

3-1-1998

# Leakage grading of inputs to CMOS logic

Mark C. Johnson

*Purdue University School of Electrical and Computer Engineering*

Dinesh Somasekhar

*Purdue University School of Electrical and Computer Engineering*

Kaushik Roy

*Purdue University School of Electrical and Computer Engineering*

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

---

Johnson, Mark C.; Somasekhar, Dinesh; and Roy, Kaushik, "Leakage grading of inputs to CMOS logic" (1998). *ECE Technical Reports*. Paper 53.

<http://docs.lib.purdue.edu/ecetr/53>

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

# LEAKAGE GRADING OF INPUTS TO CMOS LOGIC

MARK C. JOHNSON  
DINESH SOMASEKHAR  
KAUSHIK ROY

TR-ECE 98-5  
MARCH 1998



SCHOOL OF ELECTRICAL  
AND COMPUTER ENGINEERING  
PURDUE UNIVERSITY  
WEST LAFAYETTE, INDIANA 47907-1285

# Leakage grading of inputs to CMOS logic

Mark C. Johnson, Dinesh Somasekhar, and Kaushik: Roy  
School of Electrical and Computer Engineering  
Purdue University, West Lafayette, Indiana, 47907-1285, USA  
Phone: (765)494-3372, (765)494-3372, (765)494-2361  
Email: {mcjohnso, somasekh, kaushik@ecn.purdue.edu}

---

'This research was supported in part by ARPA (F33615-95-C-1625), NSF CAREER award (9501869-MIP), IBM, Rockwell, AT&T/Lucent, and ASSERT program (DAAH04-96-1-0222).

## **Abstract**

Supply voltages and threshold voltages continue to be aggressively scaled down in order to obtain power reduction, performance improvement, and increasing integration density. This leads to leakage current becoming a much more significant component of power than it has been in the past. We have previously shown that substantial leakage reduction can be achieved in single  $V_t$  circuits by turning off sticks of transistors. A theoretical model was also derived which predicts the quiescent leakage current and the idle time required to reach quiescent levels. In this report, we will review the leakage estimation model, outline a method for evaluating the leakage associated with an input vector, and use the model to identify inputs which minimize leakage in a variety of test cases.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Quiescent Leakage of a Transistor Stack</b>	<b>2</b>
<b>3</b>	<b>Transient Leakage Behavior</b>	<b>4</b>
<b>4</b>	<b>Validation of the Model</b>	<b>7</b>
<b>5</b>	<b>Selecting Standby Mode Input Vectors</b>	<b>8</b>
<b>6</b>	<b>Summary</b>	<b>13</b>
<b>7</b>	<b>Acknowledgments</b>	<b>14</b>

## List of Figures

1	Four transistor stack: schematic, layout, and typical quiescent voltages. . . . .	3
2	Leakage savings ratio vs. stack height. . . . .	5
3	Discharge currents in a transistor stack. . . . .	6
4	Simulated vs. predicted quiescent leakage. . . . .	9
5	Simulated vs. predicted settling time. . . . .	10

# 1 Introduction

Estimation and control of leakage in CMOS circuits is a problem already familiar to designers of dynamic memory and  $I_{ddq}$  tests. However, deep submicron devices, low operating voltages, and low power dissipation requirements now make this an important problem for most classes of CMOS logic design, even static logic gates. Extremely low supply voltages require low transistor threshold voltages to maintain performance. Lowering the threshold has the side effect of making the transistors more difficult to turn off. The resulting leakage currents increase power dissipation even when a circuit is idle. In this work, we are evaluating the use of input vectors selected in such a way as to minimize leakage. This technique holds promise as a supplement to existing power management techniques such as gating of clocks and latching of inputs to circuits which are idle.

We focus on control of subthreshold leakage. Diode leakage, though dominant in long channel high  $V_t$  devices, is negligible in comparison to subthreshold leakage in low  $V_t$  submicron devices [5]. Gate induced drain leakage (GIDL) may become a greater concern in the future in deep submicron devices. GIDL is largest when  $V_{DS}$  is relatively high and the gate is reverse biased.

Subthreshold leakage in a MOS transistor can be controlled by biasing the transistor in such a way that the gate becomes reverse biased relative to the source [4]. This effect can be achieved by inserting a transistor between the power rail and the remaining circuit as in [9] and [4]. However, an extra transistor is not always necessary. We have previously shown in [2] that

the self reverse biasing effect can be exploited using the transistor stacks already present in most CMOS logic gates through appropriate selection of input vectors. Halter and Najm [1] observed the input dependence of leakage from simulation of randomly selected inputs. They then proposed the use of modified registers to allow circuits to be forced to a low leakage state. We will show that one can select low leakage input vectors based on our model for leakage in stacks of transistors.

## 2 Quiescent Leakage of a Transistor Stack

Consider the pull down network of a four input NAND gate (Figure 1). If all four transistors are turned off for a sufficiently long time, the circuit will reach a state where the leakage through each transistor is equal and the voltage across each transistor will settle to a value within an order of magnitude of  $kT/q$ . The reverse bias between the gate and source of a transistor is equal to the sum of the drain-source voltages across the transistors below. Had only one transistor been turned off, the reverse bias to the gate of that transistor would be virtually zero (since  $R_{ON} \ll R_{OFF}$ ). Given a typical subthreshold slope of 85mV/dec, the leakage with all transistors off is less than one tenth of the leakage of a single transistor.

Using the BSIM [8] transistor model, we have derived a model to predict quiescent voltage levels and leakage current in a stack of transistors. We use the following simplification of the subthreshold current equation.

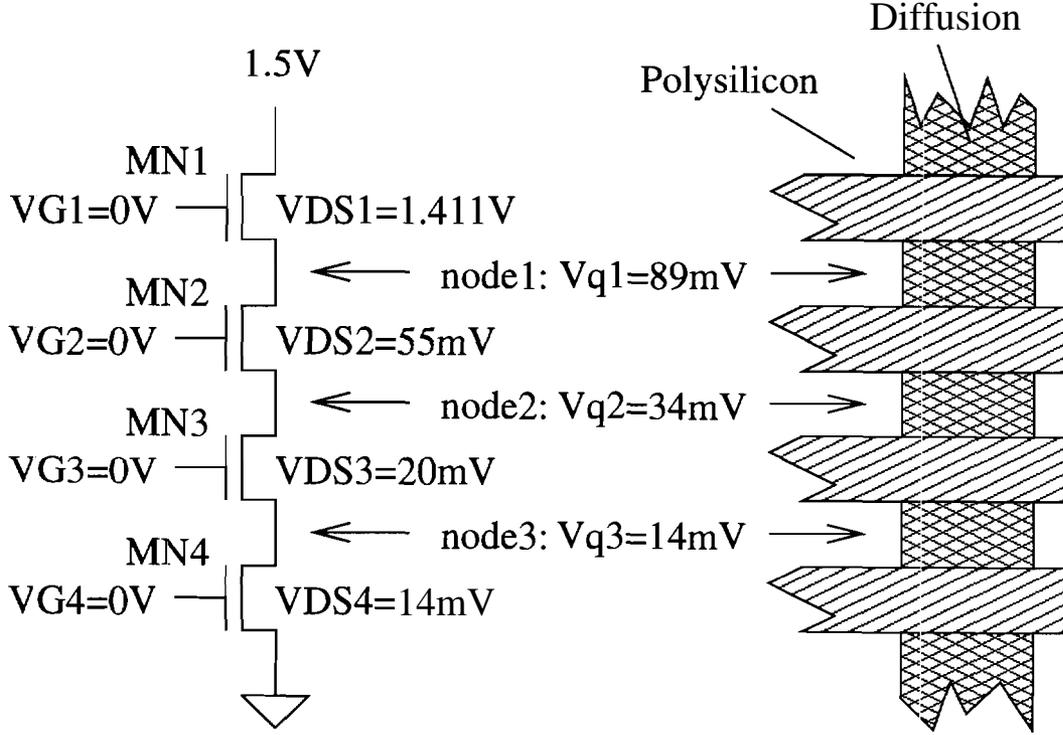


Figure 1: Four transistor stack: schematic, layout, and typical quiescent voltages.

$$I_{subth} = A \times e^{\frac{1}{n\nu_T}(V_G - V_S - V_{TH0} - \gamma' \times V_S + \eta \times V_{DS})} \times (1 - e^{\frac{-V_{DS}}{\nu_T}}) \quad (1)$$

where  $A = \mu_0 C'_{ox} \frac{W}{L_{eff}} (\nu_T)^2 e^{1.8} e^{\frac{-\Delta V_{TH}}{n\nu_T}}$ .  $V_{TH0}$  is the zero bias threshold voltage.  $\nu_T$  is the thermal voltage  $\frac{kt}{q}$ . The body effect for small values of  $V_S$  is very nearly linear. It is represented by the term  $\gamma' V_S$ , where  $\gamma'$  is the linearized body effect coefficient.  $\eta$  is the DIBL coefficient.  $C_{ox}$  is the gate oxide capacitance.  $\mu_0$  is the zero bias mobility.  $n$  is the subthreshold

swing coefficient of the transistor.  $\Delta V_{TH}$  accounts for variations in threshold voltage from one transistor to another.

Our first step in determining leakage is to calculate internal node voltages. Only transistors which are turned off are considered. Transistors which are turned on can be treated as short circuits. If the short circuited transistor is at the top of the stack,  $V_{TH}$  must be subtracted from  $V_{DD}$ . Equation 2 gives us the voltage across the second transistor from the top as a function of  $V_{DD}$ . This assumes  $V_{DD} \gg V_{q1}$ . Otherwise the  $V_{DD}$  term must be replaced by  $V_{DSq1}$ . Equation 3 gives the voltage of the  $i^{th}$  transistor in terms of the  $(i - 1)^{th}$  transistor. These equations were obtained by equating the subthreshold current through each transistor.

$$V_{DSq2} = \frac{n\nu_T}{(1 + 2\eta + \gamma')} \ln\left(\frac{A_1}{A_2} e^{\frac{\eta V_{DD}}{n\nu_T}} + 1\right) \quad (2)$$

$$V_{DSq_i} = \frac{n\nu_T}{(1 + \gamma')} \ln\left(1 + \frac{A_{i-1}}{A_i} (1 - e^{\frac{-1}{\nu_T} V_{DSq_{i-1}}})\right) \quad (3)$$

Once voltages have been determined, we can compute the leakage current using equation 1. Equation 4 expresses the leakage savings ratio as a function of N, the number of transistors which are turned off. Figure 2 plots this function to illustrate the diminishing return with increasing N.

$$S(N) = e^{\frac{1}{n\nu_T} (1 + \eta + \gamma') \sum_{i=2}^N V_{DSq_i}} \quad (4)$$

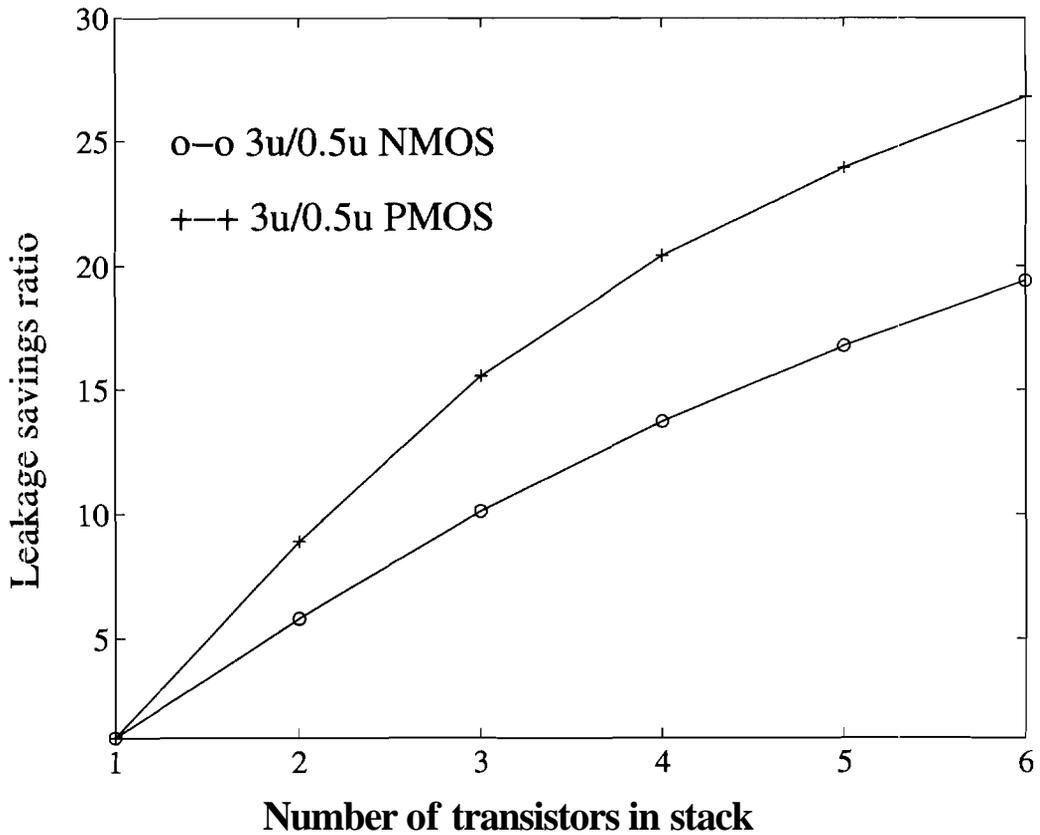


Figure 2: Leakage savings ratio vs. stack height.

### 3 Transient Leakage Behavior

When a stack of two or more transistors are turned off, the time required for voltages and currents to settle to quiescent levels is large and can vary over a wide range. We observe settling times ranging from a few microseconds up to hundreds of milliseconds. Nevertheless, we are able to derive a simplified model which tracks detailed simulation results quite well.

The worst case settling time for a transistor stack occurs when the internal

nodes are charged to the maximum possible voltage ( $V_{DD} - V_T$ ) just before all transistors are turned off. This maximizes the charge that must be dissipated by subthreshold currents before reaching quiescent levels. Coupling between the gates and internal nodes cause the internal nodes to bootstrap to a slightly lower voltage when the transistors are switched off.

Figure 3, obtained by simulation, illustrates the manner in which a stack of four transistors discharges to quiescent levels. Each curve in the figure indicates the instantaneous net discharge of current from a single internal node. Initially,  $V_{GS}$  is strongly reverse biased for all but the bottom transistor. Consequently, the bottom node must almost completely discharge before the next higher node starts to discharge. This behavior repeats itself one node at a time until all internal nodes have discharged to quiescent levels. This behavior allows us to estimate the settling time of each internal node separately and add the results together to get total settling time.

We estimate the worst case settling time as follows. Quiescent voltage levels  $V_{q_i}$  are calculated as described earlier. Another prerequisite is to calculate the voltage  $V_{boot_i}$  to which each internal node is bootstrapped after all transistors have been turned off. This is determined by the node capacitance (diffusion and possibly some interconnect) and the coupling of each gate to internal nodes (primarily gate-diffusion overlap). Details of the bootstrapping calculation are given in [2]. Given the internal node capacitance  $C(V_i)$  and discharge current  $I_{dis}(V_i)$  as a function of node voltage, we can determine the increment of time  $dt$  required for voltage to drop by  $dV_i$  as  $dt = -\frac{C_i(V_i)}{I_{dis}(V_i)}dV_i$ . Integrating this expression from  $V_{boot_i}$  down to  $V_{q_i}$ , we get the following expression for the settling time of node  $i$ . To get a closed

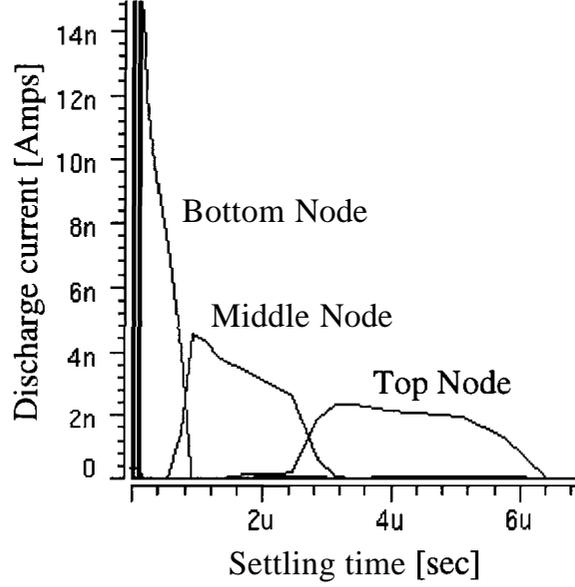


Figure 3: Discharge currents in a transistor stack.

form solution, we assume that  $C_i = C_i(V_{q_i})$  (i.e., we assume that  $C_i$  does not change as the node voltage drops).

$$t_{dis_i} = \frac{n C_i L_{eff}}{\mu_0 C_{ox} W \nu_T e^{1.8\eta}} \times \frac{1}{e^{\frac{1}{n\nu_T}((1+\gamma'+\eta)V_{q_{i+1}}+V_{TH0})}} \times \left( e^{\frac{-\eta V_{q_i}}{n\nu_T}} - e^{\frac{-\eta V_{boot_i}}{n\nu_T}} \right) \quad (5)$$

## 4 Validation of the Model

Model predictions of leakage current, leakage savings ratio, and settling time were compared to HSPICE simulation results for 64 different transistor

stacks with randomly selected design parameters and operating conditions. The parameters that were allowed to vary were the following: temperature ( $-50$  to  $150^{\circ}C$ ), number of transistors in the stack turned off (2 to 4 transistors),  $V_{TH_0}$  (from approximately 0.20V to 0.60V), supply voltage (from 1.2V to 1.8V), and transistor width (from  $2\mu$  to  $10\mu$ ).

Figure 4 compares predicted and simulated quiescent leakages for stacks of NMOS transistors. Similar results, not shown, were obtained for randomly selected stacks of PMOS transistors. Figure 5 compares predicted and simulated settling time estimates for stacks of NMOS transistors. Simulated settling time was taken to be the time required for leakage to settle to within 10% of the quiescent leakage level. The horizontal axis of each graph corresponds to a range of model predictions. The vertical axis corresponds to the range of values extracted from simulation results. Each data point identifies a model prediction and the corresponding simulation result.

These and all other simulation results were obtained using HSPICE with the BSIM 1 model for a 0.5u MOSIS process. The available MOSIS models do not include measured subthreshold characteristics, so we have estimated the subthreshold swing and related parameters from threshold voltage parameters, using the technique derived by Kang et. al. [3]. A subthreshold slope of approximately 86mV/decade was estimated and incorporated into the 0.5u BSIM model. In order to approximate the behavior of low threshold high leakage devices, we modify the flat band voltage parameter ( $V_{FB0}$ ).

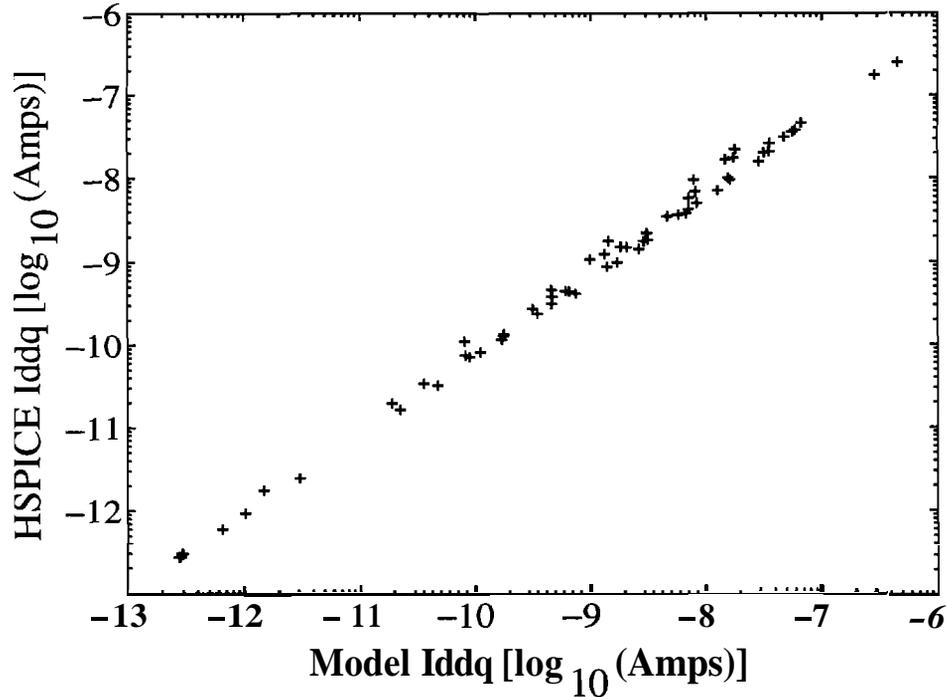


Figure 4: Simulated vs. predicted quiescent leakage.

## 5 Selecting Standby Mode Input Vectors

One promising application of the leakage estimation model is in the selection of minimum leakage standby mode input vectors. In general, the circuits of interest will not consist of a single transistor stack. The following procedure generalizes the leakage model for most CMOS circuits. For each input vector to be evaluated, identify those transistors which are turned on and replace them by a short circuit. Remove any paths which are parallel to a short circuit. Split the remaining network into a set of disjoint leakage paths.

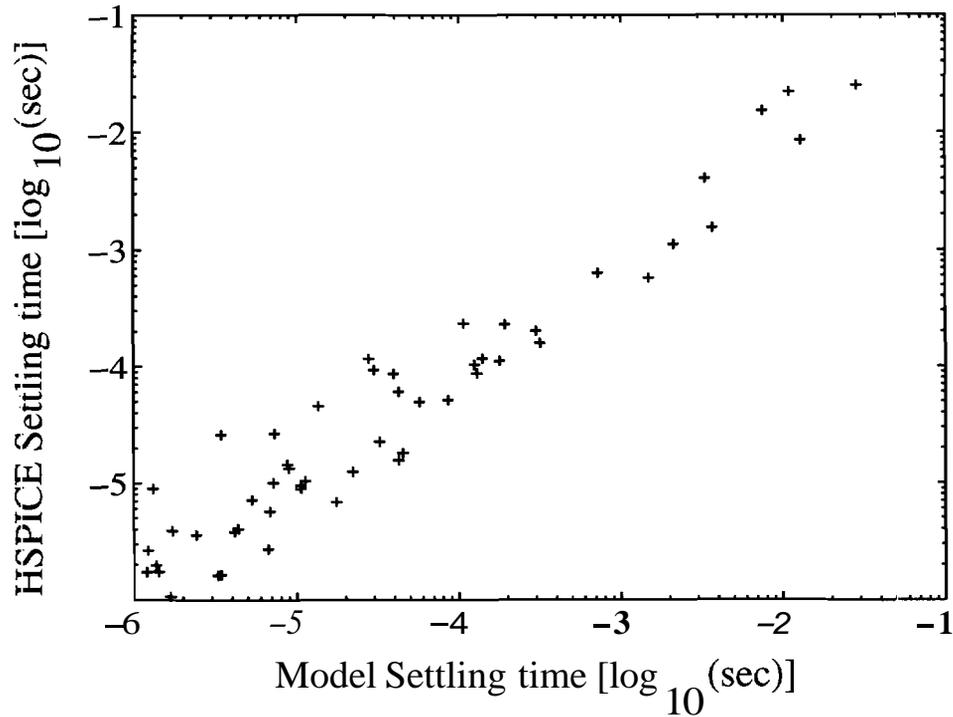


Figure 5: Simulated vs. predicted settling time.

Apply the leakage model to each stack and accumulate the total leakage.

For small circuits, this method can be used to evaluate all possible input vectors. For larger circuits, it is more practical to evaluate all input vectors for smaller subcircuits and then use this information to select input vectors that minimize total leakage.

Table 1 summarizes the results obtained for a variety of circuits. For each circuit, the leakage estimate is reported along with the simulated leakage for the highest and lowest leakage input vector. In some test cases, the model and the simulation did not identify the same input vector as best

(or worst). In such cases, either input vector was acceptable since each vector provided a similar degree of leakage savings. We found that the most significant source of this deviation was found to be variation in threshold voltage with respect to width, due to narrow channel effect.  $V_{TH}$  deviation ranged from approximately  $+10mV$  at  $W = 2\mu m$  to  $-10mV$  at  $W = 16\mu m$ . In a few cases where leakage was dominated by a single wide transistor, the threshold value was corrected to obtain a more accurate leakage estimate.

Table 1: LEAKAGE DEPENDENCE ON INPUT VECTOIRS

Circuit Description	Model $I_{ddq}$ [nA]	HSPICE $I_{ddq}$ [nA]	Comments
4 input NAND input=0000	0.72	0.60	Best
1111	23.2	24.1	Worst
3 input NOR input: 111	0.13	0.13	Best
000	29.9	29.5	Worst
Full Adder (Mirror [10]) A, B, $C_i = 111$	7.5	7.8	Best
001	56.0	62.3	Worst
4 Bit Ripple Add A, B = 0000	102.6	91.3	Best ( $C_i = 0$ )
A, B = 1111	102.6	94.0	Best ( $C_i = 1$ )
A, B = 0101	258.9	282.9	Worst ( $C_i = 1$ )
8 Bit Carry Select A = B = 11...	259.0	246.2	Best ( $C_i = 1$ )
A = B = 01...	690.4	759.6	Worst ( $C_i = 1$ )
4 Bit MCC (dynamic) $CLK = 1$	16.8	13.5	Best (inputs=1)
$CLK = 0$	15.6	15.9	Best (inputs=0)
$CLK = 0$	49.7	55.3	Worst (inputs=0)
1 Bit Generate/ Propagate CLK, A, B = 111	17.1	12.2	Best
CLK, A, B = 000	14.0	14.0	Best
$CLK$ , A, B = 011	25.2	25.4	Worst
1 Bit Sum (for MCC) G, P, C = 110	13.9	10.1	Best
G, P, C = 001	13.1	11.6	Best
G, P, C = 000	25.7	21.8	Worst
G, P, C = 001	24.9	26.1	Worst
4 Bit MCC Adder CLK = 1	154.4	126.6	Best (inputs=1)
CLK = 0	144.4	134.4	Best (inputs=0)
CLK = 0	198.8	190.4	Worst (inputs=1)

Table 2: LEAKAGE MODEL PARAMETER VALUES

Parameter	NMOS	PMOS
	value	value
Channel Length [pm]	0.5	0.5
Temperature [degC]	50	50
$I_{off}$ [ $nA/\mu m$ ]	3.32	0.967
$V_{TH_0}$ [V]	0.25	0.26
n (subthreshold slope coefficient)	1.44	1.44
$\eta$ (DIBL coefficient) [mV/V]	45	60
$\gamma'$ (linearized body effect) [V/V]	0.24	0.11

Table 2 lists the parameter values used to calculate model predictions of leakage.

## 6 Summary

In this paper, we have described and demonstrated how a model for sub-threshold leakage in transistor stacks can be used to evaluate the effect of input vectors on circuit leakage. This information can then be used to select low leakage standby mode input vectors as a supplement to existing power management techniques.

For a variety of test cases, the ratio of worst case to best case leakage varied from as little as 1.5 for a dynamic Manchester Carry Chain (MCC) based adder, up to 227 for a three input NOR gate. The test cases demonstrate that our ability to minimize leakage depends on how well we can control the

state of a majority of the possible leakage paths in a circuit. The NOR gate is trivially easy to control. However, in the MCC adder, a low leakage state for one portion of the circuit leads to high leakage states in other portions of the circuit.

## 7 Acknowledgments

Thanks go to Mamoon Hamid for many informative simulations and to Vivek De, Yibin Ye, Shekhar Borkar, Siva Narendra, and Ali Keshavarzi at Intel for many helpful discussions.

## References

- [1] J. P. Halter and F. Najm, "A gate-level leakage power reduction method for ultra-low-power CMOS circuits," In Proceedings, IEEE Custom Integrated Circuits Conference, 1997, pp. 475-478.
- [2] M. C. Johnson, D. Somasekhar, and K. Roy, "A model for leakage control by MOS transistor stacking," Technical Report TR-ECE 97-12, Purdue University, School of Electrical and Computer Engineering, December 1997.
- [3] S.-W. Kang, K.-S. Min, and K. Lee, "Parametric expression of subthreshold slope using threshold voltage parameters for MOSFET statistical modeling," IEEE Transactions on Electron Devices, vol. 43, no. 9, pp. 1382-1386, 1996.
- [4] Takayuki Kawahara et al., "Subthreshold current reduction for decoded-driver by self-reverse biasing," IEEE Journal of Solid-State Circuits, vol. 28, no. 11, pp. 1136-1143, Nov. 1993.
- [5] A. Keshavarzi, K. Roy, and C. Hawkins, "Intrinsic  $I_{DDQ}$ : Origins, reduction, and applications in deep sub- $\mu$  low-power CMOS IC's," In Proceedings IEEE International Test Conference, 1997.
- [6] T. Kobayashi and T. Sakurai, "Self-Adjusting Threshold-Voltage Scheme (SATS) for Low-Voltage High-Speed Operation," In Proceedings IEEE Custom Integrated Circuits Conference, 1994, pp. 271-274.

- [7] J. M. Rabaey, *Digital Integrated Circuits*, Upper Saddle River, NJ: Prentice Hall, 1993, p. 520.
- [8] B.J. Sheu, D.L. Scharfetter, P.K. Ko, and M.C. Jeng, BSIM: Berkeley short-channel IGFET model for MOS transistors. *IEEE Journal Solid-State Circuits (USA)*, vol. 22, no. 4, pp. 558-566, 1987.
- [9] S. Shigematsu, S. Mutoh, Y. Matsuya, Y. Tanabe, J. Yamada, "A 1-V high-speed MTCMOS circuit scheme for power-down applications," In *IEEE Journal Solid-State Circuits (USA)*, vol. 32, no. 6, pp. 861-869, June 1997.
- [10] N. H. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, 2nd ed., Reading, MA: Addison-Wesley, 1996, p. 394.