad (26)$$

where $\alpha_{\rm btbt}$, $\gamma_{\rm btbt}$, and $\beta_{\rm btbt}$ are the fitting parameters. Value of m controls the complexity of the model: $m = 1$ or 2 gives an accurate fitting. A higher value of m will improve the accuracy, but it will also make the model more complex for simulation. We have used $m = 2$ for simulations in our present work. The fitting parameters are found to be less sensitive to the variation in doping profile for a transistor of a particular channel length. The parameters can be extracted using the following simple procedure:

$$\begin{aligned} \lambda(V_{\rm dd}) &= \exp\left(\frac{\alpha_{\rm btbt} V_{\rm dd}^{1+m}}{\beta_{\rm btbt} + \gamma_{\rm btbt} V_{\rm dd}^{m}}\right) - 1 \\ &= \frac{I_{\rm BTBTmeasured}(V_d = V_{\rm dd}, V_s = V_g = V_b = 0)}{I_{\rm BTBTcomputed}(V_d = V_{\rm dd}, V_s = V_g = V_b = 0)} \\ &\quad \text{assume, } \frac{V_{\rm dd}^{1+m}}{\beta_{\rm btbt} + \gamma_{\rm btbt} V_{\rm dd}^{m}} = 1 \\ &\quad \text{and select } \alpha_{\rm btbt} \text{ to satisfy the above equation.} \\ &\quad \text{select } \beta_{\rm btbt} \text{ and } \gamma_{\rm btbt} \text{ such that they satisfy} \\ &\quad \frac{V_{\rm dd}^{1+m}}{\beta_{\rm btbt} + \gamma_{\rm btbt} V_{\rm dd}^{m}} = 1 \\ &\quad \text{and match the simulated measured result for} \\ &\quad \text{different values of } V_{\rm app}. \end{aligned} \quad (27)$$

Gate correction factor δ_g can be calculated from the actual BTBT value at low and high gate bias.

The final expression for the total BTBT current is given in (28) at the bottom of this page. The parameters, namely, ξ_{side_i} and ξ_{bot_i}, can be evaluated following the procedure discussed above. Fig. 13 shows a comparison plot of the analytical results with the simulated results from MEDICI for NMOS devices with $L_{\rm eff} = 25$ nm ($V_{\rm dd} = 0.7$ V) and 50 nm ($V_{\rm dd} = 0.9$ V) and different doping profiles. It shows that for the analytical result follows very closely the simulated result for substrate bias in the range of $-V_{\rm dd}/2$ to $+V_{\rm dd}/2$. However, deviation is observed at very high forward and reverse bias.

The evaluation of the junction BTBT current in PMOS is more involved. Fig. 14 shows the variation of the junction BTBT current with $V_{\rm db}$ of a PMOS and an NMOS device of $L_{\rm eff} = 25$ nm and the PMOS has a doping

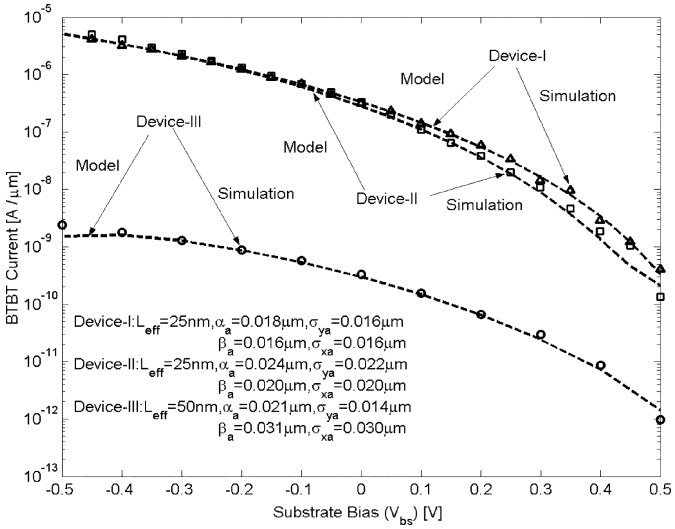


Fig. 13. Variation of BTBT current with substrate bias. Comparison of analytical result with simulated data from MEDICI for N MOS transistor with $L_{\rm eff} = 25$ nm and doping profile: $\alpha_a = 0.018\ \mu$m, $\sigma_{ay} = 0.016\ \mu$m $\beta_a = 0.016\ \mu$m, $\sigma_{\rm ax} = 0.020\ \mu$m. $V_{\rm dd} = 0.7$ V at 25 nm, and $V_{\rm dd} = 0.9$ V at 25 nm.

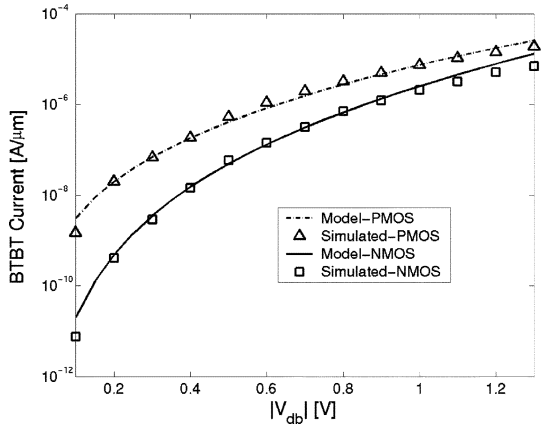


Fig. 14. Comparison of PMOS and NMOS junction BTBT current considering complementary doping profile. Both the NMOS and PMOS are of $L_{\rm eff} = 25$ nm.

profile which is just complementary to the NMOS profile (i.e., $N_a(x,y)|_{\rm PMOS} = N_a(x,y)|_{\rm NMOS}$ with N_a representing n-type doping in PMOS and p-type doping in NMOS; $N_{\rm sd}(x,y)|_{\rm PMOS} = N_{\rm sd}(x,y)|_{\rm NMOS}$ with $N_{\rm sd}$ representing p-type doping in PMOS and n-type doping in NMOS). From Fig. 14 it can be observed that the current in PMOS is considerably higher than that in NMOS. The plausible explanations for this behavior are summarized as follows.

1) The effective density of states for holes (N_V) is less than that of electrons (N_C) [12]. Hence, a higher acceptor or p-type doping concentration in PMOS compared to NMOS (in PMOS p-type region is source/drain, which has higher doping compared to the p-type region in NMOS, which is the halo region) results in a higher built-in potential and thereby larger tunneling current.
2) The doping variation in the p-type region in PMOS (source/drain) has a steeper slope and a higher value compared to those of the p-type region in NMOS (halo). This results in a higher local electric field at the different points in the p-type region in PMOS than in the p-type region in NMOS. A higher local electric field increases the generation rate of holes in the p-type region and, hence, a larger current.
3) Value of $V_{\rm bi}(x,y)$ at any point (x,y) in the junction, is higher in PMOS compared to NMOS. Hence, for the same applied bias, the tunneling volume in PMOS is larger than that in NMOS. This also results in a higher current in PMOS.

The derived analytical model captures the global nature of the difference between N_C and N_V [using (22)]. However, the model does not account for the increase in the generation rate and tunneling volume, because of the extreme local nature of these phenomena. To empirically accounts for these effects we introduce a fitting parameter "p" which multiplies the p-type doping density in (18) and (22) (i.e., $N_{\rm aside}$ in NMOS and $N_{\rm dside}$ in PMOS). The parameter "p" is equal to "1" in NMOS and it is greater than "1" in PMOS. Fig. 14 shows that with this parameter the analytical result follows the simulated result very closely for PMOS ($p = 1.4$ in this particular example).

B. Modeling Subthreshold Current ($I_{\rm ds}$)

In the "off" state of a device ($V_{\rm gs} < V_{\rm th}$), the current flowing from the drain to the source of a transistor is known as the subthreshold current. The subthreshold current flowing through a transistor is given by [5]

$$I_{\rm sub} = \frac{w_{\rm eff}}{L_{\rm eff}}\mu\sqrt{\frac{q\varepsilon_{si}N_{\rm cheff}}{2\Phi_s}}v_T^2 \times \exp\left(\frac{V_{\rm gs}-V_{\rm th}}{nv_T}\right)\left(1-\exp\left(\frac{-V_{\rm ds}}{v_T}\right)\right) \quad (29)$$

where $N_{\rm cheff}$ is the effective channel doping, Φ_s is the surface potential, n is the subthreshold swing, and v_T is the thermal

$$I_{\rm btbt} = \sum_{i={\rm drain,source}} (I_{side\_i} + I_{bottom\_i})(1-\delta_G V_G)\left(\exp\left(\frac{\alpha_{\rm btbt}V_{ib}^{1+m}}{\beta_{\rm btbt}+\gamma_{\rm btbt}V_{ib}^m}\right)-1\right)$$
$$I_{side\_i} = w_{\rm eff}A\frac{\xi_{side\_i}}{E_g^{1/2}}V_{ib}\exp\left(-\frac{BE_g^{3/2}}{\xi_{side\_i}}\right)$$
$$I_{bottom\_i} = w_{\rm eff}A\frac{\xi_{bottom\_i}}{E_g^{1/2}}V_{ib}\exp\left(-\frac{BE_g^{3/2}}{\xi_{bottom\_i}}\right) \quad (28)$$

voltage given by kT/q. Using the charge-sharing model and following the procedure given in [15]–[17], the threshold voltage can be expressed as

$$V_{\rm th} = V_{\rm FB} + (\Phi_{s0} - \Delta\Phi_s) + \gamma\sqrt{\Phi_{s0} - V_{\rm bs}}\left(1 - \lambda\frac{X_d}{L_{\rm eff}}\right) + \Delta V_{\rm NWE} \quad (30)$$

where $V_{\rm FB}$ is the flat-band voltage, Φ_{s0} is the zero bias surface potential, γ is the body factor, $C_{\rm ox} = \varepsilon_{\rm sio2}/t_{\rm ox}$ is the oxide capacitance, X_d is the depletion layer thickness, λ is a fitting parameter (≈ 1) and $\Delta V_{\rm NWE}$ is the narrow-width correction factor given in [17]. $\Delta\Phi_s$ is the reduction of the surface potential (Φ_s) of short channel devices from its zero bias value due to SCEs like DIBL and $V_{\rm th}$ roll-off [15]–[17]. Different parameters in the above model depend on the effective channel and source/drain doping [15]–[17]. We have evaluated the effective channel ($N_{\rm cheff}$) and source/drain doping ($N_{\rm sdeff}$) considering the Gaussian approximation of the 2-D doping profile [given in (1) and (2)] as given as follows:

$$N_{\rm sdeff} = \frac{1}{\Delta_{\rm sd}}\int\int_{\Delta_{\rm sd}} N_{\rm sd}(x,y)dxdy = \frac{A_{\rm sd}}{\Delta_{\rm sd}}\int_{x=X_j}^{x=L_{\rm gate}/2+L_{\rm sd}} \Gamma_{\rm xsd}(x)dx\int_{y=0}^{y=Y_j} K_{\rm ysd}(y)dy \quad (31)$$

where $\Delta_{\rm sd} = (L_{\rm overlap} + L_{\rm sd})Y_j$ is S/D area, $L_{\rm overlap}$ is the gate and the S/D overlap length and $L_{\rm sd}$ is the S/D length as shown in Fig. 7.

$$N_{\rm cheff} = \frac{1}{\Delta_{\rm ch}}\int\int_{\Delta_{\rm ch}} N_a(x,y)dxdy + N_{\rm sub} = \frac{A_p}{\Delta_{\rm ch}}\int_{x=-L_{\rm eff}/2}^{x=+L_{\rm eff}/2} \Gamma_{xa}(x)dx\int_{y=0}^{y=X_d} K_{ya}(y)dy + N_{\rm sub} \quad (32)$$

$\Delta_{\rm ch} = L_{\rm eff}X_d$ is the area of the channel region which is under the influence of gate. To calculate the effective doping X_d is assumed to be α_a since most of the depletion charge is confined in the region $y = 0$ to $y = \alpha_a$. Since we have considered the Gaussian approximation of the doping profile (instead of approximating it as step profile as described in [17]), we have been able to capture the effect of change in the doping profile more accurately. However, our major contribution in the subthreshold current modeling is the consideration of the quantum effect which is discussed below.

The model described in (29) and (30) considers the SCE, body effect, and narrow channel effect. However, it does not consider the effect of quantum mechanical confinement of electron/holes in the depletion/inversion region of the device. We have applied the correction due to the quantum mechanical effect to the threshold voltage expression given in (30). In scaled devices, due to high electric field at the surface (ξ_s) and high substrate doping, the quantization of inversion-layer electron energy modulates $V_{\rm th}$. Quantum-mechanical behavior of the electrons increases $V_{\rm th}$, thereby reducing the subthreshold current, since more band bending is required to populate the lowest subband, which is at a energy higher than the bottom of the conduction band (Fig. 15). When ξ_s is higher than

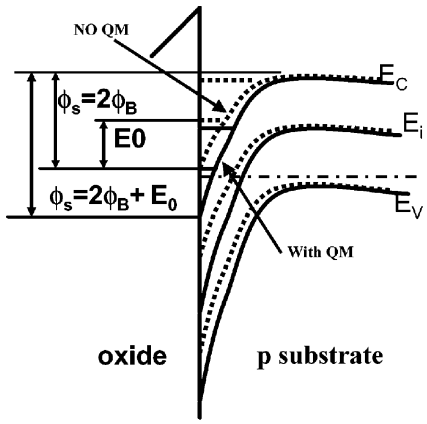


Fig. 15. Effect of electron energy quantization in the substrate on the threshold voltage of transistor.

10^6 V/cm, electrons occupy only the lowest subband. In that case, the quantization effect can be modeled as an increase in threshold voltage by an amount $\Delta V_{\rm QM}$, given by [5]

$$\Delta V_{\rm QM} = \left(1 + \frac{3t_{\rm ox}}{X_d}\right)\left(\frac{E_0}{q} - \frac{\rm kT}{q}\ln\left(\frac{8\pi q m_d \xi_s}{h^2 N_C}\right)\right) \quad (33)$$

where E_0 is the lowest subband energy given by [5] $E_o = \left[(3hq_s\xi_s/4\sqrt{2m_x})(3/4)\right]^{2/3}$, N_C is the effective conduction band density of states, m_x is the quantization effective mass of electron and m_d is the density of states effective mass of electron. It can be observed that with an increase in the surface electric field E_0 increases and hence $\Delta V_{\rm QM}$ increases. With technology scaling, first, due to higher doping and second, due to higher oxide field (scaling of oxide thickness) surface electric field increases resulting in an increase in E_0. Hence, the quantum effect becomes extremely important in nanoscaled devices. To match the simulated result, the theoretically calculated $\Delta V_{\rm QM}$ value is multiplied by an empirical factor ($\theta(V_{\rm bs}) = a_q + b_q V_{\rm bs}$). Hence, the final threshold voltage expression considering the effect of quantum correction is given by:

$$V_{\rm th} = V_{FB} + (\Phi_{s0} - \Delta\Phi_s) + \gamma\sqrt{\Phi_{s0} - V_{\rm bs}}\left(1 - \lambda\frac{X_d}{L_{\rm eff}}\right) + \Delta V_{NWE} + \Delta V_{QM}. \quad (34)$$

The simplified model shows reasonable match with the simulated result from MEDICI under substrate and drain bias variation (Fig. 16) with and without quantum correction. Substantial reduction in the subthreshold current is observed if the the quantum correction is applied.

C. Modeling Gate Direct Tunneling Current ($I_{\rm gate}$)

Gate direct tunneling current is due to the tunneling of electrons (or holes) from the bulk silicon and source/drain (S/D) overlap region through the gate oxide potential barrier into the gate [5]. The direct tunneling current density is expressed as [5]

$$J_{DT} = A_g\left(\frac{V_{\rm ox}}{T_{\rm ox}}\right)^2 \exp\left(\frac{-B_g\left(1 - \left(1 - \frac{V_{\rm ox}}{\phi_{\rm ox}}\right)^{3/2}\right)}{\frac{V_{\rm ox}}{T_{\rm ox}}}\right)$$

$$\text{where } A_g = \frac{q^3}{16\pi^2\hbar\phi_{\rm ox}} \text{ and } B_g = \frac{4\sqrt{2m^*}\phi_{ox}^{3/2}}{3\hbar q}. \quad (35)$$

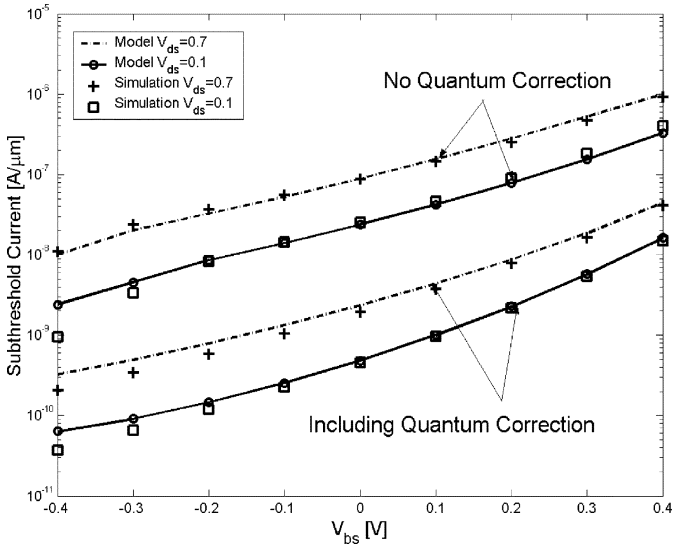


Fig. 16. Variation of subthreshold leakage with substrate bias ($V_{\rm bs}$) and drain bias ($V_{\rm ds}$) for NMOS transistor $N_{\rm ref}$.

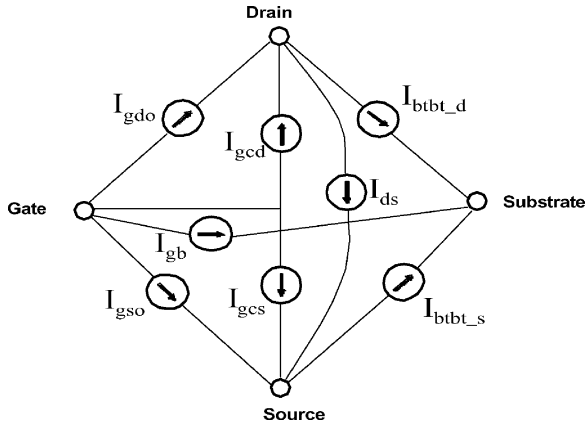


Fig. 17. SCS model of a transistor.

where $J_{\rm DT}$ is the direct tunneling current density, V_{ox} is the potential drop across the gate oxide, $\phi_{\rm ox}$ is the barrier height of the tunneling electron, m^* is the effective mass of an electron in the conduction band of silicon and $T_{\rm ox}$ is the oxide thickness. The tunneling current increases exponentially with decrease in the oxide thickness and increase in the potential drop across the oxide. Major components of gate tunneling in a scaled MOSFET device are [9], [10]: 1) gate to S/D overlap region current [edge direct tunneling (EDT)] components ($I_{\rm gso}$ & $I_{\rm gdo}$); 2) Gate to channel current ($I_{\rm gc}$), part of which goes to source ($I_{\rm gcs}$) and rest goes to drain ($I_{\rm gcd}$); and 3) Gate to substrate leakage current ($I_{\rm gb}$). We have used the model described in [9], [10] [with the effective channel and S/D doping density obtained from (31) and (32)] to model the different component of the gate current.

D. Modeling Total Leakage Current ($I_{\rm overall}$)

The overall leakage in a device is the summation of the three major leakage components. We can model the overall leakage ($I_{\rm overall}$) as

$$I_{\rm overall} = I_{\rm btbt} + I_{\rm sub} + I_{\rm gate}. \tag{36}$$

Hence, for leakage estimation we have modeled the device as a combination of voltage controlled current sources SCS model as shown in Fig. 17. Based on (22) the junction BTBT current is modeled as two current sources, one between drain and substrate (I_{btbt_d}) controlled by $V_{\rm db}$ and another between source and substrate (I_{btbt_s}) controlled by $V_{\rm sb}$. Each component of gate leakage is modeled as a current source. $I_{\rm ds}$ represents the subthreshold current. The SCS model of the transistor can be effectively used to calculate the overall leakage in a circuit. This model can also be effectively used to describe the SPICE model of a transistor.

E. Effect of Temperature on Different Components of Leakage Current

It has been discussed in Section II that different leakage components show different temperature dependence. The models of the leakage components introduced in the last three sections can be effectively used to estimate the leakage components at different operating temperatures of the device. Subthreshold current increases exponentially with temperature due to: 1) reduction in threshold voltage and 2) increase in thermal voltage (v_T) (23). The gate tunneling current is almost insensitive to temperature variation since the electric field across the oxide does not strongly depend on temperature (35). The junction BTBT current increases with temperature due to the narrowing of band gap at higher temperatures. The band gap [$E_g(T)$] at a temperature T is given by [12]

$$E_g(T) = E_g(0) - \frac{\alpha_T T^2}{(T+\beta_T)} \tag{37}$$

where $E_g(0)$ limiting value of band gap at 0 K and equal to 1.17 eV for Si. α_T and β_T are parameters with values 4.73×10^{-4} and 636, respectively, for silicon [12]. Due to the band-gap narrowing, BTBT increases with temperature (22). Fig. 18 shows the variation of each leakage component with temperature in an NMOS transistor ($L_{\rm eff} = 25$ nm). The results from the analytical model follow the numerical results very closely. It should be remembered that the analytical models does not use any temperature dependent fitting parameters, rather it tries to account for the impact of temperature on the different physical parameters (e.g., band-gap, intrinsic-carrier concentration etc.). The absence of the temperature dependent fitting parameters makes the model extremely efficient in estimating the leakage components at different temperature.

F. Effect of Variation in Doping Profile on Different Components of Leakage Current

It is observed in Section II (Fig. 3) that doping profile has a strong effect on the leakage components, particularly on subthreshold leakage and the junction tunneling leakage. With the fitting parameters extracted only for a particular doping profile, the derived analytical model for the junction BTBT shows a very close match with the simulated result under the variation in A_p, β_a, and α_a. The gate leakage is found to very insensitive to the variation in doping profile. The analytical model for the subthreshold current also follows the simulation result (both with and without the quantum correction). A better match can be obtained by multiplying the $N_{\rm cheff}$ obtained from (32) by a set of

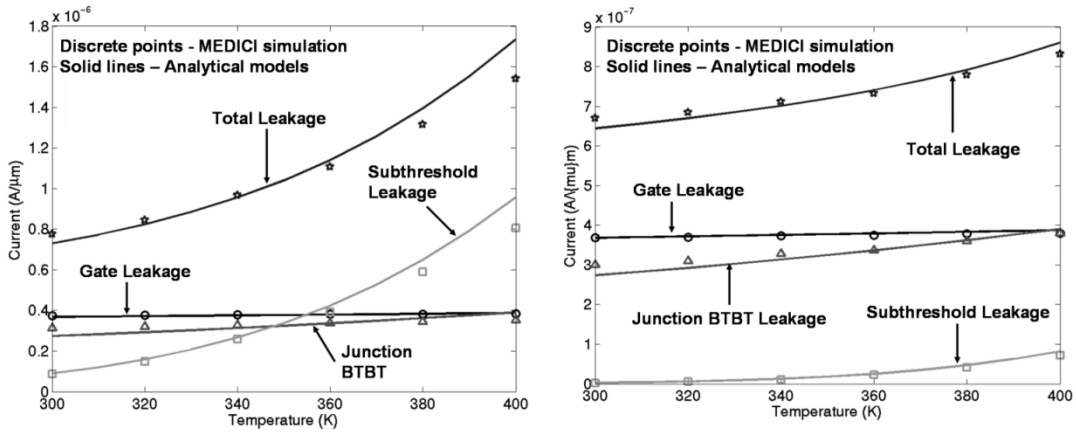


Fig. 18. Variation of different leakage component with temperature for NMOS device with $L_{\rm eff} = 25$ nm ($N_{\rm ref}$). The lines denote the results from the analytical model and the discrete points denotes the simulated value from MEDICI (a) without and (b) with quantum correction.

three linear functions which depend on A_p, β_a, and α_a as shown below

$$N_{\rm cheff}^{\rm mod} = N_{\rm cheff}\left(a_1\frac{\beta_a}{\beta_{\rm ref}}+b_1\right)\left(a_2\frac{\alpha_a}{\alpha_{\rm ref}}+b_2\right)\left(a_3\frac{A_p}{A_{\rm pref}}+b_3\right) \tag{38}$$

where $N_{\rm cheff}^{\rm mod}$ is the modified effective doping density for channel; a_1, a_2, a_3, b_1, b_2, and b_3 are empirical parameters and $\alpha_{\rm ref}, \beta_{\rm ref}$, and $A_{\rm ref}$ are the reference values α, β, and A_p. The modified effective doping density ($N_{\rm cheff}^{\rm mod}$) obtained from (38) is used to evaluate the depletion region width and threshold voltage. It can be observed from Fig. 19, that analytical models accurately describe the nature of the effect of doping variation on each of the current components. Moving the position of the lateral peak of the halo away from the center of the channel, i.e., by increasing β_a (for simplicity we kept β_a/σ_a constant) increases the junction leakage but reduces the subthreshold leakage (due to a stronger "Halo" doping) (Fig. 19a). Moving the position of the vertical peak of the doping (i.e., the position of the peak of the retrograde doping) away from the surface (i.e., increasing α_a) increases the subthreshold leakage as the doping in the channel reduces. As the position of the retrograde peak approaches the bottom junction (i.e., $\alpha_a \rightarrow Y_j$) the junction BTBT increases (marginally). However, increasing α_a further (i.e., for $\alpha_a > Y_j$ results in a reduction in the junction BTBT leakage (Fig. 19b). However, the overall sensitivity of the junction leakage to the position of the retrograde peak is low. Increasing the peak halo doping also increases the "Halo" strength thereby increasing the junction leakage and reducing the subthreshold leakage (Fig. 19c). It is observed that, the gate leakage has a very weak sensitivity to the variation in the doping profile. It should be noted that the proposed analytical model for the junction tunneling does not have any doping dependent or doping related fitting parameter, which makes it extremely useful in analyzing the impact on doping variation on the junction tunneling. Although we need a set of doping related fitting parameters for the subthreshold current, but with the introduction of these parameters the analytical model provides a excellent match with the simulated result (both with and without quantum correction). This feature makes the proposed models very useful in the leakage estimation and analysis for nanoscaled devices, since leakage in these devices is a very strong function of doping profile.

G. Impact of Substrate-Bias on Leakage Components

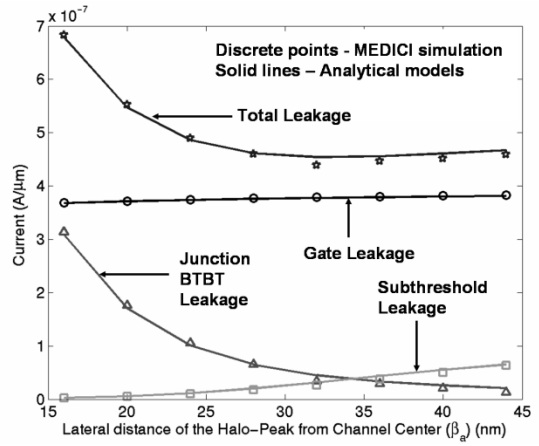
The application of reverse substrate bias is widely used for reducing subthreshold leakage in submicron devices. It is observed in Fig. 16 that application of reverse substrate bias reduces the subthreshold current. However, it is observed in Fig. 13 that application of a reverse substrate bias exponentially increases the junction tunneling leakage. The gate leakage is almost insensitive to substrate-bias variation. Application of substrate bias only affects the gate-to-body tunneling component of gate leakage. However, this component is much less compared to the gate-to-channel or overlap tunneling component of gate leakage. Hence, the net effect of substrate bias on the gate leakage is negligible. Fig. 20 shows the impact of substrate bias on individual leakage component and the total leakage for an NMOS ($N_{\rm ref}$) transistor of $L_{\rm eff} = 25$ nm for different doping. The analytical models closely match the simulation result. From Fig. 20. it is observed that the optimum substrate bias (the substrate bias at which the total leakage is minimum) depends on the relative magnitude of the subthreshold and junction tunneling leakage and the optima occurs at a value where the two current are almost equal.

V. Modeling of Leakage in Logic Gates

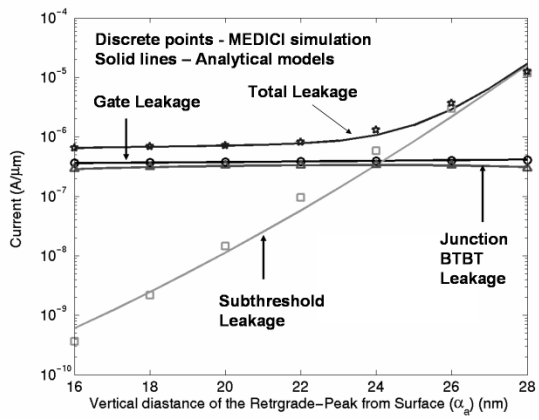
The SCS model of the transistor can be effectively used to calculate the overall leakage in a circuit. Fig. 21 shows the circuit containing two series connected NMOS transistors and the equivalent SCS model. To calculate the overall leakage, we have to solve the KCL at the intermediate node INT. From Fig. 21, the node equation at INT is given by

$$I_{\rm ds1} + I_{\rm gcs1} + I_{\rm gso1} - I_{BTBT\_s1} = I_{\rm ds2} - I_{\rm gdo2} - I_{\rm gcd2} + I_{BTBT\_d2}. \tag{39}$$

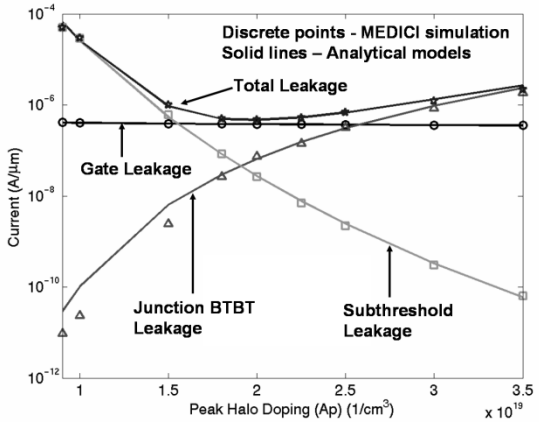
In circuits involving more than one such node, we will have a set of simultaneous equations that need to be solved. The overall leakage in the circuit can be defined as the sum of all currents



(a)



(b)



(c)

Fig. 19. Effect of variation in doping profile on different leakage component for NMOS device of $L_{\rm eff} = 25$ nm. Continuous lines denotes the results from analytical model and the discrete points denotes the MEDICI simulation result. (a) Variation in Lateral distance of Halo-peak from channel center (β_a), (b) Variation in vertical distance of retrograde-peak from surface (α_a), and (c) Variation in peak halo doping (A_p).

collected at the ground node. Hence, the overall leakage in a CMOS circuit consists of the gate currents of the PMOS and NMOS transistors for which the gate is at "0," junction BTBT currents of the NMOS transistors (assuming $V_{\rm bulk} = 0$ for NMOS and $V_{\rm bulk} = V_{\rm dd}$ for PMOS) and the current through the source of all NMOS for which the source is connected to ground. Thus, the overall leakage ($I_{\rm leakage}$) can be given by

$$I_{\rm leakage} = \sum_{\substack{\rm NMOS+PMOS \\ V_{\rm gk}="0"}} I_{\rm gk} + \sum_{\rm NMOS} I_{\rm BTBTk} + \sum_{\substack{\rm NMOS\ with\ source \\ \rm connected\ to\ ground}} I_{\rm sourcek} \quad (40)$$

where $I_{\rm gk}$ is the gate current , $I_{\rm BTBTk}$ is the junction BTBT current and $I_{\rm sourcek}$ is the source current through the k-th transistor. $I_{\rm sourcek}$ can be given by

$$I_{\rm sourcek} = \begin{cases} I_{\rm dsk}, & \text{if } V_{\rm gk} = 0 \\ I_{\rm dsk} + I_{\rm gsok} + I_{\rm gcsk}, & \text{if } V_{\rm gk} = 1 \end{cases}. \quad (41)$$

$I_{\rm gk}$ is the sum of the components of the gate currents in the kth transistor and is given by

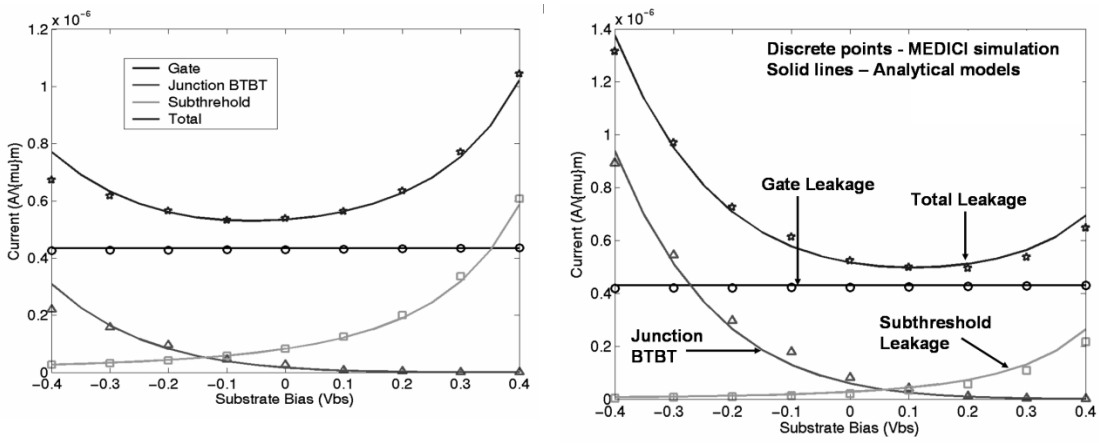
$$I_{\rm gk} = |I_{\rm gsok} + I_{\rm gdok} + I_{\rm gcsk} + I_{\rm gcdk} + I_{\rm gbk}|. \quad (42)$$

A numerical equation solver (SCS solver) is written in MATLAB to solve the set of simultaneous equations in a circuit and to determine the overall leakage under a specific input condition. Fig. 22 shows the comparison of the evaluated result and simulated result in MEDICI for a stack of 2 NMOS transistors ($N_{\rm ref}$), at normal temperature (without quantum correction). The evaluated results match the simulated results closely. However, instead of using the SCS solver in MATLAB, one can describe the circuit in SPICE and replace each transistor in the circuit by the equivalent SCS model. In that case, SPICE can be used as the nonlinear equation solver.

SCS solver can be used to evaluate the leakage components of basic gates. Figs. 23 and 24 show the different leakage components of INVERTER, NAND, and NOR gates (designed with $N_{\rm ref}$ and $P_{\rm ref}$) at normal ($T = 300$ K) and high temperature ($T = 400$ K) (with and without the quantum correction). It is observed that the overall leakage increases considerably with the temperature. At normal temperatures gate leakage dominates the subthreshold leakage and the junction BTBT leakage, whereas later two are high at higher temperatures. Also, application of the quantum correction reduces the subthreshold current considerably. Fig. 25 shows the leakage of the two-input NAND gate at different technology generations, namely, $L_{\rm eff} = 25$ nm, $L_{\rm eff} = 50$ nm and $L_{\rm eff} = 90$ nm. With technology scaling, the gate and the junction BTBT leakage becomes the dominant leakage components. The solver can easily be extended to handle other logic gates.

A. Estimation of Leakage in a Transistor Stack and the Stacking Effect

Turning "off" more than one transistor in a stack of transistors forces the intermediate node (say INT in Fig. 21) voltage to go to a value higher than zero [1], [7]. This causes a negative $V_{\rm gs}$, negative $V_{\rm bs}$ (more body effect) and reduced $V_{\rm ds}$ (less DIBL) in the top transistor, thereby reducing the subthreshold current flowing through the stack [1], [7]. This effect, known as the "stacking effect," has been used to reduce the subthreshold leakage in logic circuits in stand-by mode [1], [7]. The estimation tool described here, effectively models stacking effect for



(a) (b)

Fig. 20. Impact of substrate bias on different leakage component for devices with different doping. (a) weaker halo $A_p = 1.8\mathrm{e}19$, (b) stronger halo $A_p = 2\mathrm{e}19$. The optimum substrate bias depends on the relative strength of junction BTBT and subthreshold current.

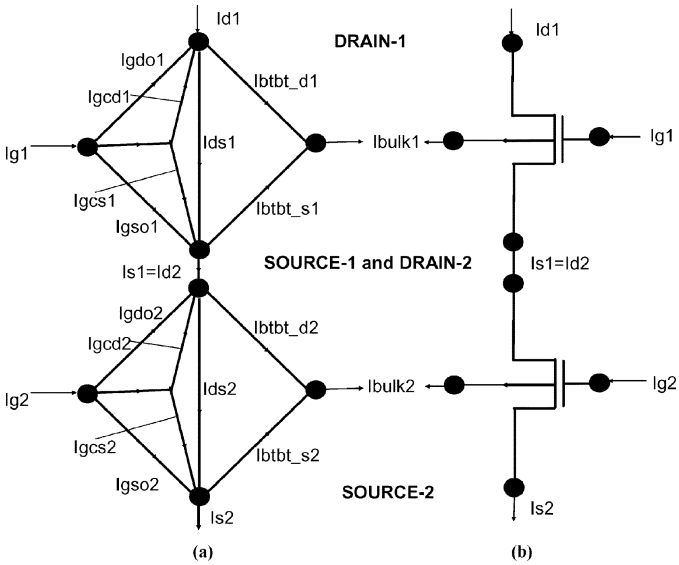


Fig. 21. Circuit configuration with SCS model for a two-transistor stack. (a) SCS model and (b) transistor-circuit diagram.

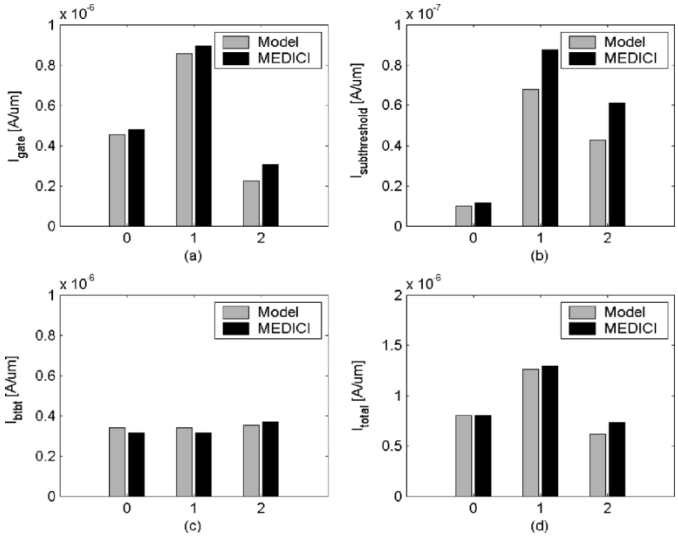


Fig. 22. Comparison of simulator (MEDICI) and model current values for a 2-transitor stack ($L_{\mathrm{eff}} = 25$ nm) for different input vectors: (a) gate, (b) subthreshold, (c) BTBT, and (d) total leakage.

the subthreshold, the gate and the BTBT leakage. Using (40), the total leakage in a stack of N-transistor can be modeled as

$$I_{\mathrm{stack}} = \sum_{\substack{k=1,\\ V_{\mathrm{gk}}='0'}}^{N} I_{\mathrm{gk}} + \sum_{k=1}^{N} I_{\mathrm{BTBTk}} + I_{\mathrm{sourceN}} \tag{43}$$

where I_{sourceN} is the source current through the Nth transistor. Using (43) the total leakage in a transistor stack can be estimated at different input vector conditions. Fig. 22 shows that the input "00" (turning "off" both transistors) produces the minimum subthreshold and BTBT leakage [BTBT leakage in fact does not depend much on stacking (Fig. 22)], however, "10" produces the minimum gate leakage condition. Hence, the input condition that minimizes the total leakage depends on the relative magnitude of the different components. In devices where the gate leakage is the dominant component, the input "10" minimizes the total leakage in a stack of two NMOS transistors as shown in Fig. 22 [3], [18].

VI. Estimation of Total Circuit Leakage

Evaluation of the leakage components of basic logic gates is used to estimate the total leakage in a gate level logic circuit. To evaluate the different leakage components in a logic circuit we have modified the leakage estimation tool described in [7]. Leakage of a logic circuit depends on primary input vectors. The primary input vectors are propagated by simulating the circuit level by level. The subthreshold (I_{Tsub}), the gate (I_{Tgate}) and the junction BTBT (I_{Tbtbt}) leakage and the overall leakage (I_{Toverall}) through the circuit is defined as the sum of the leakage

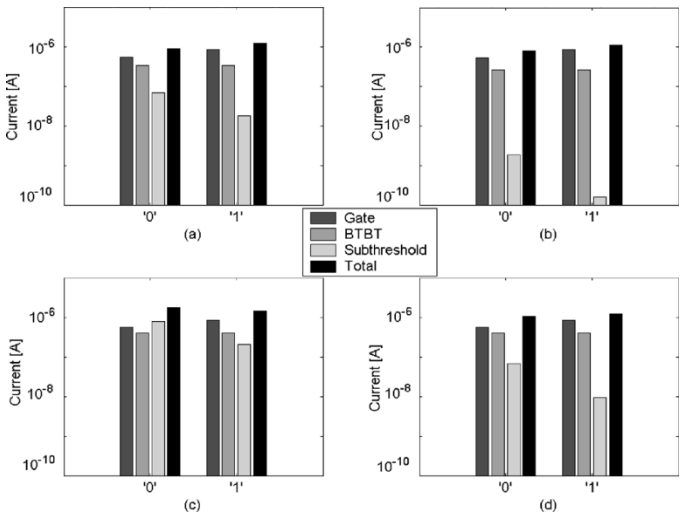


Fig. 23. Leakage of an INVERTER with input "0" and "1." (a) $T = 300$ K and no quantum correction. (b) $T = 300$ K and with quantum correction. (c) $T = 400$ K and no quantum correction. (d) $T = 400$ K and with quantum correction ($L_{\rm eff} = 25$ nm and $V_{\rm dd} = 0.7$ V).

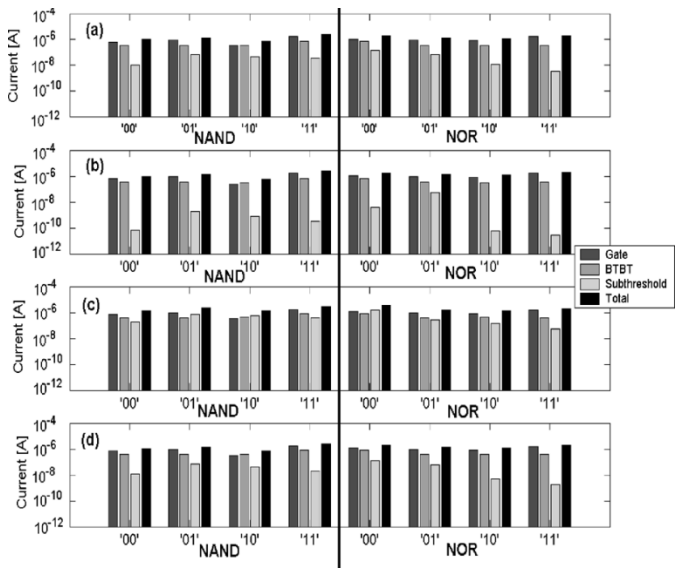


Fig. 24. Leakage of a two-input NAND and NOR gate with different input. (a) $T = 300$ K and no quantum correction. (b) $T = 300$ K and with quantum correction. (c) $T = 400$ K and no quantum correction. (d) $T = 400$ K and with quantum correction. $L_{\rm eff} = 25$ nm and $V_{\rm dd} = 0.7$ V).

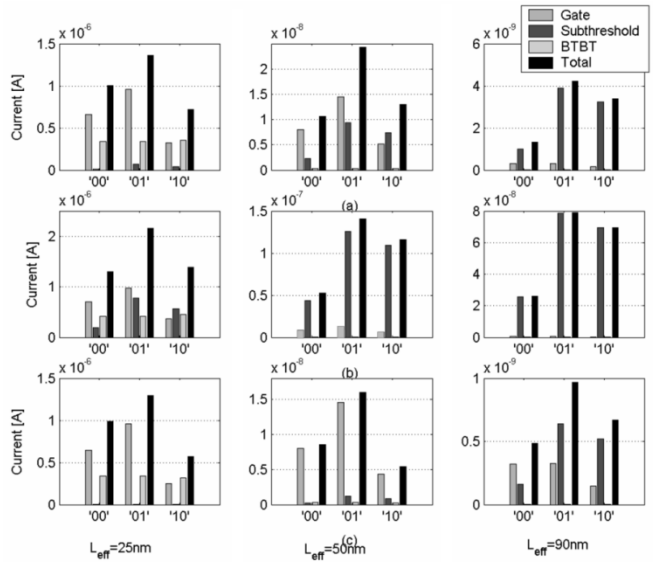


Fig. 25. Variation of leakage in a two-input NAND gate with input vector. (a) $T = 300$ K and no quantum correction. (b) $T = 400$ K and no quantum correction. (c) $T = 300$ K and including quantum correction. ($V_{\rm dd} = 0.7$ V at 25 nm, $V_{\rm dd} = 0.9$ V at 50 nm, $V_{\rm dd} = 1.2$ V at 90 nm,).

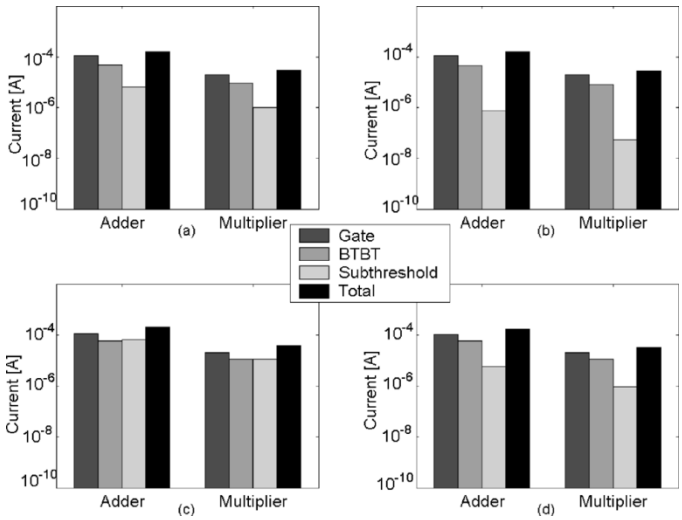


Fig. 26. Average leakage of an 8-bit adder and a 2-bit array multiplier. (a) $T = 300$ K and no quantum correction. (b) $T = 300$ K and with quantum correction. (c) $T = 400$ K and no quantum correction. (d) $T = 400$ K and with quantum correction. ($L_{\rm eff} = 25$ nm and $V_{\rm DD} = 0.7$ V).

through each of the basic gates present in the circuit and is given by

$$I_{\rm Tsub} = \sum_{k=\text{all gate}} I_{\rm ksub}, \quad I_{\rm Tgate} = \sum_{k=\text{all gate}} I_{\rm kgate}$$
$$I_{\rm Tbtbt} = \sum_{k=\text{all gate}} I_{\rm kbtbt}$$
$$I_{\rm Toverall} = \sum_{k=\text{all gate}} I_{\rm koverall} = I_{\rm Tsub} + I_{\rm Tgate} + I_{\rm Tbtbt}. \tag{44}$$

The summation of the leakage of the individual gate gives an *approximate* estimation of the total leakage of the logic circuit. The leakage of a gate can be modified because of its loading by other logic gates (i.e., because of the presence of the fanouts). The estimation tool described in [7] does not consider this effect. However, an exact value of the leakage can be obtained by solving the full circuit using a transistor level circuit simulator like SPICE and representing each transistor using the described SCS model.

A. Results

The leakage estimation tool is used to estimate the total leakage in complex logic circuits, under different primary input vectors. Fig. 26 shows the different leakage components along with the total leakage of an 8-bit ripple carry adder and a 2-bit array multiplier circuit (designed using NAND, NOR, and INVERTER gate with 25-nm NMOS ($N_{\rm ref}$) and PMOS ($P_{\rm ref}$)

transistors) averaged over a large number of primary input vectors. The average leakage ($I_{\rm avg}$) is defined as

$$I_{\rm avg} = \frac{\sum_{i=1}^{N} I_{\rm Toverall}({\rm IN}_i)}{N} \tag{45}$$

where, $I_{\rm Toverall}({\rm IN}_i)$ is the overall leakage due to input IN_i and N is the total number of applied input vectors. The leakage is evaluated at both normal ($T = 300$ K) and high ($T = 400$ K) temperatures and with and without quantum correction. The result shows that on the average gate leakage is the dominant component of the total leakage. However, at higher temperature the contributions of the subthreshold and the junction BTBT are increased.

VII. Summary and Conclusion

In this paper, we have developed a compact model for the total leakage in a transistor as the summation of subthreshold, junction BTBT and gate leakage. We have developed a circuit level analytical model (SPICE compatible) for the junction BTBT leakage in bulk-MOSFET considering 2-D doping profile. The model shows a very good match with the results from MEDICI device simulator for both PMOS and NMOS. We have also modeled and studied the impact of the quantum correction on the subthreshold leakage. The models have been effectively used to study the behavior of different leakage components under temperature, doping profile and substrate bias variation. It has been shown that for leakage estimation the transistor can be modeled as a combination of voltage controlled current sources (SCS model), where, each current source describes a leakage mechanism. SCS model can be used to describe a transistor in SPICE circuit simulator. We have developed a computer-aided design tool to estimate the total leakage in CMOS circuits based on the SCS model. The described method for leakage estimation is based on the knowledge of the transistor geometry, 2-D doping profile and operating temperature. The physical nature of the models makes them extremely powerful in: 1) the study of the effect of substrate bias variation and estimation of optimal substrate bias; 2) the study of the effect of the doping profile variation and optimization of doping profile to not only minimize leakage in a device but more importantly leakage in a logic gate and circuit; and 3) the estimation of the impact of variation in different process parameters on the leakage of a device and circuits. The proposed model can be very effectively used to accurately estimate leakage in nanoscaled CMOS logic circuits.

Appendix

A. Estimation of Location of Junction in X and Y Direction (i.e., X_j and Y_j)

The X_j and Y_j can be estimated solving the following equations:

$$N_{\rm sd}(X_j, y = 0) = N_a(X_j, y = 0)$$
$$N_{\rm sd}(x = x_{\max}, Y_j) = N_a(x_{\max}, Y_j).$$

Hence, X_j can be obtained as follows:

$$\begin{aligned} N_{\rm sd}(X_j, 0) &= A_{\rm sd}\Gamma_{\rm xsd}(X_j)K_{\rm ysd}(0) \\ &= A_{\rm sd}\exp\left(\frac{-(X_j-\beta_{\rm sd})^2}{\sigma_{\rm xsd}^2}\right) \\ N_a(X_j, 0) &= A_p\Gamma_{xa}(X_j)K_{ya}(0) \\ &= A_p\exp\left(\frac{-(X_j-\beta_a)^2}{\sigma_{xa}^2}\right)\exp\left(-\frac{\alpha_a^2}{\sigma_{ya}^2}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} & A_{\rm sd}\exp\left(\frac{-(X_j-\beta_{\rm sd})^2}{\sigma_{\rm xsd}^2}\right) \\ &= A_p\exp\left(\frac{-(X_j-\beta_a)^2}{\sigma_{xa}^2}\right)\exp\left(-\frac{\alpha_a^2}{\sigma_{ya}^2}\right) \\ {\rm or,} & \frac{(X_j-\beta_a)^2}{\sigma_{xa}^2} - \frac{(X_j-\beta_{\rm sd})^2}{\sigma_{\rm xsd}^2} \\ &= \ln\left(\frac{A_p}{A_{\rm sd}}\right) - \frac{\alpha_a^2}{\sigma_{ya}^2}. \end{aligned}$$

The above equation is a quadratic in X_j which can be solved analytically.

Similarly Y_j can also be solved as follows.

$$\begin{aligned} N_{\rm sd}(x_{\max}, Y_j) &= A_{\rm sd}\Gamma_{\rm xsd}(x_{\max})K_{\rm ysd}(Y_j) \\ &= A_{\rm sd}\Gamma_{\rm xsd}(x_{\max})\exp\left(\frac{-(Y_j)^2}{\sigma_{\rm ysd}^2}\right) \\ N_a(x_{\max}, Y_j) &= A_p\Gamma_{xa}(x_{\max})K_{ya}(Y_j) \\ &= A_p\Gamma_{xa}(x_{\max})\exp\left(\frac{-(Y_j-\alpha_a)^2}{\sigma_{ya}^2}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} & A_{\rm sd}\Gamma_{\rm xsd}(x_{\max})\exp\left(\frac{-(Y_j)^2}{\sigma_{\rm ysd}^2}\right) \\ &= A_p\Gamma_{xa}(x_{\max})\exp\left(\frac{-(Y_j-\alpha_a)^2}{\sigma_{ya}^2}\right) \\ {\rm or,} & \frac{(Y_j-\alpha_a)^2}{\sigma_{ya}^2} - \frac{Y_j^2}{\sigma_{\rm ysd}^2} \\ &= \ln\left(\frac{A_p\Gamma_{xa}(x_{\max})}{A_{\rm sd}\Gamma_{\rm xsd}(x_{\max})}\right). \end{aligned}$$

The above equation is a quadratic in Y_j which can be solved analytically.

B. Evaluation of the Effective Junction Doping

The $N_{\rm aside}$ in (20) can be obtained using the following procedure:

$$\begin{aligned} N_{\rm aside} &= \frac{1}{|y_2-y_1|}\int_{y_1}^{y_2} N_{\rm aside}(X_j, y)dy \\ &= \frac{1}{|y_2-y_1|}\int_{y_1}^{y_2} A_p\Gamma_{xa}(X_j)K_{ay}(y)dy \\ &= \frac{A_p\Gamma_{xa}(X_j)}{|y_2-y_1|}\int_{y_1}^{y_2}\exp\left(\frac{-(y-\alpha_a)^2}{\sigma_{ay}^2}\right)dy \end{aligned}$$

using

$$t = \frac{(y-\alpha_a)}{\sigma_{ay}}, \text{we get}$$
$$= \frac{A_p \Gamma_{xa}(X_j)}{|y_2 - y_1|}\sigma_{ay}\int_{t_1}^{t_2} \exp\left(-t^2\right) dt$$

where

$$t_1 = \frac{(y_1-\alpha_a)}{\sigma_{ay}} \text{ and } t_2 = \frac{(y_2-\alpha_a)}{\sigma_{ay}}.$$

The last integration is in the form of error function and can be evaluated as

$$I = \int_{t_1}^{t_2} \exp\left(-t^2\right) dt$$

Case-1:

$$t_1 \text{ and } t_2 > 0$$
$$I = \left[\int_0^{t_2} \exp\left(-t^2\right) dt - \int_0^{t_1} \exp\left(-t^2\right) dt\right]$$
$$= \frac{\sqrt{\pi}}{2}\left(erf(t_2) - erf(t_1)\right)$$

where

$$erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp\left(-t^2\right) dt$$

Case-2:

$$t_1 < 0 \text{ (say } t_1 = -t_0) \text{ and } t_2 > 0$$
$$I = \left[\int_0^{t_2} \exp\left(-t^2\right) dt + \int_0^{t_0} \exp\left(-t^2\right) dt\right]$$
$$= \frac{\sqrt{\pi}}{2}\left(erf(t_2) + erf(t_0)\right)$$
$$= \frac{\sqrt{\pi}}{2}\left(erf(t_2) + erf(-t_1)\right)$$

Case-3:

$t_1 > 0$ and $t_2 < 0$ (say $t_2 = -t_0$)
can be modified into Case-2. However, this case will not occur since we select $t1 < t2$ while forming the expression for I

Case-4:

$$t_1 < 0 \text{ (say } t_1 = -t_a)$$
$$\text{and } t_2 < 0 \text{ (say } t_1 = -t_b)$$
$$I = \int_{-t_a}^{-t_b} \exp\left(-t^2\right) dt = \int_{t_b}^{t_a} \exp\left(-t^2\right) dt$$
$$= \left[\int_0^{t_a} \exp\left(-t^2\right) dt - \int_0^{t_b} \exp\left(-t^2\right) dt\right]$$
$$= \frac{\sqrt{\pi}}{2}\left(erf(t_a) - erf(t_b)\right)$$
$$= \frac{\sqrt{\pi}}{2}\left(erf(-t_1) - erf(-t_2)\right).$$

C. Estimation of the Effective Source/Drain and Channel Doping for Subthreshold Current Computation

$$N_{\text{sdeff}} = \frac{1}{\Delta_{\text{sd}}}\int\int_{\Delta_{\text{sd}}} N_{\text{sd}}(x,y)dxdy$$
$$= \frac{A_{\text{sd}}}{\Delta_{\text{sd}}}\int_{x-X_j}^{x=L_{\text{gate}}/2+L_{\text{sd}}} \Gamma_{\text{xsd}}(x)dx \int_{y=0}^{y=Y_j} K_{\text{ysd}}(y)dy$$
$$= \frac{A_{\text{sd}}}{\Delta_{\text{sd}}}\left[\int_{x-X_j}^{x=\beta_{\text{sd}}} \exp\left(\frac{-(x-\beta_{\text{sd}})^2}{\sigma_{\text{xsd}}^2}\right) dx + \int_{x=\beta_{\text{sd}}}^{x=L_{gate}/2+L_{\text{sd}}} 1dx\right]\int_{y=0}^{y=Y_j} \exp\left(\frac{-(y)^2}{\sigma_{\text{ysd}}^2}\right) dy$$

$$\text{Let } t_x = \frac{(x-\beta_{\text{sd}})}{\sigma_{\text{xsd}}} \Rightarrow t_{x1} = \frac{(X_j-\beta_{\text{sd}})}{\sigma_{\text{xsd}}}, t_{x2} = 0;$$
$$t_y = \frac{y}{\sigma_{\text{ysd}}} \Rightarrow ty1 = 0,\ ty2 = \frac{Y_j}{\sigma_{\text{ysd}}}$$
$$N_{\text{sdeff}} = \frac{A_{\text{sd}}}{\Delta_{\text{sd}}}\left[\frac{1}{\sigma_{\text{xsd}}}\int_{t_{x1}}^{t_{x2}} \exp\left(-t_x^2\right) dt_x + \left(\frac{L_{gate}}{2} + L_{\text{sd}} - \beta_{\text{sd}}\right)\right]\int_{y=0}^{y=Y_j} \exp\left(-t_y^2\right) dt_y.$$

The above integrals are evaluated using the method described in the previous section. A similar approach can be used to estimate the effective channel doping.

REFERENCES

[1] K. Roy, S. Mukhopadhyay, and H. Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicron CMOS circuit," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, Feb. 2003.

[2] A. Keshavarzi, K. Roy, and C. F. Hawkins, "Intrinsic leakage in deep submicron CMOS ICs-measurement-based test solutions," *IEEE Trans. VLSI Sysyt.*, vol. 8, no. 6, pp. 717–723, Dec. 2000.

[3] S. Mukhopadhyay, C. Neau, T. R. Cakici, A. Agarwal, C. H. Kim, and K. Roy, "Gate leakage reduction for scaled devices using transistor stacking," *IEEE Trans. VLSI Sysyt.*, vol. 11, no. 4, pp. 716–730, Aug. 2003.

[4] "Well-Tempered" Bulk-Si NMOSFET Device Home Page. Microsystems Technol. Lab., Mass. Inst. Technol., Cambridge. [Online]. Available: http://www-mtl.mit.edu/Well/

[5] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. New York: Cambridge Univ. Press, 1998.

[6] W. Shiue, "Leakage power estimation and minimization in VLSI circuits," in *Proc. ISCAS*, 2001, pp. 178–181.

[7] Z. Chen, M. Johnson, L. Wei, and K. Roy, "Estimation of standby leakage power in CMOS circuits considering accurate modeling of transistor stacks," in *Proc. ISLPED*, 1998, pp. 239–244.

[8] R. M. Rao, J. L. Burns, A. Devgan, and R. B. Brown, "Efficient techniques for gate leakage estimation," in *Proc. ISLPED*, 2003, pp. 100–103.

[9] K. Cao, W.-C. Lee, W. Liu, X. Jin, P. Su, S. K. H. Fung, J. X. An, B. Yu, and C. Hu, "BSIM4 gate leakage model including source drain partition," in *Tech. Dig. IEDM*, 2000, pp. 815–818.

[10] BSIM4.2.1 MOSFET Model. BSIM Group, Univ. California, Berkeley. [Online]. Available: http://www-device.eecs.berkeley.edu/~bsim3/

[11] *MEDICI: Two-Dimensional Semiconductor Device Simulation Program*, AVANT! Corp., Fremont, CA, 2000.

[12] R. Pierret, *Advanced Semiconductor Fundamentals*. Reading, MA: Addison-Wesley, 1989, vol. VI, Modular Series on Solid States Devices.

[13] Z. Lee, M. B. McIlrath, and D. A. Antoniadis, "Two-dimensional doping profile characterization of MOSFET's by inverse modeling using characteristics in the subthreshold region," *IEEE Trans. Electron. Devices*, vol. 46, no. 8, pp. 1640–1649, Aug. 1999.

[14] R. Pierret, *Semiconductor Device Fundamentals*. Reading, MA: Addison-Wesley, 1996.

[15] Z. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFET's," *IEEE Trans. Electron. Devices*, vol. 40, no. 1, pp. 86–95, Jan. 1993.

[16] X. Zhou, K. Y. Lim, and D. Lim, "A general approach to compact threshold voltage formulation based on 2-D numerical simulation and experimental correlation for deep-submicron ULSI technology development," *IEEE Trans. Electron. Devices*, vol. 47, no. 1, pp. 214–221, Jan. 2000.

[17] D. Fotty, *MOSFET Modeling with SPICE*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997.

[18] S. Mukhopadhyay and K. Roy, "Accurate modeling of transistor stacks to effectively reduce total standby leakage in nano-scale CMOS circuits," in *Dig. Tech. Papers Symp. VLSI Circuits*, 2003, pp. 53–56.

Saibal Mukhopadhyay (S'99) was born in Calcutta, India. He received the B.E. degree in electronics and telecommunication electrical engineering from Jadavpur University, Calcutta, India, in 2000. He is currently pursuing the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN.

He was an intern with the IBM T. J. Watson Research Lab, Yorktown Heights, NY, in the summer of 2003 and 2004, in the High Performance Circuit Design Department. His research interests include analysis and design of low-power and robust circuits using nanoscaled CMOS and circuit design using double gate transistors.

Mr. Mukhopadhyay received the IBM Ph.D. Fellowship Award for 2004–2005. He received the Best Paper Award at the 2004 International Conference on Computer Design.

Arijit Raychowdhury (S'00) received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta, India, in 2001. He is currently pursuing the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN.

He has worked as an Analog Circuit Designer with Texas Instruments India. His research interests include device/circuit design for scaled silicon and nonsilicon devices.

Mr. Raychowdhury has received academic excellence awards in 1997, 2000 and 2001 and the Messner Fellowship from Purdue University in 2002. He received the "Best Student Paper" Award at the 2003 IEEE Nanotechnology Conference .

Kaushik Roy (S'83–M'83–SM'95–F'02) received the B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana–Champaign, in 1990.

He was with the Semiconductor Process and Design Center, Texas Instruments, Dallas, TX, where he worked on FPGA architecture development and low-power circuit design. He joined the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor. He has published more than 200 papers in refereed journals and conferences, holds five patents, and is a coauthor of a book on Low Power CMOS VLSI Design (New York: Wiley, 2000). His research interests include very large scale integration design/computer-aided design with particular emphasis on low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing.

Prof. Roy received the National Science Foundation Career Development Award in 1995, the IBM Faculty Partnership Award, the AT&T/Lucent Foundation Award, and Best Paper Awards at the 1997 International Test Conference and 2000 International Symposium on Quality of IC Design. He is on the editorial board of the IEEE DESIGN AND TEST, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and the IEEE TRANSACTIONS ON VLSI SYSTEMS. He was the Guest Editor for the Special Issue on Low-Power VLSI in the IEEE in 199) and the IEEE TRANSACTIONS ON VLSI SYSTEMS in June 2000.