Optimal Body Bias Selection for Leakage Improvement and Process Compensation Over Different Technology Generations

Cassondra Neau and Kaushik Roy
Purdue University Department of Electrical & Computer Engineering
1285 Electrical Engineering Building
West Lafayette, IN 47907 USA
{crotty, kaushik}@ecn.purdue.edu

ABSTRACT

We present techniques to determine the optimal body bias (forward or reverse) to minimize leakage current and compensate process variations in scaled CMOS technologies. A circuit trades off sub-threshold leakage with band-to-band tunneling leakage at the source/drain junctions to determine the optimal substrate bias for different technology generations and under process variations. Using optimal body bias results in 43% and 42% savings in leakage for predictive 70nm and 50nm NMOS devices, respectively. This technique also reduces the effects of die-to-die and intra-die process variations in transistor length and supply voltage by 43% and 60%, respectively, in 50nm NMOS devices, resulting in improved yield.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles – *advanced technologies, algorithms implemented in hardware, VLSI (very large scale integration).*

General Terms

Design, Theory.

Keywords

CMOS scaling, leakage current, leakage components, band-to-band tunneling, process variation, process compensation, substrate bias, body bias.

1. INTRODUCTION

The desire for higher transistor densities and faster devices drives the trend of CMOS device scaling. As the supply voltage (VDD) is reduced along with device dimensions, the threshold voltage (V_{th}) must be commensurately reduced to maintain the desired performance improvement. This leads to a large standby or "off" current (I_{OFF}) that is consumed even though no logic operations are being performed. Reverse substrate (or body) bias in the "off"-state is one leakage reduction technique that has been successfully employed to reduce I_{OFF} [1]. By selectively applying reverse body bias (RBB) in the "off"-state, the threshold voltage is raised, reducing the sub-threshold leakage in the "off"-state without sacrificing performance in the "on"-state. But the scalability of this technique has been called into question [2,3]. The problem with RBB in ultra-small technologies is an increase in the short channel effect (SCE). RBB increases drain-induced barrier lowering (DIBL) and with highly doped substrates leads to significant band-to-band tunneling (BTBT) current at the source/drain junctions. These current components can eliminate any power-saving benefits from reverse body bias and even increase leakage in future technology generations. In addition, a fixed RBB leads to an increased sensitivity to process variations.

It has recently been proposed that forward body bias (FBB) be used for microprocessors in the active mode while applying no body bias (NBB) or RBB in the standby mode [4]. The FBB in the active mode improves performance and reduces sensitivity to variations in V_{th}, gate length, oxide thickness, and channel doping in the active mode. But the optimal bias condition (RBB, NBB, or FBB) for standby mode leakage minimization depends on the particular technology employed and is sensitive to process variations. Furthermore, it has been shown in [5] that correctly applying body bias reduces the impact of die-to-die and within die parameter variations. Thus, applying the optimal body bias leads to both minimum leakage current and improved yield.

In this paper we propose a simple circuit to determine the optimal off-state body bias. By trading off the sub-threshold leakage and the source/drain junction BTBT leakage, the circuit finds the lowest leakage bias condition for a wide range of bulk MOSFET technologies, taking into account the process parameters of the particular die or region of the die. Section 2 presents the effect of body bias on the various leakage components that are particularly important with scaling (sub-threshold, BTBT, and gate leakage). There are other leakage components in CMOS devices such as gate-induced drain leakage and punchthrough current, but the three components described in section 2 are the most significant ones for the normal modes of device operation. Section 3 derives the optimal ratio of these leakage components and presents our leakage trade-off circuit. Sections 4 and 5 present the results of applying the optimal body bias in terms of leakage minimization and process compensation, respectively. Finally, section 6 provides some conclusions from this work.

ISLPED'03, August 25–27, 2003, Seoul, Korea.

2. EFFECT OF BODY BIAS ON LEAKAGE COMPONENTS

Understanding the relative importance of key leakage components in scaled technologies and how each of these components is affected by body bias is a necessary prerequisite for developing a body bias based techniques to limit leakage.

2.1. Sub-threshold Leakage

Sub-threshold current is the weak inversion conduction current that flows between the source and drain of a MOSFET when the gate voltage is below V_{th}. Due to low threshold voltages, sub-threshold leakage dominates the off-state leakage of current MOSFETs. In modern short-channel devices, the depletion region of the drain interacts with that of the source near the surface of the channel to lower the source potential barrier. This effect is referred to as Drain-Induced Barrier Lowering (DIBL) and is responsible for a reduction in threshold voltage at high drain biases, resulting in increased sub-threshold leakage. DIBL is reduced in modern MOS devices by the insertion of highly doped regions in the substrate near the source and drain regions called halo implants. Halo implants also limit V_{th} roll-off, the reduction of threshold voltage with reduced channel length. Sub-threshold leakage is also modified by the body effect. Reverse biasing the substrate to source junction of a MOSFET widens the bulk depletion region. This increases the threshold voltage and thereby reduces the sub-threshold leakage. This threshold voltage increase with reverse body bias is known as the body effect. Forward biasing the substrate to source junction has the opposite effect on the depletion region and thus increases sub-threshold leakage. The sub-threshold leakage of a MOS device, taking into account weak inversion, DIBL, and the body effect, has been modeled as [6]:

$$I_{subth} = A \times e^{\frac{1}{m v_T}\left(V_G - V_S - V_{th0} - g' \times V_{app} + h \times V_{DS}\right)} \times \left(1 - e^{\frac{-V_{DS}}{v_T}}\right) \qquad (1)$$

where

$$A = m_0 C'_{ox} \frac{W}{L_{eff}} (v_T)^2 e^{1.8} e^{\frac{-\Delta V_{th}}{h v_T}} \qquad (2)$$

V_{th0} is the zero bias threshold voltage, and $v_T = kT/q$ is the thermal voltage. The body effect for small values of source to bulk voltages is represented by the term $\gamma' V_{app}$ in (1), where γ' is the linearized body effect coefficient and V_{app} is the applied reverse body bias. η is the DIBL coefficient, C_{ox} is the gate oxide capacitance, μ_0 is the zero bias mobility, and m is the sub-threshold swing coefficient of the transistor. ΔV_{th} is a term introduced to account for transistor-to-transistor leakage variations.

2.2. Souce/Drain Junction Band-to-Band Tunneling Leakage

MOS transistors have reverse biased *pn* junctions from the drain/source to the well. The reverse biased *pn* junctions give rise to minority carrier diffusion/drift current near the edge of the depletion region. This *pn* junction reverse bias leakage is a function of junction area and doping concentration. But if both the *n*- and *p*- regions are heavily doped (as in MOSFETS using heavily doped shallow junctions and halo doping to limit short-channel effects), band-to-band tunneling (BTBT) dominates the *pn* junction leakage [7]. In steeply graded junctions, the high electric field across the reverse-biased *pn* junction causes electrons to tunnel from the valence band of the *p*-region to the conduction band of the *n*-region as shown in Figure 1 [7]. This tunneling current density is given by [7]:

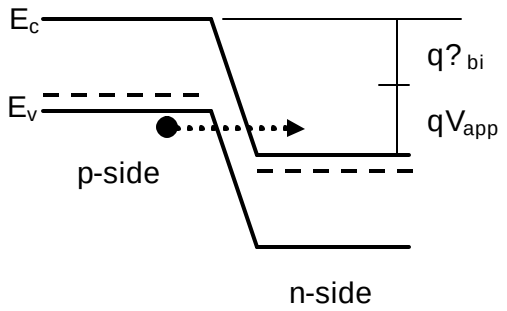

Figure 1. BTBT Electron tunneling from valence band of the *p*-side to conduction side of the *n*-side of a reverse-biased *pn*-junction.

$$J_{BTBT} = A \frac{E V_{app}}{E_g^{1/2}} exp\left(-B \frac{E_g^{3/2}}{E}\right)$$
$$A = \frac{\sqrt{2m^*}q^3}{4p^3\hbar^2}, \text{ and } B = \frac{4\sqrt{2m^*}}{3q\hbar} \qquad (3)$$

where m^* is effective mass of electron; E_g is the energy band-gap; V_{app} is the applied reverse bias; E is the electric field at the junction; q is the electronic charge; and $\hbar$ is $1/(2\pi)$ times Plank's constant. Assuming a step junction, the electric field at the junction is given by [7]:

$$E = \sqrt{\frac{2qN_aN_d(V_{app} + V_{bi})}{e_{si}(N_a + N_d)}} \qquad (4)$$

where N_a and N_d are the doping in the p and n side, respectively; e_{si} is the permittivity of silicon; and V_{bi} is the built in voltage across the junction. In scaled devices, high doping concentrations and abrupt doping profiles cause significant BTBT current through the drain-well junction.

As the reverse bias (V_{app}) is increased, the band-to-band tunneling current increases rapidly since V_{app} appears in E, the electric field at the junction. The exact dependence on V_{app} depends on the doping profile in the substrate. As stronger halo implants are used and as the halo implants are located closer to the source/drain region, the band-to-band tunneling current increases more rapidly with reverse bias. Forward bias reduces the BTBT current, but too much forward bias will cause excessive *pn* junction leakage as the junction becomes more weakly reverse biased.

2.3. Gate Leakage

The reduction of gate oxide thickness coupled with the resultant high electric field across the oxide results in significant tunneling through the gate oxide in scaled devices. The equation governing direct tunneling current density is [8]:

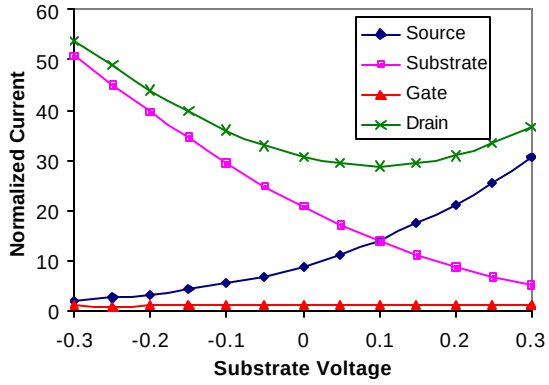

Figure 2. Effect of substrate bias on leakage components for a 70 nm predictive technology.

$$J_{DT} = AE_{ox}^2 \exp\left\{-\frac{B[1-(1-\frac{V_{ox}}{f_{ox}})^{3/2}]}{E_{ox}}\right\} \qquad (5)$$

where

$$A = \frac{q^3}{16p^2\hbar f_{ox}} \quad \text{and} \quad B = \frac{4\sqrt{2m^*}f_{ox}^{3/2}}{3\hbar q} \qquad (6)$$

and V_{ox} is the voltage drop across the oxide, ϕ_{ox} is the barrier height for electrons in the conduction band, and E_{ox} is the field across the oxide. There are three components of gate leakage: I_{gd} is the gate leakage between the gate and the drain, I_{gb} is the gate leakage between the gate and the substrate; and I_{gs} is the gate leakage between the gate and the source. For a transistor with 0 at the gate and VDD at the drain, gate leakage is dominated by I_{gd}.

As shown in (5) and in figure 2, gate leakage is far less sensitive to applied body bias than BTBT or sub-threshold current. Therefore applying optimal body bias will have a negligible effect on the gate leakage. Although gate leakage is becoming increasingly important, it will have to be controlled with other techniques such as high-κ dielectrics.

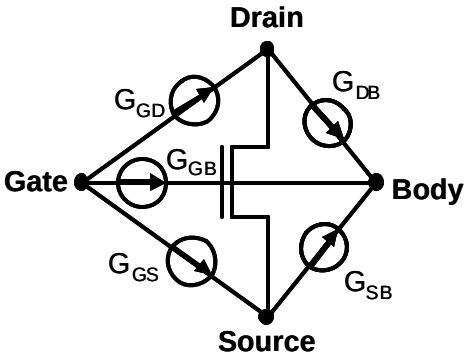

Figure 3. BSIM3 device with voltage-dependent current sources added for gate leakage and band-to-band tunneling currents.

3. OPTIMAL BODY BIAS FOR LEAKAGE REDUCTION

Figure 2 is the result of a BSIM3 simulation for a 70nm BPTM [9] NMOS device augmented with voltage controlled current sources to include the effects of gate leakage and BTBT (Figure 3). The gate leakage is modeled after the BSIM4 [10] gate leakage equations and the BTBT leakage at the source/drain to body junction is fit to the results of device simulation in Taurus [11]. Figure 2 shows that the off-state leakage is minimized at the body bias for which the sum of the leakage through the source (primarily sub-threshold leakage) and the leakage through the body (primarily BTBT) is minimized. If the body is more forward biased there will be excessive sub-threshold leakage adding to the total leakage; If the body is more reverse biased there will be excessive BTBT, also increasing the total leakage.

The location of this minimum leakage value is highly technology dependent. Figure 4 is the result of Taurus device simulation for 50nm (25nm Lmet) NMOS devices with different doping profiles. It shows that, even within a technology generation, the location of the minimum leakage body bias depends highly on the doping profile. This is true even for doping profiles of devices with nearly identical threshold voltages (18 & 19 in figures 5 & 6). The explanation for this variation is the relative contribution of sub-threshold leakage and source/drain junction BTBT leakage varies with doping profile.

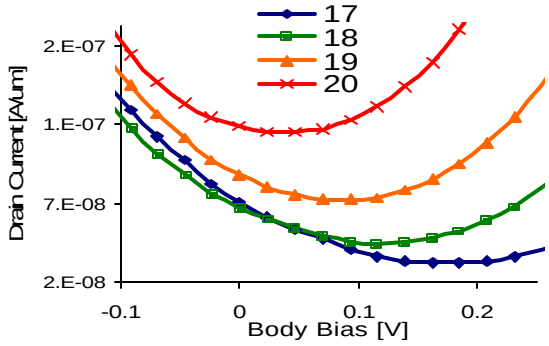

Figure 4. The relationship between substrate bias and total drain leakage for predictive 50 nm NMOS devices with different doping profiles. The profiles differ only in depth of peak halo doping concentration (17-20 nm below the oxide/silicon interface)

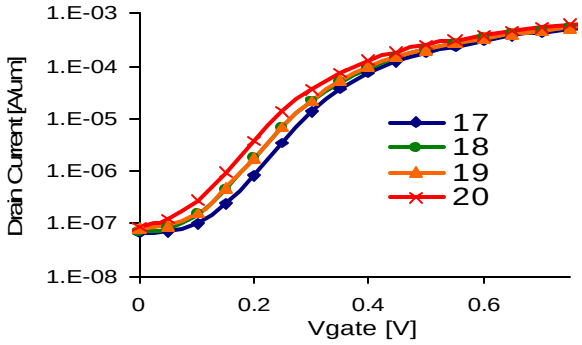

Figure 5. IV curves for doping profiles used in figure 4.

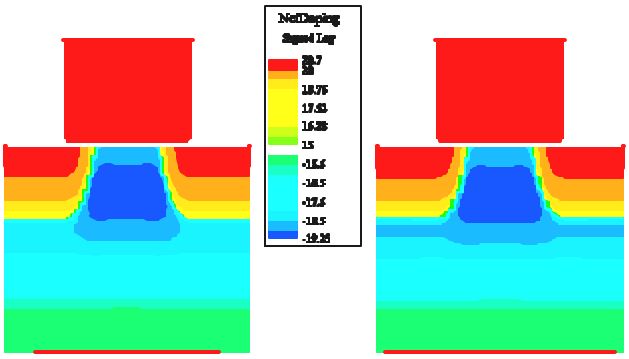

Figure 6. Doping profiles 18 (left) & 19 (right). In doping profile 18 (19), the peak halo doping location is 18nm (19nm) below the oxide-silicon interface.

From (1) and (3) it is theoretically possible to determine the substrate bias that will result in the minimum sub-threshold plus source/drain junction BTBT current, hence, the minimum overall leakage current. From these equations the rate of increase of sub-threshold current with increased body bias and the rate of decrease of BTBT current can be calculated and used to determine the ratio of sub-threshold current to BTBT current that results in a minimum overall leakage current. However, due to the complexity of calculating the electric field at a junction other than a step junction and the presence of process variations, it is impossible to accurately determine the precise ratio for minimum leakage in a real process technology.

To simplify the analysis, for a specific technology the substrate dependence of (3) can be approximated by the exponentially decaying function:

$$I_{BTBT} \approx A_b e^{-B_b V_B} \qquad (5)$$

where A_b and B_b are technology dependent constants and V_B is the applied substrate voltage. For the predictive 70nm NMOS device considered in figure 2, this approximation results in no more than 4% error in band-to band tunneling for -0.2 = V_B = 0.2.

Likewise, by combining terms in (1), the sub-threshold current's dependence on substrate voltage can be written as the exponentially increasing function:

$$I_{subth} = A_s e^{B_s V_B} \qquad (6)$$

where

$$A_s = A \times e^{\frac{1}{mv_T}(V_G - V_S - V_{th0} + h \times V_{DS})} \times \left(1 - e^{\frac{-V_{DS}}{v_T}}\right), \text{ and } B_s = \frac{g'}{mv_T}$$

The total leakage is the sum of the leakage contributions from gate leakage, band-to-band tunneling, and sub-threshold leakage ($I_{leak}=I_{gate}+I_{subth}+I_{BTBT}$). The substrate voltage for which the total leakage is minimized occurs when $\delta I_{leak}/\delta V_B=0$. Since $\delta I_{gate}/dV_B \approx 0$, the minimum leakage condition occurs when $B_s I_{subth} = B_b I_{BTBT}$. Therefore the ideal ratio of sub-threshold to band-to-band tunneling current is B_b/B_s.

For any technology this leakage ratio can be calculated. For the 70nm technology we are considering, $B_b/B_s = 0.75$. Therefore if gate leakage is negligible, the minimum leakage occurs when 43% of the leakage is sub-threshold leakage and 57% of the leakage is band-to-band tunneling. If gate leakage is non-negligible, the sub-threshold and band-to-band tunneling percentages will be smaller, but the ratio between them will still equal B_b/B_s.

The values of B_b/B_s varied with technology generation and doping profile, but were always in the range $0.5 = B_b/B_s = 2.0$. Therefore if the ratio is unknown, it can be approximated as 1.0. If the gate leakage is negligible, this value corresponds to 50% sub-threshold leakage and 50% band-to-band tunneling. Since both types of leakage are exponentially related to the applied body bias, this approximation results in only minor error, as shown in the following analysis.

The simple circuit shown in figure 7 is used to determine the point at which BTBT is half of the total leakage current. This circuit mirrors the leakage current through a single off NMOS transistor (N1) in a stack of off NMOS transistors (N2 & N3). The leakage from node A through N1 includes sub-threshold leakage through the source, BTBT through the substrate, and gate leakage from the drain to the gate of N1. The leakage from node B through N2 includes a significantly reduced sub-threshold leakage component through the source, BTBT through the substrate, and gate leakage from the drain to the gate of N2. The sub-threshold leakage through N2 is significantly less than the sub-threshold leakage through N1 due to the stack effect [12]. Assuming that the gate leakage is negligible, the leakage through N1 is the total off-state leakage (sub-threshold + BTBT) and the leakage through N2 is primarily BTBT leakage. Because the width of P1 is twice that of P2, the voltages at nodes A and B will be equal if the leakage through N1 (total leakage) is twice the leakage through N2 (BTBT leakage). If V(A)>V(B), the BTBT leakage is greater than half the total drain leakage and the substrate voltage should be increased to minimize "off"-state leakage. If V(B)>V(A), BTBT is less than half of the total leakage (sub-threshold leakage dominates), and the substrate voltage should be decreased to minimize the leakage.

The leakage in this circuit will reflect the process variation in this region of the die and the effect of ambient temperature on the leakage components.

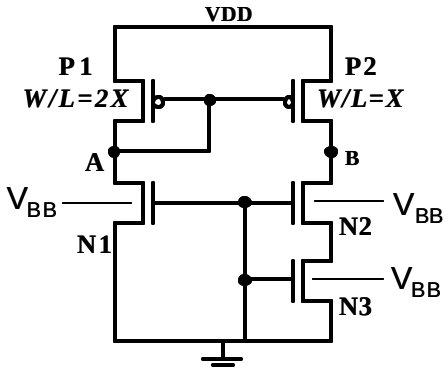

Figure 7. Current mirror circuit to determine the body bias for which source/drain junction BTBT is half of the total NMOS leakage.

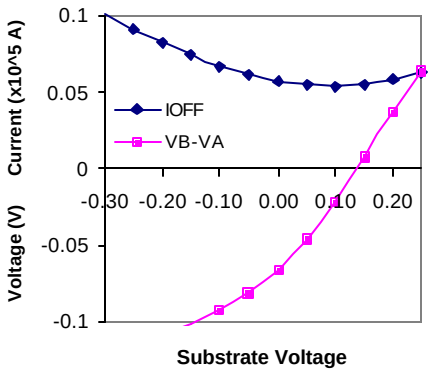

Figure 8. Results of body bias selecting circuit. The substrate voltage for which VB-VA=0 corresponds to the body bias that results in minimum leakage.

This circuit was simulated using the modified BPTM device shown in figure 3. As shown in figure 8, the BTBT current is half of the total leakage at a substrate voltage of 0.135V for the 70nm technology. The actual minimum overall leakage occurs at a substrate voltage of 0.100V. This is only 35 mV off of the best substrate bias using the ratio of 2:1 for the PMOS devices. If the PMOS devices had been sized with the ratio 1.75:1 (using B_b/B_s = 0.75), this error would have been eliminated. But even using the 2:1 PMOS sizing ratio gives a nearly ideal body bias that results in only 3% more leakage than the ideal case.

4. LEAKAGE REDUCTION WITH OPTIMAL BODY BIAS

To determine the leakage savings, device simulations were run in Taurus for predictive 70nm (44nm Lmet) and 50nm (25nm Lmet) NMOS transistors at 25°C and 70°C. The device structures were based on the profiles in [13] and the guidelines in the 2001 International Technology Roadmap for Semiconductors [14]. Tables 1 and 2 show that applying the optimal body bias (as determined by the ratio between sub-threshold and BTBT currents) results in a 43% and a 42% savings in leakage current (compared to the zero body bias case) at room temperature in 70nm and 50 nm devices, respectively. At 70°C, leakage reductions are 55% and 14% for 70nm and 50nm devices, respectively. Thus optimal body bias results in leakage savings both at room temperature, which is important for long periods in the idle mode and at elevated temperatures that occur during the active mode of operation. In addition, since forward bias results in the minimum leakage current for this 50nm device, the on-current is also improved for the 50nm device by applying the optimal body bias.

Table 1. 70 nm NMOS device simulation. Current values are normalized to the current of a 70nm device with Vgate=0 and Vsubstrate=0 at 25°C.

Temp (°C)	V_B (V)	I_{OFF} (normalized)	I_{ON} (normalized)
25	0	1	97115
25	-0.16	0.57	91005
70	0	5.14	120673
70	-0.20	2.30	118269

Table 2. 50 nm NMOS device simulation. Current values are normalized to the current of a 50nm device with Vgate=0 and Vsubstrate=0 at 25°C.

Temp (°C)	V_B (V)	I_{OFF} (normalized)	I_{ON} (normalized)
25	0	1	3478
25	0.15	0.55	3992
70	0	2.51	4044
70	0.09	2.15	4286

5. PROCESS COMPENSATION WITH OPTIMAL BODY BIAS

Because the leakage monitoring circuit is subject to the same process variations as the region of the die on which it is located, the leakage monitoring circuit will adjust the body bias value according to these variations. By partitioning the die into regions that are small enough to have similar process variations within that region, the leakage monitor will compensate for within-die as well as die-to-die process variations.

To determine the effects of process variations on the leakage reduction, two parameters were varied in the sample 70nm and 50nm device technologies. The gate length was varied by +/- 10% of the nominal gate length and the supply voltage was varied by +/- 0.1 V. Since these process variations affect sub-threshold and source/drain junction BTBT current differently, the optimal body bias is also affected by these changes. For example, figure 9 shows that substrate current (BTBT) is much more sensitive than the source (sub-threshold) current to variations in supply voltage. This results in a large change in optimal body bias with variations in supply voltage.

The results of applying optimal body bias to 50nm and 70nm devices with process variations are shown in table 3. Leakages for the smallest devices are reduced by 41% and 39% and leakages for the highest supply voltages are reduced 26% and 51%. Figure 10 shows how the distribution of leakage values is affected by applying the optimal body bias condition. In this figure, the channel lengths were assumed to be Gaussian distributed with the mean equal to the nominal gate length of 50nm and a sigma of 2.5nm. The resultant leakage values were fitted to a Gaussian distribution. Both the mean leakage value and the standard deviation were reduced by 41%. Similar results are obtained for variation in the doping profile and supply voltage. By reducing both the nominal leakage value and the spread of leakage values with process variations more devices meet the maximum leakage criteria. Furthermore, the leakage current is more consistent which is beneficial for device testing.

Table 3. Normalized off-current at 27° C for nominal NMOS transistors, transistors with gate lengths of +/- 10%, and supply voltage variation of 0.1V. Leakage values are normalized to a nominal device (70nm or 50nm) at V_B=0.

Device	V_B	Nom	L_{MIN}	L_{MAX}	V_{MIN}	V_{Max}
70 nm	0	1	1.66	0.77	0.87	1.21
70 nm	best	0.57	0.98	0.38	0.37	0.89
50 nm	0	1	1.74	0.66	0.45	1.98
50 nm	best	0.59	1.07	0.45	0.35	0.97

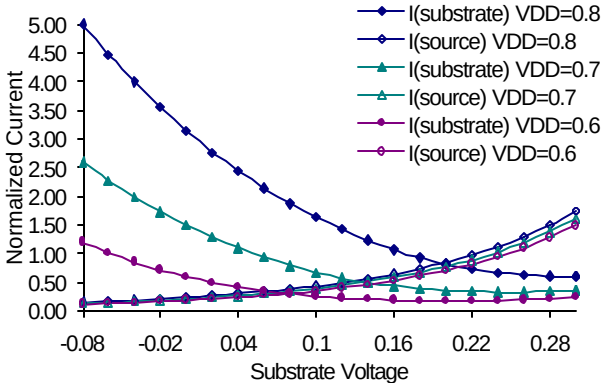

Figure 9. Sensitivity of substrate (BTBT) current and source (sub-threshold) currents with variations in supply voltage in 50 nm NMOS devices.

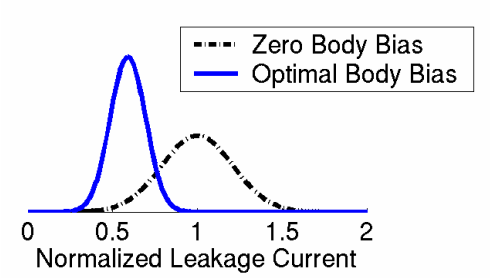

Figure 10. Leakage distribution improvement with optimal body bias. The distribution is assumed to be Gaussian, but the relative means and spreads of the distributions are experimentally determined

6. CONCLUSION

The paper presents a scheme to reduce leakage current and leakage variations in scaled technologies by applying optimum body bias to transistors. By monitoring the relative contribution of sub-threshold and the source/drain junction band-to-band tunneling leakage current, a leakage monitoring circuit determines the optimal value of body bias for the circuit under consideration. The monitor automatically adjusts this value, according to the process variations of the devices, to apply the optimal body bias. This body bias improves the nominal and worst-case leakages as well as the spread of leakages caused by the process variations and is applicable to different technology generations. Thus proper monitoring of the leakage components results in lower leakage currents and higher yields.

7. ACKNOWLEDGEMENTS

This work was funded in part by the Semiconductor Research Corporation PhD Fellowship, Giga-scale Silicon Research Center, DARPA PACC program, and by Intel and IBM Corporations.

8. REFERENCES

[1] T. Kuroda, et. al., "A 0.9-V, 150-MHz, 10-mW, 4 mm2, 2-D Discrete Cosine Transform Processor with Variable Threshold-Voltage (VT) Scheme," IEEE JSSC, vol. 31, No. 11, Nov. 1996, pp. 1770-1779.

[2] S-F. Huang, et al., "Scalability and Biasing Strategy for CMOS with Active Well Bias," 2001 Symp. on VLSI Tech. Digest of Technical Papers. pp. 107-108.

[3] A. Keshavarzi , et al., "Effectiveness of Reverse Body Bias for Low Power CMOS Circuits," Proc. of 8th NASA Symposium on VLSI Design, 1999, pp. 2.3.1-2.3.9.

[4] A. Keshavarzi, et al., "Forward Body Bias for Microprocessors in 130nm Technology Generation and Beyond," 2002 Symposium on VLSI Technology Digest of Technical Papers. pp. 312-315.

[5] J. Tschanz, et. al., "Adaptive Body Bias for Reducing Impacs of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage," 2002 ISSCC Digest of Technical Papers, pp 477-479.

[6] V. De, et. al., "Techniques for Leakage Power Reduction,"Design of High-Performance Microprocessor Circuits, Piscataway, NJ: IEEE Press, 2001, pp. 48-52.

[7] Y. Taur and T.H. Ning, Fundamentals of Modern VLSI Devices, New York, USA: Cambridge University Press, 1998.

[8] K. Schuegraf and C. Hu, "Hole Injection SiO2 Breakdown Model for Very Low Voltage Lifetime Extrapolation," IEEE Trans. on Electron Devices, vol.41, May 1994, pp. 761-767.

[9] Berkeley Predictive Technology Model, U.C. Berkeley, http://www-devices.eecs.berkeley.edu/~ptm/

[10] BSIM4. BSIM Group, U.C. Berkeley, http://www-devices.eecs.berkeley.edu/bsim3/bsim4.html

[11] Taurus-Device Simulator. Synopsys, 2002.

[12] Z. Chen, et. al. "Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks," Proc. of ISLPED, 1998, pp. 239-244.

[13] "Well-Tempered" Bulk-Si NMOS Device Home Page, Microsystems Technology Laboratory, MIT, http://www-mtl.mit.edu/Well/

[14] International Technology Roadmap for Semiconductors 2001 Edition, Semiconductor Industry Association, http://public.itrs.net