# Adaptive Power Control Technique on Power-Gated Circuitries

Wei-Chih Hsieh, *Student Member, IEEE*, and Wei Hwang, *Fellow, IEEE*

*Abstract*—An adaptive power control (APC) system on power-gated circuitries is proposed. The core technique is a switching state determination mechanism as an alternative of critical path replicas. It is intrinsically tolerant of process, voltage, and temperature (PVT) variations because it directly monitors the behavior of VDDV node. The APC system includes a multi-mode power gating network, a voltage sensor, a variable threshold comparator, a slack detection block, and a bank of bidirectional shift registers. By dynamically configuring the size of power gating devices, an average of 56.5% unused slack resulted from worst case margins or input pattern change can be further utilized. A 32–64 bit multiply-accumulate (MAC) unit is fabricated using UMC 90-nm standard process CMOS technology as a test vehicle. The measurement results of test chips exhibit an average of 12.39% net power reduction. A 7.96× leakage reduction is reported by power gating the MAC unit. For the 32-bit multiplier of MAC, the area and power overhead of proposed APC system are 5% and 1.08%, respectively. Most of the overhead is contributed by power gating devices and their control signal buffers.

*Index Terms*—Power control, power gating, switching state determination mechanism.

## I. INTRODUCTION

POWER issues continue to be critical challenges for the integrated circuits [1] as technology shrinking goes on. The power constraints nowadays are actually application-oriented. Energy-constrained applications require minimum energy consumption whereas performance-oriented applications pursue good power efficiency.

For energy-constrained applications such as emerging wireless sensor network or implantable medical electronics, performance is usually not a concern. Long device lifetime is desired because of the difficulty of a battery recharge or replacement. Ultra dynamic voltage scaling [2] technique which lowered the supply voltage down to subthreshold region of the transistor greatly reduced the power consumption. Moreover, techniques to determine the minimum energy point had been presented to address the demand of these energy-critical applications, either through schemes of energy slope tracking [3] or directly computing energy consumption [4]. These techniques scaled the

voltage to where the total amount of dynamic and leakage energy reached a minimum.

On the other hand, good power efficiency is demanded by performance-oriented applications to perform as many functions as possible for a certain amount of energy. Dynamic voltage frequency scaling (DVFS) techniques [5]–[10] were widely used for digital circuits to reduce the power consumption, especially under wide workload variations. Conventional DVFS techniques used critical path replica to track the circuit delay and to provide reference for adjusting voltage and frequency. Ring oscillator [5] was suggested to be a simple but efficient replica scheme to track gate delay variations. Except gate delay variations, interconnection delay and rise/fall delay variations have become more significant as the chip size grows. A more complicated delay synthesizer [6] was therefore proposed to track all these variations.

For the critical path replica technique, the effect of the process, voltage, and temperature (PVT) variations is not negligible. Among PVT variations, process variations have become more significant in advancing nanometer technology. The process variations can be divided into three categories, which are die-to-die, within-die (WID), and random (Ran) variations [8]. The critical path replica scheme suffers severely from WID and random variations because there is no way to detect these variations in other locations of the chip. Spatial environmental variations such as local power supply noise and thermal effect possess the same issue. Circuit wearout mechanisms such as NBTI and hot-electron degradation also have a growing importance on circuit yield. Adding delay margins on the critical path replica is straightforward for the technique to accommodate worst case WID, Ran, environmental variations, and transistor aging effects. However, required margins are getting more significant than ever with technology scaling. The increased margins directly limit the effectiveness of voltage scaling adopting critical path replica schemes.

Except using a fixed critical path replica, some other works dealt with variations by precharacterized data stored in lookup tables (LUT). The data in LUT was either used to program the critical path replica [9] or to instruct the system settings [10] in response to supply drops and thermal variations. A typical 5 × 5 matrix LUT implemented by register banks was reported to occupy about $0.4 \times 0.4\ \text{mm}^2$ [9]. The area overhead is the major drawback of the LUT scheme.

In addition to voltage scaling, power gating techniques are also widely used in low power digital circuits. Energy overhead, power/ground noise, information loss, speed degradation, and leakage reduction are main issues of the power gating technique. To deal with the energy overhead of the power gating device (PG), fine-grained configurations of PGs on small blocks
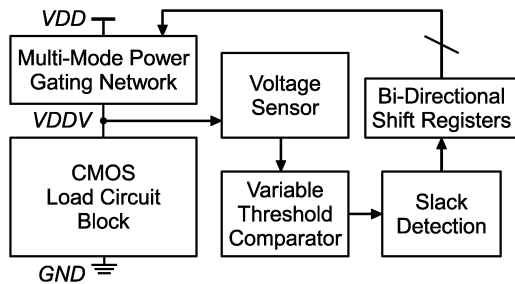
Fig. 1. Block diagram of proposed adaptive power control system.



Fig. 2. Simulation setup of inverters with PG to demonstrate speed control of PG.

were suggested [11], [12]. Charge recycling between gate node of PG and the virtual supply node [11] also helped to reduce the minimum sleep time required to compensate for the energy overhead. To suppress the power/ground noise induced by rush current at the wakeup, the wakeup timings of distributed PGs were proposed to be skewed [12]. To overcome information loss issue, data retention mode operation [13] was presented to retain the logic state. Choosing the size of PG is a tradeoff between standby leakage and speed degradation. A sizing strategy considering timing criticality and temporal currents [14] was presented to find the minimum size of PG under the speed penalty constraint. The standby leakage can be further suppressed using super cutoff technique to forward bias the gate-source voltage of PG. An automatic gate biasing technique was developed to find the optimal gate biasing voltage to maximize leakage reduction [15]. Overall, these previous works mainly contributed to cut off the power supply to reduce leakage during standby period. Although it had been reported that smaller power gating device would reduce the maximum operating frequency as well as dynamic power [11], none worked on dynamically controlling circuit speed by the power gating device.

To overcome the drawbacks of critical path replica scheme and to extend the usage of the power gating device, an adaptive power control (APC) technique which was first proposed in [16] is enhanced in this work. Section II briefly describes the overall architecture of the proposed APC system. Then the fundamentals of the APC system are presented in Section III. The speed control ability of the power gating device is introduced first, following by the derivation of the proposed switching state determination mechanism to determine the completion of switching event. The extension of the proposed mechanism for complex circuit blocks and the mechanism's tolerance of PVT variations are also given. In Section IV, the circuit operation of the proposed APC system is introduced as well as the effect of the variations on the implemented circuit. Section V shows the system overview, the simulation and test chip measurement results, and some more discussions about the proposed APC system. Finally, Section VI concludes this paper.

## II. ADAPTIVE POWER CONTROL SYSTEM ARCHITECTURE

Fig. 1 shows the block diagram of the proposed APC system. The proposed APC technique determines the completion of switching events and identifies the unused slack. The circuit speed is then altered to utilize the unused slack.

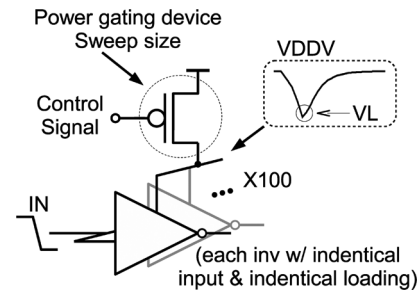The multi-mode power gating network (MPGN) that consists of power gating devices is used to control the speed of the load circuit. The control signals of the MPGN are stored in the bidirectional shift registers. The voltage sensor and the variable threshold comparator implement the switching state determination mechanism. They determine the completion of switching events from the behavior of the virtual supply node (VDDV). The completion of switching events can be considered the delay information of the load circuit. With the delay information, the slack detection block identifies the existence of the unused slack. Based on the degree of slack depletion, the MPGN is configured between modes dynamically to alter the circuit speed. The control loop responds every cycle to utilize the unused slack and to reduce power consumption under different operation conditions without harming the speed specification.

## III. SYSTEM FUNDAMENTALS

Two fundamentals of the proposed APC system are circuit speed control and the detection of completion of switching events. In this section, the speed control ability of the power gating devices is first investigated. Then the switching state determination mechanism, which is the core of the proposed APC system, is developed.

### A. Power Gated Speed Control

The concept of speed control through power gating device is inspired by an on-chip digital power supply [17]. However, instead of concerning a stable supply voltage, the behavior of the voltage and the current in the presence of power gating device is analyzed in this work.

Assuming that a $p$-type power gating device is used, the power network is split into a permanent power connected to the power supply and a virtual power network (VDDV) that drives the circuits. The voltage drop on VDDV node during switching can be considered the supply noise to the logic gates. The peak magnitude depends on the current drained by the load circuit and on the size of the power gating device. It had been demonstrated that timing distortion is more dependent on the average of supply voltage than on the peak of supply noise [18], [19]. Larger peak voltage drop on VDDV node implies lower average supply voltage of logic gates, which in turn stretches the gate delay. Therefore, sizing power gating device should be able to control the circuit delay without a physical variable voltage supply.

The simulation environment of inverters with PG shown in Fig. 2 is used to demonstrate the relationship of circuit speed versus PG size. UMC 90-nm standard process CMOS technology is used throughout this work. Only normal $V_{\text{th}}$ (about
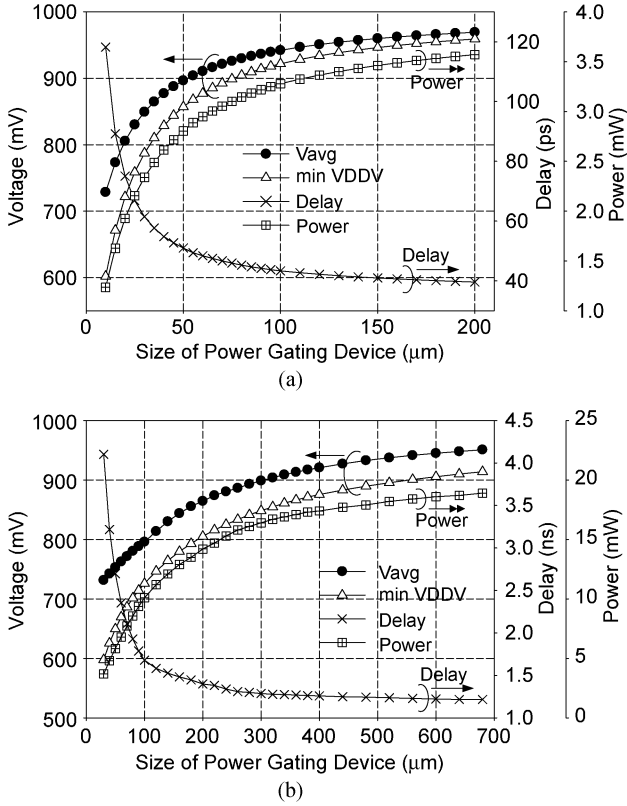
Fig. 3. Delay, power, minimum, and average voltage on VDDV node versus PG size (a) of inverters and (b) of 32-bit multiplier.



Fig. 4. Current characteristic of a 16-bit multiplier with ideal power supply.

250 mV) devices are used. The nominal power supply ($V_{\mathrm{DD}}$) is 1 V. $P$-type power gating device is used. In order to enlarge the effect, 100 identical inverters are used in parallel for the experiment. All the inverters share one identical input which is buffered by other inverters with enough driving strength to emulate the real logic behavior. Every inverter under test is loaded with an identical inverter. The experiment is also conducted on a 32-bit multiplier with similar environment as Fig. 2. The PG size is swept from 10 to 200 $\mu$m for inverter case and from 30 to 700 $\mu$m for multiplier case to observe the delay and power change.

The complete switching here is defined when the output of the inverter (or the last switching output bit of the multiplier) reaches 90% of the power supply, which is 0.9 V in this work. The switching window, which is the delay, is defined as the time from 50% of input falling to the completion of switching. The average of VDDV is calculated during the switching window. Only the output rising case is examined because the $p$-type PG affects the rising output only.

The results are shown in Fig. 3, including the minimum VDDV, the average of VDDV, the delay, and the power. Both for the inverter and the multiplier cases, the minimum and the average VDDV reduce along with the size of PG. The delay increases with smaller PG size, whereas the power decreases. These results support that PG size will affects the average supply voltage. Results also confirm that the circuit delay is dependent on the average supply voltage. In other words, by configuring PG size, the circuit speed and hence the circuit power can be controlled.
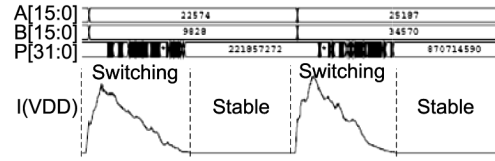
Note that the delay increases exponentially when the PG size is smaller than 50 and 100 $\mu$m for inverter case and multiplier case, respectively. It is a result of that the VDDV node is still lower than 0.9 V when the last output bit finishes switching. The rest of the recorded delay time is used to charge the VDDV node as well as the output node of the circuit. It takes longer to charge the nodes to 0.9 V especially when PG size is small.

### B. Switching State Determination Mechanism

Mathematical models were widely used to estimate the circuit delay [18], [20] and served well as the basics of static timing analysis (STA). But delay information, or the 90% $V_{\mathrm{DD}}$ criterion, is hard to acquire dynamically. In this work, instead of calculating the circuit delay, a switching state determination mechanism is proposed.

The concept of the proposed mechanism is to distinguish between switching and stable states of the circuit [16]. Fig. 4 illustrates the drained current change of a 16-bit multiplier with ideal power supply. Obviously, only leakage current exists in the stable state. Therefore, by monitoring the change of drained current, the circuit state can determined. Note that in the presence of PG, the charging period after switching described in Section III-A should be considered to be in the stable state. The current through PG is linearly related to the drain-source voltage ($V_{\mathrm{DS}}$) of PG. So the already existing PG adequately satisfies the requirement of monitoring the change of drained current.

A CMOS inverter is first analyzed here for simplicity. Fig. 5 depicts the switching response of an inverter with a $p$-type PG on top. The VDDV node exhibits a fall-then-rise behavior, supplying different amounts of current for the inverter. The proposed switching state determination mechanism discards conventional determination threshold such as 90% $V_{\mathrm{DD}}$ to determine the completion of switching event. Instead, it is defined at point B as in Fig. 5 for proposed mechanism. Point B is not a fixed point. It is chosen when $V_{\mathrm{DS}}$ of P1 equals to a predefined value ($V_{\mathrm{DS0}}$). After point B, drain current through PG keeps decreasing as VDDV and internal nodes being charged up. If the drain current of PG can be measured to be smaller than that at point B, the completion of switching event can be claimed. Note that the drain current of P1 is always equal to that of PG. Therefore, using the drain current of P1 at point B, the claim can be expressed as

$$I_{D,P1@B} \geq I_{D,PG,\text{after } B}. \tag{1}$$

Both PG and P1 are in linear region here. By adopting alpha power model [21], [22], $I_{D,P1@B}$ and $I_{D,PG}$ can be written as

$$I_{D,P1@B} = \frac{W}{L}\mu_{\mathrm{eff}}C_{\mathrm{OX}}V_{\mathrm{DS0}}\left(VV - V_{\mathrm{th}} - \frac{1}{2}V_{\mathrm{DS0}}\right) \tag{2}$$
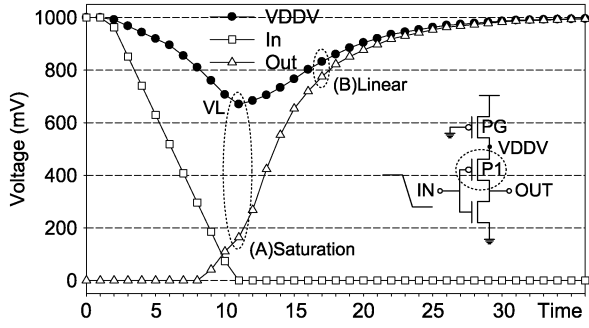
Fig. 5. Inverter switching response with power gating device on top.



Fig. 6. Graphical solution of inequality (7).

$$I_{D,PG} = \frac{W}{L}\mu_{\text{eff}}C_{\text{OX}}(V_{\text{DD}} - VV) \times \left(\frac{1}{2}V_{\text{DD}} - V_{\text{th}} + \frac{1}{2}VV\right) \tag{3}$$

where $(W/L)$ is the channel width-to-length ratio, $C_{\text{OX}}$ is the gate oxide capacitance per unit area, $\mu_{\text{eff}}$ is the effective mobility, $V_{\text{DS0}}$ is the predefined value of the proposed mechanism, $V_{\text{th}}$ is the threshold voltage of pMOS transistor, and $VV$ is the transient value of VDDV node. Note that $VV$ in (2) represents the value of VDDV at point B, whereas $VV$ in (3) stands for that of VDDV *after* point B. These two variables should be different. However, the "equal" property in (1) is mainly concerned. So the same $VV$ variable is used in (2) and (3) for simplicity. From [22]

$$\frac{W}{L}\mu_{\text{eff}}C_{\text{OX}} = \frac{I_{D0}}{V_{D0}\left(V_{\text{DD}} - V_{\text{th}} - (1/2)V_{D0}\right)} \tag{4}$$

where $V_{D0}$ is the drain saturation voltage when $V_{\text{GS}} = V_{\text{DD}}$. $I_{D0}$ is a modified drive current parameter of the transistor, differing with the transistor geometry such as width and length. Therefore, (1) can be rewritten using (2)–(4) after simplification as

$$I_{D0C}V_{\text{DS0}}\left[VV - V_{\text{th}} - (1/2)V_{\text{DS0}}\right] \geq I_{D0PG}(V_{\text{DD}} - VV) \\ \times \left[(1/2)V_{\text{DD}} - V_{\text{th}} + (1/2)VV\right]. \tag{5}$$

$I_{D0C}$ is the drain current parameter of the circuit, which is P1 in this case. And $I_{D0PG}$ is a similar parameter of the power gating device.

On the other hand, the largest voltage drop on VDDV node occurs at point A, as shown in Fig. 5. It is recorded as $VL$. Note that $VL$ is not a constant value either. It varies with different widths of PG, different load circuits and different operation conditions. The drain currents through PG and P1 are equal at point A. P1 is in saturation region whereas PG is always in linear region. Here the output node of the inverter is assumed not being charged yet. Therefore, $V_{\text{DS}}$ of P1 equals $VL$. Again using alpha
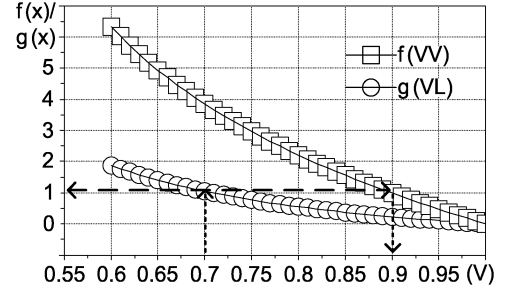
power model [21], [22] with channel length modulation property, the current equality at point A can be expressed after simplification as

$$I_{D0C}\left(\frac{VL - V_{\text{th}}}{V_{\text{DD}} - V_{\text{th}}}\right)^{\alpha}(1 + \lambda VL) = \frac{I_{D0PG}(V_{\text{DD}} - VL)}{V_{D0}\left(V_{\text{DD}} - V_{\text{th}} - \frac{1}{2}V_{D0}\right)} \\ \times \left(\frac{1}{2}V_{\text{DD}} - V_{\text{th}} + \frac{1}{2}VL\right) \tag{6}$$

where $\lambda$ is the channel length modulation parameter. Substituting $I_{D0C}$ into (5), the inequality can be simplified as

$$\left(\frac{V_{\text{DD}} - V_{\text{th}}}{VL - V_{\text{th}}}\right)^{\alpha} \cdot \frac{V_{\text{DD}} - VL}{(1 + \lambda VL)} \cdot \frac{\frac{1}{2}V_{\text{DD}} - V_{\text{th}} + \frac{1}{2}VL}{V_{D0}\left(V_{\text{DD}} - V_{\text{th}} - \frac{1}{2}V_{D0}\right)} \\ \geq \frac{(V_{\text{DD}} - VV)\left(\frac{1}{2}V_{\text{DD}} - V_{\text{th}} + \frac{1}{2}VV\right)}{V_{\text{DS0}}\left(VV - V_{\text{th}} - \frac{1}{2}V_{\text{DS0}}\right)}. \tag{7}$$

Note that transistor geometry related parameters, such as $I_{D0C}$ and $I_{D0PG}$, are eliminated. This suggests that the proposed switching state determination mechanism is independent of the geometry of the power gating devices or load circuits. It depends only on $VV$ and $VL$, which is the dynamic behavior of the virtual power supply node.

Graphical solution shown in Fig. 6 is used to obtain direct relationship between $VL$ and $VV$ in (7). Two sides of (7) are considered two individual functions with $VL$ and $VV$ as variables. Left hand side of (7) is plotted as $g(VL)$ whereas right-hand side of (7) is $f(VV)$ in Fig. 6. $V_{\text{DS0}}$ in (7) is set to 100 mV. By choosing an arbitrary $VL$, a minimum $VV$ can be obtained from Fig. 6 to satisfy (7). Repeat the process with different $VL$ until enough data point pairs of $VL$ and $VV$ are acquired. With these data pairs, a quadratic fitted inequality which serves as a good approximation of inequality (7) can be shown as

$$VV \geq 0.0902 + 1.7137 \cdot VL - 0.8109 \cdot VL^2. \tag{8}$$

The unit is Volt. Since the "equal" property is concerned, the value of $VV$ when the "equal" property is satisfied is defined as $V_{\text{det}}$. Therefore, for any lowest VDDV value $(VL)$ captured during switching, VDDV $(VV)$ must rises higher than $V_{\text{det}}$ to satisfy inequality (7) or (8) to claim a completion of the switching event.

## C. Mechanism Extension to Complex Circuits

The proposed switching state determination mechanism in [16] and previous section is based on a simple inverter. Here the proposed mechanism is modified to be suitable for complex circuit blocks. A complex CMOS circuit block can be considered a series of inverters with different sizes. These conceptual inverters do not switch at the same time. A virtual signal wave front can be imagined that propagates in the circuit block. The cells touched by the wave front start to switch. As the wave front advances from the input nodes, the number of switching cells gradually increases. After reaching a maximum, the number of switching cells reduces and converges at the output nodes. $VL$ is defined in the same way as in Section III-B, where maximum switching cells induce a largest voltage drop on VDDV node. And the point B is defined when the difference between VDDV node and the last output switching bit reaches $V_{DS0}$. The cells being charged after point B are those with logic high outputs. Unlike the single inverter case, the number of switching cells at point A differs from that being charged after point B. As a result, the $I_{D0C}$ parameters in (5) and (6) are different for complex circuit blocks. Since only parts of the cells are switching at point A, only a fraction of $I_{D0C}$ should be considered. An "$m$" factor is introduced to represent this fraction, where $0 < m \leq 1$. Therefore, (6) can be rewritten as

$$m \cdot I_{D0C} \left( \frac{VL - V_{th}}{V_{DD} - V_{th}} \right)^{\alpha} (1 + \lambda VL)$$
$$= \frac{I_{D0PG}(V_{DD} - VL)}{V_{D0}\left(V_{DD} - V_{th} - \frac{1}{2}V_{D0}\right)} \left( \frac{1}{2}V_{DD} - V_{th} + \frac{1}{2}VL \right) \quad (9)$$

and the inequality becomes

$$\frac{1}{m} \left( \frac{V_{DD} - V_{th}}{VL - V_{th}} \right)^{\alpha} \cdot \frac{V_{DD} - VL}{(1 + \lambda VL)} \cdot \frac{\frac{1}{2}V_{DD} - V_{th} + \frac{1}{2}VL}{V_{D0}\left(V_{DD} - V_{th} - \frac{1}{2}V_{D0}\right)}$$
$$\geq \frac{(V_{DD} - VV)\left(\frac{1}{2}V_{DD} - V_{th} + \frac{1}{2}VV\right)}{V_{DS0}\left(VV - V_{th} - \frac{1}{2}V_{DS0}\right)}. \quad (10)$$

If $m = 1$, it is equivalent to the single inverter case in Section III-B. Fig. 7 illustrates the $V_{det} - VL$ relationships with different $m$ parameters. Smaller $V_{det}$ can be observed for smaller value of $m$. Note that $m$ can also generally represent the percentage of $I_{D0C}$ being considered at every specific time point. Hence the extended inequality (10) can be used to determine the complete switching of cells at each point. The value of VDDV at that particular point is considered as its $VL$ variable. And a required $V_{det}$ can be calculated through (10) or its fitted relationship.

To find the $m$ parameter for a complex circuit block, the switching timing window is needed for each cell in the block. Similar to [14], *Synopsys PrimeTime* [23] is used to obtain the timing windows.

A 32-bit multiplier using standard cell library provided by UMC is synthesized to illustrate the procedure. The cells in the library are not all single gates. Some cells actually consist of multiple single gates to provide a specific function. Internal nodes of these cells may rise even when the output nodes fall. Considering only the falling window as in [14] is not appropriate in this case. Therefore, the union of rising and falling window is
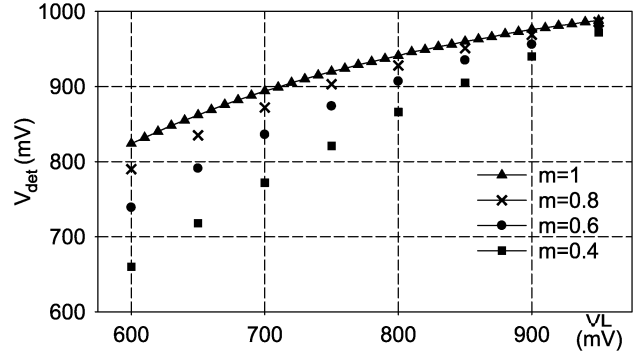


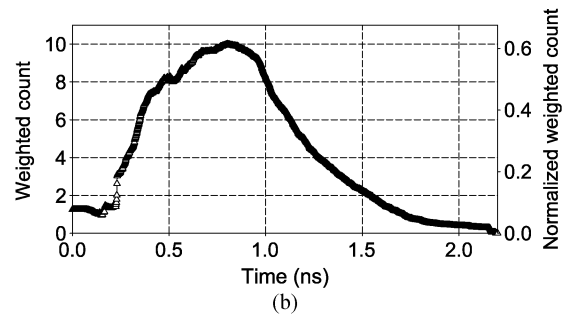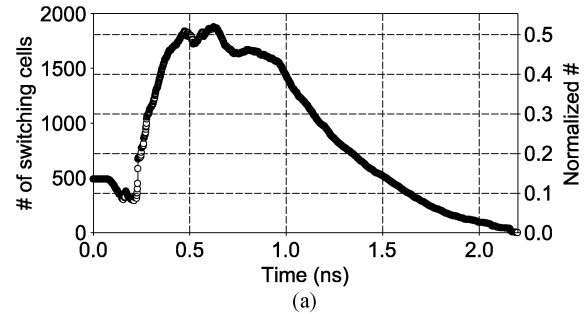Fig. 7. $V_{det} - VL$ relationships with different $m$ parameters.



Fig. 8. $m$ parameter (a) from number of switching cells and (b) with power weightings.

used as the switching window of the cell. Note that even if the output remains the same, there might still be internal switching for a multi-gate cell. So the switching window is defined by input transition timings.

Fig. 8(a) shows the number of switching cells at every time step. Total cell count is 3614 for the synthesized 32-bit multiplier. Normalized switching cell count is also depicted using the axis on the right. So based on Fig. 8(b), $m$ can be chosen as 0.52 for (10).

However, cells in the library are all different with different $I_{D0}$. Besides, both rising and falling cases are included. But the input states as well as output states are unknown in runtime. Hence, the probabilities of switching states are adopted as in [24]. In addition, power parameters which can be found from the databook of the library are proportional to $I_{D0}$. So the probabilities together with the power parameters are used as weightings of cells. The weighting of a cell can be expressed as

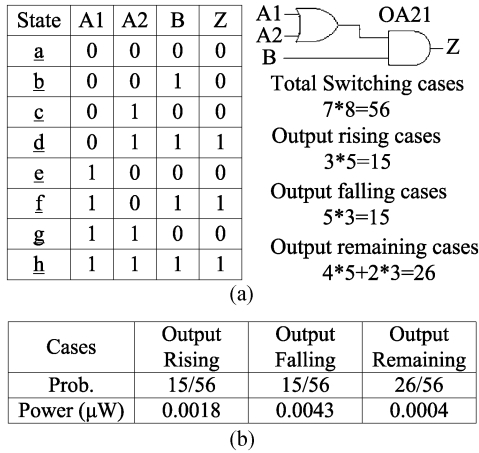$$W_{cell} = \sum_i \text{Prob}_i \times P_i. \quad (11)$$

| State | A1 | A2 | B | Z |
|-------|----|----|----|----|
| a | 0 | 0 | 0 | 0 |
| b | 0 | 0 | 1 | 0 |
| c | 0 | 1 | 0 | 0 |
| d | 0 | 1 | 1 | 1 |
| e | 1 | 0 | 0 | 0 |
| f | 1 | 0 | 1 | 1 |
| g | 1 | 1 | 0 | 0 |
| h | 1 | 1 | 1 | 1 |

A1 ⎤
A2 ⎦ OR → AND → Z
B

OA21

Total Switching cases
7*8=56

Output rising cases
3*5=15

Output falling cases
5*3=15

Output remaining cases
4*5+2*3=26

(a)

| Cases | Output Rising | Output Falling | Output Remaining |
|-------|---------------|----------------|------------------|
| Prob. | 15/56 | 15/56 | 26/56 |
| Power (µW) | 0.0018 | 0.0043 | 0.0004 |

(b)

Fig. 9. Illustration of how to find the weighting. (a) Switching states computation and (b) parameters for weighting computation.



Fig. 10. $V_{\mathrm{det}} - VL$ relationships under 10% supply voltage variations.

$i$ is over output rising, output falling, and output remaining cases. $\mathrm{Prob}_i$ and $P_i$ are the probability and power parameters of state $i$, respectively.

An OA21 cell in the library is used to demonstrate the procedure of obtaining the weighting. It is a typical multi-gate cell with an OR gate followed by an AND gate. The switching states are computed as in Fig. 9(a). As previously mentioned, the switching timing window is defined by an input state transition for the cell. For example, if state a is the initial state, there are seven other states that can be switched to. And there are eight initial states for OA21 cell. So the total switching cases will be $7 * 8 = 56$. The number of switching cases of output rising, falling, and remaining can be computed in the same manner. The probabilities and the power parameters of three switching types are shown in Fig. 9(b). Then the weighting of OA21 cell can be calculated from (11) as 0.00182.

All the cell weightings of the 32-bit multiplier are computed in the same manner. Then a more practical weighted switching cell count can be obtained as shown in Fig. 8(b). Note that the maximum of the normalized result is about 0.615 with weightings, which is higher than that without weightings. Therefore, $m$ parameter should be extracted from the weighted results of the circuit block to prevent underestimation.

### D. Mechanism's Tolerance of Variations

The same experiment environment as Fig. 2 is used to analyze the proposed switching state determination mechanism's tolerance of variations. For simplicity, the developed mechanism (10) with $m = 1$, which is equivalent to (7), is examined here.

UMC 90-nm CMOS TT corner technology at room temperature is assumed for voltage variation analysis. Fig. 10 shows the $V_{\mathrm{det}} - VL$ relationships with respect to 10% supply voltage variations. There are three sets of curves in Fig. 10. Each set represents a supply voltage condition. The solid symbol curve in each set stands for the mathematical results of (7) where as the hollow curve shows the simulated results. $V_{\mathrm{det}}$ of simulated results is obtained from the value of VDDV when the $V_{\mathrm{DS}}$ of the pMOS of the inverter equals the 0.1 V definition.
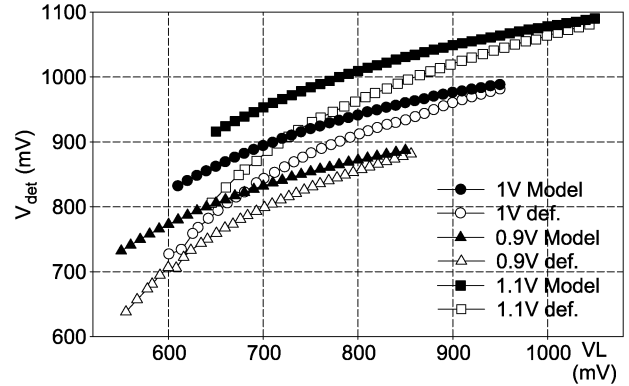
The simulation results agree with the presented model that lower $VL$ requires lower $V_{\mathrm{det}}$ to claim the completion of the switching event. The data also reveals that supply voltage variations incur a noticeable shift on the $V_{\mathrm{det}} - VL$ relationship. The shift is approximately equal to the amount of supply voltage variations. By taking 10% supply voltage variations into account in (7), i.e., altering the value of VDD correspondingly, the supply voltage variations can be compensated as in Fig. 10.

Note that the quadratic fitted inequality (8) is used to compute data points for the curve labeled "1 V Model" in Fig. 10. As for the other two "Model" curves, their own fitted inequalities are recalculated, though not shown here, by putting corresponding VDD values into (7). The model curves for the following figures are plotted in the same way.

A gap can be observed in Fig. 10 between the model and the simulated results. This gap comes from the assumption made in (6) that at point A the output node is not being charged yet. This assumption makes the proposed model assert later than the actual occurrence of the definition of the completion (0.1 V $V_{\mathrm{DS}}$). It also contributes to the proposed mechanism's tolerance of some variations which will be discussed later.

Fig. 11(a) shows simulated delays along with $V_{\mathrm{det}} - VL$ relationships of (7) at 1 V. There are three delay curves in the figure. The first one is the same 90% VDD metric as in Fig. 3. The second one is measured at the time when the value of VDDV satisfies the required $V_{\mathrm{det}}$ from the proposed model. The third one is measured at the time when $V_{\mathrm{DS}}$ of the pMOS of inverter satisfies the 0.1 V $V_{\mathrm{DS}}$ definition. Lower $VL$ implies that smaller PG is used, though not shown in the figure. These curves are simulated using the same set of power gating device sizes. It can be observed that model delays are always larger than definition delays. This is consistent with that the proposed model asserts later than actual occurrence of the definition. Compared to 90% delays, model delays are a little larger when $VL$ is high because of higher $V_{\mathrm{det}}$ requirement to assert completion. But in lower $VL$ cases, proposed model shows better efficiency on determining the circuit delay. For the clarity of the figure, only delay curves acquired from $V_{\mathrm{det}}$ satisfaction and $V_{\mathrm{DS}}$ definition are shown in the following.

Delay curves for three supply voltages are shown in Fig. 11(b). When supply voltage changes, there is a significant shift on $VL$ for the same PG size. And the delay also changes. Note that supply voltage compensation for (7) is adopted as
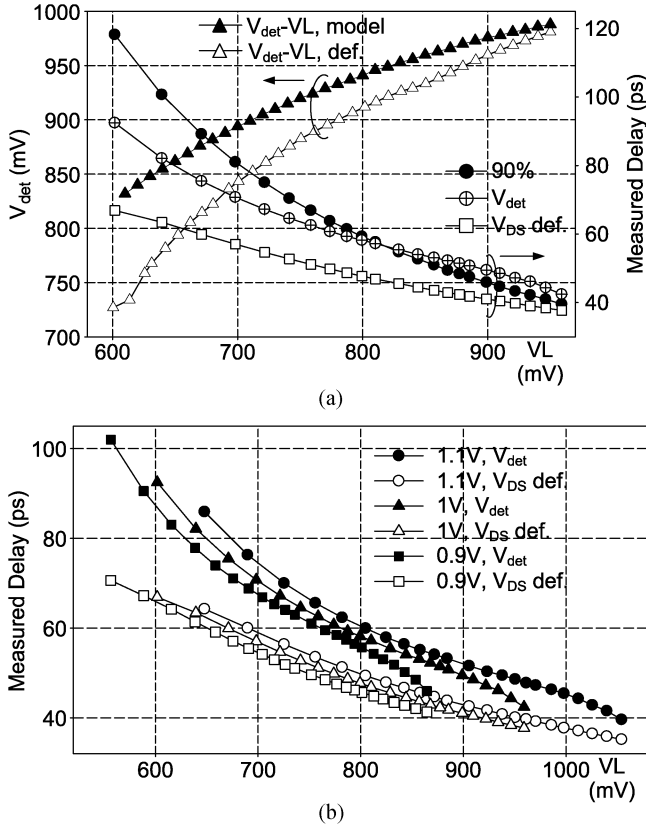
Fig. 11. Simulated delays (a) along with $V_{\text{det}} - VL$ relationships at 1 V and (b) under supply voltage variations.

mentioned to determine the delay. This compensation does help the proposed model to track the delays under supply voltage variations, as shown in the figure.

Different temperatures, from 125 °C to −25 °C, are simulated to demonstrate the effect of temperature variations on $V_{\text{det}} - VL$ relationships. TT corner, 1 V supply voltage is used here. A gap can still be observed between the model curve of (8), which is the quadratic fitted result of (7), and the simulated curves as in Fig. 12(a). The simulated relationships from 0.1 V $V_{\text{DS}}$ definition of different temperatures almost overlap each other. It means that the $V_{\text{det}} - VL$ relationship shift induced by temperature variations is very small.

Measured delays under different temperatures are shown in Fig. 12(b). The temperature variations also induce a small $VL$ shift for the same PG size that is due to the transistor parameter change under temperature variations. The delays from model determination track the delay variations well which can be observed from the figure. Note that the same model from (7) is used under different temperatures, no modification is applied. Therefore, temperature variations are tolerated by the proposed mechanism intrinsically.

Monte Carlo (MC) simulations are conducted using the transistor variation model provided by UMC from their measured data of physical devices. Transistor mismatch variations and the process variations are included in this MC model. Therefore the effects of random WID variations on proposed mechanism are examined.

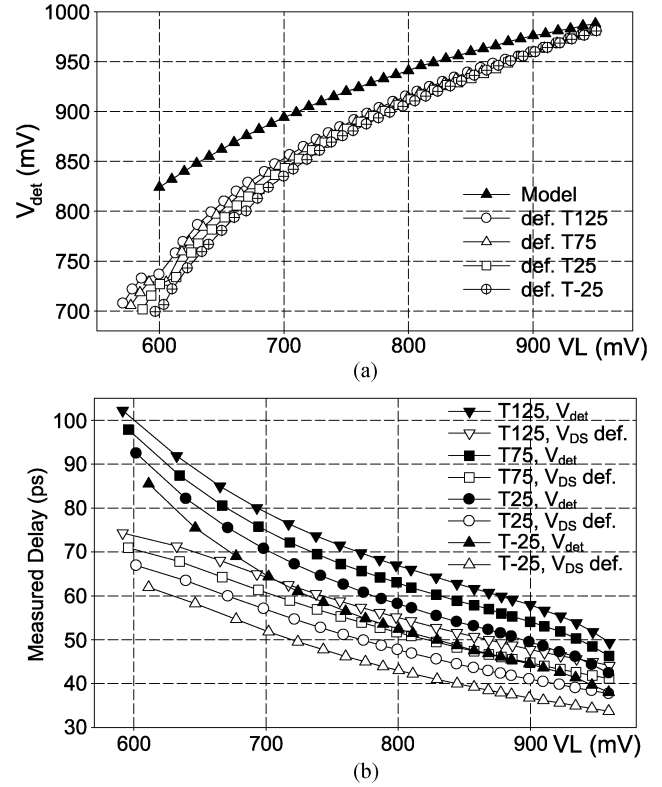The $V_{\text{det}} - VL$ relationship curves are shown in Fig. 13(a), whereas measured delay curves are plotted in Fig. 13(b). The



Fig. 12. On temperature variations. (a) $V_{\text{det}} - VL$ relationships and (b) measured delays ($T$ unit: °C).
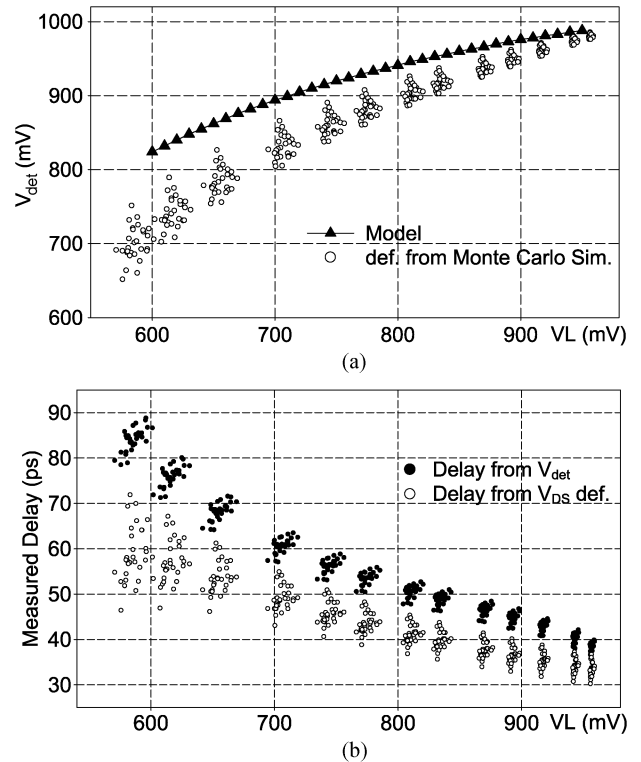


Fig. 13. On process variations. (a) $V_{\text{det}} - VL$ relationships and (b) measured delays.

simulation results are obtained under 1 V supply voltage and room temperature. Groups of MC simulations can be observed in the figure. Each group represents a setting of PG size. There
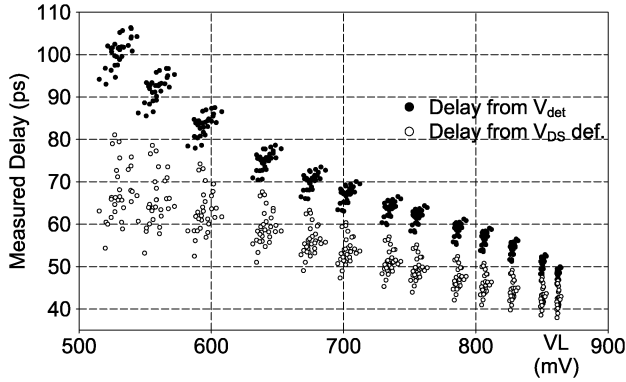
Fig. 14. Measured delays under worst case 0.9 V supply voltage, 125 °C temperature, and random process variations.



Fig. 15. (a) Multi-mode power gating network and (b) bidirectional shift registers.

are 30 MC iterations in each group. The random WID variations contribute to the variations of $V_{\mathrm{det}} - VL$ relationships and the measured delays. In Fig. 13(a), MC data points of the $V_{\mathrm{det}} - VL$ relationships spread around the nominal case point in each group. The measured delays also show a similar spread. It can be observed from Fig. 13(b) that the spread of model delays is smaller than that of definition delays. This is because that $V_{\mathrm{det}}$ of the model is a computed result from $VL$, where as the 0.1 V $V_{\mathrm{DS}}$ definition is fully influenced by random WID variations. Overall, determined delays from unmodified fitted model (8) track the definition delays. Therefore, the random WID variations are again tolerated by the proposed mechanism as shown in Fig. 13.

A worst case scenario is also presented in Fig. 14. MC simulations with random WID variations are conducted under 0.9 V supply voltage and 125 °C temperature. As can be seen in the figure, the measured delays are higher in the worst case scenario than previous cases. The model determined delays track definition delays with a gap in the figure as expected.

To summarize, the $V_{\mathrm{det}} - VL$ relationship is the most sensitive to the supply voltage variations. So the supply voltage variations should be taken good care of when putting proposed switching state determination mechanism into practice. The statistics show that the temperature and random WID variations are intrinsically tolerated by proposed mechanism. There are also circuit aging variations such as NBTI and hot electron degradation. It is not always able to separate these effects [25]. However, it was suggested that all the variation sources can be translated into an effective variation in threshold voltage [26]. In other words, these variations directly affect the transistors and in turn the behavior of the VDDV node. Proposed switching state determination mechanism has an advantage of monitoring the VDDV node. Therefore it has a good capability to tolerate and adapt to variations.

## IV. APC CIRCUITS

In this section, the circuit implementation of the proposed APC system and the circuit performance under variations are described.
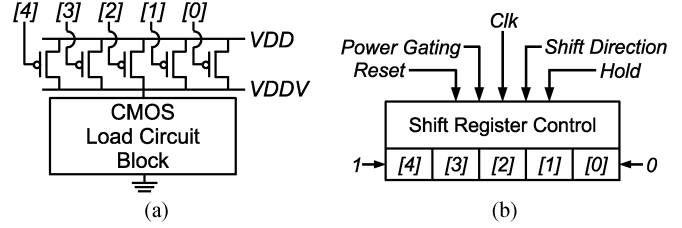
### A. Circuit Implementation

The multi-mode power gating network (MPGN) is implemented as a parallel-connected pMOS power switch network as shown in Fig. 15(a). The control signals of MPGN are stored in a bank of bidirectional shift registers as in Fig. 15(b). MPGN can be configured to have multiple modes with different numbers of pMOS turned on, causing different circuit delay as described in Section III-A. The number of modes can be chosen arbitrarily whereas there are five modes in this demonstration. More modes will provide finer control on the circuit delay with higher control power overhead.

The power gating devices are sized by approximating the circuit delay versus average supply voltage. For an inverter, the delay can be approximated as

$$C * V_{\mathrm{det}}/I_{D,\mathrm{avg}}. \tag{12}$$

$C$ is the output capacitance. For every $VL$, $V_{\mathrm{det}}$ can be obtained from (7) or (10) whereas $I_{D,\mathrm{avg}}$ can be calculated from alpha power model [21], [22]. $VL$ is used instead of actual average supply voltage of the inverter for simplicity. Corresponding delay can therefore be acquired through (12). Many sets of delay-$VL - I_D$ relationship can be calculated through the same procedure.

Table I lists five selected sets of the relationship by choosing desired delay. Note that the delay is presented as being normalized to original delay without the power gating device. The $I_D$ ratio is the normalized drain current with respect to the maximum $I_D$. The increase of delay is a result of using smaller power gating device as mentioned. The last two columns in the table depict the configuration of the MPGN for a 32-bit multiplier. The actual $I_D$ values in Table I come from the $I_D$ ratio multiplied by the maximum $I_D$. Total PG width required for each mode is calculated using $I_D$ values and corresponding $VL$ parameters. Then the individual PG size is simply the difference of total width between each mode. In this demonstration the maximum $I_D$ is acquired from a simulation with an ideal power supply. However, the accurate maximum $I_D$ value should be obtained from extensive static/dynamic analysis on the target circuit.

The voltage sensor and the variable threshold comparator in Fig. 16 implement the switching state determination mechanism described in Sections III-B and III-C. The voltage sensor consists of two diode-connected transistors, MP1 and MP2, a current pulling leg formed by four nMOSs, and a precharge pMOS. The capacitor in the figure is not a physical one but

TABLE I
DELAY-$VL - I_D(\%)$-W RELATIONSHIPS

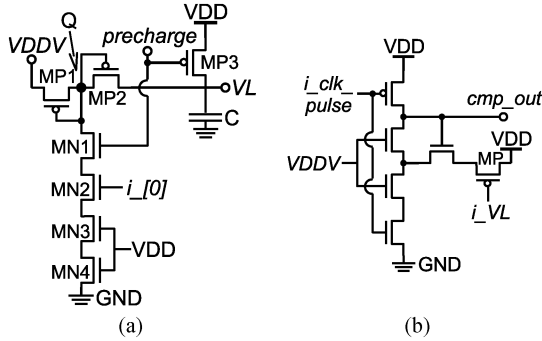| Delay | $VL$ | $V_{det}$ | $I_D$ ratio | $I_D(mA)$ | $\sum W(\mu m)$ |
|-------|------|-----------|-------------|-----------|-----------------|
| 1.1 | 0.95 | 0.986 | 0.9042 | 29.7467 | 1039 |
| 1.25 | 0.88 | 0.970 | 0.7775 | 25.5813 | 394 |
| 1.45 | 0.81 | 0.946 | 0.6597 | 21.7051 | 224 |
| 1.65 | 0.75 | 0.919 | 0.5657 | 18.6119 | 154 |
| 1.9 | 0.68 | 0.881 | 0.4641 | 15.2701 | 106 |



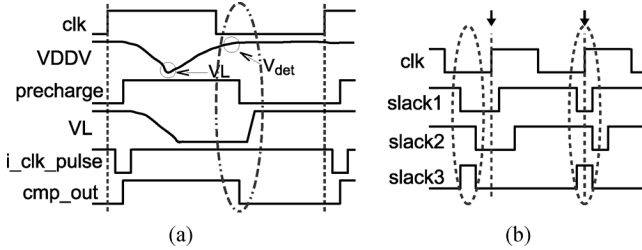Fig. 16. Schematics of (a) the voltage sensor and (b) the variable threshold comparator.



Fig. 17. Waveform illustrations of (a) the operation of the voltage sensor and the variable threshold comparator and (b) the principle of slack detection.

depicts the parasitic capacitance at $VL$ node. The variable threshold comparator is modified from the Schmitt Trigger. The operation of these two blocks to determine the completion of switching events can be illustrated by the example waveforms in Fig. 17(a).

The current pulling leg in the voltage sensor is used to remove the charges from internal node $Q$. Long length transistors are adopted to weaken the pull down strength in order not to burn unnecessary power. Node $Q$ will always be a threshold voltage drop lower than VDDV because of the diode-like behavior of MP1 whenever the voltage sensor is activated. $VL$ node is fully charged during precharging phase and will be discharged through MP2 and the current pulling leg in the evaluation phase. When VDDV and $Q$ are falling during circuit switching before reaching their lowest values, $VL$ will be a threshold voltage higher than $Q$ theoretically due to MP2's diode like behavior. Therefore, VDDV and $VL$ are equal in this period. Once VDDV and $Q$ start to charge up, the voltage drop across MP2 (seeing from $VL$ to $Q$-direction) will be smaller than a threshold voltage or even be negative. Therefore, MP2 is turned

off and $VL$ node holds its previous value which is the lowest value of VDDV as shown in Fig. 17(a).

The output of the variable threshold comparator ($cmp\_out$) is precharged by a pulsed clock at the beginning of every clock cycle. In the evaluation phase, $cmp\_out$ is determined by the contention between pull down network controlled by transient VDDV and the current pushing leg controlled inversely by $VL$. A pMOS, MP, is used in the pushing leg to have a consistent pushing strength when evaluating rising VDDV. The nMOS in the pushing leg has minor effect since it is fully turned on, unlike the partially on MP. As mentioned, the switching state determination mechanism suggests that lower $VL$ requires lower $V_{det}$. If MP is controlled directly by $VL$, lower $VL$ will make it stronger that results in higher required VDDV. It is contrary to the proposed mechanism. Therefore, an inverted $VL$ ($i\_VL$) by a skewed inverter is used to control MP. Once VDDV rises higher than $V_{det}$ as depicted in Fig. 17(a), the comparator output asserts by a falling transition. The assertion implies that the satisfaction of (7) or (10) has been verified. In other words, it claims the completion of a circuit switching.

Note that there will be a slight difference between actual lowest VDDV value and captured $VL$ due to circuit imperfection of the voltage sensor. Therefore, the skewed inverter used to produce $i\_VL$, the pull down network, and the current pushing leg in the comparator are sized to compensate for this mismatch. To reduce unnecessary power, precharge signal will be set to low right after the comparator assertion to disable the voltage sensor.

The example waveforms shown in Fig. 17(b) can be used to illustrate the concept of the slack detection. slack1 and slack2 are two delayed versions of $cmp\_out$. The same as $cmp\_out$, only falling transitions are meaningful. The margin between $cmp\_out$ and slack1 is designed to tolerate the APC circuit variations. The proposed criterion of slack depletion requires that the clock rising edge lies inside the timing window created by slack1 and slack2. It implies that the completion of the switching event, which is indicated by $cmp\_out$, is close to the next clock rising edge. In other words, the unused slack is depleted. In Fig. 17(b), two marked clock rising edges exhibit two examples. The first rising edge is later than both slack1 and slack2, i.e., unused slack exists. The second one locates between slack1 and slack2 that represents a slack depletion. The implementation of the slack detection block is through a combination of simple logic gates. The timing window is designed to be 180 ps and is translated into a pulse signal slack3. By capturing slack2 and slack3 values at the clock rising edge, the location of the clock rising edge can be determined.

Based on the slack detection result, the shift registers will modify the control signals of MPGN correspondingly. A "1" will be shifted in rightward to turn off one more power switch while unused slack exists. A "0" will be shifted in leftward to turn on one more switch if there is not enough slack left. A hold state can hold the control word without change. A reset state will set all control bits to "0" to turn on all the power switches. In power gating state, all the power switches are turned off by setting all control bits to "1". The inverse of the last bit of the control word ($i\_[0]$) is used to disable the voltage sensor in power gating state as shown in Fig. 16(a).
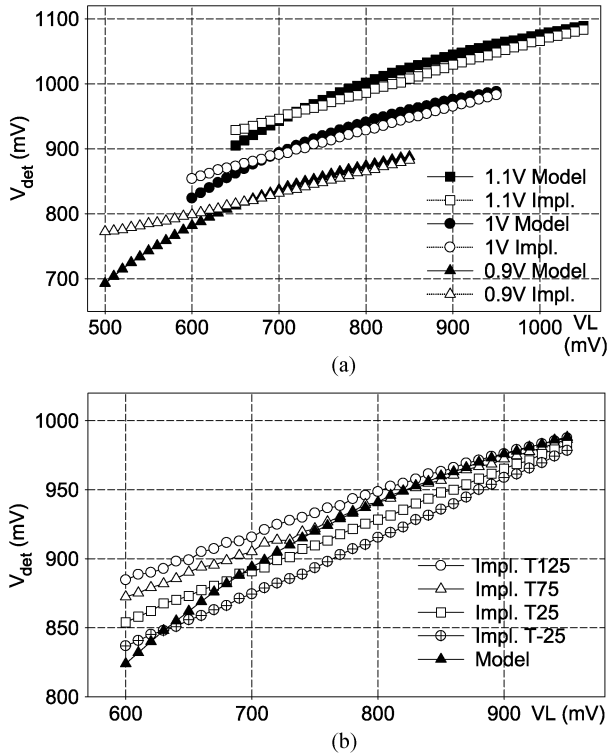
Fig. 18. Simulation results of circuit implementation of switching state determination mechanism (a) under voltage variations and (b) under temperature variations ($T$ unit: °C).
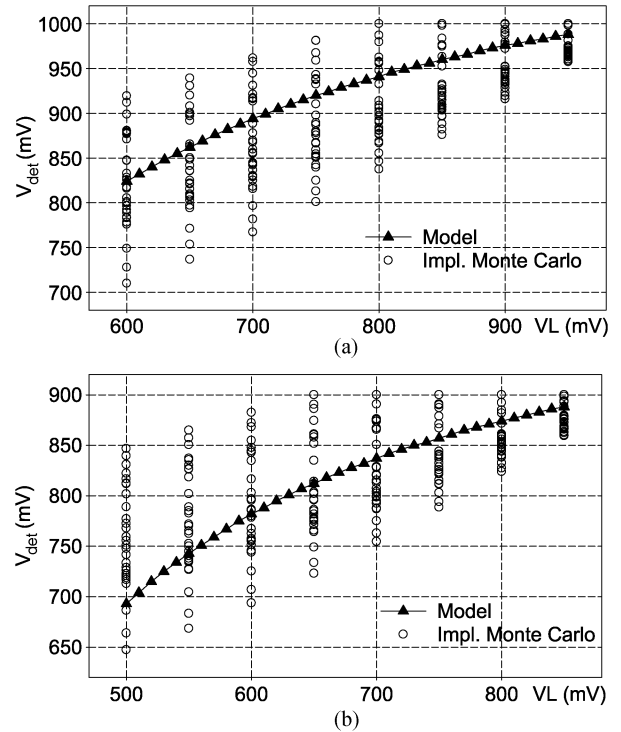


Fig. 19. Monte Carlo simulation results of circuit implementation of switching state determination mechanism (a) under WID and MISMATCH variations and (b) under worst case scenario.

### B. Circuit Performance Analysis

The performance of circuit implementation of the switching state determination mechanism is examined here. The implemented voltage sensor and the variable threshold comparator are added into the simulation environment in Fig. 2, receiving the VDDV as the input of the voltage sensor. The value of the VDDV when the comparator asserts is recorded as the $V_{\det}$ of the implemented circuit. The simulation results are shown in Fig. 18 including voltage and temperature variations.

The imperfection of implemented circuit can be observed from the figure. The implemented $V_{\det} - VL$ relationships are more linear than the model curve. They are higher than the model curves when $VL$ is low. The implemented relationship under voltage variations exhibits a shift whose value is approximately equal to the corresponding 10% supply voltage variations. Therefore, the nature of circuit response to voltage change adequately supports the compensation of voltage variations as mentioned in Section III-D. The implemented curves under temperature variations show almost equal shift every 50 °C. The circuits become slower at higher temperature. So the implemented $V_{\det} - VL$ relationship gets higher as expected. Overall, the implementation curves track the model curves well for most of the cases.

Monte Carlo simulation results of the implemented voltage sensor and the comparator with respect to WID and mismatch variations at 1 V supply and room temperature are shown in Fig. 19(a). As clearly shown in the figure, these process variations induce great variations on the $V_{\det} - VL$ relationships of implementation. A worst case scenario, at 0.9 V supply and

125 °C temperature, is also examined shown in Fig. 19(b). The maximum amount of the implemented curve that is lower than model curve is smaller in worst case because of the raise of implemented curve under 0.9 V supply voltage as in 18(a).

The variations of the implemented $V_{\det} - VL$ relationships are mainly due to the nMOS-pMOS combination in the variable threshold comparator. The opposite variations of nMOS and pMOS result in large variations of the $V_{\det}$ value for a certain $VL$. However, as shown in Section III-D, the model asserts completion always later than the actual occurrence of the definition of completion. So if the implemented curves can be always higher than the model curve under all variations, the correctness of the proposed state determination mechanism can be ensured. Based on the analysis in Fig. 19, the implemented curve is about 115 mV lower than the model curve at most. For inverters in Fig. 2 with smallest PG size, VDDV node takes about 50 ps to rise 115 mV at the end of switching. Therefore, 50 ps (or 100 ps to be safe) timing margin should be added into the slack detection block, which is the margin between $cmp\_out$ and slack1, to compensate for the imperfection of circuit implementation. This timing margin can also cover the imperfection shown in Fig. 18.

## V. SYSTEM EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, The simulation and chip measurement results of the proposed APC system are presented with test vehicles, including a 16-bit multiplier and a 32–64 multiply-accumulate (MAC) unit. The integration of the proposed technique into design flow is described. In addition, some more remarks on the proposed technique are made.
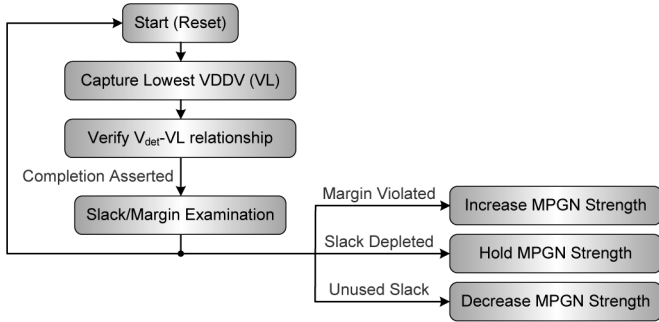
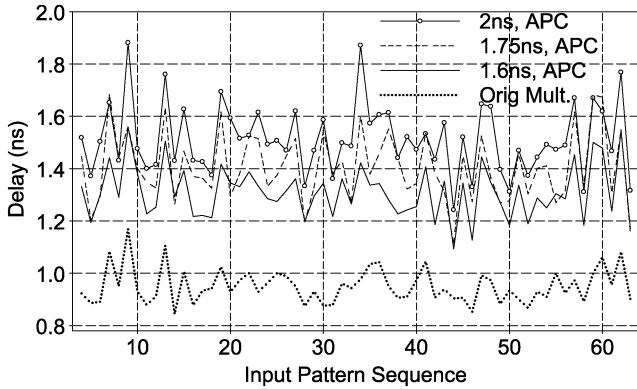Fig. 20. Flow chart of the proposed adaptive power control system.



Fig. 21. Delay patterns of a 16-bit multiplier.



Fig. 22. Test vehicle structure of the test chip.

### A. APC System Overview

Fig. 20 shows the functional flow chart of the proposed APC system. The system resets at the beginning of every clock cycle. The voltage sensor captures lowest VDDV during switching as $VL$. Different $VL$ imply different thresholds (i.e., different $V_{det}$) of the variable threshold comparator as a result of its structure. Careful sizing of the comparator is required to track the "equal" property of the $V_{det} - VL$ relationship as close as possible. The comparator asserts to indicate the completion of the switching event as soon as VDDV exceeds $V_{det}$. At the end of the clock cycle, the slack detection block examines the slack and margin information. The evaluation loop then restarts from the beginning. Meanwhile, the system will decrease, increase or hold the MPGN strength to alter the circuit speed according to three possible slack states. By repeating this procedure, the APC system can make the target circuit utilize the unused slack. Note that the APC system can respond to any environmental variations. Therefore, the minimum power consumption can be achieved under different operation conditions without harming the speed specification.

### B. Simulation Results

The proposed APC system is first analyzed with a 16-bit multiplier. Its "$m$" parameter is determined as described in Section III-C. Timing requirement of 1.87 ns is reported by the STA tool considering worst case variations. Different cycle time cases are simulated. One is 2 ns which is larger than reported 1.87 ns. The other two are 1.75 and 1.6 ns which are shorter than 1.87 ns. The simulated delays are shown in Fig. 21 for 60 consecutive input sequences. Typically using a shorter timing constraint
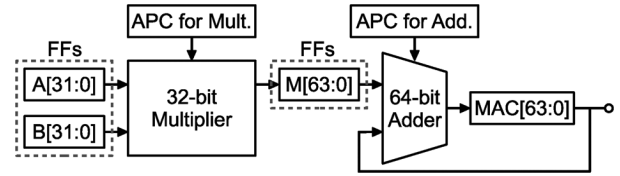
than that reported by STA tool is not desired. The circuit may have timing violations under worst case condition. However, the circuit here is analyzed under nominal condition which is 1 V supply, 25 °C, and TT corner. The original delay of the 16-bit multiplier is 0.957 ns on average as in the figure. A large amount of unused slack exists. So two shorter timing constraints are simulated to demonstrate the slack utilization performance. The average delays in 2, 1.75, and 1.6 ns cycle time cases with APC system activated are 1.524, 1.41, and 1.317 ns, respectively. To evaluate the performance of APC under different cycle time cases, the slack utilization is defined as

$$U_{slack} = \frac{slack\_without\_apc - slack\_with\_apc}{slack\_without\_apc}. \quad (13)$$

By this definition, the average slack utilization in 2, 1.75, and 1.6 ns cases are 54.8%, 57.7%, and 56.9%, respectively. The approximately 40% remaining slack results from the system circuits' delay and the added timing margins. These results show that the proposed APC system achieves a good slack utilization in different cycle time cases while not violating the timing constraints.

### C. Test Chip Implementation and Measurement Results

A test chip has been designed and implemented to demonstrate the proposed adaptive power control system. The test vehicle is a 32–64 MAC unit as shown in Fig. 22. The MAC unit is segmented into three sections, the multiplier, the adder, and the flipflop (FF) block. The FF block uses standard power supply. Both the multiplier and the adder are attached with their own APC systems. These two APC systems operate independently. Therefore, the adder and the multiplier are controlled separately according to their own operation conditions.

The layout view of the test chip is shown in Fig. 23(a). UMC 90-nm standard process CMOS technology including standard cell library and the custom design kit is used for the implementation. A linear feedback shift register bank is used as an internal pattern generator because of limited PAD numbers. An unmodified MAC unit is also fabricated for function comparison. There are four APC relevant blocks as depicted in the figure. These blocks are implemented by full-custom layout. Except APC relevant blocks, all the MAC units and the other test circuits are routed using standard cell library with *Cadence SoC Encounter* [27] for automatic place and route (P&R).

As can be observed from the figure, the area overhead of the APC system is quite small. The total area overhead including MPGN and the control part of the APC system is about 10% and 5% for the adder and the multiplier, respectively. The area overhead is mainly contributed by the MPGN and their control signal buffers. The area of the control part is fixed whereas
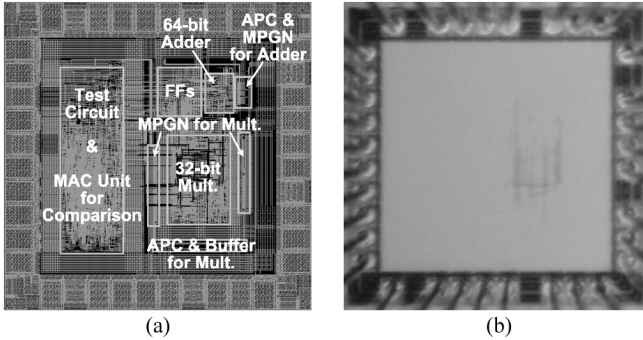
Fig. 23. Test chip implementation. (a) Layout view and (b) photo of fabricated chip with bonding wires.



Fig. 24. Measured power consumption comparison of test chips.

TABLE II
MEASURED POWER NUMBERS OF TEST CHIPS (mW)

| (mW) | | Mult. | APC of mult. | Adder | APC of adder | FFs |
|---|---|---|---|---|---|---|
| Chip 1 | Original | 9.38 | - | 0.4016 | - | 2.06 |
| | Proposed | 7.69 | 0.1046 | 0.3599 | 0.1489 | 2.10 |
| Chip 2 | Original | 8.83 | - | 0.3915 | - | 2.00 |
| | Propoed | 7.46 | 0.0914 | 0.3645 | 0.0897 | 2.00 |
| Chip 3 | Original | 9.23 | - | 0.3939 | - | 1.91 |
| | Proposed | 7.89 | 0.0986 | 0.3721 | 0.0956 | 1.93 |
| Chip 4 | Original | 9.32 | - | 0.3892 | - | 1.96 |
| | Proposed | 7.63 | 0.1048 | 0.3444 | 0.0969 | 1.99 |
| Chip 5 | Original | 10.14 | - | 0.4116 | - | 1.96 |
| | Proposed | 7.99 | 0.1055 | 0.3681 | 0.1126 | 1.97 |

the size of MPGN varies for different loading and speed requirement. So the area overhead ratio is larger for a smaller block such as an adder. Note that for most of advanced digital circuits, power gating devices including their control signal buffers are already implemented in the circuit to support the sleep/standby low power mode. What the proposed technique does is to extend the usage of the power gating devices. Therefore, the actual extra area overhead, which comes from the control part of the APC system, is 7% for the adder and 1% for the multiplier.

The fabricated chip is photographed with bonding wires as shown in Fig. 23(b). 4 ns cycle time is used for measurement since the fastest period that the chip still functions correctly is 3.6 ns. Five test chips are measured. The results are listed in Table II and reported as power numbers per block of the MAC. The power numbers are also formatted as bar charts shown in Fig. 24. The average net power reduction ratio is 12.39% as a result of slack utilization. The active power numbers of the APC system are around 100 $\mu$W no matter what the attached target circuit is.

If considering the 32-bit multiplier only, the average power reduction is 16.5% including only 1.08% power overhead. Note that in Table II the power numbers of the FFs are slightly larger when the APC systems are applied. The reason is that no level converter is implemented in the test chip. All the signals are bounded by FFs. The FFs will generate full-swing signals for the next stage because they use the standard supply. Although the
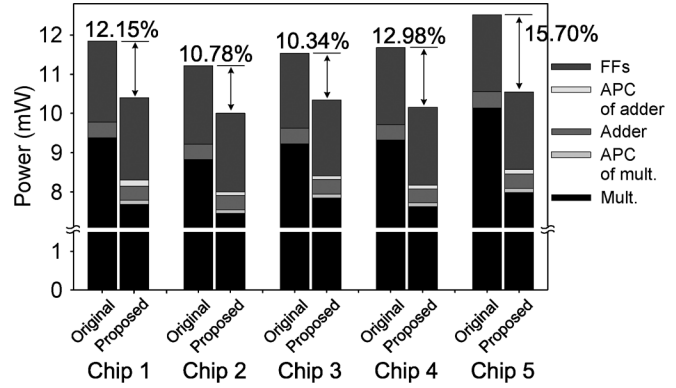
FFs can still capture the signals correctly, these signals which are not full swing ones will induce short circuit currents at the input stages of the FFs. The original power consumption of the adder is comparable to the active power of the APC system. The power reduction number of the adder cannot even compensate for APC system's power overhead. Therefore, the proposed APC system should be applied on the circuit whose power consumption is at least an order larger. The leakage current in the sleep state is also measured. The APC system cuts off the power and has an average of 7.96$\times$ leakage reduction (from 156.2 to 19.5 $\mu$A on the 32-bit multiplier on average).

### D. Design Flow Integration

The proposed adaptive power control system can be integrated into standard cell based design flow easily. There is no modification required on the main architecture. The designer only needs to identify the target functional block for power reduction. The state-of-the-art P&R tools are able to attach power gating devices on the target block. The only modification to form a MPGN as in the proposed technique is to group the power gating devices with different control signals. Sizing of each group can be integrated into current EDA tools by adopting the sizing strategy presented in Section IV-A or any other works describing delay penalties. The control part of the APC system can be constructed as a parameterized IP block. The "$m$" parameter mentioned in Section III-C and the number of MPGN modes are two main parameters. The control signals of the MPGN can then be connected from the control part of the APC system. Note that the proposed adaptive power control system operates automatically and independently. Therefore, in a large chip with multiple function blocks, the APC system can be applied repeatedly. These procedures are much easier today as a result of the highly automated digital design flow.

### E. Remarks on Proposed Technique

In this work, all the derivation and the implementation are based on $p$-type power gating devices. However, it should be noted that all the procedures can be easily migrated to using $n$-type power gating devices. Point A is now defined as the maximum voltage rise of the virtual ground node, whereas point B is defined that $V_{DS}$ of the nMOS equals 0.1 V. Then by using $n$-type drain current descriptions throughout the derivation of the switching state determination mechanism as

in Section III-B, an $n$-type counterpart of the determination model can be obtained. All the concepts described in this work can be migrated in a similar way.

One of the concerns when using power gating device is the power/ground noise induced by rush current at the wakeup. Techniques to reduce the power noise such as skewing the wakeup times of power gating devices [12] or adding an extra bypass power line for wakeup [28] had been presented. As for this work, the mitigation of power noise can also be supported though not expressed in the paper. When waking up the circuit from power gating state, the power gating devices in the MPGN can be designed to turn on sequentially instead of turning on all at the same time. Therefore, similar power noise reduction effect can be achieved as the method of wakeup time skewing [12].

For complex circuit blocks, there are still gates switching after the occurrence of $VL$. But it can be observed from Fig. 8(b) that the weighted count, which is the equivalent "$m$", drops rapidly after reaching the maximum. For those switching later than the occurrence of $VL$, lower $V_{det}$ are required. These $V_{det}$ are even lower than the required $V_{det}$ for $VL$, which means that their completion assertions are earlier than that of $VL$ point. Therefore, the APC system can focus only on the determination of $VL$ point.

The determination of the "$m$" parameter described in Section III-C is based on the switching probability of the target circuit. If the estimation of the switching probability diverges from the actual behavior of the target circuit, the quality and correctness of power control will be affected. Simulations on the 32-bit multiplier of the MAC with intentionally different "$m$" are conducted to evaluate the effect of inaccurate "$m$". The results show that if "$m$" is overestimated, i.e., a larger "$m$", the power reduction ratio is reduced. In the worst case, the reduced power cannot even compensate for the power overhead of the APC system. On the other hand, if "$m$" is underestimated, the power reduction number may looks better. However, the multiplier will violate the timing constraints in the worst case because of too early assertion of the complete switching by the APC system. In the development of "$m$" in Section III-C, the switching probability is computed from the accumulation of switching windows of all the gates. So the proposed determination of "$m$" in this work is a little overestimated since not every gate switches under every input pattern.

As mentioned in Section V-C, no level shifter is implemented in the test chip since the cell library doesn't contain such cells. If level shifters are adopted in front of the FFs, the short circuit current of the FFs can be eliminated. Prior simulations show that the power increase because of level shifters is approximately equal to the reduction of FFs from eliminating short circuit currents. On the other hand, the delay overhead of level shifters is approximately equal to the propagation delay increase of FFs because of non full-swing signals. There is no substantial benefit from the point of view of power or delay. If the level shifter is desired, it can still be applied and powered by VDDV and standard $V_{DD}$ as its low and high supply, respectively.

In the proposed APC system, the MPGN is configured based on slack information of previous cycle. In other words, the slack information of pattern $N$ is used as an expectation for pattern $N + 1$. The slack variations are expected to be accommodated by the 40% remaining slack as Fig. 21 suggests. However, in some extreme cases, the increased delay that results from a short delay pattern followed by a very long delay pattern will exceed the remaining slack and cause error calculation. Error prevention or correction techniques such as those in [29]–[31] can be used to maintain the correctness. The extra power overhead is 1.2% as reported in [31]. The concept in [30] showed that with a little increased error rate, the power consumption can be further reduced. Therefore, the proposed APC technique can be configured a little more aggressive to further reduce the power consumption, e.g., lower the $V_{det}$ requirement than (8). The increased error rate is handled by the error correction techniques. Then the extra power overhead of error correction techniques can be compensated.

## VI. CONCLUSION

An adaptive power control system on power-gated circuitries is proposed in this work. The core concept is the switching state determination mechanism which serves as an alternative of critical path replica scheme. It is intrinsically tolerant of PVT variations due to its nature of directly monitoring circuit behavior on VDDV node. The APC system consists of a multi-mode power gating network, a voltage sensor, a variable threshold comparator, a slack detection block, and a bank of bi-directional shift registers. By dynamically configuring the size of the power gating devices, the circuit speed can be altered to utilize unused slack. Simulations show that the proposed APC system achieves an average of 56.5% slack utilization rate under different cycle time cases.

A 32–64 bit MAC unit is implemented as a test vehicle, using UMC 90-nm standard process CMOS technology. The measurement results of test chips show an average of 12.39% net power reduction. The area overhead of proposed APC system is 5% for the 32-bit multiplier of MAC. The control part of the system only contributes 1%, whereas the power gating devices and their control signal buffers contribute rest of the area overhead. The power overhead is only 1.08% for the 32-bit multiplier of the MAC unit. And a 7.96× leakage reduction is also reported by power gating the MAC unit.

## REFERENCES

[1] M. Horowitz, D. Stark, and E. Alon, "Digital circuit design trends," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 757–761, Apr. 2008.

[2] B. H. Calhoun and A. P. Chandrakasan, "Ultra-dynamic voltage scaling (UDVS) using sub-threshold operation and local voltage dithering," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 238–245, Jan. 2006.

[3] Y. K. Ramadass and A. P. Chandrakasan, "Minimum energy tracking loop with embedded DC-DC converter enabling ultra-low-voltage operation down to 250 mV in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 256–265, Jan. 2008.

[4] Y. Ikenaga, M. Nomura, Y. Nakazawa, and Y. Hagihara, "A circuit for determining the optimal supply voltage to minimize energy consumption in LSI circuit operations," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 911–918, Apr. 2008.

[5] T. D. Burd, T. A. Pering, A. J. Stratakos, and R. W. Brodersen, "A dynamic voltage scaled microprocessor system," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1571–1580, Nov. 2000.

[6] M. Nakai, S. Akui, K. Seno, T. Meguro, T. Seki, T. Kondo, A. Hashiguchi, H. Kawahara, K. Kumano, and M. Shimura, "Dynamic voltage and frequency management for a low-power embedded microprocessor," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 28–35, Jan. 2005.

[7] K. J. Nowka, G. D. Carpenter, E. W. MacDonald, H. C. Ngo, B. C. Brock, K. I. Ishii, T. Y. Nguyen, and J. L. Burns, "A 32-bit PowerPC system-on-a-chip with support for dynamic voltage scaling and dynamic frequency scaling," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1441–1447, Nov. 2002.

[8] M. Eireiner, S. Henzler, G. Georgakos, J. Berthold, and D. Schmitt-Landsiedel, "In-situ delay characterization and local supply voltage adjustment for compensation of local parametric variations," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1583–1592, Jul. 2007.

[9] M. Elgebaly and M. Sachdev, "Variation-aware adaptive voltage scaling system," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 5, pp. 560–571, May 2007.

[10] J. Tschanz, N. S. Kim, S. Dighe, J. Howard, G. Ruhl, S. Vangal, S. Narendra, Y. Hoskote, H. Wilson, C. Lam, M. Shuman, C. Tokunaga, D. Somasekhar, S. Tang, D. Finan, T. Karnik, N. Borkar, N. Kurd, and V. De, "Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2007, pp. 292–293.

[11] S. Henzler, T. Nirschl, S. Skiathitis, J. Berthold, J. Fischer, P. Teichmann, F. Bauer, G. Georgakos, and D. Schmitt-Landsiedel, "Sleep transistor circuits for fine-grained power switch-off with short power-down times," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2005, pp. 302–303.

[12] K. Usami, T. Shirai, T. Hashida, H. Masuda, S. Takeda, M. Nakata, N. Seki, H. Amano, M. Namiki, M. Imai, M. Kondo, and H. Nakamura, "Design and implementation of fine-grain power gating with ground bounce suppression," in *Proc. Int. Conf. VLSI Des.*, Jan. 2009, pp. 381–386.

[13] S. Kim, S. V. Kosonocky, D. R. Knebel, K. Stawiasz, and M. C. Papaefthymiou, "A multi-mode power gating structure for low-voltage deep-submicron CMOS ICs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 7, pp. 586–590, Jul. 2007.

[14] A. Ramalingam, B. Zhang, A. Devgani, and D. Z. Pan, "Sleep transistor sizing using timing criticality and temporal currents," in *Proc. Asia South Pacific Des. Autom. Conf. (ASPDAC)*, Jan. 2005, pp. 1094–1097.

[15] A. Valentian and E. Beigne, "Automatic gate biasing of an SCCMOS power switch achieving maximum leakage reduction and lowering leakage current variability," *IEEE J. Solid-State Circuits*, vol. 43, no. 7, pp. 1688–1698, Jul. 2008.

[16] W.-C. Hsieh and W. Hwang, "In-situ self-aware adaptive power control system with multi-mode power gating network," in *Proc. IEEE Int. Syst.-on-Chip Conf. (SOCC)*, Sep. 2008, pp. 215–218.

[17] M. Meijer, J. P. de Gyvez, and R. Otten, "On-chip digital power supply control for system-on-chip applications," in *Proc. Int. Symp. Low Power Electron. Des. (ISLPED)*, Aug. 2005, pp. 311–314.

[18] M. Saint-Laurent and M. Swaminathan, "Impact of power-supply noise on timing in high-frequency microprocessors," *IEEE Trans. Adv. Packag.*, vol. 27, no. 1, pp. 135–144, Feb. 2004.

[19] Y. Ogasahara, T. Enami, M. Hashimoto, T. Sato, and T. Onoye, "Validation of a full-chip simulation model for supply noise and delay dependence on average voltage drop with on-chip delay measurement," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 10, pp. 868–872, Oct. 2007.

[20] M. Alioto and G. Palumbo, "Impact of supply voltage variations on full adder delay: Analysis and comparison," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 12, pp. 1322–1335, Dec. 2006.

[21] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[22] K. A. Bowman, B. L. Austin, J. C. Eble, X. Tang, and J. D. Meindl, "A physical alpha-power law MOSFET model," *IEEE J. Solid-State Circuits*, vol. 34, no. 10, pp. 1410–1414, Oct. 1999.

[23] Synopsys, Mountain View, CA, "PrimeTime User Guide: Fundamentals," B-2008.12, 2008.

[24] D. Lee, D. Blaauw, and D. Sylvester, "Runtime leakage minimization through probability-aware optimization," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 10, pp. 1075–1088, Oct. 2006.

[25] A. Drake, R. Senger, H. Deogun, G. Carpenter, S. Ghiasi, T. Nguyen, N. James, M. Floyd, and V. Pokala, "A distributed critical-path timing monitor for a 65 nm high-performance microprocessor," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2007, pp. 398–399.

[26] M. H. Abu-Rahma and M. Anis, "A statistical design-oriented delay variation model accounting for within-die variations," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 27, no. 11, pp. 1983–1995, Nov. 2008.

[27] Cadence, San Jose, CA, "SoC encounter," 06.20-p006_1, 2006.

[28] K. Kawasaki, T. Shiota, K. Nakayama, and A. Inoue, "A sub-$\mu$s wake-up time power gating technique with bypass power line for rush current support," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1178–1183, Apr. 2009.

[29] M. Eireiner, S. Henzler, G. Georgakos, J. Berthold, and D. Schmitt-Landsiedel, "In-situ delay characterization and local supply voltage adjustment for compensation of local parametric variations," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1583–1592, Jul. 2007.

[30] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "A self-tuning DVS processor using delay-error detection and correction," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 792–804, Apr. 2006.

[31] S. Das, C. Tokunaga, S. Pant, W.-H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull, and D. T. Blaauw, "RazorII: In situ error detection and correction for PVT and SER tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.

**Wei-Chih Hsieh** (S'10) was born in TaoYuan, Taiwan in 1981. He received the B.S. degree from the Department of Electronics Engineering, National Chiao Tung University (NCTU), HsinChu, Taiwan, in 2004. He is currently pursuing the Ph.D. degree in electronics engineering from the Institute of Electronics, National Chiao Tung University.

His research interests include power management techniques and digital-assisted mixed-signal circuit design.

**Wei Hwang** (F'01) received the B.Sc. degree from National Cheng Kung University, Taiwan, the M.Sc. degree from National Chiao Tung University, Taiwan, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1970 and 1974, respectively.

From 1975 to 1978, he was an Assistant Professor with the Department of Electrical Engineering, Concordia University, Montreal, QC, Canada. From 1979 to 1984, he was an Associate Professor with the Department of Electrical Engineering, Columbia University, New York, NY. From 1984 to 2002, he was a Research Staff Member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. In 2002, he joined National Chiao Tung University (NCTU), Hsinchu, Taiwan, as the Director of Microelectronics and Information Systems Research Center until 2008, where he currently holds a Chair Professor with the Department of Electronics Engineering. During 2003 to 2007, he served as Coprincipal Investigator of National System-on-Chip (NSoC) Program, Taiwan. From 2005 to 2007, he also served as a Senior Vice President and Acting President of NCTU, respectively. He is the coauthor of the book *Electrical Transports in Solids-With Particular Reference to Organic Semiconductors* (Pergamon Press, 1981), which has been translated into Russian and Chinese. He has authored or coauthored over 200 technical papers in renowned international journals and conferences, and holds over 150 international patents (including 65 U.S. patents).

Prof. Hwang was a recipient of several IBM Awards, including 16 IBM Invention Plateau Invention Achievement Awards, 4 IBM Research Division Technical Awards, and was named an IBM Master Inventor. He was also a recipient of the CIEE Outstanding Electrical Engineering Professor Award in 2004 and Outstanding Scholar Award from the Foundation for the advancement of Outstanding Scholarship from 2005 to 2010. He has served as the General Chair of 2007 IEEE SoC Conference (SOCC 2007) and 2007 IEEE International Workshop on Memory Technology, Design and Testing (MTDT 2007). Currently, he is serving as a Supervisor of IEEE Taipei Section for 2007 to 2010. He has served several times in the Technical Program Committee of the ISLPED, SOCC, and A-SSCC. He has also served as the General Chair of 2007 IEEE SoC Conference (SOCC 2007) and 2007 IEEE International Workshop on Memory Technology, Design and Testing (MTDT 2007). Currently, he is serving as Founding Director of Center of Advanced Information Systems and Electronics Research (CAISER) of University System of Taiwan, the Director of ITRI and NCTU Joint Research Center, and a Supervisor of IEEE Taipei Section.