

1-1-2003

Modeling and Estimation of Total Leakage in Scaled CMOS Logic Circuits

Saibal Mukhopadhyay

Arijit Raychowdhury

Kaushik Roy

Follow this and additional works at: <http://docs.lib.purdue.edu/ecetr>

Mukhopadhyay, Saibal ; Raychowdhury, Arijit ; and Roy, Kaushik, "Modeling and Estimation of Total Leakage in Scaled CMOS Logic Circuits" (2003). *ECE Technical Reports*. Paper 133.
<http://docs.lib.purdue.edu/ecetr/133>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Modeling and Estimation of Total Leakage in Scaled CMOS Logic Circuits

By

Saibal Mukhopadhyay,
Arijit Raychowdhury,
and
Kaushik Roy.

School of Electrical and Computer Engineering,
1285 Electrical Engineering Building,
Purdue University,
West Lafayette, IN 47907-1285.

This research is supported in part by SRC, GSRC, Intel and IBM Corporations.

List of Figures

Figure 1: Variation of different leakage components with (a) technology generation and oxide thickness; and (b) doping profile. “Doping-1” has a different halo profile than “Doping-2”

Figure 2: Leakage estimation steps.

Figure 3: Architecture of the device.

Figure. 4: 2-D Gaussian profile for (a) channel and (b) source-drain region.

Figure 5: Physical picture of valence band electron tunneling in a reversed bias p-n junction.

Figure. 6: “rectangular junction” approximation.

Figure 7. “step junction” approximation.

Figure 8: Variation of BTBT current with substrate bias. (a) Comparison of analytical result with simulated data from MEDICI for N MOS transistor with $L_{\text{eff}} = 25\text{nm}$ and doping profile: $\alpha_a = 0.018\mu\text{m}$, $\sigma_{ay} = 0.016\mu\text{m}$ $\beta_a = 0.016\mu\text{m}$, $\sigma_{ax} = 0.020\mu\text{m}$. (b) Variation of error.

Figure 9: Variation of BTBT current with substrate bias for different devices.

Figure 10: Variation of subthreshold leakage with substrate bias (V_{bs}) and drain bias (V_{ds}) for NMOS transistor N_{ref} (a) Without and (b) With quantum correction.

Figure 11: Variation of leakage components with temperature.

Figure 12: Sum of Current Source model of a transistor.

Figure. 13: Circuit configuration with SCS model for a 2-transistor stack, (a) SCS model, (b) transistor-circuit diagram.

Figure 14: Comparison of simulator (MEDICI) and model current values for a 2-transistor stack for different input vectors: (a) gate, (b) subthreshold, (c) BTBT and (d) total leakage.

Figure 15: Leakage of an INVERTER with input ‘0’ and ‘1’. (a) $T=300\text{K}$ and no quantum correction, (b) $T=300\text{K}$ and with quantum correction, (c) $T=400\text{K}$ and no quantum correction, (d) $T=400\text{K}$ and with quantum correction

Figure 16: Leakage of a 2-input NAND and NOR gate with different input. (a) $T=300\text{K}$ and no quantum correction, (b) $T=300\text{K}$ and with quantum correction, (c) $T=400\text{K}$ and no quantum correction, (d) $T=400\text{K}$ and with quantum correction.

Figure. 17: Illustration of loading effect of an inverter.

Figure 18: Percentage change in the (a) output voltage and (b) leakage components in an inverter due to loading.

Figure 19: Average leakage of an 8-bit adder and a 2-bit array multiplier. (a) $T=300\text{K}$ and no quantum correction, (b) $T=300\text{K}$ and with quantum correction, (c) $T=400\text{K}$ and no quantum correction, (d) $T=400\text{K}$ and with quantum correction.

ABSTRACT

Dramatic increase of subthreshold, gate and reverse biased junction band-to-band-tunneling (BTBT) leakage in scaled devices, result 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 scaled devices has been developed. Current models have been developed based on the exact device geometry, 2-D doping profile and operating temperature. A circuit level model of junction BTBT leakage (which is unprecedented) has been developed. Simple models of the subthreshold current and the gate current have been presented. Here, for the first time, 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 25nm effective length) at the room and at the elevated temperatures.

ABSTRACT

Dramatic increase of subthreshold, gate and reverse biased junction band-to-band-tunneling (BTBT) leakage in scaled devices, result 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 scaled devices has been developed. Current models have been developed based on the exact device geometry, 2-D doping profile and operating temperature. A circuit level model of junction BTBT leakage (which is unprecedented) has been developed. Simple models of the subthreshold current and the gate current have been presented. Here, for the first time, 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 25nm effective length) at the room and at the elevated temperatures.

1. 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 drastically with technology scaling [1]. 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]. The threshold voltage (V_{th}) scaling and the V_{th} reduction due to Short Channel Effects (SCE) [1], result in an exponential increase the subthreshold current. The oxide thickness scaling, required to maintain reasonable SCE immunity, results in a considerable direct tunneling current through the gate insulator of the transistor [1], [2]. In scaled devices, the higher substrate doping density and the application of the “halo” profiles (used to reduce SCE) [2] cause significantly large BTBT current through the reverse biased drain-substrate and source-substrate junctions. In the small devices each of the different leakage components increases resulting in a dramatic increase of the overall leakage. The magnitudes of each of these components depend strongly on the device geometry (namely, channel length, oxide thickness and transistor width) and the doping profiles as shown in Fig. 1.

Different leakage current components in the devices vary differently with varying temperature. Subthreshold and

BTBT leakage show a strong dependence of temperature, whereas gate leakage is relatively insensitive to temperature variations. Since digital VLSI circuits usually operate at elevated temperatures, estimation of the various leakage components and the total leakage in devices and circuits is

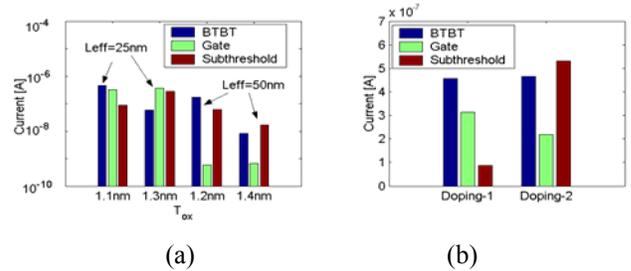


Figure 1: Variation of different leakage components with (a) technology generation and oxide thickness; and (b) doping profile. “Doping-1” has a different halo profile than “Doping-2”

necessary both at room and elevated temperatures.

In this paper we have developed a methodology for accurately estimating the total leakage of a logic circuit for different primary input vectors, based on the knowledge of, (a) the device geometry, (b) the exact 2-D doping profile of the device and (c) the operating temperature. Although, a number of previous work are reported on the estimation of leakage in logic circuits [3], [3], [4] but they have only considered the subthreshold leakage. However, as shown in Fig. 1, gate and BTBT leakage are also becoming extremely important and thus cannot be neglected for estimation of total leakage. We have developed a compact circuit level model of BTBT leakage in a MOSFET with halo [2] and retrograde doping [2]. To the best of our knowledge it is unprecedented. A simple and reasonably accurate model of the subthreshold current has been developed based on the exact 2-D doping profile. Here, for the first time, we have evaluated the direct impact of quantization of the electron energy in the substrate [2], on the leakage in logic circuits. We have used the gate leakage model presented in [5], [6]. Finally, 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. A method for calculating the total leakage of a logic circuit by adding the individual leakage contribution of its constituent gates is also proposed. 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 a multiplier.

2. LEAKAGE ESTIMATION STEPS:

In scaled devices leakage is strongly dependent on transistor geometry, doping profile (Fig. 1) and temperature. Hence, accurate estimation of total leakage of a logic circuit starts with the accurate description (device geometry, doping profile) of the transistor used to fabricate the circuit and the operating temperature. The steps followed to estimate the total leakage are shown in Fig. 2. The following sections elaborate each of the steps shown in the Fig. 2. First, the leakages for a device are modeled. Based on the model the leakage current of basic gates are calculated. The leakage of the basic gates are used to calculate the leakage of a logic circuit. The outputs of the estimation tool are the subthresholded, 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. 2.

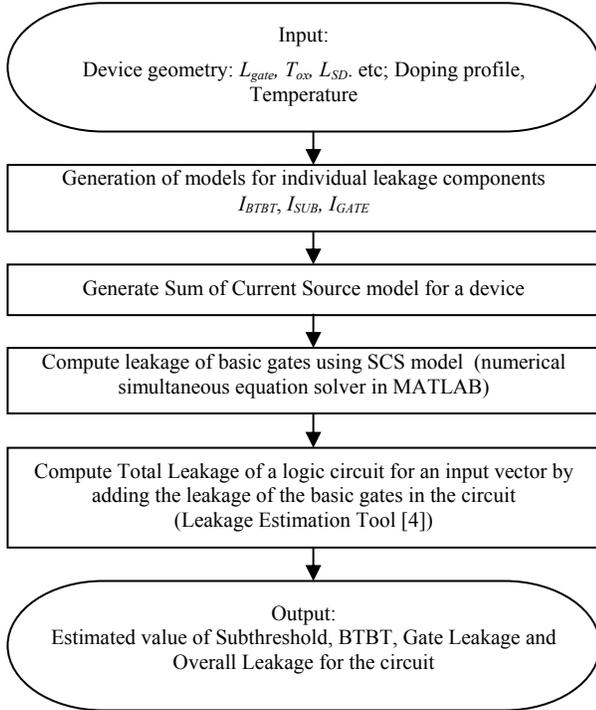


Figure 2: Leakage estimation steps.

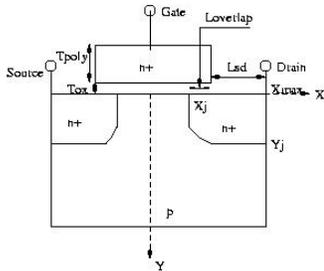


Figure 3: Architecture of the device

3. MODELING LEAKAGE COMPONENTS

This section represents the general approach used to formulate the model for the BTBT, subthreshold and gate leakage, in a MOSFET. The formulation, developed for NMOS transistors, can be easily extended to PMOS transistors. Device structures with Gaussian-shaped 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

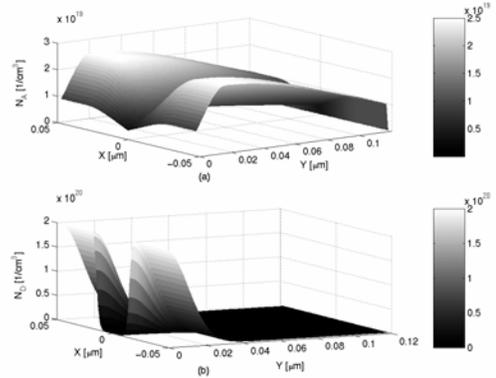


Figure 4: 2-D Gaussian profile for (a) channel and (b) source-drain region.

middle of the channel) is shown in Fig. 3 [7]. The 2-D Gaussian doping profile in the channel ($N_a(x,y)$) and S/D ($N_{sd}(x,y)$) can be represented as [7],[8]:

$$\begin{aligned}
 & x > 0, \\
 & N_{(a/sd)}(x,y) = A_{(p/sd)} \Gamma_{x(a/sd)}(x) K_{y(a/sd)}(y) + N_{SUB} \\
 & \text{where, } K_{y(a/sd)}(y) = \exp\left(\frac{-(y - \alpha_{a/sd})^2}{\sigma_{y(a/sd)}^2}\right) \\
 & \text{and } \left[\begin{aligned} & \Gamma_{x(a/sd)}(x) = \exp\left(\frac{-(x - \beta_{a/sd})^2}{\sigma_{x(a/sd)}^2}\right); & 0 \leq x \leq \beta_{(a/sd)} \\ & = 1; & x > \beta_{(a/sd)} \end{aligned} \right] \quad (1)
 \end{aligned}$$

where, suffix a and sd represents channel and S/D region respectively. A_p and A_{sd} represent the peak “halo” and S/D doping respectively. N_{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 , α_{sd} ($=0$), β_a and β_{sd} control the positions and

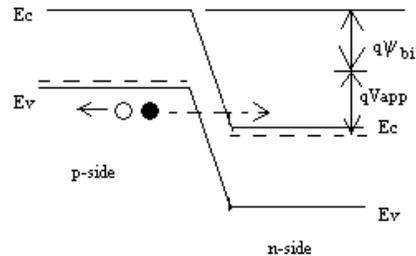


Figure 5: Physical picture of valence band electron tunneling in a reversed bias p-n junction.

σ_{ya} , σ_{xa} and σ_{ysd} , σ_{xsd} control the variances of the Gaussian profiles in channel and S/D regions [7], [8]. Unless otherwise specified in this paper we have used NMOS (N_{ref}) and PMOS (P_{ref}) transistors with $L_{eff}=25nm$, $W_{eff}=1\mu m$ and channel doping profile $\alpha_a=0.018\mu m$, $\sigma_{ya}=0.016\mu m$, $\beta_a=0.016\mu m$, $\sigma_{xa}=0.020\mu m$ and S/D profile from [8]. The Fig. 4 shows the nature of the doping profiles for the device N_{ref} .

3.1. Modeling Band-to-band leakage current (I_{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 hole in the p-region) as shown in Fig. 5 [2]. From Fig. 5, it is evident that for such tunneling to occur the total voltage drop across the junction (applied reverse bias (V_{app}) + built-in voltage(ψ_{bi})) must be more than the band-gap. Since silicon is an indirect band gap semiconductor, the BTBT current in silicon involves the emission or absorption of phonon(s), [2]. The tunneling current density through a silicon p-n junction is given by [2]:

$$J_{b-b} = A \frac{EV_{app}}{\Sigma_g^{1/2}} \exp\left(-B \frac{\Sigma_g^{3/2}}{E}\right) \quad (2)$$

$$A = \frac{\sqrt{2m^*} q^3}{4\pi^3 \hbar^2}, \text{ and } B = \frac{4\sqrt{2m^*}}{3q\hbar}$$

where, m^* is effective mass of electron, Σ_g is energy band-gap, E 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 BTBT current in the MOSFET is the sum of the currents flowing through the drain-substrate and source-substrate junctions and is given by:

$$I_{BTBT} = w_{eff} \int_l J_{b-b}(x,y) dl \Big|_{drain} + w_{eff} \int_l J_{b-b}(x,y) dl \Big|_{source} \quad (3)$$

where, w_{eff} is the effective width, $J_{b-b}(x,y)$ is the current density at a point (x,y) at the junction. For a symmetric device the current expressions for the drain and the source

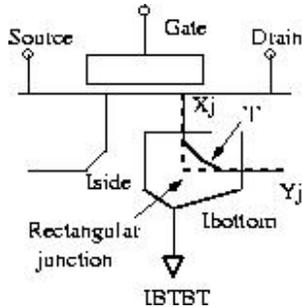


Figure 6: “rectangular junction” approximation

junctions will be identical. Hence, we have considered only one junction for deriving the model. The integration in (3) has to be done along the junction line ‘ l ’ (obtained by solving $N_a(x,y)=N_d(x,y)$) (Fig. 6) within the tunneling region i.e for all values (x,y) for which $(V_{app} + \psi_{bi}(x,y)) > E_g/q$. This integration cannot be done analytically. However, a very accurate estimate of the total current can be achieved analytically by using a “rectangular junction” approximation

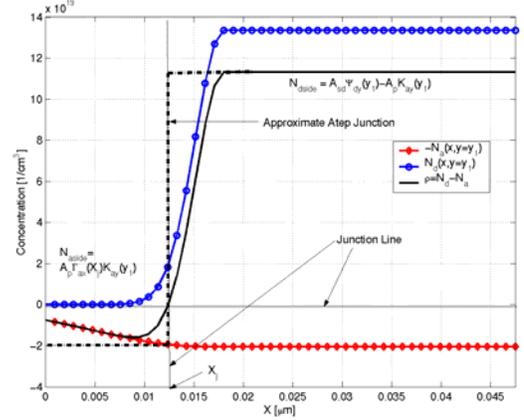


Figure 7. “step junction” approximation.

as shown in Fig. 6. Using this approximation the total current through a junction is given by:

$$I_{BTBTdrain} = I_{side} + I_{bottom}$$

$$= w_{eff} \int_{side} J_{b-b}(X_j, y) dy + w_{eff} \int_{bottom} J_{b-b}(x, Y_j) dx \quad (4)$$

where, X_j is the position of the side junction and Y_j is position of the bottom junction (Fig. 6). Here, we present 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:

$$I_{side} = w_{eff} \int_{y_1}^{y_2} J_{b-b}(X_j, y) dy = \int_{y_1}^{y_2} A \frac{E(X_j, y) V_{app}}{\Sigma_g^{1/2}} \exp\left(-\frac{B \Sigma_g^{3/2}}{E(X_j, y)}\right) dy \quad (5)$$

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

$$I_{side} = w_{eff} |y_2 - y_1| \bar{J}_{b-bside} = \int_{y_1}^{y_2} A \frac{E_{side} V_{app}}{\Sigma_g^{1/2}} \exp\left(-\frac{B \Sigma_g^{3/2}}{E_{side}}\right) dy \quad (6)$$

where, E_{side} is given by:

$$E_{side} = \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} E(X_j, y) dy \quad (7)$$

where, $E(X_j, y)$ is the electric field at the junction of the differential diode of length dy . Using the depletion approximation [10], the junction field is given by:

$$E(X_j, y) = \int_{xp}^{X_j} \frac{q}{\epsilon_{si}} \rho(x, y) dx = \int_{xp}^{X_j} \frac{q}{\epsilon_{si}} [N_{sd}(x, y) - N_a(x, y)] dx \quad (8)$$

where, ϵ_{si} is the permittivity of silicon and xp is the edge of the depletion region at the p-side (i.e. in the substrate). The exact evaluation of the electric field needs a treatment similar to the one given in [10]. However, for a non-uniform 2-D profile the expression become too complicated to be solved analytically. It can be observed from Fig. 7, that, for practical values of the doping profiles, the junction can be assumed as a step junction with doping at the p side ($N_{aside}(X_j, y)$) and n side ($N_{dside}(X_j, y)$) given by:

$$\begin{aligned} N_{aside}(X_j, y) &= N_a(X_j, y) = A_p \Gamma_{xa}(X_j) K_{ya}(y) \\ N_{dside}(X_j, y) &= N_{sd}(x = \beta_{sd}, y) - N_a(x = \beta_a, y) \\ &= A_{sd} \Gamma_{xsd}(x = \beta_{sd}) K_{ysd}(y) - A_p \Gamma_{ya}(x = \beta_a) K_{ya}(y) \\ &= A_{sd} K_{ysd}(y) - A_p K_{ya}(y) \end{aligned} \quad (9)$$

With this assumption the electric field ($E(X_j, y)$) at the junction and the built-in potential ($\psi_{bi}(X_j, y)$) can be computed as [10]:

$$E(X_j, y) = \sqrt{\frac{2qN_{aside}(X_j, y)N_{dside}(X_j, y)(V_{app} + \psi_{bi}(X_j, y))}{\epsilon_{si}(N_{aside}(X_j, y) + N_{dside}(X_j, y))}} \quad (10)$$

$$\psi_{bi}(X_j, y) = \frac{kT}{q} \ln \left(\frac{N_{aside}(X_j, y)N_{dside}(X_j, y)}{n_i^2} \right) \quad (11)$$

where, k is the Boltzmann's constant, T is the operating temperature in Kelvin and n_i is the intrinsic carrier concentration. $E(X_j, y)$ obtained from (11) can be used in (8) to determine E_{side} . However, the analytical evaluation of that

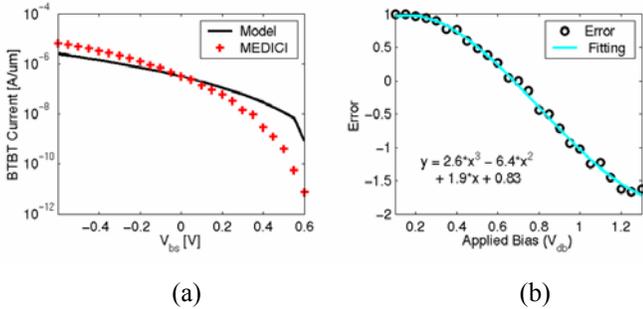


Figure 8: Variation of BTBT current with substrate bias. (a) Comparison of analytical result with simulated data from MEDICI for N MOS transistor with $L_{eff} = 25nm$ and doping profile: $\alpha_a=0.018\mu m$, $\sigma_{av}=0.016\mu m$, $\beta_a=0.016\mu m$, $\sigma_{ax}=0.020\mu m$. (b) Variation of error

integration is difficult. To simplify the derivation, keeping the essential information of the electric field, we defined the average field as:

$$E_{side} = \sqrt{\frac{2q\bar{N}_{aside}\bar{N}_{dside}(V_{app} + \bar{\psi}_{bicide})}{\epsilon_{si}(\bar{N}_{aside} + \bar{N}_{dside})}} \quad (12)$$

This is the field at the junction of the p-n junction with p-side and n-side doping equal to \bar{N}_{aside} and \bar{N}_{dside} respectively. \bar{N}_{aside} is given by:

$$\begin{aligned} \bar{N}_{aside} &= \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} N_{aside}(X_j, y) dy \\ &= \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} A_p \Gamma_{xa}(X_j) K_{ay}(y) dy \\ &= \frac{A_p \Gamma_{xa}(X_j)}{|y_2 - y_1|} \int_{y_1}^{y_2} \exp\left(\frac{-(y - \alpha_a)^2}{\sigma_{ay}^2}\right) dy \\ &= \frac{A_p \Gamma_{xa}(X_j)}{|y_2 - y_1|} \sigma_{ay} \int_{t_1}^{t_2} \exp(-t^2) dt \text{ using } t = \frac{(y - \alpha_a)}{\sigma_{ay}}, \\ t_1 &= \frac{(y_1 - \alpha_a)}{\sigma_{ay}} \text{ and } t_2 = \frac{(y_2 - \alpha_a)}{\sigma_{ay}}. \end{aligned} \quad (13)$$

\bar{N}_{dside} can also be obtained similarly and given by:

$$\bar{N}_{dside} = \frac{1}{|y_2 - y_1|} \int_{y_1}^{y_2} N_{dside}(X_j, y) dy \quad (14)$$

The built-in potential $\bar{\psi}_{bi}$ is obtained by [10]:

$$\bar{\psi}_{bicide} = \frac{kT}{q} \ln \left(\frac{N_{aside} N_{dside}}{n_i^2} \right) \quad (16)$$

The lateral junction depth X_j and vertical junction depth Y_j are found by solving following equations:

$$N_{sd}(X_j, y = 0) = N_a(X_j, y = 0) \quad (17)$$

$$N_{sd}(x = x_{max}, Y_j) = N_a(x_{max}, Y_j)$$

For simplicity the whole side junction is assumed to be tunneling (i.e. $y_l=0$ and $y_2 = Y_j$). For bottom junction $x_l=X_j$ and $x_2 = x_{max}$.

Using expressions from (13)-(17), into (12) E_{side} (and similarly E_{bottom}) can be obtained. E_{side} (and E_{bottom}) can be used in (6) to obtain $J_{b-bicide}$ (similarly $J_{b-bbottom}$). If $(V_{app} + \bar{\psi}_{bicide}) < \Sigma_g/q$, then no tunneling occurs and $J_{b-bicide}$ is zero (similar argument holds for $J_{b-bbottom}$). Hence, the total BTBT current in the drain junction is given by:

$$I_{BTBTdrain} = w_{eff} |y_2 - y_1| \bar{J}_{b-bicide} + w_{eff} |x_2 - x_1| \bar{J}_{b-bbottom} \quad (19)$$

$$J_{b-side} = A \frac{E_{side} V_{app}}{\Sigma_g^{1/2}} \exp\left(-\frac{B \Sigma_g^{3/2}}{E_{side}}\right);$$

when $(V_{app} + \psi_{b-side}) < \Sigma_g/q$
 $= 0$ otherwise

$$J_{b-side} = A \frac{E_{side} V_{app}}{\Sigma_g^{1/2}} \exp\left(-\frac{B \Sigma_g^{3/2}}{E_{side}}\right);$$

when $(V_{app} + \psi_{b-side}) < \Sigma_g/q$
 $= 0$ otherwise

For a 25nm transistor, the comparison of the analytical model given in (10) and the simulated data from MEDICI [11] shows close match for small reverse and forward substrate bias (Fig. 8). However, deviations are observed at high forward (i.e. low V_{app}) and reverse (i.e. event higher V_{app}) substrate bias. At high V_{app} , the average electric field (calculated using 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_{app} . In the low bias region, reduction of V_{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_{app} (i.e. high forward substrate bias). Also, at high gate voltage, (a) small increase in the potential near the substrate side of the side junction and (b) non-negligible 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 parameter (a_0), and function ($\lambda(V_{app})$) 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:

$$I_{BTBTcorrected} = a_0 I_{BTBTdrain} (1 - \lambda(V_{app})) (1 - \delta_g V_G) \quad (20)$$

where a_0 is the zero substrate bias multiplication factor defined as the ratio of the actual BTBT (measured/simulated) current and the analytical value at zero substrate bias and $\lambda(V_{app})$ is an empirical function (for drain-substrate junction $V_{app}=V_{db}$ and for source-substrate junction $V_{app}=V_{sb}$). From experiments it was found that a cubic function gives a good fit of the simulated result (Fig. 8). Hence, the fitting function can be written as $\lambda(V_{app}) = c_3 V_{app}^3 + c_2 V_{app}^2 + c_1 V_{app} + c_0$. The coefficient can be calculated by measuring the actual

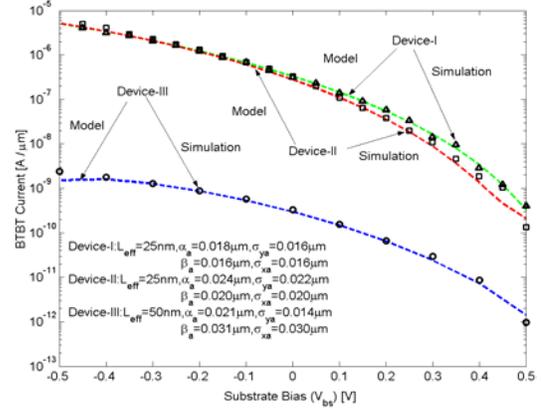


Figure 9: Variation of BTBT current with substrate bias for different devices.

BTBT and analytical currents at different V_{app} and using the relation:

$$\lambda(V_{app}) = \frac{I_{analytical}(V_{app}) - I_{measured}(V_{app})}{I_{analytical}(V_{app})} \quad (21)$$

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 by:

$$I_{BTBT} = \sum_{i=drain,source} (I_{side_i} + I_{bot_i}) (1 - \delta_g V_G) (1 - \lambda(V_{ib}))$$

$$I_{side_i} = w_{eff} A \frac{E_{side_i}}{\Sigma_g^{1/2}} V_{ib} \exp\left(-\frac{B \Sigma_g^{3/2}}{E_{side_i}}\right) \quad (22)$$

$$I_{bot_i} = w_{eff} A \frac{E_{bot_i}}{\Sigma_g^{1/2}} V_{ib} \exp\left(-\frac{B \Sigma_g^{3/2}}{E_{bot_i}}\right)$$

The parameters, namely, E_{side_i} , E_{bot_i} , can be evaluated following the procedure discussed above. Fig. 9 shows a comparison plot of the analytical result with the simulated results from MEDICI for devices with $L_{eff}=25nm$ ($V_{dd}=0.7V$) and $50nm$ ($V_{dd}=0.9V$) and different doping profiles. It shows that, for analytical result follows very closely the simulated result for substrate bias in the range of $-V_{dd}/2$ to $+V_{dd}/2$. However, deviation is observed at very high forward and reverse bias.

3.2. Modeling subthreshold current (I_{ds}):

In the “off” state of a device ($V_{gs} < V_{th}$) the current flowing from the drain to the source of a transistor is known as the subthreshold current. This current is due to the diffusion of the minority carriers through the channel. The subthreshold current flowing through a transistor is given by [2],

$$I_{sub} = \frac{w_{eff}}{L_{eff}} \mu \sqrt{\frac{q \epsilon_{si} N_{cheff}}{2 \Phi_s}} v_T^2 \exp\left(\frac{V_{gs} - V_{th}}{n v_T}\right) \left(1 - \exp\left(\frac{-V_{ds}}{v_T}\right)\right) \quad (23)$$

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

given by kT/q . To obtain the effect of the 2-D Gaussian profile described in (1),(2) on the subthreshold current we developed a simplified model for both the subthreshold current and the threshold voltage following the procedure given in [12] and [13]. Using charge sharing model the threshold voltage can be expressed as [12], [13]:

$$V_{th} = V_{FB} + \Phi_s + \gamma \sqrt{\Phi_{s0} - V_{bs}} \left(1 - \lambda \frac{X_d}{L_{eff}} \right) \quad (24)$$

where, V_{FB} is flat-band voltage, $\Phi_{s0} = 2v_T \ln \left(\frac{N_{cheff}}{n_i} \right)$ is zero

bias surface potential, $\gamma = \frac{\sqrt{2q\epsilon_{si}N_{cheff}}}{C_{ox}}$ is body factor, C_{ox}

$= \epsilon_{sio2} / t_{ox}$ is oxide capacitance, $X_d = \sqrt{\frac{2\epsilon_{si}}{qN_{cheff}}} \sqrt{\Phi_{s0} - V_{bs}}$

is depletion layer thickness and λ is a fitting parameter (≈ 1). The surface potential (Φ_s) of short channel devices is reduced from its zero bias value due to short channel effects like DIBL and V_{th} roll-off. It is given by:

$$\begin{aligned} \Phi_s &= \Phi_{s0} - \Delta\Phi_s \\ \Delta\Phi_s &= \frac{v_T \ln \left(\frac{N_{sdeff} N_{cheff}}{n_i^2} \right) - \Phi_{s0} + 0.5V_{ds}}{\left[\cosh \left(\frac{L_{eff}}{\sqrt{(\epsilon_{si} t_{ox} X_d) / (\eta \epsilon_{sio2})}} \right) \right]} \end{aligned} \quad (25)$$

where, η is another fitting parameter which is usually close to one [12]. The narrow width of the transistor also modulates the threshold voltage of the transistor. In case of local oxide isolation gate MOSFET this effect can be modeled as an increase in V_{th} by an amount ΔV_{NWE} given by [9]:

$$\Delta V_{NWE} = \frac{\pi q N_{cheff} X_d^2}{2C_{ox} w_{eff}} = 3\pi \frac{t_{ox}}{w_{eff}} \phi_s \quad (26)$$

In scaled devices, due to high electric field at the surface (E_s) and high substrate doping, the quantization of inversion-layer electron energy modulates V_{th} . Quantum-mechanical behavior of the electrons increases V_{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. When E_s is higher than 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 ΔV_{QM} , given by [2]:

$$\Delta V_{QM} = \left(1 + \frac{3t_{ox}}{X_d} \right) \left(\frac{\Sigma_0}{q} - \frac{kT}{q} \ln \left(\frac{8\pi q m_d E_s}{h^2 N_C} \right) \right) \quad (27)$$

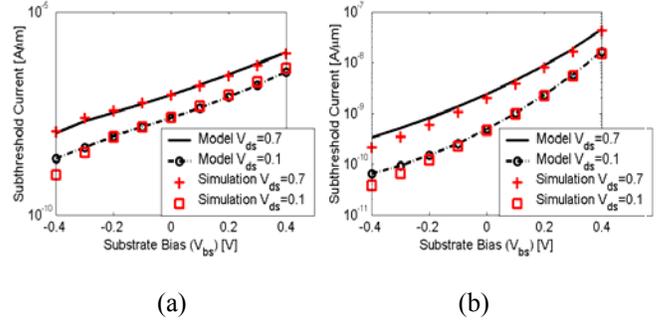


Figure 10: Variation of subthreshold leakage with substrate bias (V_{bs}) and drain bias (V_{ds}) for NMOS transistor N_{ref} (a) Without and (b) With quantum correction.

where, Σ_0 is the lowest subband energy given by[2]

$$\Sigma_0 = \left[\frac{3\hbar q_s E_s}{4\sqrt{2m_x}} \frac{3}{4} \right]^{2/3}, N_C$$

is effective conduction band density of states, m_x is quantization effective mass of electron and m_d is density of states effective mass of electron. To match the simulated result, the theoretically calculated ΔV_{QM} value is multiplied by an empirical factor ($\theta(V_{bs})$).

The effect of 2D Gaussian profile is used to calculate the effective channel and S/D doping as shown below:

$$\begin{aligned} N_{sdeff} &= \frac{1}{\Delta_{SD}} \iint_{\Delta_{SD}} N_{sd}(x, y) dx dy \\ &= \frac{A_{sd}}{\Delta_{SD}} \int_{x=X_j}^{x=L_{gate}/2+L_{sd}} \Gamma_{xsd}(x) dx \int_{y=0}^{y=Y_j} K_{ysd}(y) dy \end{aligned} \quad (28)$$

where, $\Delta_{SD} = (L_{overlap} + L_{sd}) Y_j$ is S/D area, $L_{overlap}$ is the gate and the S/D overlap length and L_{sd} is the S/D length as shown in Fig. 3.

$$\begin{aligned} N_{cheff} &= \frac{1}{\Delta_{ch}} \iint_{\Delta_{ch}} N_a(x, y) dx dy + N_{sub} \\ &= \frac{A_p}{\Delta_{ch}} \int_{x=-L_{eff}/2}^{x=L_{eff}/2} \Gamma_{xa}(x) dx \int_{y=0}^{y=X_d} K_{ya}(y) dy + N_{sub} \end{aligned} \quad (29)$$

$\Delta_{ch} = L_{eff} X_d$ 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 $y = \alpha_a$. The simplified model shows reasonable match with the simulated result from MEDICI under substrate and drain bias variation (Fig. 10) with and without quantum correction. Substantial reduction in the subthreshold current is observed using the quantum correction.

3.3. Modeling Gate Direct Tunneling Current (I_{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 [2]. The direct tunneling current density is expressed as [2]:

$$J_{DT} = A_g (V_{ox}/T_{ox})^2 \exp\left(\frac{-B_g (1 - (1 - V_{ox}/\phi_{ox})^{3/2})}{V_{ox}/T_{ox}}\right) \quad (30)$$

$$A_g = \frac{q^3}{16\pi^2 \hbar \phi_{ox}} \text{ and } B_g = \frac{4\sqrt{2m^*} \phi_{ox}^{3/2}}{3\hbar q}$$

where J_{DT} is direct tunneling current density, V_{ox} is potential drop across oxide, ϕ_{ox} is barrier height of tunneling electron, m^* is the effective mass of an electron in the conduction band of silicon. and T_{ox} is the oxide thickness. The tunneling current increases exponentially with decrease in the oxide thickness and increase in the potential drop across oxide. Major components of gate tunneling in a scaled MOSFET device are [5]: (1) Gate to S/D overlap region current (Edge Direct Tunneling (EDT)) components (I_{gso} & I_{gdo}), (2) Gate to channel current (I_{gc}), part of which goes to source (I_{gcs}) and rest goes to drain (I_{gcd}), (3) Gate to substrate leakage current (I_{gb}). Accurate modeling of each of the components is based on the following equation [5],[6]:

$$J_{DT} = A_g \left(\frac{T_{oxref}}{t_{ox}}\right)^{ntox} \left(\frac{V_g V_{aux}}{t_{ox}^2}\right) \exp(-B t_{ox} (\alpha - \beta |V_{ox}|)(1 + \gamma |V_{ox}|)) \quad (31)$$

where, T_{oxref} is the reference oxide thickness at which all parameters are extracted, $ntox$ is a fitting parameter (default 1) and V_{aux} is an auxiliary function that approximates the density of tunneling carriers and available states. We have used the current models from [5],[6] with the effective channel and S/D doping density obtained from (27) and (28).

3.4: Effect of Temperature on Different

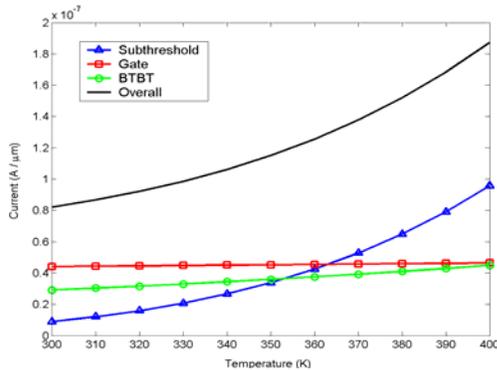


Figure 11: Variation of leakage components with temperature.

Components of Leakage Current:

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 increase of temperature. Since tunneling probability of an electron through a potential barrier does not depend directly on temperature, gate and band-to-band tunneling is expected to be less sensitive to temperature variations. However, increase of temperature reduces the band-gap of silicon, which is the barrier height for tunneling in BTBT. Hence, BTBT is expected to increase with temperature. Thus different leakage components show different temperature dependence. The models of the leakage components introduced in the last three sub-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 (a) reduction in threshold voltage and (b) increase in thermal voltage (v_T) (23). The gate tunneling current is almost insensitive to temperature since the electric field across the oxide does not strongly depend on temperature (30),(31). Band-to-band tunneling current increases with temperature due to narrowing of band-gap at higher temperature. The Band gap ($\Sigma_G(T)$) at a temperature T is given by [14],

$$\Sigma_G(T) = \Sigma_G(0) - \frac{\alpha_T T^2}{(T + \beta_T)} \quad (32)$$

where, $\Sigma_G(0)$ limiting value of band gap at 0 K and equal to 1.17 eV for Si. α_T and β_T are fitting parameters with values 4.73×10^{-4} and 636 respectively for silicon [13]. Due to the band-gap narrowing, BTBT increases with temperature (22). Fig. 11 shows the variation of each leakage component with temperature in an NMOS transistor ($L_{eff}=25\text{nm}$) using the models introduced in the last section. It is observed that, at room temp ($T=300\text{K}$) gate leakage and BTBT dominates over subthreshold current, while at elevated temperatures subthreshold leakage is the dominant component of overall leakage.

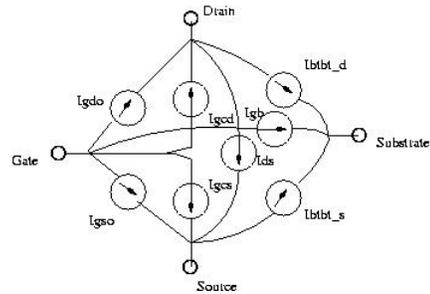


Figure 12: Sum of Current Source model of a transistor.

4. MODELING OVERALL LEAKAGE

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

$$I_{overall} = I_{BTBT} + I_{sub} + I_{gate} \quad (33)$$

Hence, for leakage estimation we have modeled the device as a combination of voltage controlled current sources as shown in Fig. 12. Based on (22) the BTBT current is modeled as two current sources, one between drain and substrate (I_{btbt_d}) controlled by V_{db} and another between source and substrate (I_{btbt_s}) controlled by V_{sb} . Each component of gate leakage described in 3.3 is modeled as a current source. I_{ds} models 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.

5. 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. 13 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. 10 the node equation at INT is given by:

$$\begin{aligned} I_{ds1} + I_{gcs1} + I_{gso1} - I_{BTBT_s1} \\ = I_{ds2} - I_{gdo2} - I_{gcd2} + I_{BTBT_d2} \end{aligned} \quad (34)$$

In circuits involving more than one such node, we will have a set of simultaneous equation that needs to be solved. The overall leakage in the circuit can be defined as the sum of all currents collected at the ground node. Hence, the overall leakage in a CMOS circuit is given by (assuming $V_{bulk}=0$ for all NMOS and $V_{bulk}=V_{dd}$ for all PMOS):

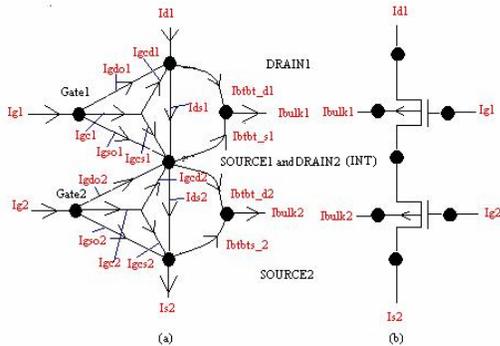


Figure. 13: Circuit configuration with SCS model for a 2-transistor stack, (a) SCS model, (b) transistor-circuit diagram.

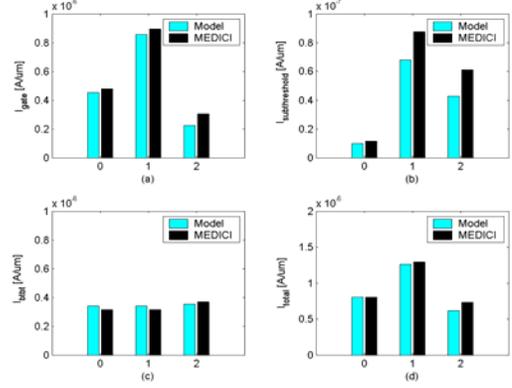


Figure 14: Comparison of simulator (MEDICI) and model current values for a 2-transistor stack for different input vectors: (a) gate, (b) subthreshold, (c) BTBT and (d) total leakage

$$I_{leakage} =$$

$$\sum_{NMOS+PMOS} (V_{dd} - V_{gk}) I_{gk} + \sum_{NMOS} I_{BTBTk} + \sum_{NMOS \text{ with source connected to ground}} I_{sourcek} \quad (35)$$

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. 14 shows the comparison of the evaluated result and simulated result in MEDICI for a stack of 2 NMOS transistors (N_{ref}), at normal temperature (without quantum correction). The evaluated results match the simulated results closely.

SCS solver can be used to evaluate the leakage components of basic gates. Fig. 15 and 16 show the different leakage components of INVERTER, NAND and NOR gates (designed with N_{ref} and P_{ref}) at normal ($T=300K$) and high temperature ($T=400K$) (with and without the quantum

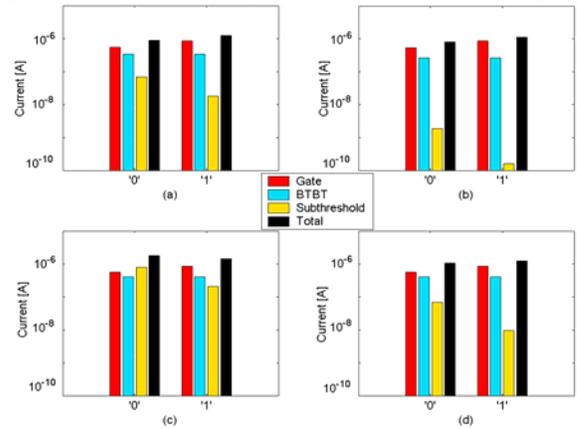


Figure 15: Leakage of an INVERTER with input '0' and '1'. (a) $T=300K$ and no quantum correction, (b) $T=300K$ and with quantum correction, (c) $T=400K$ and no quantum correction, (d) $T=400K$ and with quantum correction

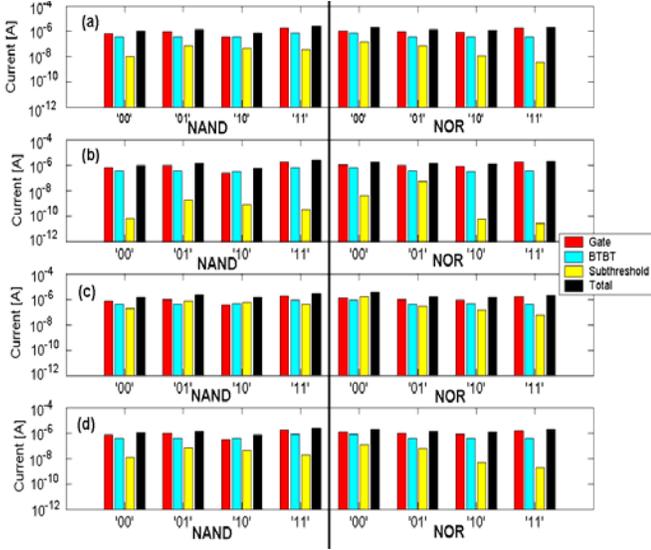


Figure 16: Leakage of a 2-input NAND and NOR gate with different input. (a) T=300K and no quantum correction, (b) T=300K and with quantum correction, (c) T=400K and no quantum correction, (d) T=400K and with quantum correction.

correction). It is observed that, the overall leakage increases considerably with the temperature. At normal temperatures gate leakage dominates the subthreshold leakage and BTBT leakage, whereas later two are high at higher temperatures. Also, application of the quantum correction reduces the subthreshold current considerably. The solver can easily be extended to handle other logic gates.

5.1. Stacking Effect

Turning “off” more than one transistor in a stack of transistors forces the intermediate node (say INT in Fig. 10) voltage to go to a value higher than zero [1], [4]. This causes a negative V_{gs} , negative V_{bs} (more body effect) and reduced V_{ds} (less DIBL) in the top transistor, thereby reducing the subthreshold current flowing through the stack considerably [1], [4]. This effect, known as the “stacking effect”, has been used to reduce the subthreshold leakage in logic circuits in stand-by mode [1], [4]. The estimation tool described here, effectively models stacking effect for subthreshold, gate and BTBT leakage. Fig. 14 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.14)), 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 gate leakage is the dominant component the input ‘10’ minimizes the total leakage in a stack of two NMOS transistors as shown in Fig. 14.*

6. 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 [4]. Leakage of a logic circuit depends on the primary input vector. The primary input vector is propagated by simulating the circuit level by level. subthreshold (I_{Tsub}), the gate (I_{Tgate}) and the BTBT (I_{Tbibt}) leakage and overall leakage ($I_{Toverall}$) through the circuit is defined as the sum of the leakage through each of the basic gates present in the circuit and is given by:

$$\begin{aligned}
 I_{Tsub} &= \sum_{k=all\ gate} I_{ksub} ; \\
 I_{Tgate} &= \sum_{k=all\ gate} I_{kgate} ; \\
 I_{Tbibt} &= \sum_{k=all\ gate} I_{kbtbt} ; \\
 I_{Toverall} &= \sum_{k=all\ gate} I_{koverall} = I_{Tsub} + I_{Tgate} + I_{Tbibt} ;
 \end{aligned} \tag{36}$$

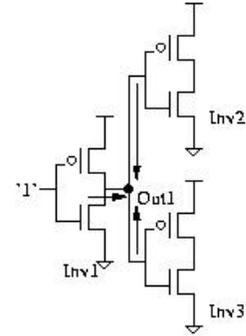


Figure 17: Illustration of loading effect of an inverter.

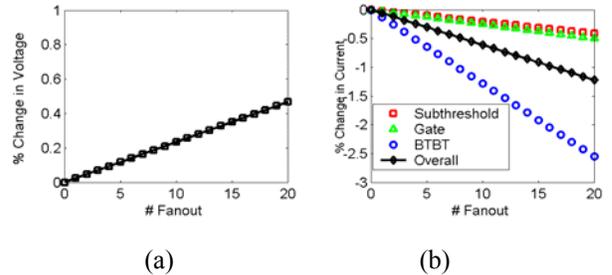


Figure 18: Percentage change in the (a) output voltage and (b) leakage components in an inverter due to loading.

6.1: Loading effect:

The estimation method using (36) neglects the change of the leakage currents of a gate due to the loading by its fanout gates. To understand how loading can modify the leakage of a gate let us consider Fig. 17, where output of an inverter is loaded by the two other inverters. Also, consider the situation where, input of the inverter INV1 is '1' and we would like to determine leakage of INV1. From, discussion in section 5 the leakage of INV1 can be found by solving KCL at output node OUT1. The equation is given by:

$$\begin{aligned} I_{ddN} &= (I_{dsN} + I_{gcdN} + I_{gdoN} - I_{BTBTdrainN}) \\ I_{ddP} &= (I_{dsP} - I_{gdoP} - I_{gcdP} + I_{BTBTdrainP}) \\ I_{ddP} + I_{ddN} &= 0 \end{aligned} \quad (37)$$

The leakage value for this condition is given in Fig. 14. However, since the output is connected to the gate of 2 other inverters, the gate leakage from these gates will also add to current at OUT1, thereby changing (37) to:

$$I_{ddP} + I_{ddN} + \sum_{i=load\ gates} I_{gate_i} = 0 \quad (38)$$

The net effect will be a change in the voltage at OUT1, which in turn will modify the leakage of INV1. To understand how the leakage of a gate varies with its loading, we studied the variation of the leakage of an inverter (say INV1) with loading. Fig. 18 shows the percentage change in the voltage at OUT1 and the leakage current of the inverter with the increase in the number of its fanouts. It is observed that, even for a fanout of 20 the leakage of the inverter remains almost constant. Hence, we can conclude that, the summation of the leakage of individual gates gives a reasonably accurate estimate of the total leakage of a circuit. However, for an exact value of the leakage one has to solve the full circuit using a transistor level circuit simulator. SPICE circuit simulator can be used to evaluate the leakage

in a circuit, by representing the transistors using the described SCS model.

6.2: Results

The leakage estimation tool is used to estimate the total leakage in complex logic circuits, under different primary input vectors. Fig. 19 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) averaged over a large number of primary input vectors. The leakage is evaluated at both normal (T=300K) and high (T=400K) 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 BTBT is increased.

7. SUMMARY and CONCLUSION

In this paper we have developed a compact model for the total leakage in a transistor as the summation of subthreshold, BTBT and gate leakage. It has been shown that for leakage estimation the transistor can be modeled as a Sum of Current Sources, 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 CAD 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 and can be effectively used to accurately estimate leakage in a scaled CMOS logic circuit.

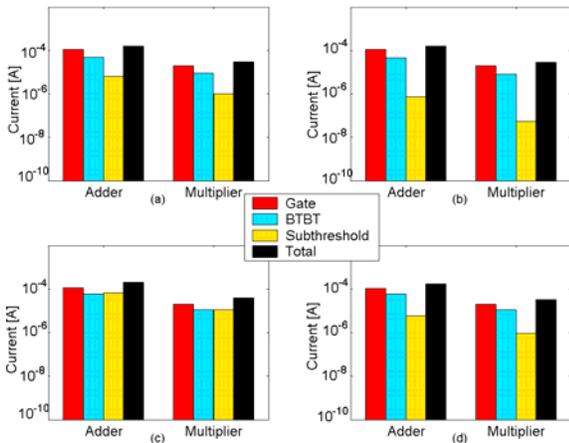


Figure 19: Average leakage of an 8-bit adder and a 2-bit array multiplier. (a) T=300K and no quantum correction, (b) T=300K and with quantum correction, (c) T=400K and no quantum correction, (d) T=400K and with quantum correction

REFERENCE

- [1] K.Roy, S.Mukhopadhyay, and H.Miamand, "Leakage current mechanisms and leakage reduction techniques in deep-submicron CMOS circuit", Proceedings of IEEE, Feb. 2003
- [2] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, 1998.
- [3] W. Shiue, "Leakage power estimation and minimization in VLSI circuits", ISCAS, 2001, pp.178-181
- [4] Z. Chen, et. al., "Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks". Int. ISLPED, 1998, pp.239-244.
- [5] K. Cao, et.al. "BSIM4 Gate Leakage Model Including Source Drain Partition", IEDM, pp. 815-818, 2000.
- [6] *BSIM4.2.1 MOSFET Model*, BSIM Group, UC Berkeley, <http://www-device.eecs.berkeley.edu/~bsim3/>
- [7] Z. Lee, et.al., "Two-dimensional doping profile characterization of MOSFET's by inverse modeling using characteristics in the subthreshold Region", IEEE Trans. on Elec. Dev., Aug. 1999, pp. 1640 –1649.
- [8] <http://www-mtl.mit.edu/Well/>
- [9] D. Fotty, MOSFET Modelling with SPICE, New Jersey, USA: Prentice Hall PTR, 1997,
- [10] R. Pierret, Semiconductor Device Fundamentals, Addison-Wesley Publishing Company,1996.
- [11] MEDICI: Two-dimensional semiconductor device simulation program, Synopsys, Inc. CA 94043
- [12] Z. Liu, et.al., "Threshold voltage model for deep-submicrometer MOSFET's", IEEE Trans. On Elec. Dev., Jan. 1993
- [13] X. Zhou, et.al., "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. On Elec. Dev., Jan. 2000
- [14] R. Pierret, Advanced Semiconductor Fundamentals, Volume VI in "Modular Series on Solid States Devices", Addison-Wesley Publishing Company,1989.