Modeling and Analysis of Loading Effect on Leakage of Nanoscaled Bulk-CMOS Logic Circuits

Saibal Mukhopadhyay, *Student Member, IEEE*, Swarup Bhunia, *Member, IEEE*, and Kaushik Roy, *Fellow, IEEE*

***Abstract*—In nanoscale complementary metal–oxide–semiconductor (CMOS) devices, a significant increase in subthreshold, gate, and reverse-biased junction band-to-band–tunneling (BTBT) leakage results in large leakage power in logic circuits. Leakage components interact with each other at the device level (through device geometry and the doping profile) and at the circuit level (through the node voltages). Due to the circuit-level interaction of the different leakage components, the leakage of a logic gate depends on the circuit topology, i.e., the number and the nature of the other logic gates connected to its input and output. In this paper, the effect of loading on a leakage of a circuit is analyzed for the first time. The authors have also proposed a method to accurately estimate the total leakage in a logic circuit from its logic-level description considering the impact of loading and transistor stacking.**

***Index Terms*—Gate leakage, junction tunneling, logic gate, subthreshold leakage.**

I. Introduction

THE AGGRESSIVE scaling of complementary metal–oxide–semiconductor (CMOS) devices in each technology generation has resulted in a significant increase in the leakage current in CMOS devices. In nanoscaled devices, three major leakage components can be identified as (Fig. 1) a subthreshold leakage, a gate leakage, and the reverse-biased drain–substrate and source–substrate junction band-to-band-tunneling (BTBT) leakage [1]–[4]. In a transistor, the relative magnitudes of these components depend on the device geometry (channel length, oxide thickness, etc.), the doping profiles, and the operating temperature [1]–[4]. On the other hand, different leakage components also depend on the terminal voltages of a transistor. Hence, in logic circuits, the different components of a leakage interact with each other through the node voltages. If the output node (e.g., N0) of a gate (e.g., G) is connected to the input of other gates ($\mathrm{G_{out1}}, \mathrm{G_{out2}}, \ldots, \mathrm{G_{outn}}$), the gate leakage from these gates change the voltage at N0 (Fig. 2). This modifies the leakage of the gates G, $\mathrm{G_{out1}}, \ldots, \mathrm{G_{outn}}$. This effect can be defined as the "loading effect." The loading effect is not important in technology generations with negligible gate tunneling leakages. However, in nanoscale devices with high gate currents, the effect of loading on the leakage of logic gates and circuits needs to be analyzed and evaluated.

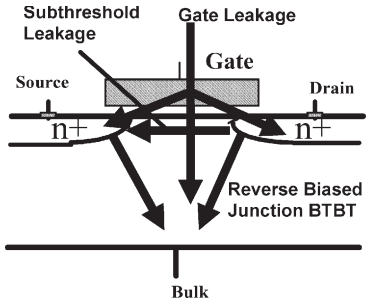

Fig. 1. Leakage components in a device.

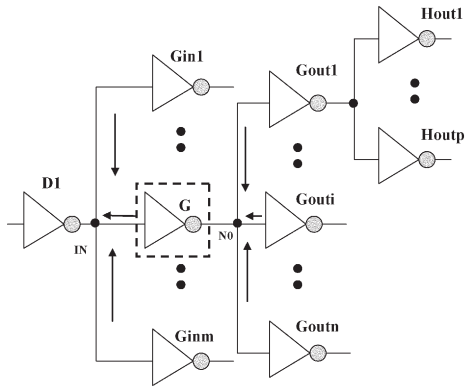

Fig. 2. Illustration of the loading effect for the logic gate G.

In this paper, we analyze the impact of the "loading effect" on the leakage of logic gates and logic circuits for the first time. We propose a method to estimate the total leakage of a logic circuit from its gate level description, considering the loading effect. Consideration of the loading effect improves the accuracy of the leakage estimation in logic circuits. Our analysis shows that the loading effect modifies the leakage of individual logic gates by $\sim$ 5%–8%. We also show that, depending on the input vector, the loading effect can either increase or decrease the leakage of a logic gate. Hence, in large circuits, the effect of loading on different logic gates may cancel out each other. Our analysis shows that the cancellation effect reduces the effect of loading in the leakage of large circuits. We further show that the loading effect strongly depends on the device design,

Manuscript received October 26, 2004; revised March 1, 2005 and May 27, 2005. This work was supported in part by the National Science Foundation (NSF), the Semiconductor Research Corporation under Contract 1078.001, International Business Machines Corporation (IBM), and Intel Corporation. An initial version of this work was presented in Design, Automation and Test in Europe, 2005. This paper was recommended by Associate Editor H. Kosina.

S. Mukhopadhyay and K. Roy are with the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: sm@ecn.purdue.edu; kaushik@ecn.purdue.edu).

S. Bhunia is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: skb21@case.edu).

Digital Object Identifier 10.1109/TCAD.2005.855934

TABLE I
DEVICE CHARACTERISTICS

(L_{gate}=50nm, T_{ox}=1nm, V_{DD}=1.0V)		
	NMOS	PMOS
On Current (μA/μm)	1060	310
Off. Current (nA/μm)	500	410
Vt (mV)	0.1	0.11
DIBL (mV/V)	100	100
Subth. Slope (mV/decade)	100	115

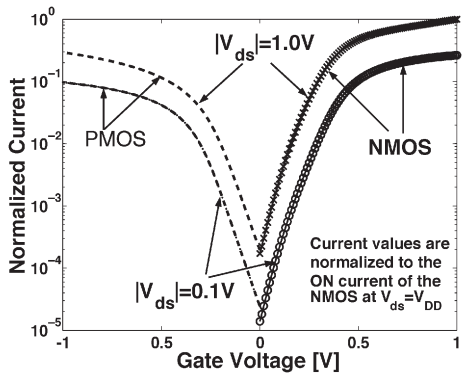

Fig. 3. Device characteristics.

temperature, and the parametric variations. In particular, we show that the loading effect significantly increases the leakage spread of logic gates under parametric variations.

II. BACKGROUND: LEAKAGE COMPONENTS IN A DEVICE

A. Device Structures and the Simulation Environment

The leakage analysis presented in this paper is based on the transistors of a 50-nm gate length (metallurgical channel length $L_{\rm eff} = 25$ nm) designed using the MEDICI device simulator [5]. The device structure and the "super halo" doping profiles given in [6] were used in designing the transistors. The different structural and electrical parameters of the NMOS and PMOS transistors (designed for equal subthreshold leakage) used in the analysis are shown in Table I. Fig. 3 shows the $I_d - V_g$ characteristics of the designed devices. The Aurora parameter extraction tool [7] was used to extract the BSIM4 [9] simulation program with integrated circuits emphasis (SPICE) model parameters of the designed devices to do the SPICE simulations. Since BSIM4 does not have models for the junction-tunneling current, we have incorporated the junction-tunneling models presented in [2] as the voltage-controlled current sources in the SPICE simulations.

B. Leakage Components

In nanometer scaled devices, the major leakage components are (Fig. 1) the subthreshold leakage, the gate direct-tunneling leakage, and the reverse bias drain–substrate and source–substrate junction BTBT leakage. It was observed that other leakage components, like the gate induced drain leakage (GIDL), the punchthrough current, etc., are not very serious for the $V_{\rm DD}$ shown in Table I. Next, we will briefly discuss the three major leakage components in a device.

1) Subthreshold Current: The subthreshold current in a transistor is due to the diffusion of the minority carriers from the source to the drain [1]–[4]. The subthreshold current in a transistor is given by [1] as

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

where $N_{\rm cheff}$ is the effective channel doping, Φ_S is the surface potential, and m is the subthreshold swing factor. It can be observed from (1) that the subthreshold current exponentially depends on the threshold voltage of a transistor [1]–[4]. In nanoscale devices, the short-channel effects (SCE) [the penetration of the drain electric field into the channel] reduce the threshold voltage, thereby increasing the subthreshold current [1]–[4]. Due to the SCE, the subthreshold current increases with an increase in the drain bias [drain-induced barrier lowering (DIBL)] and the reduction in the channel length ($V_{\rm th}$-roll off) [1]–[4] and [8]. In sub-100-nm transistors, the quantization of electron energy in the channel increases the threshold voltage due to a high oxide field, thereby lowering the subthreshold current [1]–[4].

2) Gate Leakage: At the ultrathin gate-oxide regime, due to the high electric field and the low oxide thickness, an electron can tunnel through the gate oxide. This is known as the direct tunneling of an electron and it results in a large gate leakage in nanoscale transistors [1]–[4], [9]. The gate tunneling current density in a MOSFET is given by

$$J_{\rm DT} = A_{\rm g}\left(\frac{V_{\rm ox}}{T_{\rm ox}}\right)^2 \exp\left(\frac{-B_{\rm g}\left(1-\left(1-\frac{V_{\rm ox}}{\phi_{\rm ox}}\right)^{\frac{3}{2}}\right)}{\frac{V_{\rm ox}}{T_{\rm ox}}}\right) \quad (2)$$

where $V_{\rm ox}$ is the potential drop across oxide, $\phi_{\rm ox}$ is the barrier height of the tunneling electron (or hole), and $T_{\rm ox}$ is the oxide thickness. $A_{\rm g}$ and $B_{\rm g}$ are physical parameters [4]. The gate current increases exponentially with an increase in the supply voltage (higher oxide field) and a reduction in the oxide thickness (lower tunneling thickness). The gate leakage in a MOSFET is composed of the following components [9]: a) the gate–source/drain overlap region current ($I_{\rm gso}$ and $I_{\rm gdo}$); b) the gate–channel current ($I_{\rm gc}$), part of which goes to the source ($I_{\rm gcs}$) and rest goes to the drain ($I_{\rm gcd}$); and c) the gate–substrate current ($I_{\rm gb}$). The overlap tunneling current ($I_{\rm gdo}$ and $I_{\rm gso}$) dominates gate leakage in the "OFF" state ($V_{\rm gs} =$ "0"), whereas the gate–channel current becomes significant in the "ON" state ($V_{\rm gs} =$ "1"). $I_{\rm gb}$ is negligible compared to the overlap and the gate–channel tunneling.

3) Junction BTBT Current: In nanoscale MOSFETs, highly doped "halo" regions are used in the substrate side of the drain–substrate and source–substrate junctions to suppress the SCE [1]–[4]. This increases the junction electric field and causes

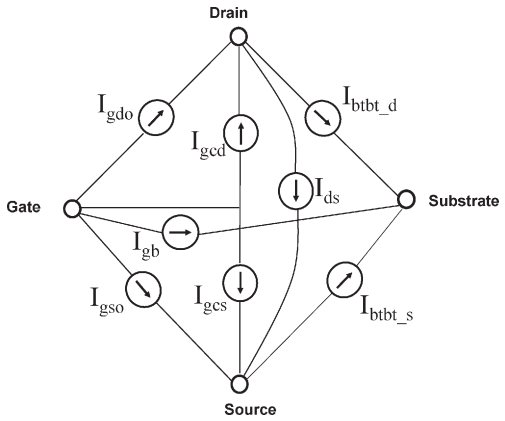

Fig. 4. Transistor as a combination of voltage-controlled current sources.

the significant tunneling of electrons from the valence band of the p-region, to the conduction band of the n-region [4]. This leakage current is known as the junction BTBT current ($I_{\rm JN}$). In the "OFF" state of a transistor [for NMOS ($V_{\rm g}=0$, $V_{\rm d}=V_{\rm dd}$, $V_{\rm s}=0$, $V_{\rm b}=0$)], the junction tunneling results in a leakage current from the drain to the substrate. The junction BTBT leakage exponentially increases with an increase in the junction doping and the supply voltage, both of which increase the junction field [1]–[4]. The junction BTBT in a MOSFET with halo doping can be expressed as [2]

$$I_{\rm JN} = WA\frac{\xi_{\rm JN}}{E_{\rm g}^{\frac{1}{2}}}V_{\rm db}\exp\left(-\frac{BE_{\rm g}^{\frac{3}{2}}}{\xi_{\rm JN}}\right)$$
$$\xi_{\rm JN} = \sqrt{\frac{2qN_{\rm halo}N_{\rm d}^{\rm s}(V_{\rm db}+V_{\rm bi})}{\varepsilon_{\rm si}\left(N_{\rm halo}+N_{\rm d}^{\rm s}\right)}} \tag{3}$$

where W is the width, $\xi_{\rm JN}$ and $V_{\rm bi}$ are the electric field and the built-in-potential at the drain–substrate junction, respectively; $V_{\rm DB}$ is the drain–substrate bias; A and B are the physical parameters (given in [4]); $E_{\rm g}$ is the band-gap; $N_{\rm halo}$ and $N_{\rm S/D}$ are the effective doping density in the substrate (halo doping) and source/drain (S/D) side of the junctions, respectively. From (4), it can be observed that the junction BTBT increases exponentially with an increase in the doping density (particularly higher halo doping) and the drain–substrate bias ($|V_{\rm DB}|$).

4) Total Leakage: The total leakage in a device is principally due to the three different leakage components mentioned above. Hence, to analyze the leakage current in CMOS logic circuits, a device can be considered as a combination of the voltage-controlled current sources (Fig. 4). In Fig. 4, each voltage-controlled current source represents a leakage mechanism. In the "OFF" state ($V_{\rm gs}=0$, $V_{\rm ds}=V_{\rm DD}$), the total leakage ($I_{\rm total}$) is determined by the subthreshold, the junction tunneling, and the gate–drain overlap leakage current [2]. In the "ON" state ($V_{\rm gs}=V_{\rm DD}$, $V_{\rm ds}=V_{\rm DD}$ or 0), the gate–channel leakage current of a device contributes to the total leakage.

C. Interdependence of Leakage Components in a Device

The subthreshold, the gate, and the junction BTBT leakage depend on each other through the device geometry (particularly oxide thickness), the doping profile, and the temperature [1]–[4]. Increasing the "halo" doping concentration increases the junction BTBT (by increasing the junction field), whereas it reduces the subthreshold current (by reducing the SCE) [1]–[4]. The gate leakage is insensitive to the halo doping concentration. Increasing the oxide thickness reduces the gate leakage (lower probability of tunneling), but increases the subthreshold leakage (higher SCE) in nanoscale transistors [1]–[4]. The oxide thickness does not strongly modify the junction BTBT leakage. The different leakage components show different temperature dependence. The subthreshold current increases exponentially with temperature, whereas the gate tunneling current is almost insensitive to a temperature variation. Due to the reduction of the band gap at a higher temperature, the junction BTBT increases (slowly) with the temperature [3].

III. Interaction of Different Leakage Components in a Circuit

In a logic circuit, different leakage components interact with each other through the internal nodes. Such an interaction changes the internal node voltages in a circuit, and hence, modifies the leakage of individual logic gates in a logic circuit.

A. Estimation of Leakage of a Logic Gate in Isolation

To estimate the leakage of a logic gate considering the interaction of different components of leakage within a logic gate (intragate interaction), we first solve Kirchoff's current law (KCL) at intermediate nodes in the gate. For example, to estimate the leakage of an inverter (in isolation), first, the voltage at the output node N0 [V(N0)] is obtained by solving

$$I_{\rm dd\_P} + I_{\rm dd\_N} = 0 \tag{4}$$

where $I_{\rm dd\_N}$ and $I_{\rm dd\_P}$ are the drain currents of the NMOS and PMOS, respectively. Using the estimated intermediate node voltages [i.e., V(N0) for the inverter], the total leakage of the gate is estimated by adding the different leakage components. For example, in the case of the inverter, the total leakage is obtained as follows:

$$\begin{aligned}&\text{Input}=0,\quad \text{Output}=1 :\Rightarrow\\ &I_{\rm INV\_0} = I_{\rm sub\_N} + I_{\rm JN\_N} + I_{\rm gdo\_N}\\ &\qquad + I_{\rm gc\_P} + I_{\rm gdo\_P} + I_{\rm gso\_P}\\ &\text{Input}=1,\quad \text{Output}=0 :\Rightarrow\\ &I_{\rm INV\_1} = I_{\rm sub\_P} + I_{\rm JN\_P} + I_{\rm gdo\_P}\\ &\qquad + I_{\rm gc\_N} + I_{\rm gdo\_N} + I_{\rm gso\_N}\end{aligned} \tag{5}$$

where $I_{\rm INV\_0}$ and $I_{\rm INV\_1}$ are the leakage currents of the inverter with input "0" and "1," respectively. $I_{\rm sub\_N/P}$, $I_{\rm JN\_N/P}$, $I_{\rm gdo\_N/P}$, $I_{\rm gc\_N/P}$, and $I_{\rm gso\_N/P}$ represent the subthreshold, the junction, the gate–drain overlap, the gate–channel, and the gate–source overlap currents, respectively (subscript N represents the NMOS and subscript P represents PMOS device). An example of the impact of the intragate interaction of leakage

components on the total leakage of a logic gate is the "stacking effect" [10] and [11]. Due to the "stacking effect," the leakage of a logic gate is a strong function of the input vectors [10] and [11]. It has been shown in [10] that for a subthreshold leakage dominated device, the minimum leakage vector for a two-input NAND gate is "00," while for a gate-leakage dominated device, it is "10."

B. Origin Estimation of the Loading Effect

The method described in the previous section considers the intragate interaction of the leakage components. However, it does not consider the interaction of the leakage components of different logic gates (intergate interaction). To understand the impact of the intergate interaction of leakage at the circuit level, let us consider the circuit shown in Fig. 2. Since the output node N0 of the gate G is also connected to the inputs of inverters $\mathrm{G_{out1}}$ to $\mathrm{G_{out}}n$, the gate leakage from these inverters is added to node N0. Hence, the KCL at node N0 is modified to

$$I_{\mathrm{dd\_P}} + I_{\mathrm{dd\_N}} + \sum_{i=1,\ldots,N} [I_{\mathrm{gate}N-\mathrm{Gouti}} + I_{\mathrm{gate}P-\mathrm{Gouti}}] = 0. \tag{6}$$

From (6), it can be observed that $V(\mathrm{N0})$ also depends on the gate leakage of the NMOS ($I_{\mathrm{gate}N-\mathrm{Gouti}}$) and the PMOS ($I_{\mathrm{gate}P-\mathrm{Gouti}}$) device of inverters $\mathrm{G_{out1}}$ to $\mathrm{G_{out}}n$. The modification of $V(\mathrm{N0})$ due to the gate leakage of the gates $\mathrm{G_{out1}}, \ldots, \mathrm{G_{out}}N$ has following effects: 1) a change in the output voltage $[V(\mathrm{N0})]$ of inverter G modifies its leakage and 2) a change in the input voltage $[V_{\mathrm{IN}-\mathrm{Gouti}} \equiv V(\mathrm{N0})]$ of inverters $\mathrm{G_{out1}}$ to $\mathrm{G_{out}}n$ modifies their leakage. Hence, the leakage of a gate depends on its input [i.e., the total gate leakage ($I_{L-\mathrm{IN}}$) of the other gates connected to its input node] and the output loading [the total gate leakage ($I_{L-\mathrm{OUT}}$) of the other gates connected to its output node]. The input ($\mathrm{LD_{IN}}$) and output ($\mathrm{LD_{OUT}}$) loading effect can be defined as the change in the leakage of a logic gate due to its input and output loading currents, respectively. $\mathrm{LD_{IN}}$ and $\mathrm{LD_{OUT}}$ are expressed as follows:

$$\mathrm{LD_{IN}}(I_{L-\mathrm{IN}}) = \frac{L_{\mathrm{G}}(I_{L-\mathrm{IN}}) - L_{\mathrm{NOM}}}{L_{\mathrm{NOM}}}$$
$$\mathrm{LD_{OUT}}(I_{L-\mathrm{OUT}}) = \frac{L_{\mathrm{G}}(I_{L-\mathrm{OUT}}) - L_{\mathrm{NOM}}}{L_{\mathrm{NOM}}} \tag{7}$$

where L_{NOM} is the nominal leakage of a gate in isolation (i.e., without any output and input loading), $L_{\mathrm{G}}(I_{L-\mathrm{IN}})$ is the leakage of the gate with input loading $I_{L-\mathrm{IN}}$, and $L_{\mathrm{G}}(I_{L-\mathrm{OUT}})$ is the leakage of the gate with output loading $I_{L-\mathrm{OUT}}$. The overall loading effect ($\mathrm{LD_{ALL}}$) depends on both input and output loading and is given by

$$\mathrm{LD_{ALL}}(I_{L-\mathrm{IN}}, I_{L-\mathrm{OUT}}) = \frac{L_{\mathrm{G}}(I_{L-\mathrm{IN}}, I_{L-\mathrm{OUT}}) - L_{\mathrm{NOM}}}{L_{\mathrm{NOM}}}. \tag{8}$$

For the logic gates with multiple inputs, there will be an $\mathrm{LD_{IN}}$ associated with each input.

C. Loading Effect in an Inverter

Fig. 5 shows the variation of $\mathrm{LD_{IN}}$ and $\mathrm{LD_{OUT}}$ of an inverter with a variation in $I_{L-\mathrm{IN}}$ and $I_{L-\mathrm{OUT}}$ at different input conditions. It can be observed that the impact of loading increases with an increase in the loading currents and it depends on the input condition of the inverter. The input loading current modifies the voltage at the input node of the inverter. If the input of G in Fig. 2 is at "0" [i.e., $V(\mathrm{IN}) = 0$ V], the gate leakage of $\mathrm{G_{IN1}}, \ldots, \mathrm{G_{IN}}n$ increases $V(\mathrm{IN})$ from 0 V. If the input is at "1" [i.e., $V(\mathrm{IN}) = V_{\mathrm{DD}}$], the gate leakage through $\mathrm{G_{IN1}}, \ldots, \mathrm{G_{IN}}n$ reduces the voltage from V_{DD}. This increases the $|V_{\mathrm{GS}}|$ of the "OFF" transistor (i.e., NMOS at input = "0" and PMOS at input = "1"), thereby increasing the subthreshold leakage of inverter G. On the other hand, it marginally reduces the gate currents of the PMOS and the NMOS by reducing their $|V_{\mathrm{GD}}|$ (PMOS and NMOS) and $|V_{\mathrm{GS}}|$ (of the PMOS at input = "0" and the NMOS at input = "1"). Since the junction leakage is a weak function of the gate voltage, the input loading has a minimal impact on the junction leakage [2]. For the output loading, V_{OUT} increases from 0 V when the output is "0" and decreases from V_{DD} when output is "1." This reduces 1) $|V_{\mathrm{DS}}|$ of the "OFF" transistor (PMOS when output = "0" and NMOS when output = "1"), thereby reducing the subthreshold leakage; 2) $|V_{\mathrm{GD}}|$ of the PMOS and NMOS, thereby reducing the gate leakage; and 3) $|V_{\mathrm{DB}}|$ of the transistor contributing to the junction BTBT (e.g., PMOS when output is "0" and NMOS when output is "1"), thereby reducing the junction leakage. Hence, due to the input loading, the subthreshold leakage increases, the gate leakage decreases, and the junction leakage remains almost constant (Fig. 5). On the other hand, due to the output loading, all the three components of the leakage decrease (Fig. 5). It can also be observed from Fig. 5 that the input loading effect ($\mathrm{LD_{IN}}$) is most pronounced in the subthreshold leakage as it changes the V_{gs} of the "OFF" transistor. Output loading ($\mathrm{LD_{OUT}}$) has the strongest impac t on the junction leakage by changing $|V_{\mathrm{DB}}|$ of the transistor, which contributes to the junction BTBT. The effect of loading is minimal on the gate leakage.

The input loading effect is more observable with an input of "0" (Fig. 5). This is due to the fact that with an input of "0," the subthreshold leakage is principally contributed by the "OFF" NMOS transistor. The "OFF" PMOS transistor determines the subthreshold leakage with an input of "1." It can be observed from Table I that the NMOS device has a better subthreshold slope compared to the PMOS device [i.e., lower m in (1)]. Hence, the voltage change in node IN (i.e., an increase in $|V_{\mathrm{gs}}|$) results in a higher increase in the subthreshold leakage of the NMOS device. On the other hand, it can be observed from Table I that the PMOS device has a (marginally) higher DIBL (i.e., higher SCE) compared to the NMOS device. Hence, the V_{ds} sensitivity of the subthreshold leakage is higher in the PMOS device compared to that of the NMOS device. Since the output loading modifies the V_{ds} of a transistor, it has a (slightly) stronger impact on the PMOS (i.e., when input = "1" and output = "0") subthreshold leakage. The impact of output loading on junction BTBT is also stronger with output "0" (PMOS determines the junction BTBT) than with output

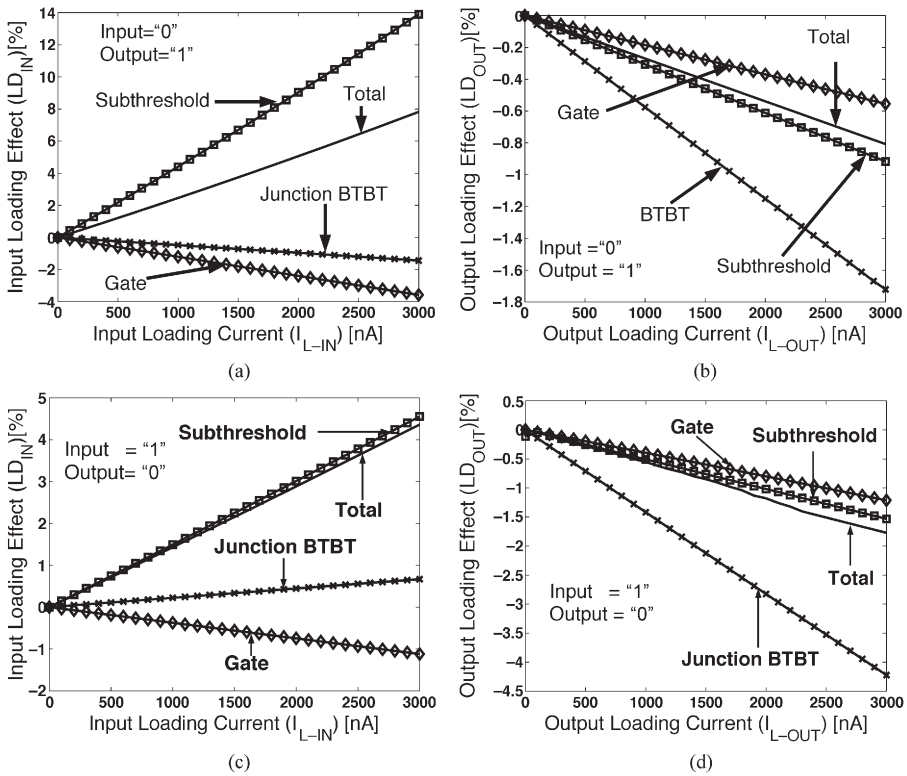

Fig. 5. Input and output loading effects with different inputs to an inverter. (a) Input loading effect for input = "0." (b) Output loading effect for input = "0." (c) Input loading effect for input = "1." (d) Output loading effect for input = "1."

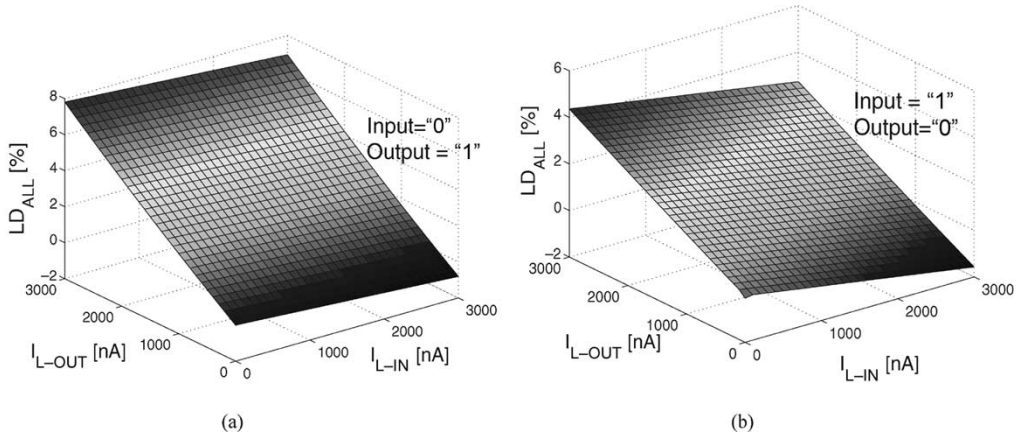

Fig. 6. Variation of leakage due to loading at both input and output with (a) input "0" and (b) input "1."

"1" (NMOS determines junction BTBT). This is due to the fact that the PMOS has a larger junction BTBT current [3]. Consequently, the effect of the output loading is higher with an output of "0." Fig. 6 shows the loading effect considering both the input and output loading (i.e., $\mathrm{LD_{ALL}}$). It can be observed that $\mathrm{LD_{ALL}}$ is normally higher with input = "0."

D. Loading Effect in a NAND Gate

The analysis of the loading effect on the two-input NAND gate shows the input vector dependence of the loading effect (Fig. 7). From Fig. 7, it can be observed that the input loading is higher if at least one of the inputs is at "0" (i.e., with vectors "00," "01," and "10"). This is because of the fact that the input loading has a stronger effect on the subthreshold leakage of an "OFF" NMOS. Due to the reduction in the subthreshold leakage by the stacking effect [10], the input loading effect is less pronounced with an input of "00" compared to that with inputs "01" or "10." As observed in the case of the inverter, the effect of output loading is higher with the output equal to "0." Moreover, depending on the input vector, the loading effect may increase or reduce the total leakage of the gate.

IV. Estimation of the Circuit Leakage Considering the Loading Effect

Traditionally, the leakage current in a circuit is calculated by determining the individual leakage values for each gate and accumulating them. This procedure is valid, assuming that the leakage current in a gate is not affected by the leakage of the

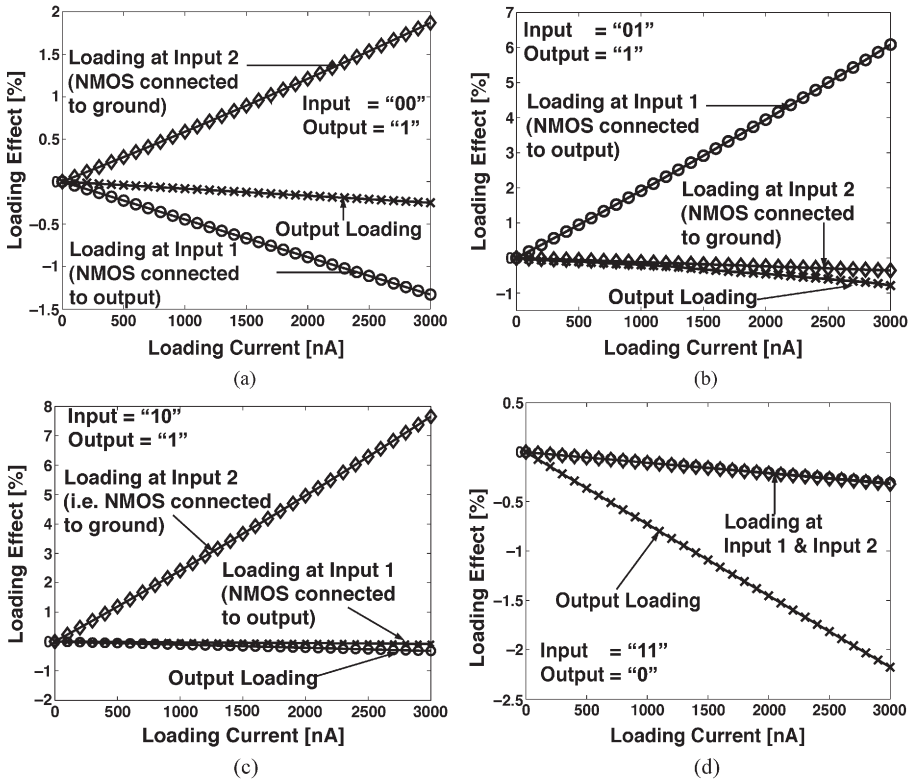

Fig. 7. Loading effects (input and output) in the total leakage of a NAND gate under different input vectors.

other gates. However, due to the loading effect, the leakage of a gate depends on the leakage of the other gates. Hence, we need to consider the propagation of the loading effect across logic gates for an accurate circuit-level estimation of the leakage current in a nanoscale CMOS.

To estimate the leakage of a circuit considering loading effect, we need to simultaneously solve a set of KCL equations with n variables, where n is the number of internal nodes of the circuit. To understand how the need for solving a large number of simultaneous equations (which is computationally very complex and expensive) can be eliminated, let us first discuss how the loading effect propagates across logic levels. Let us consider the output loading effect for the simple circuit shown in Fig. 2. From Fig. 2, it can be observed that leakage of gate $\mathrm{G_{out1}}$ is determined by the leakage of gates $H_{\mathrm{out1}}, \ldots, H_{\mathrm{out}p}$ (i.e., the "output loading effect" on $\mathrm{G_{out1}}$). Similarly, the leakage of inverter $\mathrm{G_{out1}}$ modifies the leakage of inverter G (the "output loading effect" on G). However, as discussed in Section III (see Fig. 5), the output loading of inverter $\mathrm{G_{out1}}$ has a weak impact on its gate leakage. Hence, the loading of inverter $\mathrm{G_{out1}}$ by inverters $H_{\mathrm{out1}}, \ldots, H_{\mathrm{out}p}$ has a negligible impact on the voltage of the node N0. Therefore, the output loading effect on inverter G depends only on the inverters connected directly to its output (i.e., $\mathrm{G_{out1}}, \ldots, \mathrm{G_{out}}n$ in Fig. 2). In other words, the propagation of the loading effect (output loading effect in this case) beyond one level is negligible. A similar argument can also be made to explain that input loading effect does not propagate strongly beyond one level.

Based on the above observation, we have developed an efficient algorithm that estimates the different components of leakage, considering the loading effect (Fig. 8). We start with a graph representing the circuit, with each vertex representing a logic gate and each edge representing a net. First, the vertices in the graph are topologically sorted [13] and the leakage values are initialized to zero. The logic values from the primary inputs

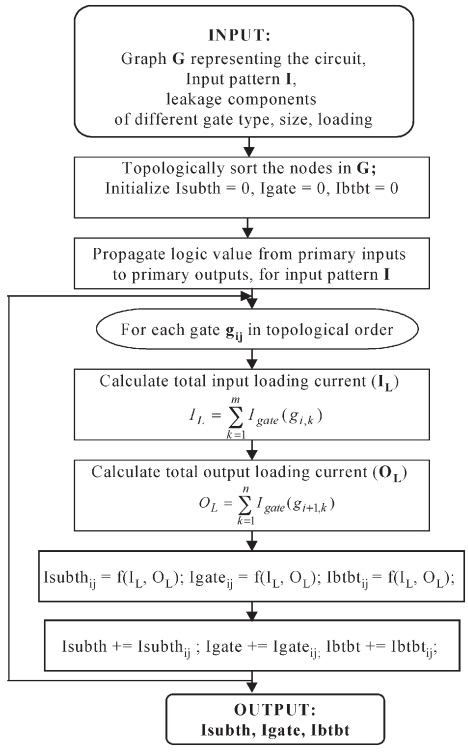

Fig. 8. Flowchart to estimate the leakage components for a circuit considering the loading effect.

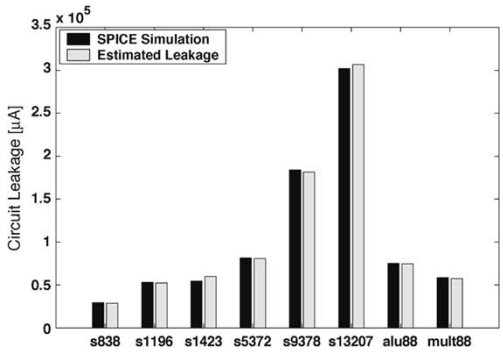

Fig. 9. Comparison of leakage estimation using the proposed procedure with results from SPICE simulation.

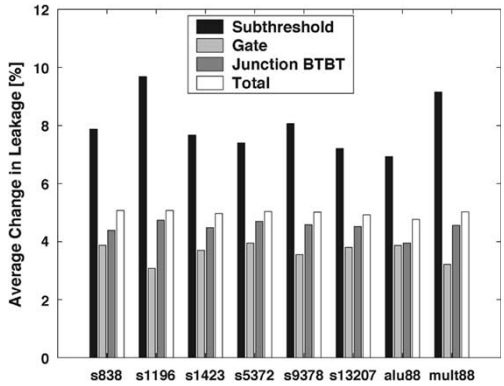

Fig. 10. Estimation of leakage using the proposed procedure showing the average change in leakage due to loading effect of over 100 random vectors.

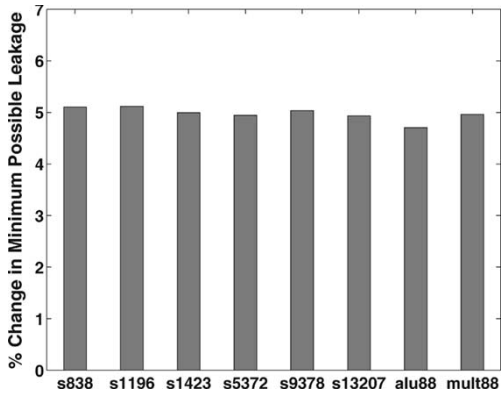

Fig. 11. Change in the minimum possible leakage for a set of benchmark circuits due to the loading effect estimated using the proposed method.

for an applied input vector are then propagated to the primary outputs. This assigns a logic state to each gate in the circuit. Next, we compute the total input and output loading current due to the gate leakage of the corresponding gates for each node in the graph in topological order. Using the computed input and output loading currents, the leakages of the individual logic gates are calculated. Finally, the leakage values of all logic gates are added to obtain the total circuit leakage.

The algorithm is implemented using the C-programming language and is tested on six ISCAS89 benchmark circuits, an 8-bit Wallace-tree multiplier (800 logic gates), and an 8-bit integer arithmetic logic unit (600 logic gates). We estimate the leakage of the circuit for 100 random input vectors at $T = 300$ °K. The leakage models for the different logic gates with and without the loading effect were obtained from a device simulation for a subthreshold leakage dominated device. Leakage values generated by the algorithm closely match results obtained from the SPICE simulations (Fig. 9), while being about 1000 times faster than SPICE in run time. The average variations in the leakage of the different circuits are shown in Fig. 10. It can be observed from Fig. 10 that the consideration of the loading effect using the proposed method improves the accuracy of the estimation of the subthreshold leakage, the gate leakage, and the junction leakage by 8%, 4.5%, and 3.6%, respectively. It was also observed that the loading effect increases the subthreshold leakage and decreases the junction and the gate leakage. Moreover, in a large circuit, the loading effect increases the total leakage of some logic gates while decreasing that of the other gates. This is due to that fact that the loading effect on the leakage of a gate depends on its input and output states. Due to these factors, the effect of loading on the total leakage of a logic circuit is reduced. The proposed method improves the estimation accuracy of the total leakage by 5%. The proposed method is also applied to estimate the input vector that results in the minimum total leakage in different circuits (required in input vector control techniques for leakage reduction [10], [14]). It is observed that the loading effect modifies the best input vector (i.e., the input vector that results in minimum leakage) and the minimum possible leakage in a logic circuit. The consideration of loading effect while selecting the best input vector improves estimation accuracy of the minimum leakage value (for varying inputs) and thus improves the leakage saving by input vector control. As shown in Fig. 11, the average change in the minimum leakage due to loading effect is about 5%.

V. Variation in the Loading Effect

The strength of the loading effect depends on the relative strengths of the gate, the subthreshold, and the junction leakages. An increase in the gate leakage increases the loading currents. A higher subthreshold and junction current increases the "effect of loading" on the leakage of a logic gate. In this section, we briefly discuss the effect of the strengths of different leakage components on the loading effect of a gate. To simplify the analysis, we present results for the inverter here. However, NAND and NOR gates also show a similar trend.

A. Effect of the Relative Strengths of Leakage Components

It has been discussed in Section II that the relative magnitudes of the different leakage components in a device strongly depend on the doping profile and the oxide thickness. We have designed three different devices with an equal total leakage, but with different individual leakage components. The subthreshold leakage, the gate leakage, and the junction leakage dominate the total leakage in devices $D_{25-\mathrm{S}}$, $D_{25-\mathrm{G}}$, and $D_{25-\mathrm{JN}}$, respectively (the total leakage is same in all three devices). Fig. 12 shows the input and output loading effect of an inverter designed with these three devices. It has been discussed earlier that the input loading has the strongest impact on the

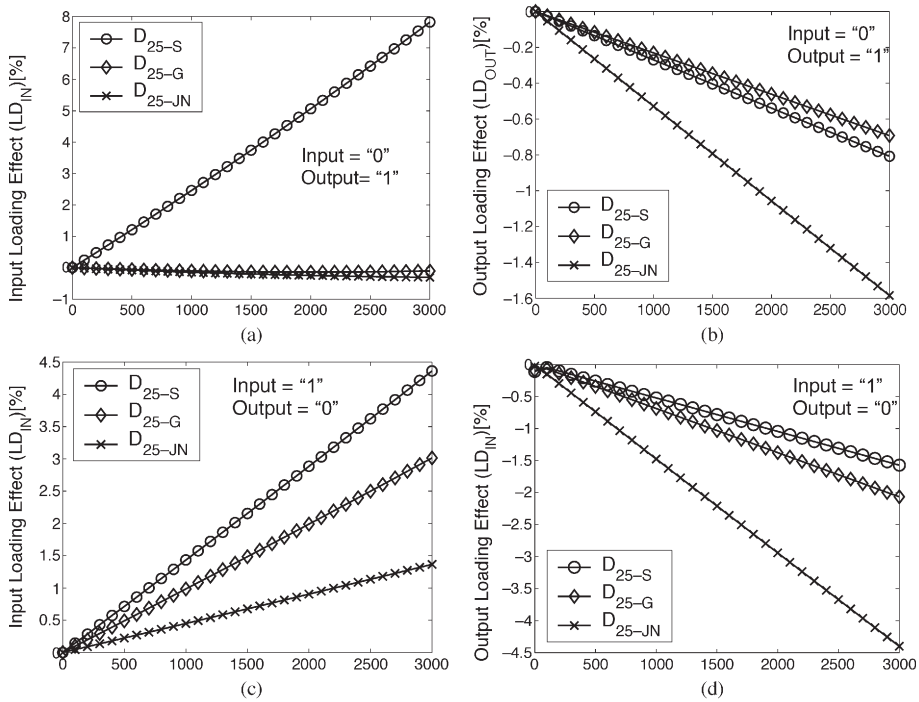

Fig. 12. Loading effects for different devices. (a) Input and (b) output loading effect with input "0" and (c) input and (d) output loading effect with input "1."

subthreshold leakage. Hence, the input-loading effect is most pronounced in the inverter designed with $D_{25-\mathrm{S}}$ (subthreshold leakage dominated device). The input loading has a weaker impact on the inverters designed with $D_{25-\mathrm{JN}}$ (a junction leakage dominated de vice) and $D_{25-\mathrm{G}}$ (gate-leakage dominated device). On the other hand, the output loading effect is most pronounced in the inverter designed with $D_{25-\mathrm{JN}}$ (since the junction BTBT is the strongest function of the output loading among the three different leakage components). In general, the loading has the least impact on the inverter designed with the gate-leakage dominated device.

B. Impact of Temperature on the Loading

It was explained in Section II that the subthreshold, the gate, and the junction BTBT current have different temperature dependencies. Hence, it is expected that the loading effect will also depend on the temperature. Since the gate leakage is a weak function of temperature [3], the "cause of the loading" does not increase significantly with temperature. But, the "effect of loading" is observed in the subthreshold and the junction leakage, which increases with an increase in the temperature. Let us consider the loading effect of inverter G in Fig. 2, with an input $= 0$ at different temperatures. An increase in the temperature exponentially increases the subthreshold leakage in a device, which has following impact on loading.

1) The subthreshold leakage of inverter G increases at a higher temperature [see (1)].
2) The contribution of the subthreshold current and the junction current of the PMOS of inverter D1 to node IN (i.e., input of G and output of D1) increases at a higher temperature. Hence, the voltage rise at the node IN (i.e., $|V_{\mathrm{GS}}|$ of the NMOS) increases, thereby further increasing the subthreshold current of inverters G, $\mathrm{G}_{\mathrm{in}1}, \ldots, \mathrm{G}_{\mathrm{inn}}$.

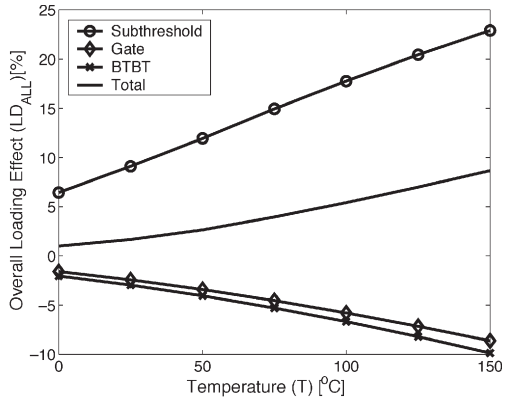

Fig. 13. Impact of the temperature on the overall loading effect $(\mathrm{LD}_{\mathrm{ALL}})$ of an inverter (input $=$ "0," output $=$ "1").

3) The increase in the input voltage of inverter G reduces its output voltage (due to larger subthreshold current of NMOS). This reduction in the output voltage reduces the gate and the junction BTBT current (as explained in Section III).

Hence, at a higher temperature, the effect of loading on all the different leakage components increases (Fig. 13). It can be observed that due to the increase in the voltage rise at node IN with an increase in temperature, the subthreshold leakage of inverter G increases, while its gate leakage and junction leakage are reduced. Hence, the loading effect on the total leakage is less sensitive to an increase in temperature (Fig. 13).

C. Effect of Parameter Variations

The variations in the process parameters [e.g., the channel length (L), the oxide thickness $(T_{\mathrm{ox}}$, the threshold voltage (V_{th}), the supply voltage (V_{DD}) etc.] result in a large variation

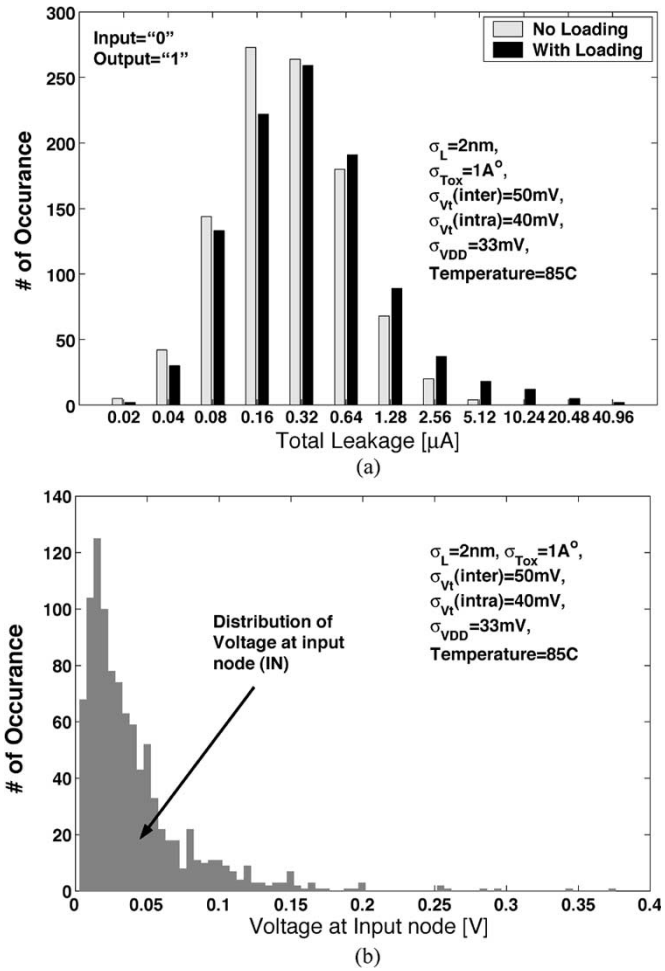

Fig. 14. Distribution of the total leakage components of an inverter (input = "0" and output = "1") with and without loading (input loading of six inverters and output loading of six inverters). (a) Distribution of the total leakage. (b) Distribution of the input node voltage.

in the leakage in transistors and in the logic gates [12]. The subthreshold leakage is very sensitive to the process variation, whereas the gate leakage and the junction BTBT leakage are less sensitive [12]. The application of random variation in L, $V_{\rm th}$, and $T_{\rm ox}$ of different transistors and in $V_{\rm DD}$ results in significant variations in the different leakage components and the total leakage (obtained through Monte Carlo simulations in SPICE for an inverter with an input = "0" and an output = "1"). The loading effect increases the spread of the distribution of the total leakage of the inverter as shown in Fig. 14(a). This is primarily due to the fact that the parameter variation modifies the gate leakage of the inverters $\rm G_{in1}, \ldots, G_{inn}$, the subthreshold, and the junction leakage of the PMOS in the inverter D1. Hence, with parameter fluctuations, the (input) loading current at the input node of the inverter G (i.e., node IN of G in Fig. 2) gets modified, resulting in a variation in the voltage at that node. Hence, due to loading effect, the parameter variations result in a significant variation in the input voltage of the inverter G, as shown in Fig. 14(b). It should be remembered that if the inverter G is considered in isolation, the input voltage is fixed at 0 V. The variation in the voltage at node IN increases the variation in the leakage of the inverter. Hence, the loading effect amplifies the effect of the process variation on leakage.

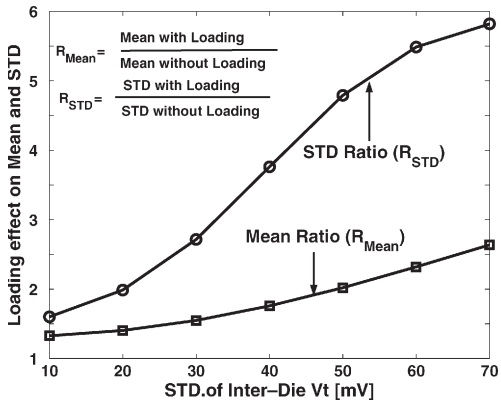

Fig. 15. Effect of loading on the mean and the standard deviation of the total leakage of an inverter (input = "0").

It was also observed that the loading effect does not strongly modify the distribution of the voltage of the output node N0 and the modification in the leakage spread is principally due to the spread in the input node voltage. Since the loading effect primarily increases the spread of the voltage at the input node (IN), the maximum modification is observed in the distribution of the subthreshold leakage. From the above discussion, it can be concluded that the mean and the standard deviation of the leakage of a gate is a function of the loading effect. The effect of loading on the mean and the standard deviation is estimated as follows:

$$R_{\rm Mean} = \frac{\rm mean\left(L(I_{L-IN}, I_{L-OUT})\right)}{\rm mean(L_{NOM})}$$
$$R_{\rm STD} = \frac{\rm std\left(L(I_{L-IN}, I_{L-OUT})\right)}{\rm std(L_{NOM})} \quad (9)$$

where $\rm mean[L(I_{L-IN}, I_{L-OUT})]$ and $\rm std[L(I_{L-IN}, I_{L-OUT})]$ represent the mean and the standard deviation of leakage with loading, $\rm mean(L_{NOM})$ and $\rm std(L_{NOM})$ represent the mean and the standard deviation of leakage without loading. The effect of loading on the mean and standard deviation increases with an increase in the parameter variation (Fig. 15). It can also be observed that the mean and the standard deviation of the leakage distribution considering loading are considerably higher than that without loading. As explained earlier, the increase in the mean and the standard deviation is principally attributed to the variation in the input node voltage when the loading effect is considered. This indicates that the worst case (considering the parameter variation) value of leakage significantly increases due to the loading effect.

VI. Conclusion

In this paper, we have analyzed the effect of loading on different leakage components of a circuit. It was observed that due to the presence of the gate leakage, the leakage of a logic gate depends on the leakage of the other gates connected to its input (input loading) and output (output loading). Our analysis shows that the loading effect modifies the leakage of a logic gate by 5%–8%. The effect of loading on the leakage of a logic gate depends on the device geometry and the doping profile, and the effect varies with temperature and parameter fluctuations. We have developed an algorithm for the fast estimation of circuit

leakage considering the loading effect. The proposed algorithm improves the accuracy of the estimation of different leakage components and the total leakage in a logic circuit. It was observed that in large circuits, depending on the applied input vector, the leakage of different logic gates moves in different directions (some increase and some decrease). Due to this cancellation effect, we observed that the net change in the overall leakage due to the loading effect reduces in large circuits. However, we observed that the leakage spread of a logic gate due to a parameter variation increases significantly due to loading effect. Hence, in nanoscale circuits with large variations in process parameters, the consideration of the loading effect is important.

REFERENCES

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

[2] S. Mukhopadhyay *et al.*, "Accurate estimation of total leakage current in scaled CMOS logic circuits based on compact current modeling," in *Proc. Design Automation Conf. (DAC)*, Anaheim, CA, 2003, pp. 169–174.

[3] ——, "Accurate estimation of total leakage in nanometer-scale bulk CMOS circuits based on device geometry and doping profile," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 24, no. 3, pp. 363-381, Mar. 2005.

[4] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[5] *MEDICI: Two-Dimensional Semiconductor Device Simulation Program*. Mountain View, CA: Synopsys Inc.

[6] "Well-Tempered," *Bulk-Si NMOSFET Device*. [Online]. Available: http://www-mtl.mit.edu/researchgroups/Well

[7] *AURORA: Device Characterization and Parameter Extraction System*. Mountain View, CA: Synopsys Inc.

[8] Z. Liu *et al.*, "Threshold voltage model for deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 40, no. 9, pp. 86–95, Jan. 1993.

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

[10] S. Mukhopadhyay *et al.*, "Gate leakage reduction for scaled devices using transistor stacking," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 4, pp. 716–730, Aug. 2003.

[11] M. Johnson *et al.*, "Models and algorithms for bounds on leakage in CMOS circuits," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 18, no. 6, pp. 714–725, Jun. 1999.

[12] S. Mukhopadhyay *et al.*, "Modeling and estimation of total leakage current in CMOS devices considering the effect of parameter variation," in *Proc. Int. Symp. Low Power Electronics and Design (ISLPED)*, Seoul, Korea, 2003, pp. 172–175.

[13] T. Cormen, C. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambrige, MA: MIT Press, 1990.

[14] A. Abdollahi, F. Fallah, and M. Pedram, "Leakage current reduction in CMOS VLSI circuits by input vector control," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 140–154, Feb. 2004.

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

He worked as an intern at the High Performance Circuit Design Department, International Business Machines Corporation (IBM) T. J. Watson Research Lab, Yorktown Heights, NY, in the summer of 2003, 2004, and 2005. His research interests include the analysis and design of low-power and robust circuits using nanoscaled CMOS and circuit design using double gate transistors.

Mr. Mukhopadhyay is the recipient of the IBM Ph.D. Fellowship Award for 2004–2005 and the "Best Paper Award" at the 2003 IEEE Conference on Nanotechnology and the 2004 International Conference on Computer Design.

Swarup Bhunia (S'00–M'05) received the B.E. degree (with honors) from Jadavpur University, Calcutta, India, in 1995, the M.Tech. degree from the Indian Institute of Technology, Kharagpur, India, in 1997, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 2005.

He has worked in the semiconductor industry on register-transfer level synthesis, verification, and low-power design for about three years. Currently, he is an Assistant Professor at the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH.

Dr. Bhunia has served the technical program committee of Design Automation and Test in Europe (2006) and in the program committee of International Online Test Symposium (2005). Currently, he is serving as a Program Committee Member of the Test Technology Educational Program. He was the recipient of the Semiconductor Research Corporation Technical Excellence Award (2005), Best Paper Award in the International Conference on Computer Design (2004), Best Paper Award in the Latin American Test Workshop (2003), and Best Paper Nomination in the Asia and South Pacific Design Automation Conference (2006).

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 (IIT), Kharagpur, India, in 1983 and the Ph.D. degree from University of Illinois at Urbana-Champaign, Urbana, in 1990.

He was with the Semiconductor Process and Design Center of Texas Instruments, Dallas, TX, where he worked on field-programmable gate array (FPGA) architecture development and low-power circuit design. He joined the Electrical and Computer Engineering Department, Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor and holds the Roscoe H. George Professor of electrical and computer engineering. His research interests include very large scale integration (VLSI) design/computer-aided design (CAD) for nanoscale silicon and nonsilicon technologies, low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing. He is the author/coauthor of more than 300 papers in refereed journals and conferences and is the coauthor of two books on *Low Power CMOS VLSI: Circuit Design* (Wiley) and *Low Voltage, Low Power VLSI Subsystems* (McGraw-Hill). He is the holder of eight patents.

Dr. Roy is the recipient of the National Science Foundation Career Development Award in 1995, the IBM Faculty Partnership Award, the ATT/Lucent Foundation Award, the 2005 Semiconductor Research Corporation (SRC) Technical Excellence Award, the SRC Inventors Award, and the Best Paper Award at the 1997 International Test Conference, the IEEE 2000 International Symposium on Quality of IC Design, the 2003 IEEE Latin American Test Workshop, the 2003 IEEE Conference on Nanotechnology, the 2004 IEEE International Conference on Computer Design, and the 2005 IEEE Circuits and System Society Outstanding Young Author Award (Chris Kim). He is a Purdue University Faculty Scholar. He is the Chief Technical Advisor of Zenasis Inc. and a Research Visionary Board Member of Motorola Labs (2002). He has been on the Editorial Board of *IEEE Design and Test of Computers,* IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS. He was a Guest Editor for a special issue on low-power VLSI in the *IEEE Design and Test of Computers* (1994) and IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS (June 2000), IEE PROCEEDINGS–COMPUTERS AND DIGITAL TECHNIQUES (July 2002).