

Adaptive Performance Compensation With In-Situ Timing Error Predictive Sensors for Subthreshold Circuits

Hiroshi Fuketa, *Member, IEEE*, Masanori Hashimoto, *Member, IEEE*, Yukio Mitsuyama, *Member, IEEE*, and Takao Onoye, *Senior Member, IEEE*

Abstract—We present an adaptive technique for compensating manufacturing and environmental variability in subthreshold circuits using “canary flip-flop (FF),” which can predict timing errors. A 32-bit Kogge–Stone adder whose performance was controlled by body-biasing was fabricated in a 65-nm CMOS process. Measurement results show that the adaptive control can compensate process, supply voltage, and temperature variations and improve the energy efficiency of subthreshold circuits by up to 46% compared to worst-case design and operation with guardbanding. We also discuss how to determine design parameters, such as the inserted location and the buffer delay of the canary FF, supposing two approaches: configuration in the design phase and post-silicon tuning.

Index Terms—Adaptive speed control, body biasing, manufacturing variability, subthreshold circuit, timing error prediction.

I. INTRODUCTION

SUBTHRESHOLD circuits are drawing the attention of designers implementing ultra-low power applications, such as sensor-node processors [1]–[3]. However, the performances of subthreshold circuits are extremely sensitive to manufacturing and environmental variability due to their exponential dependence on threshold voltage (V_{th}) and supply voltage, preventing them from being widely used. Therefore, a traditional “worst-case” design with guardbanding is inefficient and an adaptive performance control is indispensable for subthreshold circuits.

Traditionally, replica circuits have been used for performance monitoring. Adaptive control techniques with a critical path replica for nominal supply voltage circuits have been presented in [4]–[6]. However, the critical path replica is inadequate for subthreshold circuits, since the delay mismatch between the replica and the actual critical path is remarkably large due to within-die random V_{th} variation. Although Chang *et al.* proposed the critical path replica technique for subthreshold circuits using a number of parallel delay lines to mitigate the

influence of within-die V_{th} variation [7], this technique does not eliminate the within-die variation completely.

In order to overcome this mismatch, *in-situ* techniques have been proposed [8]–[12]. “Razor I” [8] and “Razor II” [9] detect timing errors in actual paths and correct the errors. In contrast, [10]–[12] presented an error predictive sensor embedded into actual paths. This sensor cannot detect timing errors but predict them. [10] and [11] used the sensor for detecting the degradation in circuit delay due to aging induced by electromigration and negative bias temperature instability (NBTI). Sato *et al.* utilized the sensor for an adaptive control under dynamic voltage and frequency scaling (DVFS) systems [12]. They called the error predictive sensor “canary flip-flop (FF)” and we use the term throughout this paper.

“Razor” and “canary FF” techniques assume that they are applied to the nominal supply voltage. In terms of employing them for subthreshold circuits, we can see the following advantages in canary FF technique compared to Razor.

- Razor technique requires a re-execution mechanism to correct timing errors. The re-execution is performed through architectural replay, which is often integrated in high-performance processors to support branch prediction [9]. However, it is impracticable for general sequential circuits and simple processors on which subthreshold circuits typically focus [13]–[15]. In contrast, canary FF predicts the occurrence of timing errors. This means that any error recovery mechanisms are not needed as long as the prediction is appropriate. Therefore, it is suitable to apply canary FF to subthreshold circuits.
- Razor FF requires the timing window of error detection just after the clock edge in order to detect a late-arriving signal as a timing error. Thus signals arriving during the timing window are considered as timing errors. This means that the timing window is equivalent to a hold time of Razor FF. For subthreshold circuits, however, path delays significantly fluctuate due to manufacturing variability. Consequently, the timing window, which is set to be large for capturing large setup-time violations, is much larger than a hold time of a normal FF and canary FF and hence Razor FF inherently suffers from severer minimum path delay constraints compared to a normal FF and canary FF. This makes design of the buffer-insertion more complicated.

While canary FF technique has the above mentioned advantages, the occurrence of timing errors cannot be completely eliminated because canary FF can only “predict” the occurrence

Manuscript received June 24, 2010; revised October 09, 2010; accepted November 30, 2010. Date of publication January 28, 2011; date of current version January 18, 2012. This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

H. Fuketa is with the Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan (e-mail: fuketa@iis.u-tokyo.ac.jp).

M. Hashimoto, Y. Mitsuyama, and T. Onoye are with the Department of Information Systems Engineering, Osaka University, Osaka 565-0871, Japan and also with the JST, CREST, Tokyo 102-0075, Japan (e-mail: hashimoto@ist.osaka-u.ac.jp; mituyama@ist.osaka-u.ac.jp; onoye@ist.osaka-u.ac.jp).

Digital Object Identifier 10.1109/TVLSI.2010.2101089

of timing errors and the prediction is not always guaranteed. Therefore, to apply canary FF technique to practical applications, the occurrence rate of timing errors must be quantitatively assured. Some systems could accept the occurrence of timing errors, when the occurrence rate is extremely low. For example, video decoding for TV and video recording for security monitoring can accept an error per day, since a small piece of image degradation in a short time is not a problem.

This paper presents the first work to apply the adaptive speed control with canary FF for subthreshold circuits and measure it on silicon. The preliminary work of this paper is presented in [16]. We demonstrate that fabricated chips operate in a subthreshold region at the adaptive speed and compensate manufacturing and environmental variability. We also reveal that the adaptive control provides much more energy-efficient operation in comparison with the worst-case operation with guardbanding. In addition, we examine the occurrence rate of timing errors, because the error rate is a key metric for the adaptive speed control with canary FF as previously mentioned. References [17] and [18] pointed out that the timing error rate and the power dissipation of the adaptive speed control with canary FF depend on the design parameters such as the inserted location and the buffer delay of canary FF. We also examine the design parameter configuration supposing the following two cases: tuning the design parameters after fabrication and determining them in the design phase and we find the most appropriate approach to determine the parameters.

In the adaptive speed control with canary FF presented in this paper, we focus on compensating delay variation due to manufacturing variability and slow delay fluctuation whose time constant is seconds to years due to such as temperature shift, aging and buttery degradation. Here, fast delay fluctuation caused by dynamic power supply noise is out of scope. Since the current is reduced exponentially in subthreshold region as the supply voltage is lowered linearly, dynamic power supply noise decreases exponentially. The sensitivity of circuit delay to the power supply noise in subthreshold region is exponential to the voltage fluctuation. Therefore, the delay fluctuation due to IR drop becomes small. On the other hand, subthreshold applications are typically powered by a battery or energy scavenging [3]. Hence, we have to pay attention to the degradation in the supply voltage level due to low battery charge or bad environmental condition.

The remainder of this paper is organized as follows. Section II describes the structure of the test chip for examining the adaptive speed control circuit with canary FF. In Section III, the measurement results of the adaptive speed control circuit are shown. Section IV discusses the configuration of the design parameters such as the inserted location and the buffer delay of canary FF. Finally, Section V concludes this paper.

II. SELF-ADAPTIVE SPEED CONTROL WITH CANARY FF

A. Overview

Fig. 1 shows an overview of self-adaptive speed control with canary FF. Canary FF, which consists of a normal FF (we call it “shadow FF”), a delay buffer and a comparator, generates a warning signal to predict the occurrence of timing errors. The adaptive speed control with canary FF works as follows. The

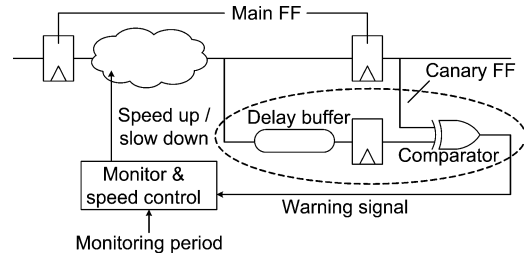


Fig. 1. Self-adaptive speed control with canary FF.

warning signal is monitored during a specified period. Once the warning signal is detected, the circuit is sped up. If no warning signals are generated during the monitoring period, the circuit is slowed down to reduce power dissipation. Consequently, the circuit operates at appropriate speed according to process, supply voltage, and temperature (PVT) variations, which enables much more energy-efficient operation than the worst-case operation with guardbanding.

There are several implementations to control the circuit speed, such as supply voltage scaling and body-biasing. Since several works have pointed out that the adaptive body-biasing technique is more efficient for subthreshold circuits [14], [19], we use body-biasing for the speed control. For the adaptive speed control, multi-level body-bias voltages are required. Although their generation is not discussed in this paper, they can be generated by, for example, digital-to-analog (D/A) converter [6].

For the purpose of the aging detection [10], [11], slowing down the circuit is not required. On the other hand, the adaptive speed control with canary FF presented in this paper makes the circuit speed slower to reduce the power dissipation. In this case, timing errors cannot be completely eliminated when the adaptive speed control is applied to normal (non-test) operations. This is because the circuit might be slowed down excessively when the paths, where canary FFs are inserted, have not been activated for a long time [17], [18]. We thus evaluate the mean time between failure (MTBF) in addition to performance.

B. Circuit Structure of Test Chip

A test circuit was designed and fabricated to demonstrate the adaptive speed control with canary FF in a 65-nm CMOS process. The structure of the test circuit is depicted in Fig. 2 and the micrograph is shown in Fig. 3. A 32-bit Kogge–Stone adder (KSA) was adopted as a circuit whose performance was controlled adaptively with body-biasing. The circuit speed is controlled digitally and we use the term “speed level” to describe how fast or slow the circuit is controlled. A higher speed level means the circuit is controlled for faster operation. $S[32]$ - $S[0]$ denote the outputs of the KSA and $S[32]$ is the most significant bit. It should be noted that the inserted “location” of canary FF indicates the output bit to which a canary FF is inserted and does not mean the physical location of canary FF in the layout.

Input patterns are generated by a linear feedback shift register (LFSR). The KSA outputs are compared to the answer to check if a timing error occurs. The answer is generated by an “always correct” adder operating at higher supply voltage.

A timer signal is asserted when the monitoring period of the warning signal elapses. In this test chip, the monitoring period is

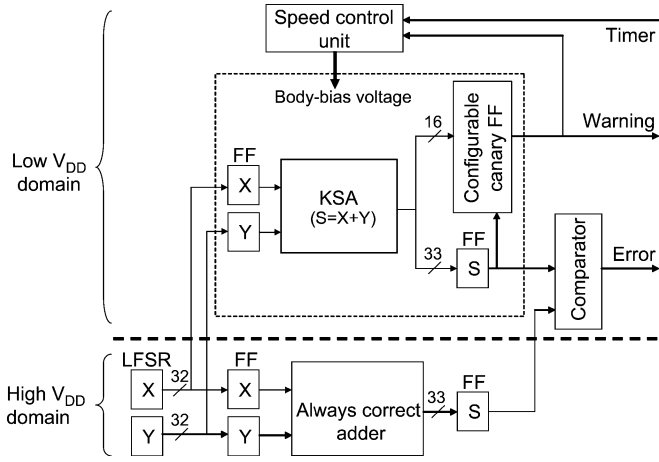


Fig. 2. Block diagram of test circuit. 32-bit KSA is controlled adaptively with configurable canary FF.

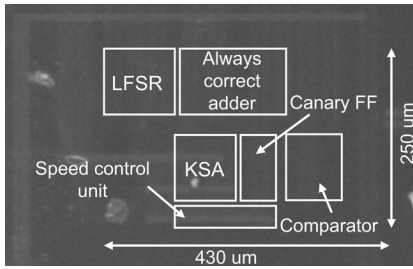


Fig. 3. Micrograph of test chip.

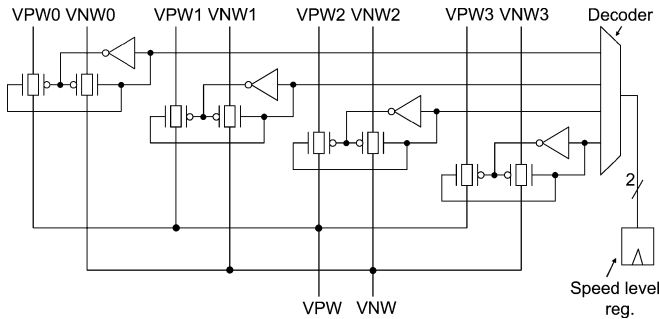


Fig. 4. Schematic of speed control unit. VPW and VNW denote body-bias voltages of KSA, main FFs, and canary FFs.

counted off-chip. When implementing the timer circuit on-chip, ultra-low power timers proposed in [20] and [21] are desirable.

The speed control unit alters by body-biasing the speed of the KSA, main FFs and canary FFs at inputs and outputs of the KSA. Fig. 4 shows a schematic of the speed control unit. VPW/VNW denotes the p-/n-well body-bias voltage of the KSA, main FFs and canary FFs. Four speed levels can be provided by applying four pairs of body-bias voltages (VPW0-3 and VNW0-3) and each body-bias voltage is supplied by external dc voltage sources.

VPW and VNW are selected from VPW0-3 and VNW0-3 according to the speed level stored in a two-bit register, that is, when the stored value in the speed level register is three, for instance, VPW3 and VNW3 are selected for the body-bias voltages. Circuit operation starts at the maximum speed level. When the timer signal is asserted, the speed control unit decrements

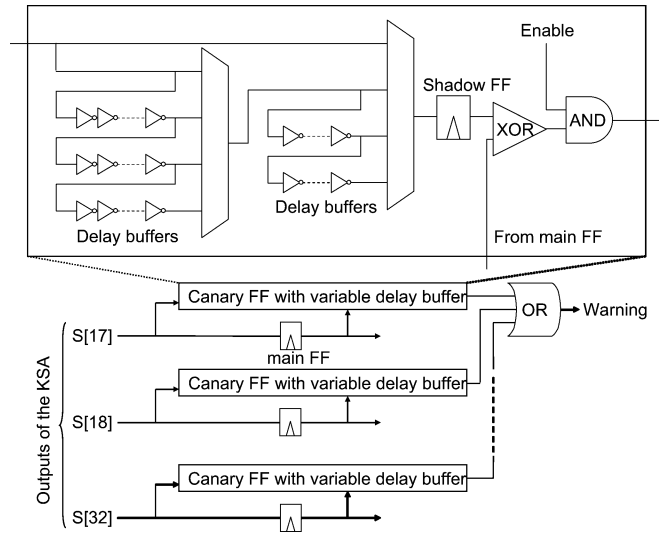


Fig. 5. Schematic of the configurable canary FF.

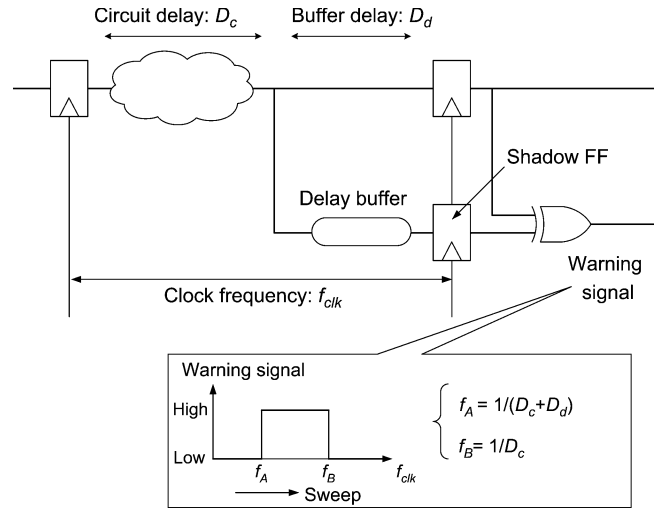


Fig. 6. Buffer delay measurement.

the speed level by one and the circuit is slowed down. In contrast, when the warning signal is asserted, the speed control unit immediately increments the speed level by one.

A “configurable” canary FF is implemented such that the inserted location and the buffer delay can be configured. Fig. 5 illustrates the configurable canary FF, which is composed of 16 canary FFs with variable delay buffers. Each canary FF inserted at S[17]–S[32] can be enabled or disabled individually.

The configured buffer delay can be measured according to the following procedure and as shown in Fig. 6.

- i) An input vector, which activates the output bit where a canary FF is inserted, is determined. D_c in Fig. 6 denotes the circuit delay when the vector is given.
- ii) The clock frequency f_{clk} is swept from a frequency which is slow enough to cause no warning signals. Clock frequencies at which warning signals occur are searched when the input vector specified in step i) is injected. f_A in Fig. 6 represents the minimum frequency at which a timing error occurs only in a shadow FF. f_B indicates the minimum frequency at which timing errors occur in

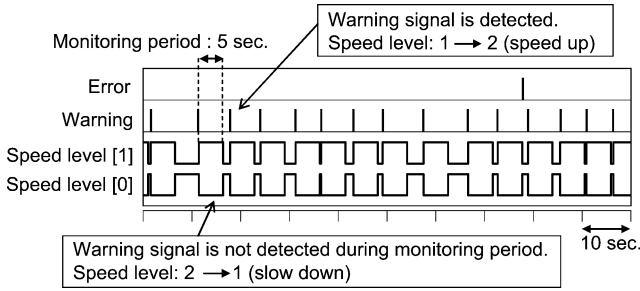


Fig. 7. Measured timing error, warning signals and transitions of speed level (2 MHz @ $V_{DD} = 0.35$ V).

both a shadow FF and a main FF. The difference between $f_A (= 1/(D_c + D_d))$ and $f_B (= 1/D_c)$ give an estimation of the buffer delay D_d .

In this adaptive speed control, canary FF might become metastable. Even in this case, main FF captures a correct value and the circuit continues to operate correctly, because the timing constraint of canary FF is always severer than that of main FF. Meanwhile, the occurrence of metastability at canary FF may cause the failure of timing error prediction, which detrimentally affects MTBF and the power dissipation.

III. MEASUREMENT RESULTS

A. Operation Example

Fig. 7 shows an operation example with a measured timing error, warning signals and speed level transitions when the circuit was controlled adaptively with a canary FF. The operation frequency and V_{DD} were 2 MHz and 350 mV. The step of body-biasing levels was set to 30 mV, which means speed level 1 corresponds to a 30-mV forward body bias (FBB) when speed level 0 is zero body bias (ZBB). The speed level was altered according to the warning signal. A timing error occurred in this example.

The scope of the adaptive speed control with canary FF is to compensate slower delay fluctuation as described in Section I. According to [13], the ambient temperature differences between a ten minute sampling interval range from -1 °C to 1 °C in almost all cases. Thus it is acceptable to set the monitoring period to the order of seconds for practical use. In the experiments discussed in this section, we set the monitoring period to 10^7 cycles, which is equivalent to five seconds at 2 MHz operation and 3.3 s at 3 MHz operation.

According to [17] and [18], MTBF increases by lengthening the monitoring period. Therefore, MTBF would become larger in the case when longer monitoring period than 10^7 cycles is set. However, too long monitoring period could deteriorate the adjustment response of the adaptive control.

B. Adaptive Compensation of Environmental Variability

Fig. 8 shows the power dissipation at various temperature conditions (25 °C–70 °C) when the operation frequency was set to 3 MHz in the following cases:

- none CT1: the circuit was controlled adaptively with a canary FF;
- none CT2: 200-mV FBB, which was the minimum body-bias for a 3-MHz operation at 25 °C, was fixedly applied;

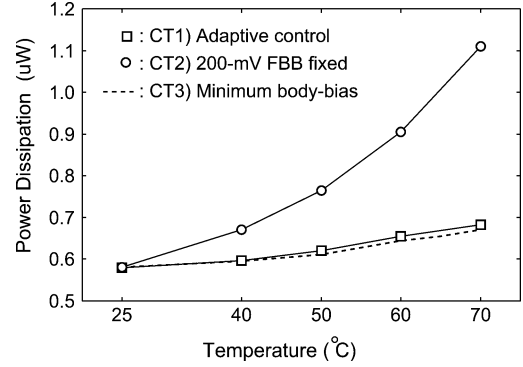


Fig. 8. Measured power dissipation at the various temperature conditions (3 MHz @ $V_{DD} = 0.35$ V). Circuit operates CT1) adaptively, CT2) with 200-mV FBB fixedly and CT3) with minimum body-bias required for 3-MHz operation at each temperature.

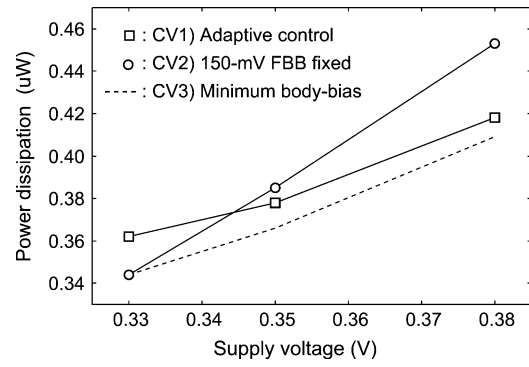


Fig. 9. Measured power dissipation at the various supply voltage (2 MHz). Circuit operates CV1) adaptively, CV2) with 150-mV FBB fixedly and CV3) with minimum body-bias required for 2-MHz operation at each voltage.

none CT3: the minimum FBB voltage required for a 3-MHz operation at each temperature was applied. In CT1, a canary FF at S[20] was enabled and its buffer delay was 130 ns at ZBB and 25 °C. The power dissipation includes those of the KSA, main FFs, speed control unit, and canary FF. The power overhead of the canary FF was estimated to be around 2% by circuit simulation. This measurement set four speed levels out of seven speed levels (ZBB—180-mV FBB) at each temperature. No timing errors were observed during 1.8×10^9 cycles at all temperature conditions.

Fig. 8 indicates that the power dissipation of CT1 is very close to that of CT3, which means optimal body-bias voltages were selected adaptively at each temperature. On the other hand, when the 200-mV FBB was fixedly applied (CT2), the power dissipation at 70 °C was 63% larger than that of CT1.

Fig. 9 shows the power dissipation at various supply voltages (0.33–0.38 V) when the operation frequency was set to 2 MHz in the following cases:

- none CV1: the circuit was controlled adaptively with a canary FF;
- none CV2: 150-mV FBB, which was the minimum body-bias required at $V_{DD} = 0.33$ V, was fixedly applied;
- none CV3: the minimum FBB voltage required for a 2-MHz operation at each supply voltage was applied.

No timing errors were observed during 1.2×10^9 cycles at each supply voltage. The power dissipation of the adaptive control

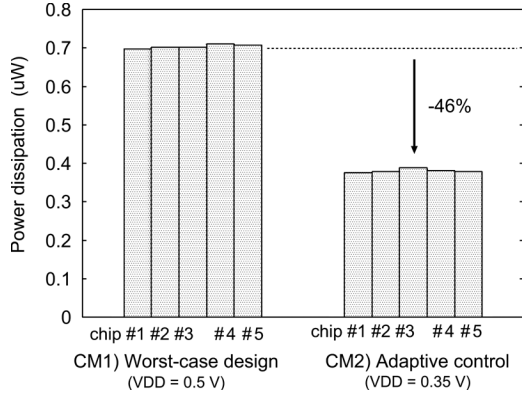


Fig. 10. Measured power dissipation when operation frequency is 2 MHz in the following cases: CM1) all chips operate at $V_{DD} = 0.5$ V, CM2) all chips operate with adaptive control at $V_{DD} = 0.35$ V.

(CV1) follows that with a minimum body-bias at each supply voltage (CV3), which means the circuit was adaptively controlled appropriately.

Figs. 8 and 9 indicate that the adaptive speed control with canary FF can compensate delay fluctuation due to temperature shift and supply voltage degradation.

C. Comparison to Operation Considering Worst Case

This section demonstrates how inefficient the worst-case design and operation for process, supply voltage and temperature are for subthreshold circuits and clarifies how beneficial the adaptive performance control is.

First, we discuss the worst-case design in terms of manufacturing variability. Assuming 2-MHz operation, the supply voltage must be 0.5 V or higher for a chip at the SS device corner, for example. In this case, all chips should operate at $V_{DD} = 0.5$ V when the traditional worst-case design with guardbanding is adopted. Fig. 10 shows the power dissipation of five chips in the following cases:

- none CM1: all chips operated at $V_{DD} = 0.5$ V, which was the minimum V_{DD} for a chip at the SS device corner;
- none CM2: all chips operated with adaptive control at $V_{DD} = 0.35$ V.

One canary FF was enabled and its location and buffer delay were determined such that no timing errors occurred during 1.2×10^9 cycles (10 min). The power dissipation with the adaptive control (CM2) was smaller than that with guardbanding (CM1) by 46%, because of lower supply voltage.

Fig. 11 shows the power dissipation when temperature is 60 °C (3 MHz @ $V_{DD} = 0.35$ V) in the following cases:

- none CVT1: body-bias voltage required for operation at the worst-case environmental condition (here, 25 °C and $V_{DD} = 0.33$ V) was fixedly applied, assuming that the body-bias voltage can be ideally obtained and given for each chip at a pre-shipment test;
- none CVT2: the circuit was controlled adaptively with a canary FF.

The power of CVT2 is 34% smaller than that of CVT1.

Even if an optimal body-bias could be given for each chip through expensive delay testing and the manufacturing variability unique to each chip could be eliminated, correct opera-

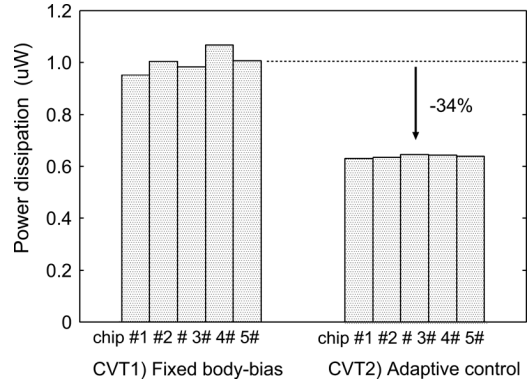


Fig. 11. Measured power dissipation when temperature is 60 °C (3 MHz @ $V_{DD} = 0.35$ V). “Fixed body-bias” denotes body-bias voltage required to operate at worst-case of environmental condition (in this example, 25 °C and $V_{DD} = 0.33$ V) is fixedly applied assuming that body-bias voltage can be ideally obtained at pre-shipment test.

tion at the worst-case environmental condition has to be assured. The body-bias selected for the worst-case is higher than needed at other environmental conditions. The adaptive speed control can select the appropriate body-bias according to the current environmental condition in addition to the manufacturing variability and hence the design with the adaptive control is much more efficient in power dissipation than the worst-case design.

IV. DESIGN PARAMETER CONFIGURATION

Section III revealed the effectiveness of the adaptive speed control with a canary FF. This section discusses how design parameters, such as the inserted location and the buffer delay of a canary FF, should be determined.

Figs. 12(a)–(c) show the measured timing error rate (MTBF) at each inserted location of a canary FF when the buffer delay is constant. MTBF at each inserted location is calculated by counting the number of timing errors for 10 min (1.2×10^9 cycles) and then dividing the number by the period. The input vector is generated by LFSR as depicted in Fig. 2. These figures indicate that the dependence of MTBF on the inserted location varies chip by chip. This is because the delay characteristic of each transistor fluctuates due to manufacturing variability. Therefore, consideration for manufacturing variability is required.

References [17] and [18] show that the timing error rate (MTBF) and power dissipation depend on the design parameters and larger power dissipation is required if MTBF is kept larger. In this section, the following two cases are examined: one in which the design parameters can be configured after fabrication and one in which the design parameters are fixedly set in the design phase and post-silicon configuration is not performed.

For practical use, the design parameters should be decided to satisfy the required timing error rate, which is thought to be much higher than that of the measurement setup described in Section III. Since it is not easy to measure such high required MTBF, this section uses simulations for higher MTBF evaluation based on the evaluation framework described in [17] and [18].

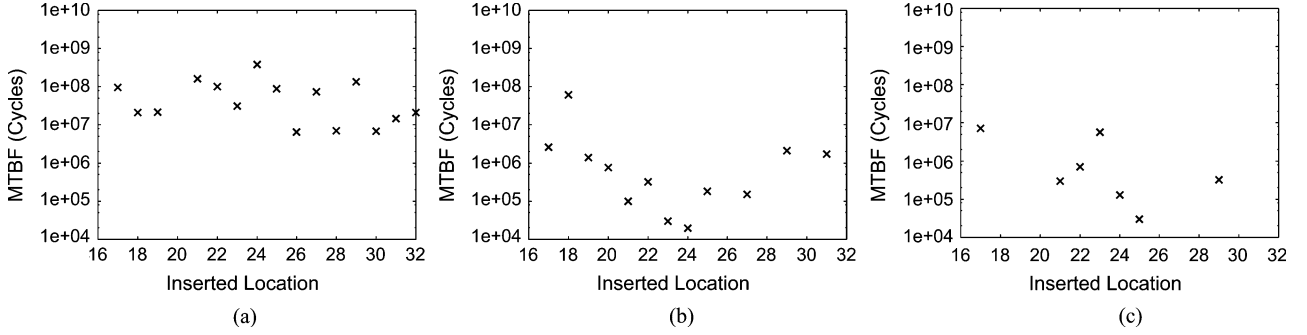


Fig. 12. MTBF at each inserted location when buffer delay is constant (2 MHz @ $V_{DD} = 0.35$ V). Values in parentheses represent inserted buffer delay at ZBB and 25 °C. Buffer delay of each chip is determined such that similar MTBF can be obtained among three chips: (a) chip A (130 ns); (b) chip B (100 ns); (c) chip C (100 ns).

We briefly introduce this framework [17], [18] (see [17] and [18] for details). The framework exploits the path activation probabilities to estimate the timing error rate and power dissipation. The occurrence probabilities of warning signals and timing errors are derived from the path activation probabilities. The speed-level transition satisfies Markov property, because the next state (speed level) is derived from only the current state and the occurrence probability of warnings. Therefore, the state (speed level) transition probabilities and the state probability of being at each speed level are calculated from the occurrence probability of warnings. Based on the state probability and the occurrence probability of timing errors, the timing error rate and power dissipation are obtained.

In this paper, we improved the evaluation framework to take manufacturing variability into account. Details of simulations based on the framework with consideration for manufacturing variability are explained in the following section.

A. Simulations Setup

In order to take manufacturing variability into account for the evaluation framework described in [17] and [18], simulations were performed according to the following procedure.

- i) For reproducing manufacturing variability, 100 chips were virtually fabricated using Monte Carlo simulations with threshold voltage (V_{th}) variation. Simulations described in this section assume that the standard deviations of within-die V_{th} variation and die-to-die V_{th} variation are 30 and 20 mV, respectively.
- ii) Path activation probabilities P_i and P_{all} are used. As defined in [18], $P_i(t, l, X)$ is the probability that at least one of the paths terminating at the i^{th} FF, whose delays are larger than t , is activated in a cycle at speed level l and $P_{all}(t, l, X)$ is the probability that at least one path in a circuit, whose delay is larger than t , is activated in a cycle at speed level l . X denotes the operating conditions such as temperature and the supply voltage and it is assumed that they are fixed to 25 °C and 0.35 V, respectively. P_i and P_{all} are expressed as histograms and to identify chips, they are written as $P_i^{[m]}$ and $P_{all}^{[m]}$ for chip m .
- iii) For chip m , circuit simulations are conducted with 10^8 random input vectors when the speed level is 0. Toggles at each output bit $S[0]$ – $S[32]$ and their delays from primary inputs are observed. $P_i^{[m]}(t, 0, X)$ is derived

by dividing the number of the toggles at $S[i]$, whose delays are larger than t , by the number of the input vectors. $P_{all}^{[m]}(t, 0, X)$ is calculated as $P_{all}^{[m]}(t, 0, X) = P_0^{[m]}(t, 0, X) \cup P_1^{[m]}(t, 0, X) \cup \dots \cup P_{32}^{[m]}(t, 0, X)$.

- iv) $P_i^{[m]}(t, l, X)$ and $P_{all}^{[m]}(t, l, X)$ at speed level l are calculated from those at $l = 0$ for simplicity as follows:

$$P_i^{[m]}(t, l, X) = P_i^{[m]}(t \cdot \gamma^{-l}, 0, X) \quad (1)$$

$$P_{all}^{[m]}(t, l, X) = P_{all}^{[m]}(t \cdot \gamma^{-l}, 0, X) \quad (2)$$

where γ is a coefficient expressing how much the circuit delay is decreased by incrementing l by one. γ is estimated to be 0.85 when the step of body-biasing levels is 30 mV.

- v) Steps iii)–iv) are repeated for virtually fabricated 100 chips and $P_i^{[m]}$ and $P_{all}^{[m]}$ for all chips are obtained.
- vi) $P_{ow}(l, X)$, the power dissipation of the KSA at each speed level, is obtained using circuit simulations for every chip.

For each virtually fabricated chip, the timing error rate is derived from P_i and P_{all} . The average power dissipation of the adaptively controlled KSA is calculated from P_{ow} . It was assumed from the simulations that the step of body-biasing levels was set to 30 mV and the monitoring period was 10^7 cycles with a 2-MHz clock frequency to conduct simulations consistent with the experiments described in Section III. In the simulations, the occurrences of metastability in main FFs and canary FFs are not considered for simplicity.

B. Post-Silicon Configuration

This section discusses the case when one canary FF is inserted and its buffer delay is adjustable after fabrication. It is assumed that the buffer delay can be configured ideally and the overhead in power dissipation is not considered.

1) *Dependence on Inserted Location*: First, we discuss how to decide the inserted location when the buffer delay can be adjusted after fabrication. The power dissipation necessary to meet an MTBF condition is used as a metric. Fig. 13 shows the simulated results of the minimum buffer delay required to make MTBF larger than 10^8 cycles at each inserted location. The power dissipation in the case when the minimum buffer

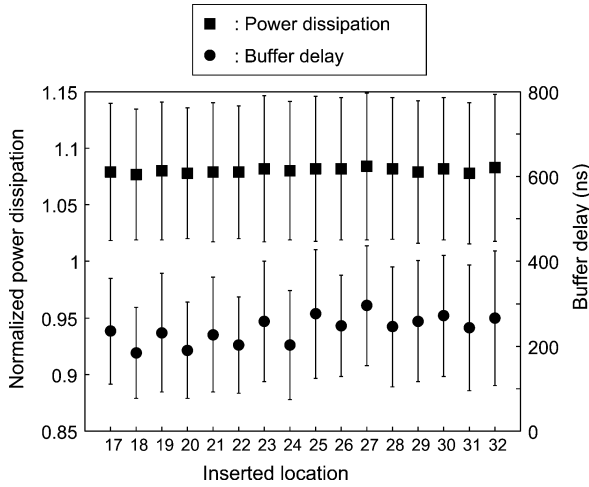


Fig. 13. Simulation results of minimum buffer delay and power dissipation such that condition $MTBF > 10^8$ cycles is satisfied (2 MHz @ $V_{DD} = 0.35$ V). Error bar at each inserted location represents ± 1 standard deviation for 100 chips. Power dissipation is normalized by that at 60-mV FBB.

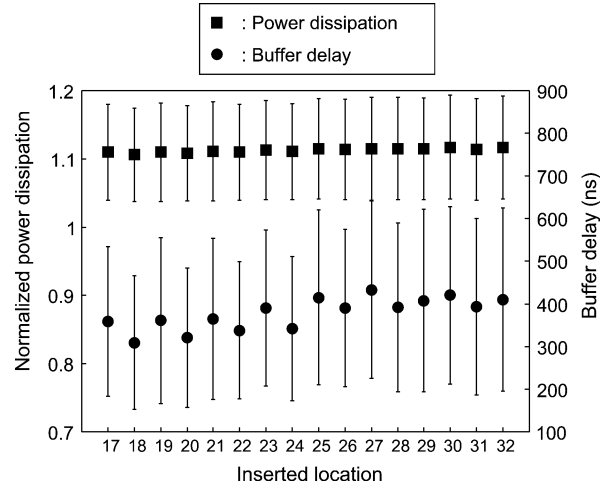


Fig. 15. Simulation results of minimum buffer delay that makes $MTBF$ larger than 10^{15} cycles and power dissipation (2 MHz @ $V_{DD} = 0.35$ V). Error bar at each inserted location represents ± 1 standard deviation for 100 chips. Power dissipation is normalized by that at 60-mV FBB.

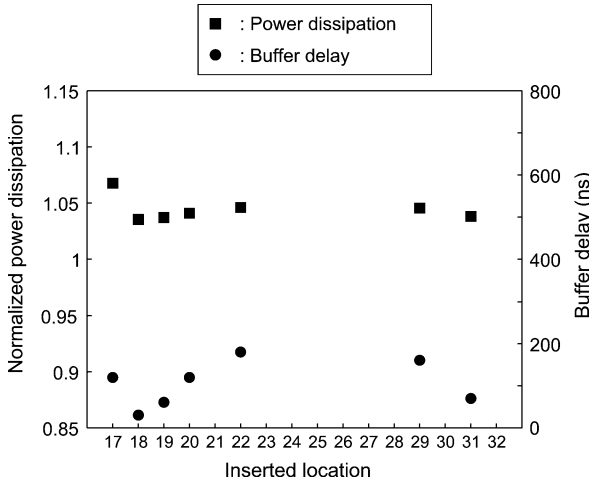


Fig. 14. Measured minimum buffer delay and power dissipation on certain chip (2 MHz @ $V_{DD} = 0.35$ V). Power dissipation is normalized by that at 60-mV FBB.

delay is inserted is also shown. The buffer delay and the power dissipation are obtained in the following procedure.

- i) The minimum buffer delay at each inserted location is determined for every virtually fabricated chip such that the condition (for example, $MTBF > 10^8$ cycles in this simulation) is satisfied.
- ii) The average power dissipation of the adaptively controlled KSA with canary FF with the minimum buffer delay is derived.
- iii) Steps i) and ii) are repeated for 100 chips and the mean (μ) and the standard deviation (σ) of the buffer delays and the power dissipations are calculated at each inserted location.

Each point in Fig. 13 denotes the mean for the buffer delay and the power dissipation and each error bar represents ± 1 standard deviation.

Fig. 14 shows the silicon measurement results of the minimum buffer delay and the power dissipation to meet the condition $MTBF > 10^8$ cycles at each inserted location. As for the in-

serted locations without dots in this figure, the condition $MTBF > 10^8$ cycles was not satisfied even with the maximum buffer delay that can be set in the configurable canary FF implemented in this chip. Both measured and simulated results indicate that the buffer delay required to satisfy the constraint varies the location by location, whereas the dependence of power dissipation on the inserted location is small. This means that the inserted location of a canary FF is not important for the KSA in terms of power dissipation when the buffer delay is tunable after fabrication.

Fig. 15 shows the minimum buffer delay when the constraint is $MTBF > 10^{15}$ cycles (15 years @ 2-MHz operation) and the power dissipation with the minimum buffer delay at each inserted location. Even in such a severer constraint, the dependence of power dissipation on the inserted location is still small.

2) *Tradeoff Relations Between Timing Error Rate and Power Dissipation:* Next, the tradeoff relation between the timing error rate ($MTBF$) and power dissipation is examined. Fig. 16 shows a comparison of the measured and simulated tradeoff relations. The power dissipation is normalized by that at a 60-mV FBB. The measured tradeoff relations are obtained by measuring the $MTBF$ s and power dissipations with various inserted locations and buffer delays for two manufactured chips, denoted as “chip X” and “chip Y” in this figure. In contrast, the simulated tradeoff relation in Fig. 16 is derived as follows to take manufacturing variability into consideration.

- i) At each inserted location from S[17] to S[32] in every virtually fabricated chip, the minimum buffer delay is determined such that the constraint $MTBF > 10^n$ ($n = 5, 6, \dots, 15$) cycles is satisfied and the power dissipation with the minimum buffer delay is derived. Consequently, 1600 power dissipations (16 inserted locations \times 100 chips) are obtained.
- ii) The mean and standard deviations of the power dissipation are calculated from the 1600 power dissipations.
- iii) Step i) and ii) are repeated with $n = 5, 6, \dots, 15$. From these steps, the tradeoff relation between $MTBF$ and power dissipation is obtained.

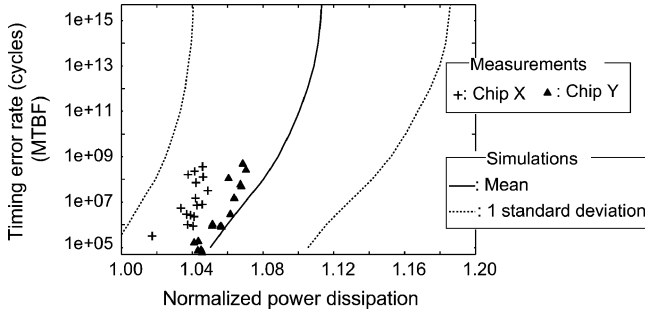


Fig. 16. Trade-off relations between timing error rate (MTBF) and power dissipation (2 MHz @ $V_{DD} = 0.35$ V). Solid line represents the mean of simulated tradeoff relations and dotted lines denote ± 1 standard deviation. Power dissipation is normalized by that at 60-mV FBB.

The measured tradeoff relations lie within ± 1 standard deviations. From the measurement results, the average power dissipation to meet the condition $MTBF > 10^8$ cycles was $0.367 \mu\text{W}$. From the prediction based on the simulated tradeoff, the power dissipation to satisfy the condition $MTBF > 10^{15}$ cycles (15 years at 2 MHz operation) is estimated to be $0.378 \mu\text{W}$. This implies that MTBF is dramatically improved by a small power overhead and the power dissipation to achieve higher MTBF is still much smaller than that with guardbanding ($0.7 \mu\text{W}$) depicted in Fig. 10.

C. Configuration in Design Phase

This section discusses the case where the buffer delay is determined in the design phase and is not tuned after fabrication. The required buffer delay varies chip by chip. Thus, to overcome this fluctuation, the following two approaches are considered: 1) to add one canary FF with the buffer delay, which is long enough to cover the worst-case condition and 2) to add multiple canary FFs. We assume that the required MTBF is larger than 10^{15} cycles. In simulations, the buffer delay is normalized by the average delay of inverters in each chip, that is, the buffer delay is expressed as the number of logic stages. In the design phase, the number of required stages of the delay buffer is determined.

1) *Insertion of One Canary FF With Longer Buffer Delay:* The procedure for obtaining the mean (μ) and standard deviation (σ) of the buffer delay at each inserted location are described in Section IV-B. First, the $\mu + 3\sigma$ buffer delay at each inserted location is calculated to evaluate the worst-case buffer delay. Fig. 17 depicts the buffer delay when the constraint is $MTBF > 10^{15}$ cycles. This result indicates that it is optimal to insert a canary FF at S[20] because the buffer delay in the inserted location is smallest. In this case, the inserted buffer delay is equivalent to 69% of the critical path delay on average. Please note that the dependence on the inserted location shown in Fig. 17 depends on a circuit to which the adaptive speed control with canary FF is applied. In case of a circuit whose delays of most paths are comparable, for example, the dependence on the inserted location becomes smaller. On the other hand, the mean of the delay variation of critical path becomes larger due to the statistical effect caused by the max operation [22] and hence the buffer delay should be determined according to a difference between the path delay where canary FF is inserted and the critical path delay with consideration for such statistical effect.

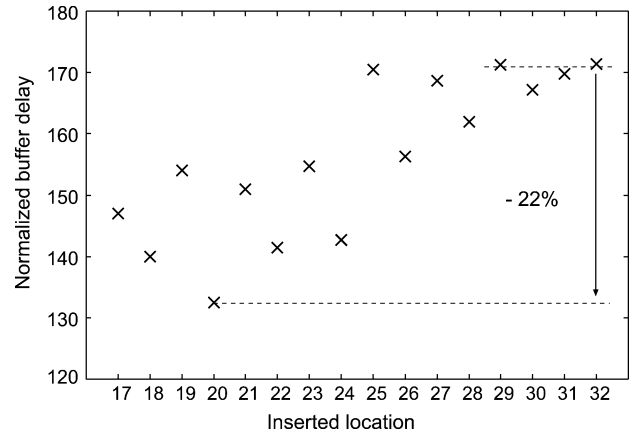


Fig. 17. Mean plus three standard deviations of buffer delay ($MTBF > 10^{15}$ cycles). Buffer delay is normalized by average delay of inverters in each chip.

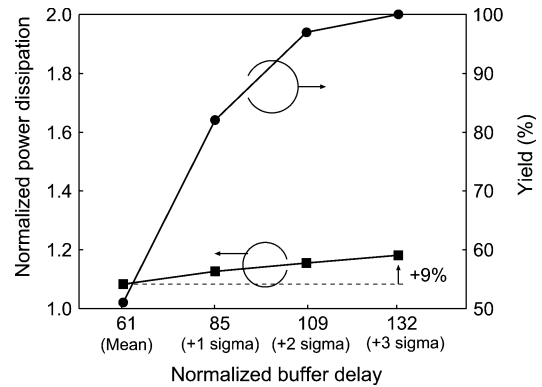


Fig. 18. Power dissipation and yield when canary FF with fixed buffer delay is inserted S[20] (2 MHz @ $V_{DD} = 0.35$ V). Yield means ratio of chips that satisfy condition $MTBF > 10^{15}$ cycles. Power dissipation is normalized by that at 60-mV FBB.

Fig. 18 shows the average power dissipation of the virtually fabricated 100 chips and the yield when a canary FF with the fixed buffer delay is inserted S[20]. The yield is defined as the ratio of the number of chips satisfying the MTBF requirement to the total number of chips. The fixed buffer delay is set to $\mu + k\sigma$ ($k = 0, 1, 2, 3$). A longer buffer delay can cover a wider process variability space and all chips in the virtually fabricated chips can meet the constraint $MTBF > 10^{15}$ cycles using one canary FF with the fixed $\mu + 3\sigma$ buffer delay. In this case, the overhead in the power dissipation is 9%.

2) *Multiple Canary FF Insertion:* Next, another approach, multiple canary FF insertion, is discussed. Multiple canary FFs are inserted as follows.

- i) Assuming that one canary FF is inserted, the mean (μ) and standard deviation (σ) of the buffer delay at each inserted location are calculated according to the procedure described in Section IV-B1.
- ii) The output bit with smaller μ is given a higher priority to the multiple canary FF insertion.

Fig. 19 plots the average power dissipation and the yield as a function of the number of canary FFs. Each canary FF fixedly has the μ buffer delay. It should be noted that this analysis is optimistic because the warning occurrence probability P_w of multiple canary FFs is calculated as the sum of P_w of each canary

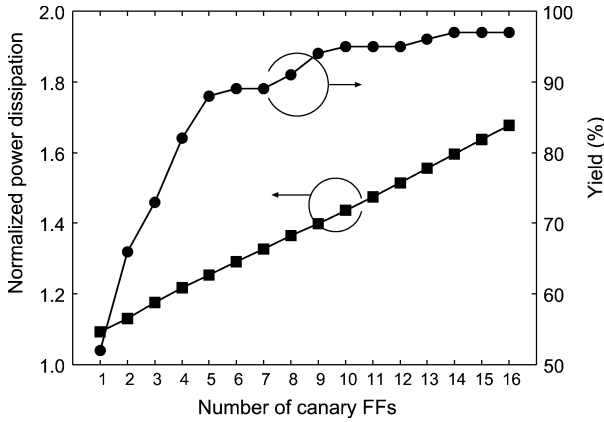


Fig. 19. Power dissipation and Yield as function of number of canary FFs whose buffer delays are set to μ (2 MHz @ $V_{DD} = 0.35$ V). Yield means ratio of chips that satisfy condition $MTBF > 10^{15}$ cycles. Power dissipation is normalized by that at 60-mV FBB.

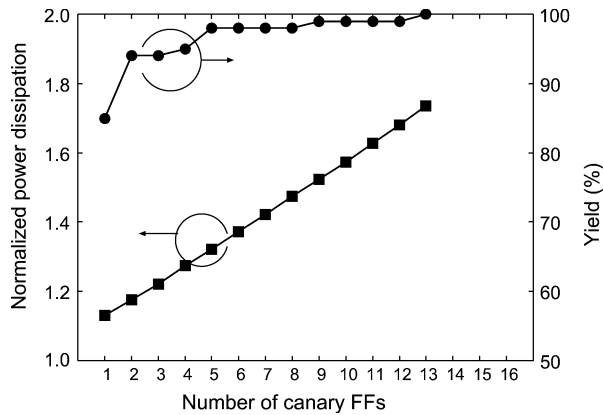


Fig. 20. Power dissipation and yield as function of number of canary FFs whose buffer delays are set to $\mu + 1\sigma$ (2 MHz @ $V_{DD} = 0.35$ V). Yield means ratio of chips that satisfy condition $MTBF > 10^{15}$ cycles. Power dissipation is normalized by that at 60-mV FBB.

FF, which actually should be expressed as a union. Even when 16 canary FFs are inserted, the yield does not reach 100%. This is because the buffer delays among the inserted location are relatively correlative (the average of the correlation coefficients are 0.61).

Fig. 20 shows the case where each canary FF has a $\mu + 1\sigma$ buffer delay. In this case, all chips of the virtually fabricated chips can satisfy the constraint $MTBF > 10^{15}$ cycles by inserting 13 canary FFs.

D. Discussion

This section compares the three cases mentioned above, i.e., 1) design parameters are ideally tunable after fabrication; 2) one canary FF with a fixed buffer delay is inserted; and 3) multiple canary FFs with a fixed buffer delay are inserted.

Fig. 21 shows the power dissipation required to achieve 100% yield for virtually fabricated 100 chips. The case of one canary FF insertion assumes that the canary FF is inserted in $S[20]$ and the fixed delay is the $\mu + 3\sigma$ buffer delay. The case of multiple canary FF insertion assumes that each canary FF has the $\mu + 1\sigma$ buffer delay.

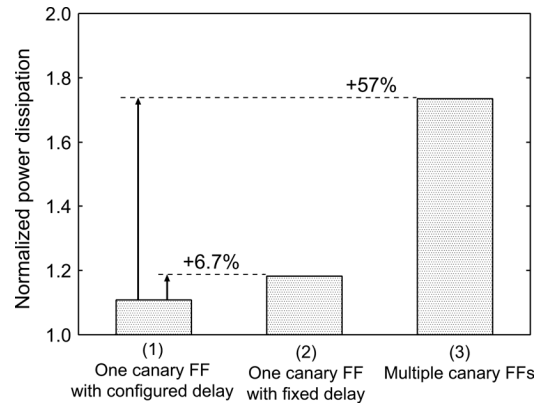


Fig. 21. Power dissipation required to achieve 100% yield when constraint $MTBF > 10^{15}$ cycles is assumed (2 MHz @ $V_{DD} = 0.35$ V): 1) one canary FF with tuned buffer delay after fabrication is inserted in $S[20]$; 2) one canary FF with fixed $\mu + 3\sigma$ buffer delay is inserted in $S[20]$; and 3) multiple canary FFs with fixed $\mu + 1\sigma$ buffer delay are inserted. Yield means ratio of chips that meet constraint. Power dissipation is normalized by that at 60-mV FBB.

This figure indicates that the power dissipation, when multiple canary FFs are inserted, is 57% larger than that of one canary FF with a tuned buffer delay, whereas the power dissipation when one canary FF with a fixed buffer delay is 6.7% larger. Since the timing error rate in the KSA is sensitive to the buffer delay, it is adequate to use the buffer delay to satisfy the MTBF requirement. On the other hand, it is not effective to insert multiple canary FFs due to the correlations of the buffer delays among the inserted locations.

In a practical design, tuning of the buffer delay of canary FF is performed during the test phase. For example, one or more test vectors are set using scan chains or other methods and then the occurrence of timing errors are examined by performing at-speed tests with the vectors. According to the occurrence of timing errors, the buffer delay of canary FF is tuned. Although it is most efficient that the buffer delay is tuned for every chip as shown in Fig. 21, its configuration cost after fabrication cannot be ignored. In addition, the buffer delay is not ideally tunable in a practical case and it is possible that the configurability of delay buffer would not be sufficient to attain the required MTBF due to manufacturing variability. The delay buffer with such wide configurability requires additional implementation cost as well as energy overhead.

Therefore, a most realistic approach could be to insert one canary FF with a buffer delay that is long enough to cover the worst-case condition in the design phase. In this case, the area overhead of canary FF was 10.5% of the whole area of the circuit, which contains the KSA and main FFs at inputs and outputs of the KSA, in the case when the delay buffers are composed of the minimum inverter.

A note here is that the increase in the power dissipation shown in Fig. 21 depends on the ratio of an energy dissipation of a combinational logic to that of FFs. When the circuit size is scaled up and the energy of the combinational logic become relatively larger, the power overhead to insert multiple canary FFs relatively decreases. Therefore, the optimum selection of canary FF insertion (single or multiple) could depend on the circuit size and structure. In addition, it is possible that the increase

TABLE I

DEPENDENCE ON STEP OF BODY-BIASING LEVEL: (A) ONE CANARY FF WITH CONFIGURED DELAY (POST-SILICON CONFIGURATION); (B) ONE CANARY FF WITH FIXED BUFFER DELAY (CONFIGURATION IN DESIGN PHASE)

Step of body-biasing	Ratio of buffer delay to critical path delay	Normalized power dissipation
30 mV	0.31	1.000
10 mV	0.25	0.972
5 mV	0.24	0.965

(a)

Step of body-biasing	Ratio of buffer delay to critical path delay	Normalized power dissipation
30 mV	0.69	1.067
10 mV	0.58	1.033
5 mV	0.54	1.023

(b)

in the circuit size reduces the gates and paths shared between the critical path and the path where canary FF is inserted. Consequently, to exploit the advantage of canary FFs, insertion of multiple canary FFs might be reasonable. The dependence of the most energy-efficient approach for the canary FF insertion on the circuit size and structure is a future work.

E. Dependence on Body-Biasing Step

In the simulations in this section above, the step of body-biasing level was set to 30 mV in order to perform the consistent simulations with the measurements described in Section III. The step of body-biasing level is also considered as one of the design parameters. Thus, this section examines the step of body-biasing level.

Table I lists the dependence of the step on the inserted buffer delay, which is normalized by the critical path delay and the power dissipation when the constraint $MTBF > 10^{15}$ cycles (2 MHz @ $V_{DD} = 0.35$ V) is given. The power is normalized by that of one canary FF with configured delay when the step of body-biasing level is 30 mV. The values of delay and power dissipation in Table I are average ones of the virtually fabricated 100 chips. As the step of body-biasing level becomes finer, the inserted buffer delay and the power dissipation are reduced. This is because more appropriate body-bias voltage is applied.

V. CONCLUSION

We presented a self-adaptive compensation technique using canary FF for subthreshold circuits. A 32-bit KSA, whose performance was controlled by body-biasing, was fabricated in a 65-nm CMOS process. Fabricated chips demonstrated that the adaptive speed control with canary FF functioned at 350 mV. Measurement results showed that the adaptive control compensated manufacturing and environmental variability and reduced power dissipation by 46% compared to traditional worst-case design. We also discussed how to determine design parameters, such as the inserted location and the buffer delay of a canary FF. Simulation results indicated that it is appropriate to adjust the buffer delay to attain higher MTBF, whereas it is not efficient to insert multiple canary FFs. One canary FF insertion with the sufficient buffer delay to cover a wider manufacturing variability space is the most practical.

ACKNOWLEDGMENT

This work was performed by the authors for STARC as part of the Japanese Ministry of Economy, Trade and Industry sponsored "Next-Generation Circuit Architecture Technical Development" program. The VLSI chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with STARC, e-Shuttle, Inc., and Fujitsu Ltd.

REFERENCES

- [1] M. Seok, S. Hanson, Y. S. Lin, Z. Y. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "The phoenix processor: A 30 pW platform for sensor applications," in *Int. Symp. VLSI Circuits Dig. Tech. Papers*, 2008, pp. 188–189.
- [2] N. Ickes, D. Finchelstein, and A. P. Chandrakasan, "A 10-pJ/instruction, 4-MIPS micropower DSP for sensor applications," in *Proc. Asian Solid-State Circuits Conf. (ASSCC)*, 2008, pp. 289–292.
- [3] G. Chen, M. Fojtik, D. Kim, D. Fick, J. Park, M. Seok, M. Chen, Z. Foo, D. Sylvester, and D. Blaauw, "Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2010, pp. 288–289.
- [4] T. Kuroda, K. Suzuki, S. Mita, T. Fujita, F. Yamane, F. Sano, A. Chiba, Y. Watanabe, K. Matsuda, T. Maeda, T. Sakurai, and T. Furuyama, "Variable supply-voltage scheme for low-power high-speed CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 454–462, Mar. 1998.
- [5] J. W. Tschanz, J. T. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis, A. P. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.
- [6] J. T. Kao, M. Miyazaki, and A. P. Chandrakasan, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov. 2002.
- [7] I. J. Chang, S. P. Park, and K. Roy, "Exploring asynchronous design techniques for process-tolerant and energy-efficient subthreshold operation," *IEEE J. Solid-State Circuits*, vol. 45, no. 2, pp. 401–410, Feb. 2010.
- [8] 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.
- [9] S. Das, C. Tokunaga, S. Pant, W. H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull, and D. Blaauw, "Razor II: 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.
- [10] T. Nakura, K. Nose, and M. Mizuno, "Fine-grain redundant logic using defect-prediction flip-flops," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2007, pp. 402–403.
- [11] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit failure prediction and its application to transistor aging," in *Proc. VLSI Test Symp. (VTS)*, 2007, pp. 277–286.
- [12] T. Sato and Y. Kunitake, "A simple flip-flop circuit for typical-case designs for DFM," in *Proc. Int. Symp. Quality Electron. Des. (ISQED)*, 2007, pp. 539–544.
- [13] S. Hanson, M. Seok, Y. S. Lin, Z. Y. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "A low-voltage processor for sensing applications with picowatt standby mode," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1145–1155, Apr. 2009.
- [14] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, "Exploring variability and performance in a sub-200-mV processor," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 831–891, Apr. 2008.
- [15] A. Wang and A. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.
- [16] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error prediction for subthreshold circuits," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2009, pp. 215–218.

- [17] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Trade-off analysis between timing error rate and power dissipation for adaptive speed control with timing error prediction," *IEICE Trans. Fund.*, vol. E92-A, no. 12, pp. 3094–3102, Dec. 2009.
- [18] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Trade-off analysis between timing error rate and power dissipation for adaptive speed control with timing error prediction," in *Proc. Asia South Pacific Des. Autom. Conf. (ASP-DAC)*, 2009, pp. 266–271.
- [19] D. Bol, D. Flandre, and J. D. Legat, "Technology flavor selection and adaptive techniques for timing-constrained 45 nm subthreshold circuits," in *Proc. Int. Symp. Low Power Electron. Des. (ISLPED)*, 2009, pp. 21–26.
- [20] L. S. Lin, D. Sylvester, and D. Blaauw, "A sub-pW timer using gate leakage for ultra low-power sub-Hz monitoring systems," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2007, pp. 397–400.
- [21] L. S. Lin, D. M. Sylvester, and D. T. Blaauw, "A 150 pW program-and-Hold timer for ultra-low-power sensor platforms," in *Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2009, pp. 326–327.
- [22] M. Hashimoto and H. Onodeva, "Increase in delay uncertainty by performance optimization," in *Proc. Int. Symp. Circuits Syst. (ISCAS)*, 2001, pp. 379–382.



Hiroshi Fuketa (S'07–M'10) received the B.E. degree from Kyoto University, Kyoto, Japan, in 2002 and the M.E. and Ph.D. degrees in information systems engineering from Osaka University, Osaka, Japan, in 2008 and 2010, respectively.

He is currently a Research Associate with the Institute of Industrial Science, the University of Tokyo, Tokyo, Japan. His research interests include ultra-low-power circuit design and variation modeling.

Dr. Fuketa is a member of IEICE and IPSJ.



Masanori Hashimoto (S'00–A'01–M'03) received the B.E., M.E., and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2001, respectively.

Since 2004, he has been an Associate Professor with the Department of Information Systems Engineering, Osaka University, Osaka, Japan. His research interests include computer-aided-design for digital integrated circuits and high-speed and low-power circuit design.

Dr. Hashimoto was a recipient of the Best Paper Award at ASP-DAC 2004. He is a member of IEICE and IPSJ. He served on the technical program committees for international conferences including DAC, ICCAD, ASP-DAC, DATE, ICCD, and ISQED.



Yukio Mitsuyama (S'97–M'02) received the B.E., M.E., and Ph.D. degrees in information systems engineering from Osaka University, Osaka, Japan, in 1998, 2000, and 2010, respectively.

He is currently an Assistant Professor with Graduate School of Engineering, Osaka University. His research interests include reconfigurable architecture and its VLSI design.

Dr. Mitsuyama was a recipient of the Best Paper Award at IEEE ISCE 2004. He is a member of IEICE and IPSJ.



Takao Onoye (S'93–M'95–SM'07) received the B.E. and M.E. degrees in electronic engineering and the Dr.Eng. degree in information systems engineering from Osaka University, Osaka, Japan, in 1991, 1993, and 1997, respectively.

He was an Associate Professor with the Department of Communications and Computer Engineering, Kyoto University, Kyoto, Japan. Since 2003, he has been a Professor with the Department of Information Systems Engineering, Osaka University.

He has published over 200 research papers in the field of VLSI design and multimedia signal processing in reputed journals and proceedings of international conferences. His current research interests include media-centric low-power architecture and its SoC implementation.

Dr. Onoye has served as a member of the CAS Society Board of Governors since 2008. He is a member of IEICE, IPSJ, and ITE-J.