01 May 2007

Accurate Macro-Modeling for Leakage Current for I_{DDQ} Test

Kyung Ki Kim

Yong-Bin Kim

Minsu Choi
Missouri University of Science and Technology, choim@mst.edu

Nohpill Park

Recommended Citation

K. K. Kim et al., "Accurate Macro-Modeling for Leakage Current for I_{DDQ} Test," *Proceedings of the IEEE Instrumentation and Measurement Technology Conference: Synergy of Science and Technology in Measurement (2007, Warsaw, Poland)*, Institute of Electrical and Electronics Engineers (IEEE), May 2007. The definitive version is available at https://doi.org/10.1109/IMTC.2007.379346

Instrumentation and Measurement
Technology Conference – IMTC 2007
Warsaw, Poland, May 1-3, 2007

Accurate Macro-modeling for Leakage Current for I_{DDQ} Test

Kyung Ki Kim[1], Yong-Bin Kim[1], Minsu Choi[2], Nohpill Park[3]

[1]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA USA
[2]Department of Electrical and Computer Engineering, University of Missouri-Rolla, Rolla, MO USA
[3]Department of Computer Science, Oklahoma State University, Stillwater, OK USA

E-mail: E-mail: kkkim@ece.neu.edu[1], ybk@ece.neu.edu[1], choim@umr.edu[2], npark@cs.okstate.edu[3]

Abstract *- This paper proposes a new precise macro-modeling for leakage current in BSIM4 65nm technology considering sub-threshold leakage, gate tunneling leakage, stack effect, and fanout effect. Using the accurate macro-model, a heuristic algorithm is developed to estimate the leakage power and generate input test pattern for minimum leakage. The algorithm applies to ISCAS85 benchmark circuits, and the results are compared with the results of Hspice. The experimental result shows that the leakage power estimation using our macro-model is within 5% difference when comparing to Hspice results.*

Keywords *- Leakage Current, Subthreshold Leakage Current, Gate Tunneling Leakage Current, I_{DDQ}, Test Pattern Generator.*

I. INTRODUCTION

Due to the continued scaling of technology and supply and threshold voltage, leakage power has become more and more significant in the power dissipation of nanoscale CMOS circuits. Therefore, testing deep sub-micron (DSM) chips with millions of transistors is a difficult challenge.

As transistor geometries are reduced, it is necessary to reduce the supply voltage to avoid electrical break down and to get the required performance. However, to retain or improve the performance it is necessary to reduce the threshold voltage (V_{th}) as well. The reduced V_{th} increases sub-threshold leakage current exponentially. To control short channel effect and increase the transistor driving strength in DSM circuits, gate-oxide thickness also becomes thinner as technology scales down. The aggressive scaling in the gate-oxide results in tunneling current through the oxide which is a strong exponential function of the oxide thickness and the voltage magnitude across the oxide [1]-[3].

Therefore, the aggressive scaling increases variation in fault-free I_{DDQ}. As fault-free and faulty I_{DDQ} distributions overlap, it is not clear to distinguish between leakage and defect current. In order to solve the problem, this paper proposes a new macro-modeling for leakage. Although some papers have been published on this modeling for sub-threshold leakage and gate tunneling leakage, they neglected the interactions between the two leakages. They also didn't consider fanout effect in the leakage current [4][5]. Therefore, better understanding and more accurate model of leakage currents are essential for successful chip testing in sub-90nm CMOS technologies.

This paper shows that the accuracy of the macro-modeling for leakage current in nanoscale CMOS circuits is improved by considering the interactions between sub-threshold leakage and gate tunneling leakage, stacking effect, and fanout effect. Then, a simple ATPG (automatic test pattern generation) based heuristic algorithm is developed using C language to estimate leakage and generate input pattern for minimum leakage in large circuits.

The remainder of this paper is organized as follows. Section II illustrates the analysis and modeling of the leakage current. Section III describes input vector generation for minimum leakage power. Results from the experimental leakage power estimation on ISCAS85 benchmark circuits are listed and compared with Hspice results in Section IV followed by conclusion in Section V.

II. LEAKAGE CURRENT ANALYSIS

A. Gate Leakage Current

Gate leakage is a current flowing (tunneling) into the gate of the transistor. With the increase in gate oxide thickness the tunneling drops exponentially, and the equation is given by

$$I_{gate\_tunneling} = (A \cdot C) \cdot (W \cdot L) e^{-B \cdot \frac{T_{ox}}{V_{gs}} \cdot \alpha}) \qquad (1)$$

where $A = q_3 / 8\pi h\phi_b$, $B = 8\pi\sqrt{2m_{ox}}\phi_b^{3/2} / 3hq$, $C = (V_{gs} / T_{ox})^2$, α is a parameter which is ranged from 1 to 0.1 depending on the voltage drop across the oxide. H is the Planck's constant, and φ_b is the barrier height for electronics/holes in the conduction/valance band.

The gate tunneling leakage has already increased to more than double the sub-threshold leakage current in the nano-scale CMOS technology. Figure 1 shows the sub-threshold and gate tunneling leakage currents produced by the NMOS and PMOS transistor. As shown in Figure 1(a), the gate tunneling current consists of four components: gate-to-channel tunneling (*Igc*), gate-to-drain edge tunneling (*Igd*), gate-to-source edge tunneling (*Igs*), and gate-to-body

tunneling (*Igb*). The magnitude of the gate tunneling is strongly dependent on the applied Vgs voltage. In case of NMOS, four possible states exist depending on the voltages of three terminals of CMOS: drain/gate/source = {1/0/0},{1/0/1},{0/0/1}, and {0/1/0}. The leakage current under the {0/1/0} state is the highest due to the strong inversion. For PMOS, the current direction and the voltages are symmetric compared with NMOS as shown in Figure 1(a). Since holes have to pass a higher barrier to tunnel, the PMOS tunneling current is less than the NMOS tunneling current [6]-[7].

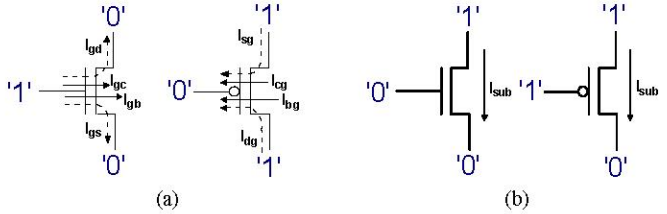

Figure 1. The gate tunneling and sub-threshold leakage current in NMOS and PMOS transistors: (a) maximum gate tunneling leakage current state, (b) maximum sub-threshold leakage current state.

B. Sub-threshold Leakage Current

Even though the transistor's gate voltage is lowed to below V_{th}, a small current still flows between the source and drain terminals as shown in Figure 1(b) [1]-[3]. The equation for the sub-threshold leakage current is given by

$$I_{sub} = I_0 e^{\frac{V_{gs}-V_{th}}{\eta kT/q}} (1 - e^{-\frac{V_{ds}}{kT/q}}) \quad (2)$$

where

$$I_0 = \mu_0 C_{ox} (W/L) (\frac{kT}{q})^2 (1 - e^{1.8}) \quad (3)$$

W and L are the transistor channel width and length, μ_0 is the low field mobility, C_{ox} is the gate oxide capacitance, k is the Boltzmann constant, q is the electronic charge, and N is the sub-threshold swing factor.

C. Stacking Effect and Fanout Effect

When there are two or more stacked off-transistors, the subthreshold leakage is reduced. This reduction depends on the choice of the input pattern during standby periods because it determines the number of OFF transistor in the stack. Turning OFF more than one transistor in a stack of transistors forces the intermediate node voltage to go to a value higher than zero. This causes a negative *Vgs*, negative *Vbs* (more body effect) and *Vds* reduction (less DIBL) in the top transistor, thereby reducing the subthreshold leakage current flowing through the stack considerably, which is known as the stack effect. Because of the transistor stack effect, the leakage current of a gate depends on its input combination. Individual CMOS gates show a variation in the leakage power based on different input patterns.

Figure 2 shows the static current paths that appears when the leakage current is considered in CMOS circuits. In the circuits, each inverter has a few paths of subthreshold and gate tunneling leakage. It is assumed that Inv2 is the DUT (Device Under Test), and the input of the Inv1 is '0'. Inv2 has three leakage components that are dependent on the fanout structures of the Inv2.

- The first component is the gate tunneling current $I_{gate\text{-}inv2}$ starting from the PMOS of Inv1.
- The second is the subthreshold leakage of the OFF state PMOS in Inv2 ($I_{sub\text{-}inv2}$).
- The last component is the gate tunneling current $I_{gate\text{-}inv3}$ staring from Inv3.

Therefore, the total leakage current is the sum of $I_{gate\text{-}inv2}$, $I_{sub\text{-}inv2}$, and $I_{gate\text{-}inv3}$. However, when a macro-model of a cell is generated, one leakage tunneling current ($I_{gate\text{-}inv3}$) should be removed not to count the leakage components twice when the total leakage currents of Inv3 are calculated. In Figure 2, $I_{gate\text{-}inv3}$ should be considered as the gate tunneling leakage of Inv3, i.e. only $I_{gate\text{-}inv2}$ and $I_{sub\text{-}inv2}$ are the leakage currents of Inv2.

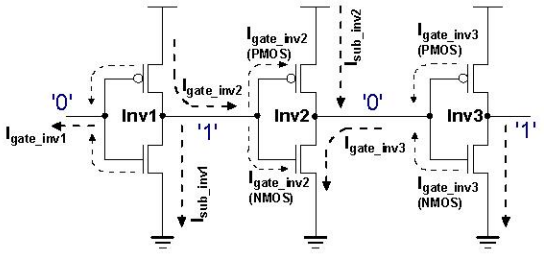

Figure 2. Leakage Current flows in nanoscale CMOS circuits.

Depending on the primary input pattern, the subthreshold leakage current and the gate tunneling are affected by the adjacent (fanins/fanouts) logic circuits. Figure 3 illustrates the dependency of the leakage current on the fanout structures. In Figure 3, the primary input is logic '1', the number of fanouts of inverter G2 is two, and the number of fanouts of inverter G3 is three. First, the current I_{gG3} is the gate tunneling leakage of inverter G3.

In this circuit, I_{gG2} and I_{gG4} are the gate tunneling leakage current of G2 and G4, respectively. The directions of the three currents converge into the input of inverter G3. The sum of gate leakage current at node N3 is a function of fanout of gate G1 and the subthreshold current of G2, G3, and G4. The "0" state voltage at node N3 increases as the fanout of G1 increases, which in turn reduces the gate leakage current of G2, G3, and G4 since the voltage between the input and output of those gates are reduced. The gate leakage current of G2, G3, and G4 is also a function of their subthreshold current since the subthreshold currents affect the voltage between the input and output of those gates. Considering these fanout effects, I_{gG3} is about one third of the gate tunneling leakage of the case where G1 has only one fanout. Consequently, the subthreshold current is influenced by the number of fanouts of the previous driver. However, the fanouts of inverter G3 cannot have a significant much effect

on the leakage current of the inverter G3. As the number of fanouts of G3 increases, the output voltage of G3 is reduced, and then the subthreshold leakage and gate tunneling leakage of G3 are reduced.

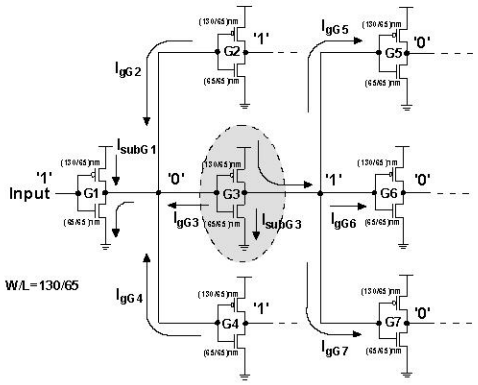

Figure 3. Fanout effect for G3 gate

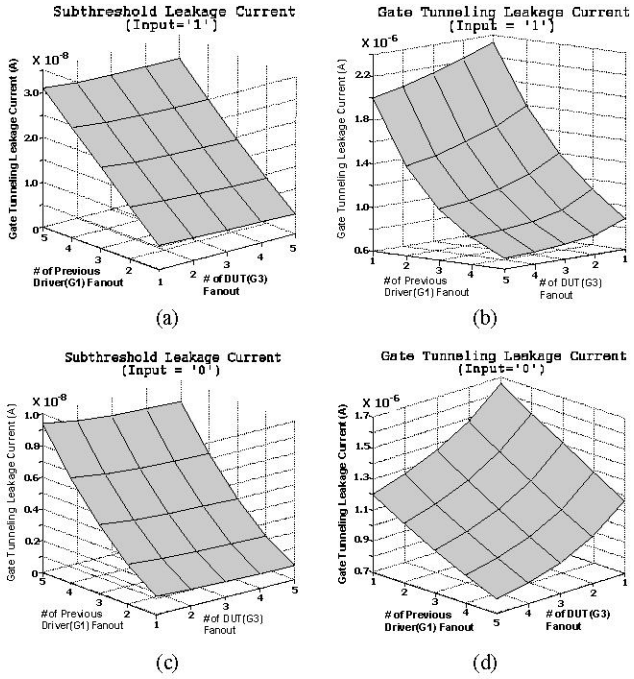

Figure 4. Leakage Current Variation due to fanout effect in BSIM4 65nm technology: (a) Subthreshold leakage current (Input='1'), (b) Gate tunneling leakage current (Input='1'), (c) Sub-threshold leakage current (Input='0'), (d) Gate tunneling leakage current (Input='0')

In summary, the total leakage of inverter G3 is affected by the fanouts of G1 and G3 and it is necessary to consider the interaction of each leakage current component in both previous stages and the next stages for an accurate leakage estimate in nanoscale CMOS circuits. However, the effects of the leakage current components beyond one logic level from the DUT are negligible.

Figure 4 presents the fanout effect on the leakage current for inverter G3 in Figure 3. The leakage currents are measured at inverter G3 in Figure 3. The number of fanouts of G1 is varied from 1 to 5, and the number of fanouts of G3 is varied from 1 to 5. It is assumed that the history effect is ignored to show the fanout effect on nanometer CMOS gates.

Figures 4 (a) and (b) show the subthreshold leakage and gate tunneling, respectively when the input of the inverter G3 is '1'. Figures 4(c) and (d) show the subthreshold leakage and gate tunneling, respectively when the input of the inverter G3 is '0'. As expected, the number of fanouts of G1 affects the leakage current. For the input '0', the fanouts of G3 have a considerable effect on the leakage current, but less than the fanouts of the previous driver.

The smallest total leakage (0.73 μA) is generated for the '1' input with five fanouts of G1 and five fanouts of G3. The highest total leakage (2.33 μA) is generated for the '0' input with one fanout of G1 and one fanout of G3. If the fanout effect is not considered to model the leakage current, the smallest total leakage is 2.26 μA under the '0' input and the largest total leakage is 3.92 μA under the '1' input.

III. MINIMUM LEAKAGE TEST PATTERN GENERATION

Based on the fanout effect in leakage current, the macro-model for each cell (inverter, nand, and nor gate) is developed based on Hspice simulation, where controlling variables are the number of fanouts, size of the cell, and input pattern considering stack effect under the fixed V_{DD}, V_{th}, T_{OX}, and temperature. Based on the accurate macro-model for cells, a heuristic approach is implemented to generate the minimum leakage test pattern. The leakage of each cell in the circuit depends on the input pattern applied to the circuits. Several techniques have been proposed to generate the input pattern for minimum leakage current and solve the NP-hard problem [8][9]. An easy way to solve the problem is to use the functional dependencies in the circuits, and the controllability of its nodes. In this paper, the methodology is improved to estimate the accurate leakage current with fast simulation time.

First, before finding the optimal input pattern to reduce the leakage power dissipation, the functional dependencies between cells should be searched, and dominated cells and conflicting cells for each cell should be listed in order of the weight function given by

$$Weight\_of\_cells(G_i) = \sum_i (MLK\_of\_conflicting\_cells(G_i)) - \sum_i (MLK\_of\_dominated\_cells(G_i)) - MLK(fanout\_of\_G_i) - MLK(G_i) \quad (4)$$

where MLK is the mean leakage of the cell that depends on input pattern and fanout effect.

Once the list is determined, one cell with the least weight function will be selected. If the cell satisfies functional constraints for minimum leakage current, the primary input patterns controlled by the cell can be determined. After finding the proper input patterns, the cell is removed from the list, and at the same time dominated cells and conflicting cells of the cell are removed from the list. The procedure is repeated until there is no cell in the list or there are only cells that are not defined. If the undefined cells are found, proper patterns have to be assigned considering conditions for low

leakage current because they have no dominated cell and no conflicting cells. The algorithm is shown in Table 1.

Table 1. Algorithm for minimum leakage test pattern generation

```
For each Node in the circuit
    List controllability
End
For each Cell
   Determine the List of dominated and conflicting Cells
   Sort the List
End
Setup Variables from Lookup Table for each Cell
While the List is not empty
   Calculate Weight Function from the Lookup Table for each Cell
   Sort the Cells using the Weight Function
   Select one Cell with the least Weight Function
   Determine Input pattern using the selected Cell
      If it doesn't break functional limitation
         Remove the selected Cell, its dominated,
         and conflicting cells from the List
      End
    Update the List
End
If undefined Input is found
     Assign a proper Input pattern to minimize leakage current.
End
```

IV. EXPERIMENTAL RESULTS

The proposed minimum leakage test pattern generator for nanometer (BSIM4 65nm) CMOS circuits has been implemented in Hspice and C language, and run on 500 MHz UltraSPARC-IIe with 500Mbyte memory. The algorithm is proved by the results from various ISCAS85 benchmark circuits. The minimum leakage current is measured in Hspice using Monte Carlo simulation with 50 repetitions in Hspice. In addition, the proposed methodology using the fanout effect is compared against the same methodology without considering the fanout effect.

Table 2 shows the summary of the results of the proposed method and other simulations. The first column shows the measured circuits, and the second column is the number of gates of the circuits. The third and fourth are the estimated leakage current without considering the fanout effect and considering the fanout effect by our proposed method, respectively. The fifth shows the error rate between our proposed method and Hspice simulation. The sixth and the seventh represent the CPU simulation time for each method. The accuracy of the proposed method is within 5% difference compared with Hspice results. In addition, the simulation time of the proposed method is much faster than that of Hspice simulation.

V. CONCLUSION

As technology scales down below 90nm, leakage current becomes a critical problem. Especially, leakage current surpassing the defect current levels renders the I_{DDQ} test invalid. In this paper, to distinguish leakage current from defect current, an accurate macro-modeling for leakage current is proposed. The proposed methodology focuses on the fanout effect based on the previous simple modeling. The simulation results show that the modeling without considering the fanout effect cannot estimate accurate leakage current nor generate the best minimum leakage pattern. Finally, the paper shows a simple heuristic algorithm for generating the minimum leakage test pattern using the proposed macro-cell model. It is developed and experimented by ISCAS85 benchmark circuits. The experiment shows that the proposed method has high accuracy (within 5 %) and efficiency compared with Hspice results.

REFERENCES

[1] Siva G. Narendra, Anantha Chandrakasan, "Leakage in Nanoscale CMOS Technologies", Springer, 2005.

[2] Kaushik Roy, et al., "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicronmeter CMOS Circuits", Proceeding of the IEEE, Vol. 91, NO. 2, pp. 305-327, Feb. 2003.

[3] A. Agarwal, S.Mukhopdhyay, et al. , "Leakage Power Analysis and Reduction: Models, Simulation and tools", IEE Proc. Computer. Digit. Tech. Vol. 152, No. 3, May 2005.

[4] Kaushik Roy, et al., "Test Consideration for Nanoscale Scale CMOS Circuits", Volum 23, Issue2. pp. 128-136, March 2006.

[5] C.P. Ravikumar, Rahul Kumar, "Divide-and-Conquer I_{DDQ} Testing for Core-based System Chips", IEEE ASP-DAC 2002, pp. 761-766, 2002.

[6] Shengqi Yang, Wayne Wolf, et al., "Accurate Stacking Macro-modeling of Leakage Power in Sub-100nm Circuits", IEEE VLSID 2005, pp. 165-170, 2005.

[7] Dongwoo Lee, et al., "Analysis and Minimization Techniques for Total Leakage Considering Gate Oxide Leakage", DAC 2003, pp. 175-180, June 2003.

[8] Yongjun Xu, Zuqing Luo, "Minimum leakage Pattern Generation Using Stack Effect", In proceeding of ASICON03, pp. 1239-1242, 2003.

[9] Feng Gao and John P. Hayes, "Exact and Heuristic Approaches to Input Vector Control for Leakage Power Reduction", IEEE ICCAD 2004, pp.527-532, 2004.

Table 2. Leakage estimation results for benchmark circuits

Circuit	# of Gates	Estimated Leakage Current (μA) (Minimum value)		Hspice Minimum Leakae (μA)	Error Rate (%)	CPU Run Time	
		without fanout effect	with fanout effect			This work (*msec*)	Hspice (*sec*)
C432	160	3.75	1.49	1.53	2.70	0.89	9.45
C499	202	10.48	6.78	7.00	3.20	5.60	40.34
C880	383	9.67	5.34	5.40	1.20	3.50	28.88
C1355	546	16.54	7.43	7.21	3.90	5.87	40.41
C1908	880	15.53	6.84	6.57	4.00	6.65	35.94
C2670	1193	18.43	11.14	10.75	3.50	9.88	64.83