IBM Research Report

Experimental Measurement of a Novel Power Gating Structure with Intermediate Power Saving Mode

Suhwan Kim
Electrical Engineering
Seoul National University
Seoul 151-744
Korea

Stephen V. Kosonocky, Daniel R. Knebel, Kevin Stawiasz
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

IBM

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Experimental Measurement of A Novel Power Gating Structure with Intermediate Power Saving Mode

Suhwan Kim[1], Stephen V. Kosonocky[2], Daniel R. Knebel[2], and Kevin Stawiasz[2]

[1] Electrical Engineering, Seoul National Univestiy, Seoul 151-744, Korea

[2] IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA

suhwan@ee.snu.ac.kr and stevekos, knebeld, stawasz@us.ibm.com

ABSTRACT

A novel power gating structure is proposed for low-power, high-performance VLSI. This power gating structure supports an intermediate power saving mode as well as a traditional power cut-off mode. To evaluate our power gating structure, we design and fabricate three different macros in 0.13 μm CMOS bulk technology. Our measurement results show that the additional intermediate power-mode allows us to cover various power-performance trade-off regimes, compared to conventional power gating structures.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles— *advanced technologies, microprocessor and microcomputers*

General Terms

Reliability Design

Keywords

clock gating, power gating, wake-up latency, inductive noise, ground bounce, system-on-a-chip (SOC) design.

1. INTRODUCTION

With the recent trend toward high-performance portable system-on-a-chip (SoC) for communication and computing, power dissipation has become a critical design constraint. It is known that supply voltage scaling is the most effective way to reduce power dissipation, especially in CMOS digital circuits. Reducing supply voltage, however, increases circuit delay which leads to decreasing threshold voltage in order to maintain performance. Unfortunately, this in turn increases leakage current dramatically due to the exponential nature of leakage current in the subthreshold regime of the transistor. Additionally, most hand-held devices are characterized by intermittent operations with long periods of idle time.

Thus, leakage current is the dominant component of total power dissipation dissipation.

In this paper, we propose a novel power gating structure to support an intermediate power-saving mode as well as a traditional power cut-off mode. To evaluate our power gating structure, we design and fabricate three macros on a test chip in 0.13 μm CMOS bulk technology. We use single-threshold devices for both logic and the sleep transistor. The measurement results of the three different macros show the potential benefits of our power gating structure, in terms of power and performance.

The remainder of this paper is organized as follows. The motivation of our works is described in Section 2. Section 3 describes our novel power gating structure with an additional intermediate power-saving mode that allows us to cover various power- performance trade-off regimes. Section 4 and 5 present our measurement results from thee different macros designed and fabricated in 0.13 μm CMOS bulk technology. Our contribution is summarized in Section 6.

2. BACKGROUND

The multi-threshold CMOS (MTCMOS) circuit, called a "power gating structure", is one of the well-known techniques for reducing leakage power in standby mode while still maintaining high speed in active mode [1, 2, 3, 4].

By turning off the sleep transistor during the sleep period, however, the virtual ground (VGND) node of the power gating structure is charged up to a steady state value close to VDD. As a result, the data in storage elements are completely lost. Extra data-recovery process steps are required and significantly degrade system performance. Additionally, as shown in Figure 1, the instantaneous discharge current through the sleep transistor operating in its saturation region creates current surges at the sleep/active mode change. Because of the self-inductance of the off-chip bonding wires and the on-chip parasitic inductance inherent to the power rails, current surges cause voltage fluctuations in the on-chip power distribution network [5, 6].

Figure 2 shows the virtual power/ground rail clamp (VRC) scheme to dynamically reduce the supply voltage across a circuit during standby mode by interrupting he power supply and ground connections of the circuit [7]. During the standby mode, PFET and NFET switches (MP and MN) are turned off by asserting low (high) to CS (/CS) and the diodes (DN and DP) clamp the supply voltage potential to a lower one. The VRC allows state retention in the storage elements and eliminates the state restoration pro-
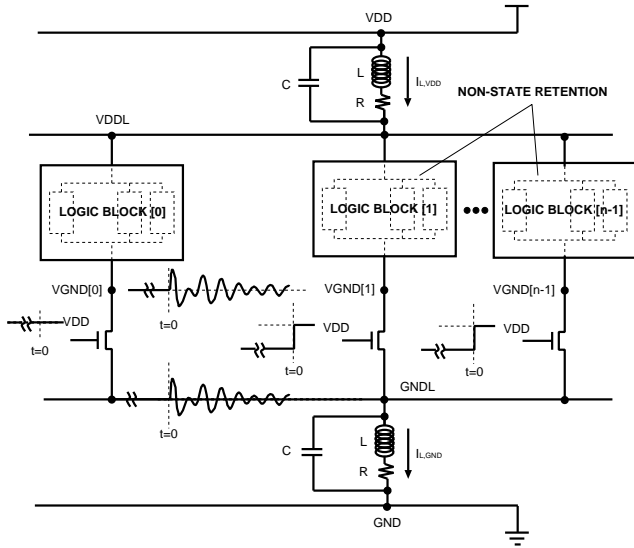
Figure 1: **Non-state retention and ground bounce noise in a system-on-a-chip (SoC) employing a conventional power gating structures to control leakage power.**

Figure 2: **Virtual power/ground rail clamp (VRC) CMOS.**

Figure 3: **Novel power gating structure showing dominant current flow in RUN/IDLE mode.**

cedures that causes overall performance degradation of a system with power gating structures. It also reduces the ground bounce noise by limiting the voltage level of the virtual ground node. The amount of leakage saving in this mode is relatively small compared to that of power cut-off mode of conventional power gating structures, however. For some hand-held mobile applications, this power saving mode by itself may not be useful. As a result, there is a demand for a power gating structure that supports a state-retention mode as well as a power cut-off mode and to cover a wide range of low-power, high-performance applications.

3. NOVEL POWER GATING STRUCTURE

To solve the problems described in the previous section, we propose a novel power gating structure in this section. By adding only a single built-in PFET to a conventional power gating structure, our power gating structure supports an additional intermediate power saving mode as well as power cut-off mode. In the intermediate mode, leakage reduction and data retention are realized and the magnitude of power supply voltage fluctuations during power-mode transitions is reduced.

Figure 3 shows the power gating structure used to decouple the ground node from the logic in RUN/IDLE mode. In this mode, the signal PG is asserted high to force the NFET transistor of power gating structure into a low resistance conducting state, while HLD is set high. In this mode, the NFET is used to short the logic circuit virtual ground to the real ground potential, allowing the full VDD supply voltage to be applied across the circuit to allow high speed operation.

Figure 4 shows the circuit in a non-state retentive mode or COLD mode. The signal PG is held low and the signal HDL is high. In this mode, the current path is cut off from GND and the voltage across the logic circuit collapses, providing maximum suppression of both gate and subthreshold leakage
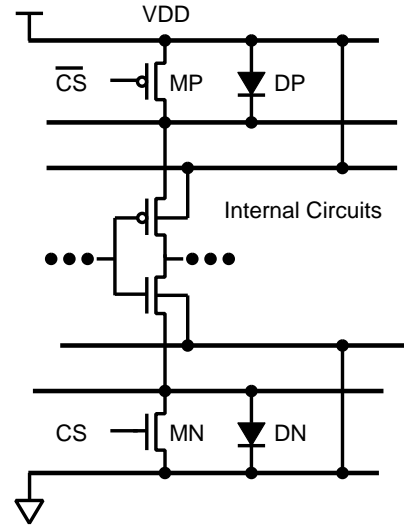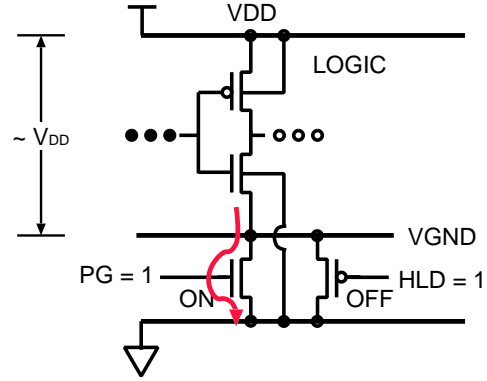
currents.

In the state retention mode or PARK mode, as shown in Figure 5, the signal PG is asserted low, and HLD is asserted low. This turns off the NFET device and the PFET device operates as a source-follower. In this mode, the virtual ground rail, VGND, is held a threshold voltage of the PFET (V_{tp}) above the ground rail. The voltage across the logic circuit is $V_{dd} - V_{tp}$, causing a reduction of gate leakage as well as subthreshold leakage in this circuit, since these leakages are dependent on the voltage applied to the devices. Further reduction in the subthreshold leakage also occurs because the logic circuit NFET body connection is connected to the real ground, while its source is raised by the PFET source-follower device. This effectively reverse body biases the logic NFET, causing an increase in the NFET threshold voltage further reducing the subthreshold leakage. In this mode, the voltage level of a virtual GND is limited by V_{tp}. As a result, state is retained and the ground bounce induced by power mode transition is smaller than that of COLD mode. To reduce the ground bounce induced by the transition from COLD to RUN/IDLE, this PARK can be used as an inter-
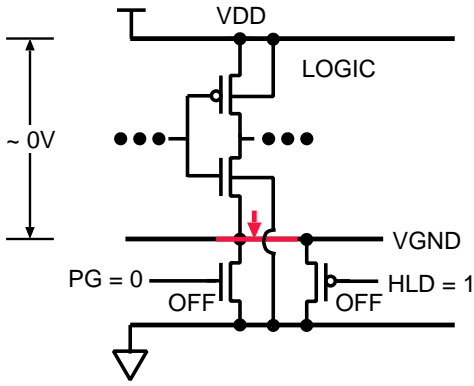
Figure 4: Novel power gating structure showing dominant current flow in COLD mode.
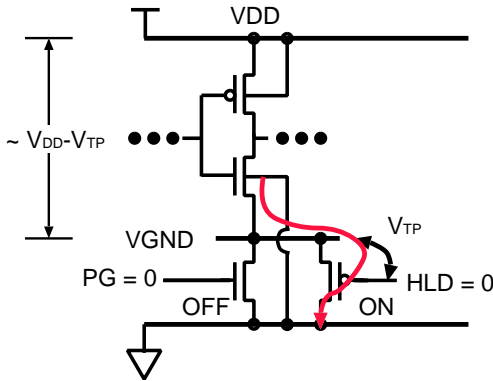

Figure 5: Novel power gating structure showing dominant current flow in PARK mode.

mediate step in the mode transition. That is, the transition sequence from COLD to RUN/IDLE is as follows: COLD, PARK, and RUN/IDLE.

4. TEST CHIP DESIGN

To demonstrate the effectiveness of our newly proposed power gating structure with intermediate power saving mode, three different macros are designed and fabricated, using 0.13um CMOS bulk technology. To minimize the process variation, the three macros are implemented on a multiple project wafer (MPW). All of the three macros include nine identical design-under-test (DUT) modules. The basic components of each DUT module consist of two linear-feedback shift registers (LFSR's), one 32-bit carry lookahead adder (CLA), and one multiple input signature register (MISR). The LFSR's are used to generate and feed a sequence of pseudo random patterns to the CLA, and the MISR is used to validate the correct operation of the DUT module, respectively. The ground nodes of the CLA and output register in the first and second macros are connected to GND through our novel power gating structure. The sleep transistor of the first macro is sized at 2.6% of the total NFET and PFET size of the CLA and output register. Similarly, the sleep transistor of the second one is sized at 1.0% of the total NFET and PFET size of the CLA and output register.

To compare the performance and power consumption of

our power gating structure with a non-power gating structure, we also implemented the third macro in which the ground nodes of the CLA and output register are directly connected to GND, without a power gating structure.
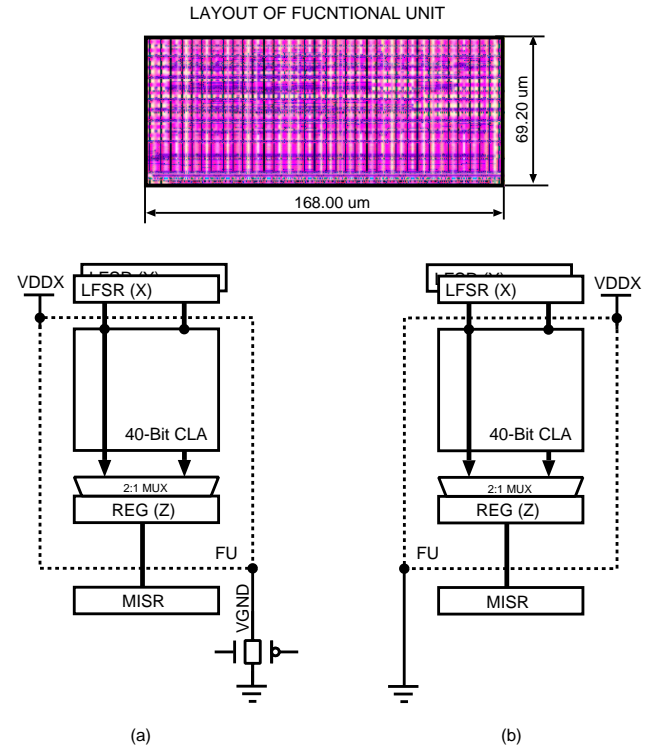

Figure 6: Block diagram of (a) DUT-TYPE A and (b) DUT-TYPE B.

In detail, a 32-bit carry lookahead adder (CLA) is implemented as the basic logic element for experimental measurement. Figure 6 shows a block diagram of the basic test structure that includes built-in self-test logic capable of at-speed testing of the adder and associated registers. The ground interrupt switch, also shown in the diagram, is applied to only the adder and associated registers. Likewise, independent voltage sources power the adder and associated register for accurate current measurement of the gated logic. The layout of the adder and register is shown at the top of the block diagram.

The block diagram and layout of Figure 7 depicts the power gate control and supply/ground rail distribution to the 3×3 array of DUT units. For the first and second macros, the power gate structures are applied individually to each DUT-TYPE A, thereby creating independent power islands (VGND0 – VGND8) with power-mode controls that allow one island to remain running while the others are put into PARK or COLD mode. It also allows us to observe that voltage disturbs due to the RLC networks representing the VDD and GND distribution are coupled to each of the other DUT elements. Furthermore, the third macro let us compare the effect of the power gating structures to a non-power gated DUT, in terms of performance and power consumption. Figure 8 shows a photograph of the hardware setup used for test and measurement of our chips.
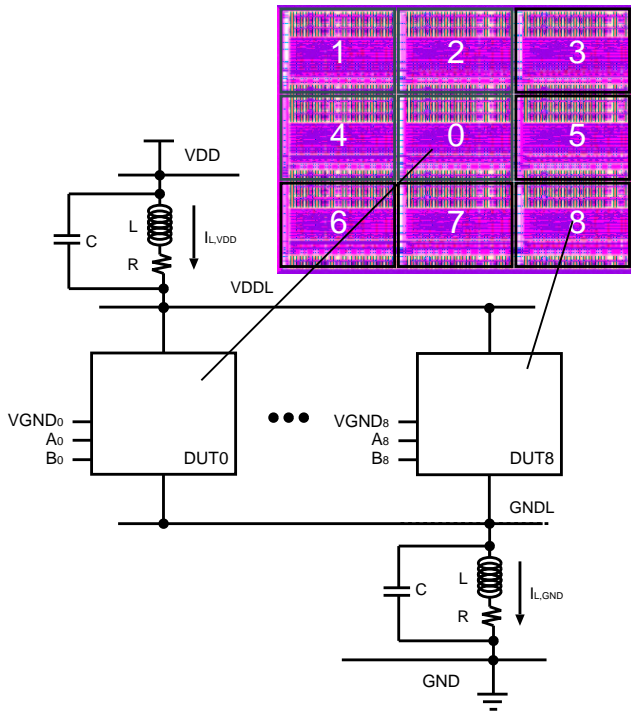
3

Figure 7: Block diagram and layout of one of macros implemented in our test chip.

5. EXPERIMENTAL RESULTS

Four independent test and measurement scenarios are developed to quantify the effectiveness of our power gating implementation and to compare performance and power consumption of the PARK and COLD modes described previously. With the power gating structure in RUN/IDLE mode, the DUT may be in either RUN mode (clocked at the highest operating frequency that does not result in any failure signature) or IDLE mode (clock is gated off). The first one measures the maximum operating frequency of the DUT and active power consumption at that frequency. For this test, DUT0 is in RUN mode and DUT1 through DUT8 are in IDLE mode. The second one compares leakage current with all DUTs in IDLE mode and is repeated for PARK and COLD modes. In the third one, the on-chip ground voltage is measured with DUT0 in IDLE mode and the remaining DUTs switching from COLD to IDLE modes and repeated for switching from PARK to IDLE modes. In this test, DUT1 through DUT8 create ground bounce noise due to switching of the power gating structures. Finally, the effect of the ground bounce on the performance of nearby logic is measured by putting DUT0 in RUN mode and switching the remaining DUTs from COLD to IDLE and PARK to IDLE modes.

In Figure 9, the performance of DUT-TYPE A is compared to our baseline, the maximum operating frequency of an identical 32-bit CLA design without the sleep transistor, across the allowed range of supply voltage. Figure 9 shows that negligible frequency degradation is measured in our hardware when the sleep transistor of DUT-TYPE A is sized at 2.6% of the total PFET and NFET size of the functional part including 32-bit CLA and corresponding output
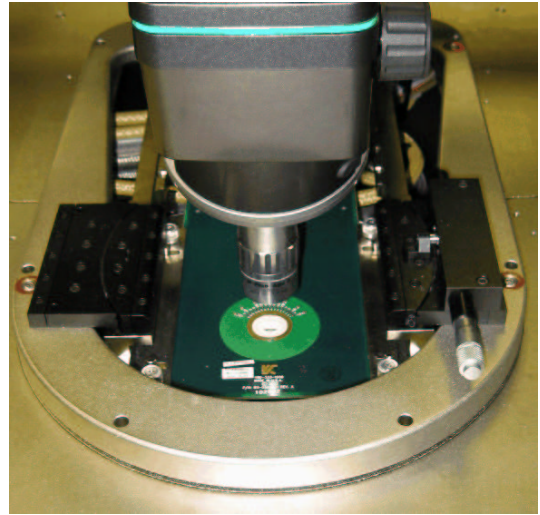

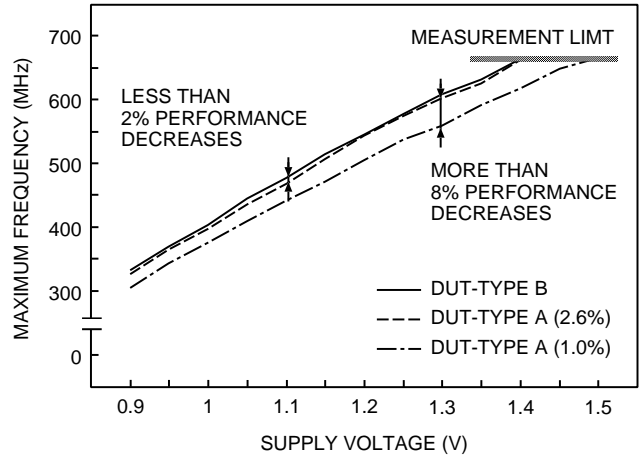**Figure 8: Hardware setup used for test and measurement of our chips.**

Figure 9: Performance as a function of supply voltage for DUT-TYPE A (2.6%), DUT-TYPE A (1.0%), and DUT-TYPE B.

registers. For a smaller sleep transistor of DUT-TYPE sized at the 1.0% of the total PFET and NFET size, frequency degrades as much as 8.25%. The kink in the upper curves at the high voltages is due to a 650 MHz clocking limit in our test and measurement setup.

Another metric for comparing power-gated structures is its average active power at a fixed frequency. This is useful for calculating total energy savings when a frequency requirement can be met by adjusting the supply voltage. Hardware measurement results shown in Figure 10 plot the increase in active power consumption traded for leakage power saving.

Leakage consumption of a macro with DUT-TYPE B is compared to that of the macro hardware with DUT-TYPE A when in PARK and COLD modes in the plot of Figure 11. At a supply voltage of 0.9V, the COLD mode reduces leakage power by $\times$ 43 compared to the leakage power consumption in IDLE mode. The effectiveness of this power supply in-
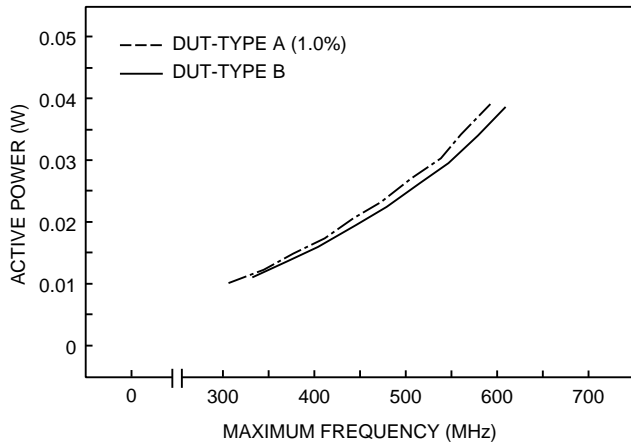
Figure 10: Active power consumption as a function of maximum operating frequency for DUT-TYPE A (1.0%) and DUT-TYPE B.
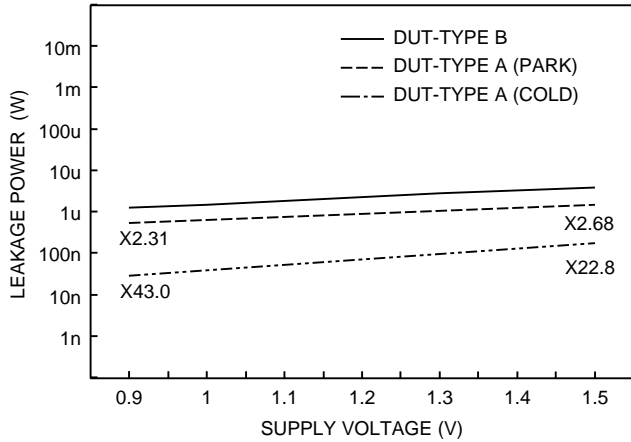

Figure 11: Leakage saving benefits of PARK and COLD modes as a function of supply voltage.

Figure 12: Measured ground bounce when the power mode is transited from COLD to IDLE mode.
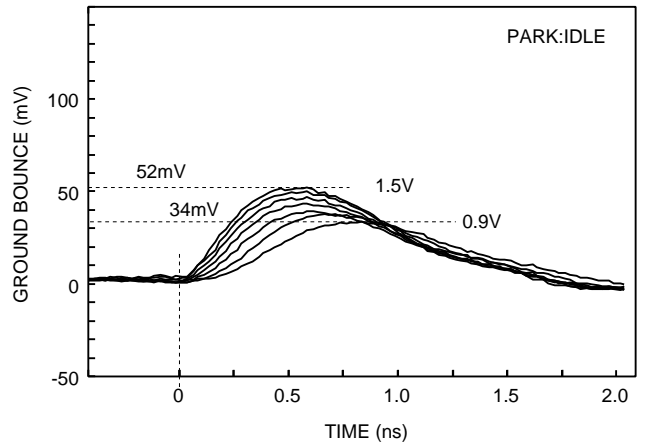
Figure 13: Measured ground bounce when the power mode is transited from PARK to IDLE mode.

terrupt is gradually reduced at higher supply voltage and is approximately $\times$ 23 at a supply voltage of 1.5V. The diode-connected PFET transistor of the power gate switch in PARK mode provides a regulating effect to the leakage reduction across the allowed voltage range. Leakage power reduction is shown to be approximately $\times$ 2.6 lower than for IDLE mode.

Figure 12 and Figure 13 show the measured off-chip ground bounce when the power modes of DUT1 – DUT8 are switched from COLD to IDLE and from PARK to IDLE and DUT0 is held in IDLE mode. The family of curves is generated by repeating the mode transition and ground rail voltage measurement with supply voltage reset in 0.1V increments between 0.9V and 1.5V. The plots of Figure 12 and Figure 13 also show that ground bounce noise during the transition from PARK to IDLE is much smaller than that in the transition from COLD to IDLE.

The effect of the on-chip ground bounce on maximum operating frequency of DUT0 is measured by performing power-mode transitions of DUT1 – DUT8 from COLD to RUN and from 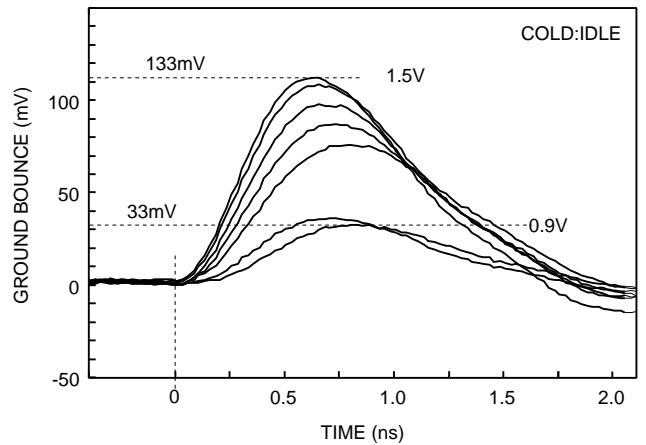PARK to RUN. For the power-transition of DUT1 – DUT8 from COLD to RUN, we turn off the NFET of a sleep transistor with $V_{GS} = 0$ and also turn off the PFET of the sleep transistor with $|V_{GS}| = 0$ and wait for 50 μs for the voltage level of all of internal nodes to be stabilized. Then, we turn on the NFET sleep transistor with V_{GS} = VDD and measure the maximum operating frequency of DUT0. The measurement is compared to the maximum operating frequency of DUT0 on the macro in which the sleep transistors of DUT1 – DUT8 are never turned off and clock are gated. This effectively excludes the ground bounce noise due to clock gating from our measurement. For the power-transition of DUT1 – DUT8 from PARK to RUN, we apply similar steps as for the transition from COLD to RUN. The bar chart of Figure 14 shows the degradation of the maximum operating frequency of DUT0 over the range between 0.9V and 1.5V. Unlike the measurement results shown in Figure 12 and Figure 13, Figure 14 shows the internal impact of power-mode transition related ground bounce on the maximum operating frequency of the CMOS logic circuits.

6. CONCLUSION

A power gating structure with two power saving modes

5
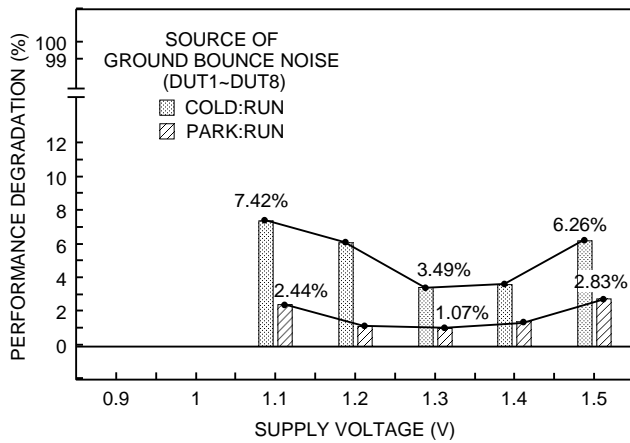
Figure 14: Effect of the ground bounce on the performance of nearby logic (DUT0).

for both high leakage reduction without state retention and intermediate leakage reduction with state retention is proposed and evaluated. Representative logic circuits with and without power gating circuit are designed and fabricated in 0.13 μm CMOS bulk technology. Measured results show when moderate area overhead is dedicated to the sleep transistor of power gating structure ($< 2.6\%$), maximum operating frequency decreases by less than 2.0%. Leakage current is dramatically reduced when the ground supply to the logic circuit is interrupted by the switch and is moderately reduced (slightly better than by a fact of two) when the built-in PFET switch is used to reduce the rail-to-rail voltage.

Ground bounce induced by switching between power modes are measured as well as its effect on performance of neighboring circuits. The intermediate leakage saving mode is shown to significantly reduce the ground bounce and effect on the neighboring circuit performance.

7. REFERENCES

[1] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamda, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE Journal of Sold-State Circuits*, vol. SC-30, pp. 847–854, Aug. 1995.

[2] J. Kao, S. Narendra, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," in *Proceedings of the Design Automation Conference*, pp. 495–500, June 1998.

[3] S. V. Kosonocky, M. Immediato, P. Cottrell, T. Hook, R. Mann, and J. Brown, "Enhanced multi-threshold (MTCMOS) circuits using variable well bias," in *Proceedings of International Symposium on Low-Power Electronics and Design*, pp. 165–169, Aug. 2001.

[4] M. Anis, S. Areibi, M. Mahmoud, and M. Elmasry, "Dynamic and leakage power reduction in MTCMOS circuits using an automated efficient gate clustering technique," in *Proceedings of the Design Automation Conference*, pp. 480–485, June 2002.

[5] S. Kim, S. V. Kosonocky, and D. R. Knebel, "Understanding and minimizing ground bounce during mode transition of power gating structure," in *Proceedings of International Symposium on Low-Power Electronics and Design*, pp. 22–25, Aug. 2003.

[6] S. Kim, S. V. Kosonocky, D. R. Knebel, K. Stawiasz, D. Heidel, and M. Immediato, "Minimizing inductive noise in system-on-a-chip with multiple power gating structures," in *Proceedings of European Solid-State Circuits*, pp. 16–18, 2003.

[7] K. Kumagai, J. Iwaki, H. Suzuki, T. Yamada, and S. Kurosawa, "A novel powering-down scheme for low Vt CMOS circuits," in *Digest of Technical Papers of IEEE Symposium on VLSI Circuits*, pp. 44–45, 1998.