

Effective Static Power Reduction Using Variation Tolerant Reconfigurable Multi mode Switches



¹ J. Sai Priyadarshini

²C. Madhu

¹ M.Tech(VLSI Design) Student, Department of ECE,S V Engineering College for Women, saipriya.15.91@gmail.com.

² Assistant Professor, Department of ECE,S V Engineering College for Women, cherukulamadhu@gmail.com

ABSTRACT: — A power-gating scheme was presented to support multiple power-off modes and reduce the leakage power during short periods of inactivity. However, this scheme can suffer from high sensitivity to process variations, which impedes manufacturability. Recently power-gating techniques are used for increasing power efficiency however, they are sensitive to process variations and having more wake up time. The proposed power gating scheme is tolerant to process variations and scalable to more than two intermediate power-off modes. The proposed design requires less design effort and offers greater power reduction and smaller area cost than the previous method. In addition, it can be combined with existing techniques to offer further static power reduction benefits. Analysis and extensive simulation results demonstrate the effectiveness of the proposed design.

Index Terms—*Leakage power, Multi-mode VTCMOS switches, Power Consumption reduction, process variation, Reconfigurable power-gating structure.*

1. INTRODUCTION:

As CHIP density increases relentlessly along Moore's law, power consumption is emerging as a major burden for Contemporary systems [1]. Dynamic energy is proportional to the square of the supply voltage. Thus, a lower voltage level yields a quadratic reduction in the energy consumption. To further reduce the dynamic power, systems-on-chip (SoCs) are partitioned into voltage islands with separate supply rail and unique power characteristics [2]–[4]. Moreover, as devices keep shrinking, the channel length shortens and the gate oxide thickness reduces, increasing the gate-induced drain leakage, the gate oxide tunneling current, and the junction leakage [5].

Many techniques have been presented in the literature for reducing static power. One common approach is to synthesize the circuit using dual-*Vt libraries*[6]. High-*Vt* cells reduce the leakage current at the expense of reduced performance; thus their use on noncritical circuit domains reduces the leakage Power considerably without affecting circuit performance. Another technique exploits the fact that the leakage power consumed by each gate strongly depends on the input vector applied at the gate. Therefore, in order to reduce static power, it controls the input vector and the internal state of the circuit during periods of inactivity. Various techniques reduce peak rush current. A special class of these techniques reduces the large current rush by using one intermediate power off mode, while the methods presented in and apply a three step wake-up process. Intermediate power-off modes overcome another limitation of power switches, i.e., the time required for recovering from the idle mode, referred to as the wake-up time.

Long wake-up time prohibits the use of power switches during short periods of inactivity In addition; there are applications that can exploit static power savings in parts of the system provided that these parts can wake up fast upon request. The long wake-up time of power switches prohibits their use in such cases too. In particular, this technique requires that the memory elements (flip-flops) are forced to specific logic values prior to the activation of a power-off mode. To address proposed a new flip-flop design (the phase-forcing flip-flop) to ensure that all internal gate nodes in the combinational logic will be forced to predictable states during the power-off mode. This new flip-flop is not available in common standard cell libraries, which limits the applicability. In addition, the zigzag topology requires that, for

each power supply, a pair of rails is distributed inside the standard cells (V_{dd} and V_{ddv} as well as V_{ss} and V_{ssv} , where V_{ddv} is the virtual V_{dd} rail and V_{ssv} is the virtual ground rail). This requirement drastically increases the area overhead.

Finally, dedicated design automation tools, which are not commonly available, are needed to support this design style. Increased overhead is also imposed by the method proposed, which requires additional power rails and extra bypass switches. The method proposed requires the intelligent placement of keepers on selected circuit lines. Besides the additional overhead, the keepers cannot be easily placed in

non regular structures. The authors proposed a structure with intermediate power-off mode, which reduces the wake-up time at the expense of reduced leakage current suppression. Similar structures were proposed. The authors extended this tradeoff between wake-up overhead and leakage power savings into multiple power-off modes. Using these techniques, instead of consuming power by remaining in the active mode during the short periods of inactivity, the circuit is put into an appropriate power-off mode (i.e., low-power state), which is determined by both the wake-up time and the length of the idle period. The longer the period of inactivity, the higher are the power savings achieved by using the most aggressive power-off mode that can be tolerated.

Even though the architecture proposed is efficient for reducing leakage power during short periods of inactivity, it has several drawbacks that limit its applicability. First, it cannot be easily extended to support more than two intermediate power-off modes and thus it cannot fully exploit the power reduction potential of the power-gating structure, especially for high-performance circuits. Second, the architecture consumes a significant amount of power, and this reduces the benefits offered by the power switches.

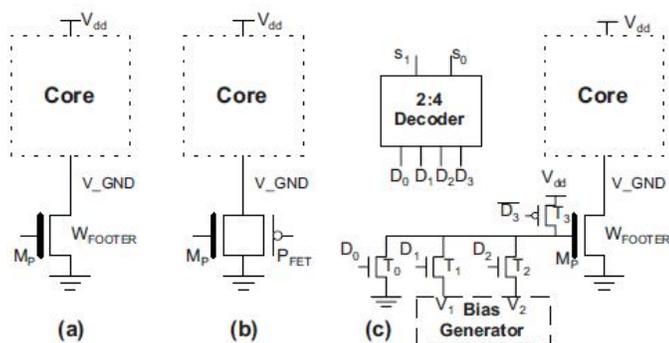


Fig 1 Various power-gating architectures.

Third, this structure is very sensitive to process variations, which can adversely affect its manufacturability and predictability. Finally, it is not easily testable, as it consists of analog components. In this paper, we present an effective body biasing architecture that has none of the above drawbacks of the architecture proposed. The proposed structure requires minimal design effort since it is very simple, and with no analog components. It is considerably smaller than the architecture proposed and offers greater power savings for similar wake-up times.

The proposed architecture is also more tolerant to process variations; thus its operation is more predictable. Finally, a reconfigurable version of the proposed architecture is also proposed, which can tolerate even greater process variations, enabling thus the utilization of the proposed architecture for newer technologies. The organization of the rest of this paper is as follows Section II presents background material to place the proposed work in an appropriate context. Section III introduces the proposed body biasing architecture, the design method, and the reconfigurable architecture. Section IV presents an evaluation of the proposed architecture, including comparisons with previous work. Finally, Section V concludes this paper.

2. MULTI-MODE POWER SWITCH TECHNIQUE

2.1. Proposed Architecture:

Fig.2 presents the proposed design. It consists of the main power switch transistor M_P and two small transistors M_0 and M_1 , each corresponding to an intermediate power-off mode (M_0 corresponds to the dream mode and M_1 corresponds to the sleep mode). Transistor M_P is a high- V_t transistor and it remains on only during the active mode. Transistors M_0 and M_1 are small low- V_t transistors that are turned on only during the corresponding power-off mode. (i.e., M_0 is turned on during the dream mode and M_1 is turned on during the sleep mode). The various modes of operation are as follows. One is active mode, snore mode, dream mode and sleep mode.

2.2 MODES:

2.1.1 Active Mode: Transistors M_P , M_0 , M_1 are on.

2.1.2 Snore Mode: Transistors M_P , M_0 , and M_1 are off as shown in Fig. 2(a). In this case, the leakage current of the

core, $I L_{core}$, is equal to the aggregate leakage current flowing through transistors M_0, M_1, M_P ($I L_{core} = I L_{M_0} + I L_{M_1} + I L_{M_P}$), which is very small (note that M_0, M_1 are small transistors and M_P is a high- V_t transistor). Thus the voltage level at V_GND is close to V_{dd} and the circuit consumes a negligible amount of energy, but the wake-up time is high.)

2.1.3 Dream Mode: Transistor M_0 is on and transistors MP and M_1 are off as shown in Fig. 2(b). In this case the current flowing through transistor M_0 (and, V_{dd} (i.e., $V_{V_GND} < V_{dd}$). Thus the aggregate current flowing through M_0, M_1 Static power consumed by the core is higher compared to the snore mode, but the wake-up time is less. Voltage level which is lower than transistor M_0 , and it sets the virtual ground node at the exact value of I_{M_0} depends on the size of an MP increases because M_0 is on ($I_{M_0} > I_{L_{M_0}}$).

2.1.4 Sleep Mode: Transistor M_1 is on, and MP, M_0 are off as shown in Fig. 2(c). Provided that transistor M_1 has larger aspect ratio than M_0 ($W_{M_1}/L_{M_1} > W_{M_0}/L_{M_0}$), the agree MP increases even more when M_1 is on (note that $I_{M_1} > I_{M_0}$). Consequently, the voltage level at the virtual ground node is further Reduced compared to the dream mode and thus the wake-up time decreases at the expense of increased power consumption ate current flowing through M_0, M_1 .

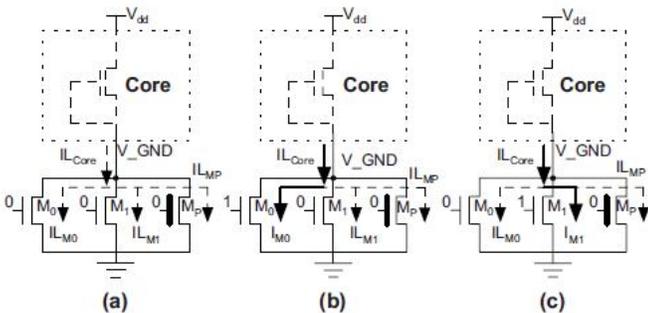


Fig 2 Proposed architecture: (a) snore mode, (b) dream mode, (c) sleep mode.

2.2. Design Method:

The correct operation of the proposed design depends on the correct sizing of transistors M_0 and M_1 . For simplicity, we model the core with a single equivalent nMOS transistor, and we consider only the sub threshold leakage current (we note that this structure is used only for the mathematical analysis in this section). Considering the very low transistor threshold voltage levels (V_{THC}) in nanometer technologies and the intermediate voltage levels at the virtual ground during the various intermediate power-off modes (excluding the

complete power-off mode), the corresponding power-off transistors M_0 and M_1 are in the linear region of operation when they are active ($V_{gs} = V_{dd}$).

This is because $V_{ds} = V_{V_GND} < V_{gs} - V_{THC} = V_{dd} - V_{THC}$, where V_{gs} and V_{ds} are the gate-source and drain-source voltages, respectively, for M_0 and M_1 . Let us consider the dream mode shown in Fig. 2(b). In this mode, transistor M_0 is on. Let V_0 V_GND be the voltage at the virtual ground node at this mode. As stated above, V_0 $V_GND < V_{dd} - V_{THC}$ (V_{THC} is the threshold voltage of the low- V_t transistors M_0, M_1) and taking into account that $V_{gs} - V_{THC} > 0$ we can deduce that M_0 operates in the linear region when it is on. Therefore, the current flowing through

M_0 is given by the following equation:

$$I_{M_0} = \mu_n C_{ox} \frac{W_{M_0}}{L_{M_0}} \left((V_{dd} - V_{THC}) V_{v_GND}^0 - \frac{(V_{V_GND}^0)^2}{2} \right) \dots(1)$$

$$I_L = I_{0e} \frac{V_{GS} - V_{th0} - \eta V_{DS} + \gamma V_{BS}}{n V_T} \left[1 - e^{-\frac{V_{DS}}{V_T}} \right] \dots(2)$$

Where I_0 is a constant, which depends on the width and length of the transistor. V_{th0} is the zero-bias threshold voltage, V_T is the thermal voltage, and n is the sub threshold swing coefficient. η is the DIBL coefficient and γ is the linearized body effect coefficient. In our case, for the diode-connected transistor that represents the core (in Fig. 2), $V_{GS} = 0, V_{DS} = V_{dd} - V_0$ $V_GND, V_{BS} = 0$, and the sub threshold leakage current of the core is calculated as

$$I_{L_{core}} = I_0^{core} e^{-\frac{V_{THC} - \eta(V_{dd} - V_{v_GND}^0)}{n V_T}} \left[1 - e^{-\frac{V_{v_GND}^0 - V_{dd}}{V_T}} \right] \dots(3)$$

The leakage current of power switch MP and transistor M_1 can be calculated in the same way. Based on Kirchhoff's current law, we can obtain $I L_{core} = I_{M_0} + I_{L_{M_1}} + I_{L_{M_P}}$. Note that $W_{M_1}/L_{M_1} > W_{M_P}/L_{M_P}$; thus $I_{L_{M_1}} > I_{L_{M_P}}$.

Therefore, the equation above is simplified to

$$I L_{core} = I_{M_0} + I_{L_{M_P}} \dots(4)$$

Substituting (1)–(3) into (4), we get the size of M_0 as

$$\frac{W_{M_0}}{L_{M_0}} = \frac{2(I_{L_{core}} - I_{L_{M_p}})}{\mu_n C_{ox} \left(2(V_{dd} - V_{THC})V_{V_GND}^0 - (V_{V_GND}^0)^2 \right)} \quad (5)$$

By using (5) we can adjust the voltage level V_0 V_{GND} to any value in the range $(0, V_{dd} - V_{THC})$ and calculate the aspect ratio of transistor M_0 . The wake-up time is calculated by the equation $T_{wake-up} = C_{total} \cdot Req$, where C_{total} is the parasitic capacitance of the virtual ground and Req is the equivalent resistance of transistor MP when it discharges the virtual ground node. Note that Req is the average resistance of MP for the conducting time duration.

Thus, the wake-up time is provided by the following equation:

$$T_{wake-up} = C_{total} \cdot \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{V_{V_GND}^0(t)}{I_D(t)} dt \quad (6)$$

or equivalently

$$T_{wake-up} = C_{total} \times \frac{1}{-V_{V_GND}^0} \int_{V_{V_GND}^0}^0 \frac{V}{I_D(V)} dV. \quad (7)$$

Since MP is in the linear region during the wake-up operation ($V_{GS} = V_{dd}$), (7) is written as follows:

$$T_{wake-up} = C_{total} \times \frac{-2I_{M_p}}{\mu_n C_{ox} W_{M_p} V_{V_GND}^0} \times \int_{V_{V_GND}^0}^0 \frac{1}{2(V_{dd} - V_{THC}) - V} dV. \quad (8)$$

The same analysis can be used for calculating the size and the related wake-up time for transistor M_1 (in that case, the voltage at virtual ground node is equal to V_1 V_{GND}). Equation (5) can be used for calculating the transistor size required to set the virtual ground rail at any particular voltage level in the range $(0, V_{dd} - V_{THC})$. Thus the extension of the design to more power-off modes is straightforward. Note that in the above analysis, we considered only the sub threshold leakage current for every device that is turned-off.

For a more accurate estimation, however, the total leakage current of the core and the power switch MP must be used in (5). During the transition from any power-off mode to the active mode, spurious glitches may occur in the internal circuit nodes that are not captured by the above simplified mathematical model. As a consequence, a detailed ramp-up analysis is needed for every circuit and every intermediate power-off mode. Note that, the wake-up time depends on the

internal state of the core since leakage current is input-pattern-dependent. As shown in Section IV, the highest wake-up time corresponds to the test vector which consumes the lowest leakage power at the power-off modes.

An upper bound of the wake-up time can be estimated by using worst case analysis which assumes that each gate receives the input combination that is the least leaky among all input combinations for this gate. However, in practice, there is no input pattern that can simultaneously drive the least leaky combination at the inputs of every gate of the circuit. For a more accurate estimation, we can apply techniques similar to those proposed, which find the input vector that minimizes static power.

Moreover, the proposed scheme can benefit from input vector control techniques to offer further static power reductions. By using the intermediate modes, the ground bounce in neighboring circuits can be reduced as well. Ground bounce is a well-known problem in MTCMOS power switched designs, which is caused by the large instantaneous current flowing during wakeup of a circuit. It is shown that the proposed method reduces ground bounce when the core wakes up from the intermediate power-off modes.

To further suppress ground bounce, various activation techniques can be exploited for the micro switches that form the main power transistor. Among them, very effective techniques include the "staircase" control of the voltage applied at the gate of the main power transistor and the daisy chain microswitch structures that can be further supported by proper sizing of the corresponding transistors.

In fact, the staircase control technique is similar to transitioning from the full power-off mode toward the power-on mode by visiting every intermediate power-off mode (the virtual ground node is gradually discharged). It is shown that the proposed technique can be combined with the staircase and the daisy-chain approaches to further reduce ground bounce.

3. EVALUATION AND COMPARISONS

In this section, we present simulation results and comparisons against other techniques presented in the literature.

3.1 Results And Comparisons Using A Large Logic Core:

The target of the first subsection is to evaluate the proposed method when it is applied to large logic cores that are comparable in size to real designs from industry. To this

end, we present simulation results on a large logic core consisting of 9 million transistors. This core consists of multiple inverters of various sizes which are driven by various input vectors. Even though it is not a real circuit, it is representative of a realistic industrial circuit in terms of static power consumption during dc operation in power-off mode. We used the 45-nm predictive technology with 1.1-V power supply.

The leakage power consumption of the core in idle mode with no power gating is equal to 10 mW. All simulations were done using the Synopsis HSpice simulation engine. We note that, because of the use of a different core with respect as well as different experimental parameters (such as the technology, the voltage setting, and the input vector), we cannot directly compare the experimental results of our method with the results presented. Therefore, we implemented both the architecture [see Fig. 1(c)] and the proposed architecture (see Fig. 2) for the aforementioned logic core. As suggested, the width of the main power switches. (Transistor denoted as *MP*).

| .sp file | Mode of operation | Power dissipation in Watts |
|----------|------------------------|----------------------------|
| Inv 1 | Normal inverter | 6.8218E-03 |
| Inv 2 | With Wfooter | 7.8290E-03 |
| Inv 3 | With transmission gate | 8.2616E-03 |
| Inv 4 | With bias network | 1.9170E-02 |
| Inv 5 | Snore mode | 2.5000E-05 |
| Inv 6 | Dream mode | 3.5311E-05 |
| Inv 7 | Sleep mode | 3.5311E-05 |
| Inv 8 | Active mode | 1.1779E-04 |

Table I: Static power dissipation

Was set equal to 12% of the total width of the n MOS transistors in the logic core. For the logic core that we used, the width/length ratio of transistor *MP* is calculated as equal to $43.2 \times 106 \text{ nm}/45 \text{ nm}$ and it is implemented as the parallel connection of a number of smaller transistors. In order to provide fair comparison, the transistor sizes in both architectures were selected in such a way as (a) to be of minimum size required and (b) to provide similar wake-up times, in both architectures. In dream and sleep mode the

power dissipation is same, as only one transistor is in on mode in the network apart from core logic. Moreover, in the proposed scheme, the sizes of transistors *M0* and *M1* have been selected in such a way as to provide the same voltage level at the virtual ground node with the scheme proposed at each power-off mode and for the same input vector. Thus, the logic core consumes the same amount of static power in both architectures at each power-off mode. For example, considering an input vector that drives the two-thirds of the transistors to logic “1” and the rest of the transistors to logic 0,” the voltage level at the V_GND node is equal to 217.1, 415.8, 541.8, and 668.5 mV at four intermediate power-off mode.

For both architectures, we assume that the voltage at the V_GND node settles to the expected value before the waking up process begins. In addition, the core is considered as fully operational after the virtual ground node is discharged to the value of 1% of V_{dd} . First, we compare both architectures in terms of area overhead measured as transistor sizes. The width of transistors *M0*, *M1* in the proposed structure is equal to 250 and 480 nm, therefore, for comparison purposes; we excluded the overhead of these transistors from the overhead of both architectures. We also exclude the decoder, as it is optional in both architectures and can be omitted (we implemented both schemes without the use of decoders). The rest of the circuitry in the proposed architecture occupies Almost one-fifth (1/4.8) of the area of the architecture and it is less than 0.0002% of the area of the core. Even though this is an estimate based on transistor sizes, it is apparent from Figs. 1(c) and 2 that the proposed architecture is much simpler.

Note that the schemes proposed support only one intermediate power-off mode, which is denoted as Dream for comparison purposes. Entries in Table II corresponding to the second intermediate power-off mode (i.e., the Sleep mode) which is not applicable for the schemes proposed, are denoted as “N/A” (not applicable). The last three rows show the number of cycles that are needed for waking up the core from each power-off mode that is supported by each method. The first two columns present the results for both a high- V_t and low- V_t parker transistor proposed. The next two columns present the results for the high- V_t and low- V_t parker transistor proposed. We assumed four different bias voltages, 0, -0.2 , -0.4 , and -0.6 V for both the high- V_t and low- V_t parker transistor proposed but the results were nearly the same (except for the sleep mode where the static power slightly

varied between 0.94 and 0.96 mW as reported in Table II). The next two columns present the results for configurations Conf. 1 and Conf. 2 that are used.

The last column presents the results for the proposed method. It is obvious that the methods proposed fail to deliver a tradeoff between wake-up time and power consumption regardless of the kind of parker transistor (high- V_t or low- V_t) or the bias voltage. Even though multiple types of these transistors and/or bias voltages are used at the same core, with an obvious impact on area overhead, they still fail to deliver a sufficient range of wake-up times. The method proposed in offers low static power consumption at the expense of very large wake-up times and increased area overhead. More importantly, similar to the method proposed, the method proposed supports only a single intermediate power-off mode.

In contrast to the proposed method offers more than one intermediate power-off mode with a wide range of wakeup times and, as will be presented later, the proposed method can easily provide even more than two intermediate power off modes—a target that is obviously unachievable for the other methods. Finally, the proposed method has the smallest used area overhead. Therefore, the proposed method better exploits the tradeoff between static power dissipation and wake-up time with much less area overhead than the rest of the methods.

4. EXPERIMENTAL RESULTS ANALYSIS

In figures a given below a Snore Mode, Dream Mode, Sleep Mode, Input/output waveforms. When input is high, output is low. The wave forms can represented in x-axis time and in y-axis voltage.

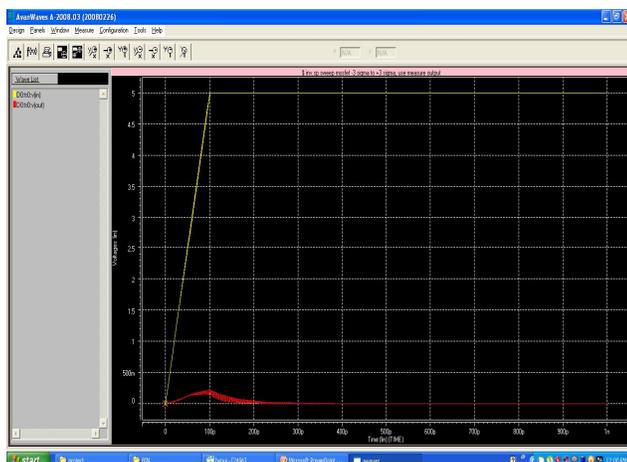


Fig4: Snore Mode Operation

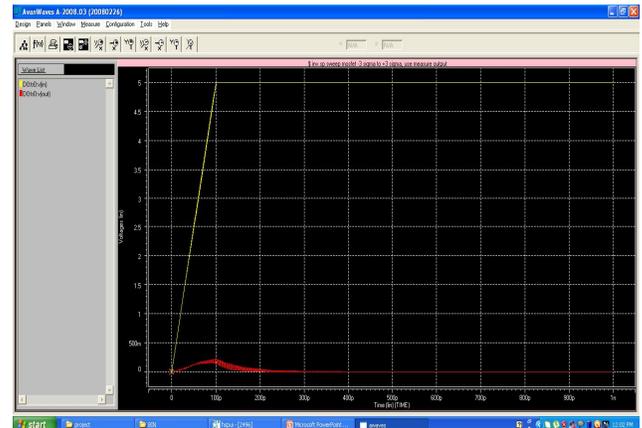


Fig5: Dream Mode Operation

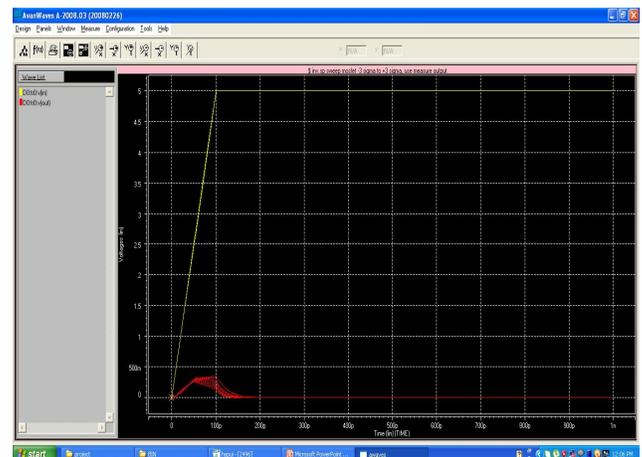


Fig6: Sleep Mode Operation

5. CONCLUSION

As chip density increases relentlessly along Moore's law, power consumption is emerging as a major burden for contemporary systems. Dynamic power is tackled nowadays by the reduction of the supply voltage level. Dynamic energy is proportional to the square of the supply voltage. Thus, a lower voltage level yields a quadratic reduction in the energy consumption. To further reduce the dynamic power, systems-on-chip (SoCs) are partitioned into voltage islands with separate supply rail and unique power characteristics. We described a new power-gating scheme that provides multiple power-off modes.

The proposed design offered the advantage of simplicity and required minimum design effort. It is very simple and all digital, and it is minimally sized since it

consists of only a single small transistor for each power off mode. Extensive simulation results showed that, in contrast to arecent power-gating method, the proposed design is robust to process variations and it is scalable to more than two poweroff modes. Moreover, it requires significantly less area and consumes much less power than the previous design. Finally, a reconfigurable version of this method can be used to increase the manufacturability and robustness of the proposed design in technologies with larger process variations.

REFERENCES

- [1] *Semiconductor Industry Association.* (2007) [Online]. Available:<http://www.itrs.net/Links/2007ITRS/Home2007.htm>
- [2] D. Lackey, P. Zuchowski, T. Bednar, D. Stout, S. Gould, and J. Cohn, "Managing power and performance for system-on-chip designs using voltage islands," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design*, Nov. 2002, pp. 195–202.
- [3] R. Puri, D. Kung, and L. Stok, "Minimizing power with flexible voltage islands," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2005, pp. 21–24.
- [4] R. Puri, L. Stok, J. Cohn, D. Kung, D. Pan, D. Sylvester, A. Srivastava, and S. Kulkarni, "Pushing ASIC performance in a power envelope," in *Proc. Design Autom. Conf.*, Jun. 2003, pp. 788–793.
- [5] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deepsubmicrometer CMOS circuits," *Proc. IEEE*, vol. 91, no. 2, pp. 305–327, Feb. 2003.
- [6] S. Idgunji, "Case study of a low power MTCMOS based ARM926 SoC: Design, analysis and test challenges," in *Proc. IEEE Int. Test Conf.*, Oct. 2007, pp. 1–10.