MTTF-Aware Design Methodology of Adaptively Voltage Scaled Circuit with Timing Error Predictive Flip-Flop

Yutaka MASUDA[†∗a)] *and* **Masanori HASHIMOTO**[†b)]*, Members*

SUMMARY   Adaptive voltage scaling is a promising approach to overcome manufacturing variability, dynamic environmental fluctuation, and aging. This paper focuses on error prediction based adaptive voltage scaling (EP-AVS) and proposes a mean time to failure (MTTF) aware design methodology for EP-AVS circuits. Main contributions of this work include (1) optimization of both voltage-scaled circuit and voltage control logic, and (2) quantitative evaluation of power saving for practically long MTTF. Experimental results show that the proposed EP-AVS design methodology achieves 38.0% power saving while satisfying given target MTTF.

key words: *adaptive voltage scaling, activation-aware slack assignment, mean time to failure, timing error predictive FF*

1. Introduction

Aggressive device miniaturization due to technology scaling has been improving the average device performance. However, circuits have become sensitive to static manufacturing variability and dynamic environmental fluctuation. Moreover, device aging, which is another temporal variation and is represented by negative bias temperature instability (NBTI) [1], [2], degrades performance gradually in the field. These static and temporal variations directly lead to circuit reliability degradation. For overcoming variabilities mentioned above, design and operating margins are given in design time and field, respectively, for ensuring correct circuit operation. However, as the performance variation becomes significant, such margin tends to be too painful for designers. Therefore, a traditional worst-case (WC) design with guard-banding is inefficient, and an adaptive performance compensation is desired.

The most effective tuning knob for post-silicon compensation is supply voltage control, and then adaptive voltage scaling (AVS) is intensively studied [3]–[7]. AVS is expected to minimize process, voltage, temperature, and aging (PVTA) margin of each chip and allocate only a little margin for the entire lifetime as illustrated in Fig. 1. The excessive conventional PVTA margin in most of the chips can be exploited as the source for power savings.

There are two AVS strategies in literature; error detec-
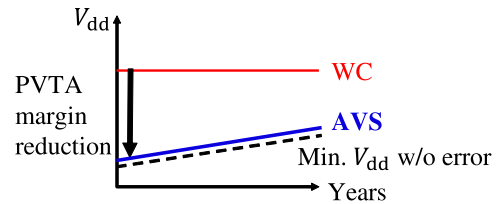
Fig. 1   Supply voltages of AVS and conventional WC designs in device lifetime. Ideal AVS minimizes PVTA margin of each chip.

tion and recovery based control (e.g., Razor [3]), and error prediction and prevention based control (e.g. canary FF [8], slack monitor [7], error predictive FF [9]∗). In both the strategies, sensors are embedded to detect/predict timing errors, and the supply voltage is controlled according to the sensor outputs. Therefore, these existing works [3]–[9] focus on where to insert sensors and how to control supply voltage and discuss the design methodology of the voltage control system. Most of the conventional works embed sensors to timing-critical paths to detect/predict setup timing violation. This strategy is reasonable since timing-critical paths can be extracted in design time with static timing analysis (STA) tools.

On the other hand, for implementing AVS systems that fully exploit run-time adaptation and eliminate the redundant margin, we have found that we should pay attention to the main logic circuit under AVS in addition to the sensing circuit. In the conventional VLSI design flow, there tends to be many critical paths since the timing slack is exploited for power and area reduction. On the one hand, we observe that inherent critical paths whose path delays cannot be reduced at all are limited. This observation suggests that activation-aware slack assignment (ASA) to the main logic circuit under AVS, which allocates larger slack to highly active paths, could improve the efficacy of the AVS and enable further power savings with extremely low error rate.

This work focuses on the error prediction based AVS (EP-AVS) and proposes a design methodology for EP-AVS circuits. The proposed methodology optimizes both the main logic under AVS and sensing circuit. In the main logic design, we perform a mean time to failure (MTTF) aware ASA [10] and estimate the MTTF of AVS circuits with a stochastic framework [11]. The MTTF-aware ASA enforces larger timing slack on the FFs that have frequent input transitions immediately before the clock edge since those FFs tend to

∗There are several names, but the sensor structure is the same.

cause setup timing errors. The number of FFs which are likely to fail is reduced in the design phase, and thus the insertion of error prediction sensors is facilitated in the EP-AVS design. As for the sensing circuit design, we propose a novel sensor insertion method that maximumly decreases the sum of gate-wise timing failure probabilities, where the timing failure probability is the joint probability of activation and timing violation probabilities. By exploiting the information on the paths with higher timing failure probability, the proposed sensor insertion makes EP-AVS efficiently monitor the timing-critical and highly-active FFs. Experimental results show that MTTF-aware main logic design is highly compatible with EP-AVS, and they mutually enhance and provide further power savings and performance improvement with margin elimination.

Main contributions of this work include (1) optimization of both main logic under AVS and sensing circuit, and (2) quantitative evaluation of power savings for practically long MTTF. To the best of our knowledge, this is the first work that optimizes both the main logic under AVS and sensing circuit under the constraint of MTTF in units of several years and demonstrates the power savings explicitly taking into account such a practically long MTTF. Figure 2 illustrates the expected power saving effects. The top black curve represents the conventional WC design that adds timing margins assuming the worst PVTA condition. The middle yellow and bottom blue curves correspond to the conventional EP-AVS without main logic optimization and the proposed EP-AVS with the ASA. The proposed EP-AVS is expected to attain a better trade-off relation between the clock period and power thanks to the main logic optimization. This power saving effects in an embedded processor and a cipher circuit will be experimentally demonstrated.

Preliminary results of voltage reduction thanks to the design methodology for EP-AVS circuits were reported in [12]. This work evaluates power saving instead taking into account the area overhead and the increase in the number of low-Vth cells by ASA. Also, we apply ASA [10], which adjusts timing slack under MTTF constraint, to the main logic for improving performance, whereas [12] applied an earlier work called critical path isolation (CPI) [13] that increases timing slacks of highly activated paths as much as possible.
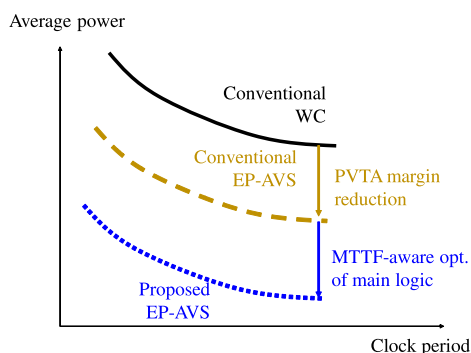
Also, this work utilizes the pre-ASA circuit optimization which is also proposed in [10]. Section 5 will show that the ASA with pre-ASA circuit optimization reduces the area even from the baseline pre-ASA circuit in a test case.
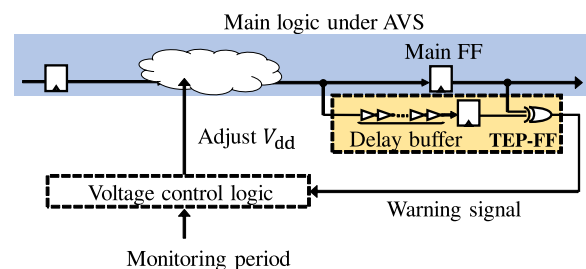
The rest of this paper is organized as follows. Section 2 describes the assumed AVS in this paper and explains the overview of the proposed design which consists of the main logic optimization and the sensing circuit optimization. Section 3 designs the main logic with referring to [10], where the pre-ASA design and ASA implementation are introduced. Section 4 explains the proposed sensor insertion methodology, which is applied to the ASA circuit designed in Sect. 3. Section 5 evaluates the trade-off between average power and the clock period of the conventional WC design, conventional EP-AVS, and the proposed EP-AVS and demonstrates the power saving effects thanks to the proposed EP-AVS. Lastly, concluding remarks are given in Sect. 6.

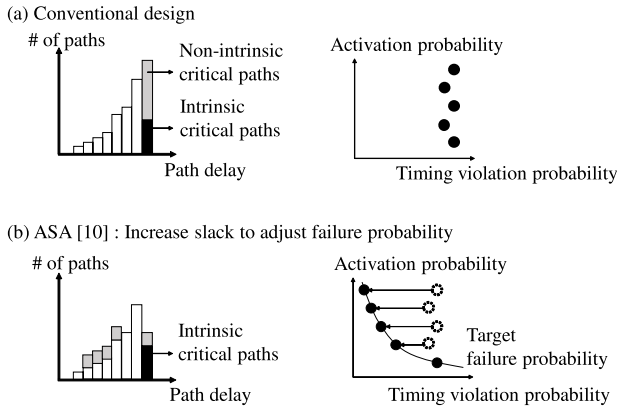## 2. Overview of Proposed Design Methodology for EP-AVS

The proposed design methodology for EP-AVS consists of the ASA for the main logic under AVS and the insertion of error prediction sensors. Section 2.1 explains the assumed EP-AVS and Sect. 2.2 formulates the design optimization problem of EP-AVS. Then, Sect. 2.3 explains the overview of the proposed design methodology.

2.1 Assumed EP-AVS

Figure 3 illustrates an EP-AVS circuit assumed in this paper. The EP-AVS circuit is composed of the main circuit, timing error predictive FF (TEP-FF) and voltage control unit. The TEP-FF consists of a FF, delay buffers, and a comparator (XOR gate), and works with the main FF. When the timing margin is gradually decreasing, a timing error occurs at the TEP-FF before the main FF captures a wrong value due to the delay buffer, which enables us to know that the timing margin of the main FF is not large enough. An error prediction signal is generated to predict the timing errors, and this signal is monitored during a specified period. Note that timing errors are predicted, not detected, which is a distinct difference from Razor [3]. Once an error prediction signal is observed, the higher supply voltage is given to reduce circuit delay. Note that clock frequency is fixed throughout this paper. If no

Fig. 2 Expected performance improvement thanks to the proposed EP-AVS design methodology.

Fig. 3 Assumed EP-AVS.

(a) Conventional design

(b) ASA [10] : Increase slack to adjust failure probability

Fig. 4 Path delay distributions (left side) and the activation probability and timing violation probability of non-intrinsic critical paths (right side) of circuits. (a) conventional design, (b) ASA [10]. ©[2018] IEEE. Reprinted, with permission, from [10].

error prediction signals are observed during the monitoring period, the circuit is slowed down for power reduction. This proactive AVS is expected to overcome the variation of the timing margin which is different in every chip and varies depending on operating condition and aging.

Figure 4 illustrates the concept of ASA which is utilized for the main logic design. The left side of Fig. 4(a) illustrates the path delay distribution of a conventionally designed circuit, and the right side shows the pair of the activation probability and timing violation probability of non-intrinsic critical paths, where the non-intrinsic critical paths are timing paths which originally had large timing slacks before the downscaling and replacement. In the conventional circuit design flow, cell instances included in non-critical paths are replaced with smaller cells and high-Vth cells for reducing power dissipation and area. Therefore, the number of non-intrinsic critical paths increases. On the other hand, this replacement decreases timing margin of the paths that go through the replaced instances and may increase the timing error occurrence probability under variations. In other words, more instances are prone to cause path delay variations.

On the other hand, ASA increases timing slacks of highly-activated non-intrinsic critical paths. The left side of Fig. 4(b) exemplifies the path delay distribution of the ASA circuit. As ASA enforces larger slacks on highly activated paths, highly-activated paths sustain timing margin even when gate delay varies. Accordingly, as shown in the right side of Fig. 4(b), timing violation probability in these paths is dramatically reduced compared to the conventional circuit, which is the main advantage of the ASA. These reductions extend MTTF and consequently save power. Here, it should be noted that ASA partially loses the power and area reduction acquired by the conventional design optimization. From this sense, we need to find a better trade-off relation between the timing error occurrence probability and power. For pursuing the better trade-off, ASA proposed in [10] adjusts failure probability of the path to target failure

probability as shown in the right side of Fig. 4(b). Here, the amount of slack increase is assigned to the minimum value that satisfies the target MTTF for reducing power and area overheads. Thanks to this assignment, the ASA can save the overhead while extending MTTF and saving power. Note that the failure probability is defined as the product of activation probability and timing violation probability of a path, and the target failure probability can be calculated from the target MTTF [10].

2.2 Problem Definition of EP-AVS Design

Based on the discussion in the Sect. 2.2, we formulate the design optimization of EP-AVS including ASA and TEP-FF insertion.

- Input
  - N_{CKT} pre-ASA candidates
- Output
  - one EP-AVS circuit
- Objective
  - Minimize : Power =
    $$\min(\text{Power}_1, \cdots, \text{Power}_{N_{\text{CKT}}})$$
- Constraints
  - $\text{MTTF}_j \geq \text{MTTF}_{\min}(1 \leq j \leq N_{\text{CKT}})$
  - $\text{Area}_{\text{ASA}_j} \leq \text{Area}_{\text{ASA}}^{\max}(1 \leq j \leq N_{\text{CKT}})$
  - $\text{Area}_{\text{TEP}_j} \leq \text{Area}_{\text{TEP}}^{\max}(1 \leq j \leq N_{\text{CKT}})$
  - $\text{NLvth}_{\text{ASA}_j} \leq \text{NLvth}_{\text{ASA}}^{\max}(1 \leq j \leq N_{\text{CKT}})$
- Variables
  - $\Delta\text{setup}_{i,j} \quad (1 \leq i \leq N_{\text{FF}}, 1 \leq j \leq N_{\text{CKT}})$
  - $B_{\text{TEP}_{i,j}} \quad (1 \leq i \leq N_{\text{FF}}, 1 \leq j \leq N_{\text{CKT}})$

The inputs of this problem are N_{CKT} pre-ASA candidates, and the output is one EP-AVS circuit in which ASA is applied and TEP-FFs are inserted. The objective of this problem is to minimize the power of the EP-AVS circuit. The EP-AVS circuit is constrained by MTTF ($\text{MTTF}_{\min}$), circuit area ($\text{Area}_{\text{ASA}}^{\max}$ and $\text{Area}_{\text{TEP}}^{\max}$), and the number of low-Vth cells ($\text{NLvth}_{\text{ASA}}^{\max}$). The variables $\Delta\text{setup}_{i,j}$ are the slacks given to FFs in j-th pre-ASA circuit by ASA, where $\Delta\text{setup}_{i,j}$ is given to the layout ECO as an intentional increase in setup time of i-th FF_i in j-th pre-ASA circuit. N_{FF} is the number of FFs in the circuit, and it is identical in all the pre-ASA circuits. When $\Delta\text{setup}_{i,j} = 0$, i-th FF_i is not included in the set of target FFs for ASA of j-th pre-ASA circuit. Thus, the number of target FFs in j-th pre-ASA circuit is expressed as the number of FFs whose $\Delta\text{setup}_{i,j}$ is larger than 0. $B_{\text{TEP}_{i,j}}$ is a binary variable, and it becomes 1 when i-th FF in j-th circuit is replaced to TEP-FF. Therefore, the number of TEP-FF in j-th pre-ASA circuit is expressed as the number of FFs whose $B_{\text{TEP}_{i,j}}$ equals to 1. Here, MTTF_j depends on $\Delta\text{setup}_{i,j}$ and $B_{\text{TEP}_{i,j}}$ and these relations are evaluated by
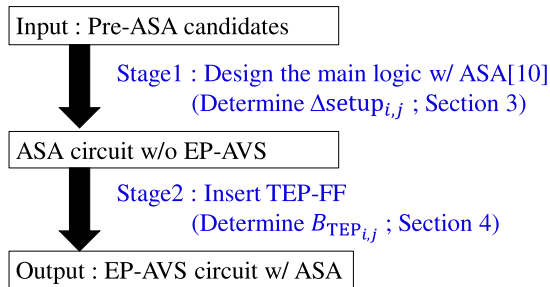
Fig. 5   Overview of the proposed design. Proposed design methodology with a two-stage procedure; (1) Design the main logic under AVS with ASA [10], (2) Insert TEP-FF.

1.   Increase FF setup time by Δsetup_i (e.g., 50 ps)
     + re-layout
2.   Restore setup time

Fig. 6   An example of FF-based ASA. ©[2018] IEEE. Reprinted, with permission, from [10].

the stochastic error rate estimation method [11]. $\text{Area}_{\text{ASA}_j}$ and $\text{NLvth}_{\text{ASA}_j}$ vary depending on $\Delta\text{setup}_{i,j}$, and $\text{Area}_{\text{TEP}_j}$ is determined by $B_{\text{TEP}_{i,j}}$.

2.3   Overview of Proposed EP-AVS

A difficulty to solve the formulated problem is the non-linear relations among MTTF_j, $\text{Area}_{\text{ASA}_j}$, $\text{Area}_{\text{TEP}_j}$, $\text{NLvth}_{\text{ASA}_j}$, $B_{\text{TEP}_{i,j}}$, and $\Delta\text{setup}_{i,j}$. Also, the evaluations of MTTF_j, $\text{Area}_{\text{ASA}_j}$, $\text{Area}_{\text{TEP}_j}$, and $\text{NLvth}_{\text{ASA}_j}$ need relatively long CPU time, and hence an explicit optimization is difficult concerning CPU time. Thus, to determine the set of $\Delta\text{setup}_{i,j}$ and $B_{\text{TEP}_{i,j}}$ efficiently, we propose a two-step procedure.

Figure 5 shows the overview of the proposed design which includes both the main logic design and sensor insertion. The proposed design methodology solves this problem with the two-stage procedure. The first stage designs the main logic under AVS using ASA [10], i.e., determines $\Delta\text{setup}_{i,j}$, and the second stage performs TEP-FF insertion, i.e., determines $B_{\text{TEP}_{i,j}}$. The following sections explain these two stages.

3.   First Stage: ASA for Main Logic

In the first stage, the ASA is performed for the main logic. Note that the design methodology of the ASA circuit is identical to [10]. This subsection briefly explains the design methodology of the ASA circuit. For the detail, please see [10].

The ASA consists of two main procedures. The first procedure prepares several pre-ASA design candidates laid out with different timing constraints, screens pre-ASA candidates using the trade-off analysis between MTTF and power, and identifies the most promising candidate that is expected to achieve the lowest power operation after ASA. Note that this candidate is given to the second stage. An important consideration in the first procedure is how to design the pre-ASA circuit to obtain the better ASA circuit. For example, [13] prepares a pre-ASA circuit that is designed at the maximum operating frequency (FMAX) and performs ASA. This pre-ASA circuit tends to include more low-Vth cells and larger cells and consequently increases dynamic and static power. On the other hand, the circuit designed at
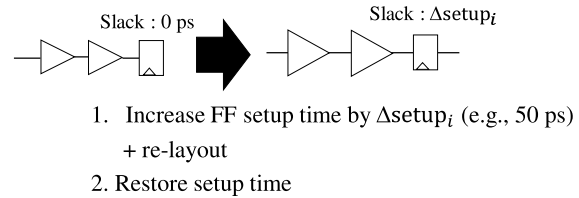
a looser frequency may be flexible to accept an additional design change in ECO compared to the FMAX design, and hence ASA may provide better optimization results. Based on the above consideration, [10] selects the most promising pre-ASA circuit with the following three steps; (1) finding the MTTF-dominant FF after ASA, (2) calculating the minimum supply voltage after ASA ($V_{\min}$), and (3) estimating the minimum power after ASA. After this candidate selection, the circuit parameter of j is fixed, and the following second step of ASA implementation will determine $\Delta\text{setup}_{i,j}$.

Then, the set of Δsetup_i, i.e., target FFs and setup slack of these FFs, are determined in the second procedure. For target FFs determination, the ASA method focuses on a gate-wise failure probability for reducing the timing failure probability and consequently improving the MTTF. Note that the gate-wise failure probability denotes how much each instance contributes to the timing error occurring in the circuit. The ASA methodology in [10] first distributes the failure probability from endpoint FF to instances at the upper stream of the FF as gate-wise failure probability. Then, this method selects target FFs by solving the covering problem of instances weighted with the gate-wise failure probability as an integer linear programming (ILP) problem. After the set of target FFs are determined, the ASA gives timing slacks for each target FF so that the failure probability of each FF is equal to or smaller than the target failure probability.

Once the set of Δsetup_i is determined, the methodology in [10] performs the FF-based ASA with two steps as shown in Fig. 6; (1) increase setup time of the target i-th FF by Δsetup_i artificially and re-layout the design as an engineering change order (ECO) process, and (2) restore the setup time for the successive analysis process. With this FF-based ASA, we enforce the paths ending at the target FF to have the slack of more than Δsetup_i.

4.   Second Stage: Sensing Circuit Insertion

In the second stage, the sensor circuit is inserted into the ASA circuit. Our sensor insertion methodology proposed in this paper considers the failure probability, which is the joint probability of the timing violation probability and the activation probability. Note that if the inserted sensors are not activated frequently enough, AVS scarcely checks the timing slack of critical paths and thus cannot adjust supply voltage appropriately. On the other hand, the failure probability driven sensor insertion helps AVS to monitor critical paths frequently, which improves the MTTF and thus contributes
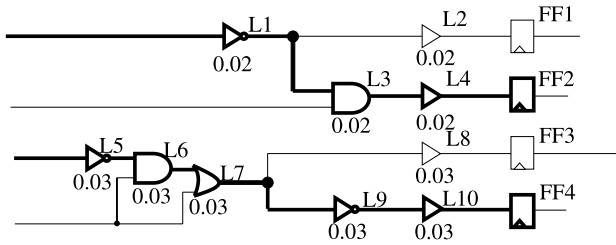
Fig. 7 FF selection strategy. ©[2018] IEEE. Reprinted, with permission, from [10].

Fig. 8 An example of failure probability calculation.

to reducing the power dissipation. Moreover, we develop a FF selection methodology that maximizes the sum of gate-wise timing failure probabilities aiming at further MTTF improvement. When we insert TEP-FFs to such FFs, inserted TEP-FFs are expected to monitor a set of gates which are likely to cause timing errors and they prevent timing errors occurring due to the delay variation in the monitoring gates. Consequently, the proposed methodology efficiently reduces the timing failure probability and extends MTTF.

Figure 7 shows a simple example, where the circuit is composed of ten combinational logic cells and four FFs. The numbers attached to each gate are the gate-wise failure probabilities, where their values are computed from the FF failure probabilities. The detailed computation will be explained later. Given the number of FFs to which ASA is applied (N_{TEP}) is 2, we can see that the most promising pairs of FFs are FF2 and FF4. When the slack times of FF2 and FF4 are increased, the slack times of L1, L3, L4, L5, L6, L7, L9, and L10 are also increased, and the expected probability of error reduction corresponds to the sum of gate-wise failure probabilities and it is 0.21 (= 0.02 + 0.02 + 0.02 + 0.03 + 0.03 + 0.03 + 0.03 + 0.03). We note that if we choose FF3 and FF4, i.e., descending order of the sum of gate-wise failure probability, the slack times of L5, L6, L7, L8, L9, and L10 are increased. In this case, the reduced failure probability is 0.18 (= 0.03 × 6), and this amount of reduction is smaller than the previous one.

Our insertion method consists of the following two steps; (1) calculating gate-wise failure probabilities, and (2) finding out a set of FFs that maximally reduces the sum of gate-wise failure probability by solving instance covering problem as an ILP problem. In the first step, our method calculates the timing failure probability of FF by multiplying the timing violation probability and the activation probability, which is illustrated in Fig. 8. The timing violation probability can be calculated by performing SSTA. The activation probability of each path is derived by associating the signal transition time in logic simulation and the path delay in STA. Note that logic simulation is just one way to obtain the activation probability and other strategies, for example, setting the transition density of the primary input between 0 and 1 and calculating the activation probability of the internal gates, can be used. Then, we compute the gate-wise failure probabilities from the failure probabilities of FFs, i.e. $P_{\text{fail\_inst}_k}$, from $P_{\text{fail\_FF}_i}$, as follows. Note that $P_{\text{fail\_inst}_k}$ will
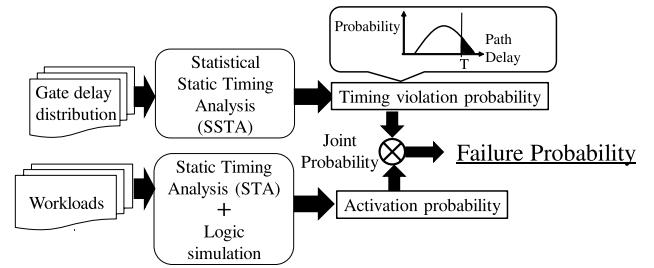
be utilized in the instance covering problem (or FF selection problem) at the second step.

$$P_{\text{fail\_inst}_k} = \max_{1 \le i \le N_{FF}} \left\{ \frac{P_{\text{fail\_FF}_i}}{\sum_{k=1}^{k_{\max}} (B_{\text{FF}_i\_\text{inst}_k})} \right\}. \tag{1}$$

The above equation assumes that individual instances included in the fan-in cone of i-th FF have the same contribution to the timing error at the FF, and hence the $P_{\text{fail\_FF}_i}$ is divided by the number of instances in the fan-in cone of i-th FF. When we need to consider the different contributions of each instance due to, for example, different intrinsic variation sensitivities of the instances themselves, we may distribute $P_{\text{fail\_FF}_i}$ to each gate wise failure probability taking into account the different sensitivities. We also note that an instance can be included in the fan-in cones of multiple FFs. For coping with this, the max operation is performed in Eq. (1).

In the second step, we propose the FF selection methodology that maximizes the sum of gate-wise timing failure probabilities. We formulate this FF selection problem as an ILP problem to derive the exact solution. Our ILP formulation is as follows:

- Input
    - one ASA circuit
- Output
    - one EP-AVS circuit
- Objective
    - Maximize : $\sum_{k=1}^{N_{\text{inst}}} (P_{\text{fail\_inst}_k} \times B_{\text{inst}_k})$
- Constraints
    - $B_{\text{inst}_k} \in \{0, 1\} \quad (1 \le k \le N_{\text{inst}})$
    - $B_{\text{TEP}_i} \in \{0, 1\} \quad (1 \le i \le N_{FF})$
    - $\sum_{i=1}^{N_{FF}} B_{\text{TEP}_i} \le N_{\text{TEP}}^{\max}$
    - $B_{\text{inst}_k} \le \sum_{i=1}^{N_{FF}} (B_{\text{TEP}_i} \times B_{\text{FF}_i\_\text{inst}_k})$
- Variables
    - $B_{\text{TEP}_i} \quad (1 \le i \le N_{FF})$

The input of this problem is the ASA circuit designed and selected in Sect. 3 and the output is one EP-AVS circuit. The number of instances in the circuit is N_{inst}. The objective of this ILP problem is to maximize the sum of ($P_{\text{fail\_inst}_k} \times$

B_{inst_k}), where $P_{\text{fail\_inst}_k}$ is the gate-wise failure probability, and it means how much k-th instance contributes to the timing error. B_{inst_k} is a binary variable, and it becomes 1 when k-th instance is included in paths ending at target FFs for TEP-FF insertion. Therefore, the sum of $P_{\text{fail\_inst}_k} \times B_{\text{inst}_k}$ represents the gate-wise failure probability reduction. In this problem, we assign binary variables B_{TEP_i}, where B_{TEP_i} becomes 1 when i-th FF is selected to target FFs for TEP-FF insertion.

The first and second constraints are given to restrict B_{inst_k} and B_{TEP_i} to binary numbers. The third constraint means that the number of target FFs for TEP-FF insertion should be equal or less than $N_{\text{TEP}}^{\max}$ and this constrains the area overhead due to TEP-FF insertion. The fourth constraint is a key constraint that defines the relation between B_{inst_k} and B_{TEP_i}. $B_{\text{FF}_i\_\text{inst}_k}$ is a binary constant which is determined by the circuit topology, and it becomes 1 when k-th instance is included in the paths ending at i-th FF. The product term of $B_{\text{TEP}_i} \times B_{\text{FF}_i\_\text{inst}_k}$ becomes 1 when both B_{TEP_i} and $B_{\text{FF}_i\_\text{inst}_k}$ are 1. B_{inst_k} becomes 0 only when the product of B_{TEP_i} and $B_{\text{FF}_i\_\text{inst}_k}$ is 0 for all the FFs. On the other hand, if k-th instance is included in the paths ending at target FFs, at least one of the products of B_{TEP_i} and $B_{\text{FF}_i\_\text{inst}_k}$ become 1. In this case, B_{inst_k} can be 1. In this ILP formulation, we are maximizing the sum of ($P_{\text{fail\_inst}_k} \times B_{\text{inst}_k}$) and hence B_{inst_k} is necessarily assigned to be 1.

We note that ILP has proven to be NP-hard [14] in general and thus the ILP may not be suitable for large-scale optimization problems due to computational cost. When the circuit size becomes larger and the CPU time is unacceptable, we need to, for example, find an approximate solution or partition the circuit into sub-circuits for problem size reduction.

5. Experimental Evaluation

This section experimentally evaluates the performance improvement from the conventional WC design to the proposed EP-AVS. Section 5.1 explains the evaluation setup. Section 5.2 shows the performance improvement results regarding power saving effects and demonstrates that the proposed EP-AVS achieves the lower supply voltage compared with two TEP-FF insertion methodologies which insert TEP-FF with (1) ascending order of FF setup slack and (2) descending order of FF failure probability.

5.1 Experimental Setup

In this work, we used the advanced encryption standard (AES) circuit and OR1200 OpenRISC processor, which is a 32-bit RISC microprocessor with five pipeline stages, as target circuits. These two circuits were designed by a commercial P&R tool with a 45 nm Nangate standard cell library. Also, standard cell memories [15] were used as external main memories in OpenRISC processor. The minimum clock period of post-layout circuits at 1.20 V in the typical PVTA conditions and the worst-case are 3,150 ps and 4,260 ps in Open-

RISC and 370 ps and 480 ps in AES, respectively. Hereafter, the target clock period is set to 4,260 ps in OpenRISC and 480 ps in AES, and then ASA optimizes the timing slack of FF/path for these target clock periods. The post-layout circuits include 23,247 combinational logic cells, 2,504 FFs, two macro cells of standard cell memory in OpenRISC, and 17,948 combinational logic cells and 530 FFs in AES, respectively. Thus, sets of N_{inst} and N_{FF} are 23,249 and 2,504 in OpenRISC, 17,948 and 530 in AES, respectively.

We used Gurobi Optimizer 7.0 to solve the ILP problem defined in Sect. 4. The solver was executed on a 2.4 GHz Xeon CPU machine under the Red Hat Enterprise Linux 6 operating system with 1 TB memory. The required CPU times were at most 7.13 seconds in AES and 1.46 seconds in OpenRISC. For calculating meaningful MTTF, practical delay variations should be considered. Our evaluation took into account the following variations.

- Dynamic supply noise, which is assumed to fluctuate between -50 mV and 50 mV by 10 mV with eleven steps.
- Manufacturing variability, which is assumed to consist of the intra-die random variation and inter-die variation. Both the intra-die random variation and inter-die variation include NMOS and PMOS threshold voltage variation of $\sigma = 30$ mV and gate length variation of $\sigma = 2$ nm.
- NBTI aging, whose model was obtained by fitting a trapping/de-trapping model [16] to the measured data in [17]. Six degradation states of 0 mV, 0.5 mV, 1 mV, 5 mV, 10 mV and 15 mV are prepared. Note that we used the NBTI degradation data with stress probability of 100% in [17] as a worst case. Therefore, our experimental setup gives the most pessimistic MTTF regarding NBTI effects. Activation probability aware analysis and optimization are included in our future work.

We performed SSTA with the following processes. First, we generated probability density functions of gate delay variability according to the assumed variations. Second, we executed sensitivity-based SSTA (e.g. [18] and [19]) with common path pessimism removal (e.g. [20] and [21]) to obtain the canonical-form expression of the timing violation probability. Third, we calculated the timing violation probability by integrating the canonical-form expression with MATLAB 2017a.

As for the workload in OpenRISC, we selected three benchmark programs (CRC32, SHA1, and Dijkstra) from MIBenchmark [22]. For each program, 30 sets of input data were prepared for MTTF estimation. Totally, we used 90 (= 3 × 30) workloads. In AES, 1,000 random test patterns were used. Figure 9 shows the distributions of activation probability in AES and OpenRISC. We can see that Open-RISC is less activated and the activation probability is widely spread, which suggests the ASA is more effective to Open-RISC. Note that in AES case, to decide the number of test patterns, we swept the number of test patterns from 1,000
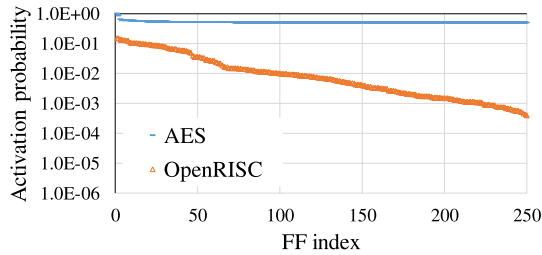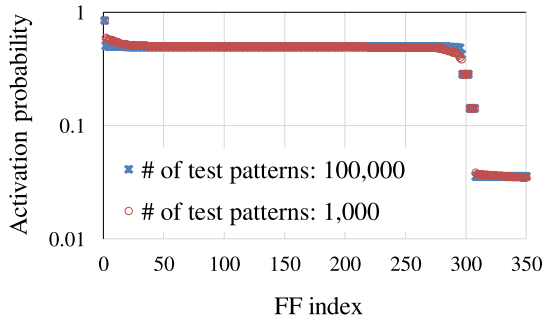
Fig. 9 Activation probability of FF.

Fig. 10 Activation probability comparison in AES. The activation probability of each FF is calculated with different numbers of test patterns.

Fig. 11 Expected minimum power after ASA in AES.

to 100,000 and compared the activation probability of each FF. Figure 10 shows the comparison results of the activation probability of each FF between two cases; (1) the number of random test patterns equals to 1,000 and (2) the number of test patterns equals to 100,000. From Fig. 10, we observed that the activation probability of each FF is similar between these two cases. Therefore, in our experiment, we set the number of test patterns to 1,000 and calculated the activation probabilities, which is included in Fig. 9. On the other hand, we think that ensuring the necessary and sufficient test patterns for the target coverage is a challenging task, and from this point of view, we have a room for improvement for test pattern selection.

We set MTTF of 1.00×10^{17} cycles, i.e., 10 years in OpenRISC and 1.6 years in AES, as $\mathrm{MTTF_{min}}$. Note that the above $\mathrm{MTTF_{min}}$ is just an example, and the proposed design methodology can cope with other constraints of $\mathrm{MTTF_{min}}$ similarly. We prepared seven supply voltages, i.e., from 1.20 V to 0.90 V with 50 mV interval, and swept clock period from 450 ps to 500 ps in AES and from 4,000 ps to 6,000 ps in OpenRISC. Note that, at each clock period, EP-AVS dynamically adjusted the supply voltage within the range from 1.20 V to 0.90 V.

With this setup, we performed ASA to both AES and OpenRISC. The number of pre-ASA candidate circuits was seven in AES, where P&R clock periods of these pre-ASA circuits were 370, 380, 400, 420, 440, 460, and 480 ps. As for OpenRISC, seven candidates with 3,150, 3,200, 3,400, 3,600, 3,800, 4,000, and 4,200 ps were given. Figure 11 shows the estimation results of the expected minimum power after ASA for each pre-ASA candidate in AES. From Fig. 11, we can see that the pre-ASA candidate designed at 460 ps is
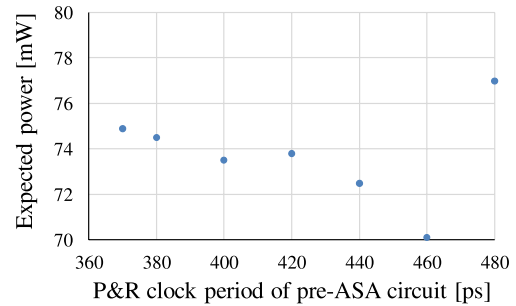
the most promising one regarding power. We then selected the pre-ASA circuit that was laid out at 460 ps. Similar to AES, we evaluated the expected minimum power of OpenRISC candidate circuits and selected the circuit laid out at 4,000 ps. Next, we performed ASA to the chosen candidates. The constraints for overhead of the area and the number of low-Vth cells by ASA were set to 0.7% and 0.0% for OpenRISC and the maximum number of target FFs for ASA, i.e., $N_{\mathrm{ASA}}^{\mathrm{max}}$ was set to 255. Similarly, we set $N_{\mathrm{ASA}}^{\mathrm{max}}$ to 255 in AES. Note that in AES, the ASA circuit achieved the smaller circuit area compared with the pre-ASA circuit design at 370 ps, which will be discussed in Sect. 5.2.2.

Next, we inserted several TEP-FFs to the voltage-scaled circuits. The constraint of area overhead for TEP-FF was set to 0.8% for both AES and OpenRISC, and the number of maximum TEP-FF ($N_{\mathrm{TEP}}^{\mathrm{max}}$) was set to 5 in AES and 20 in OpenRISC, respectively. When inserting TEP-FF, we need to determine the number of delay buffers for each TEP-FF. In this work, we inserted the delay buffers whose delay were comparable to the delay variation caused by 50 mV supply noise, where this number of 50 mV corresponds to one level decrement of the supply voltage.

The MTTF and average supply voltage under PVTA variation were evaluated by the stochastic MTTF estimation framework [11]. In our experiment, the monitor period for EP-AVS was swept from 10^6 to 10^{13} clock cycles. Here, the monitor period of 10^6 cycles means, if no error prediction signals are outputted for 10^6 cycles, the supply voltage is decreased. The minimum monitor period, i.e., 10^6 cycles, is about 3 ms in OpenRISC and 0.5 ms in AES, respectively, and it is longer than the response time of the fast transient voltage regulator, e.g., 1.6 μs in [23].

5.2 Evaluation Results

This subsection first shows power savings thanks to the proposed EP-AVS, and then examines the effectiveness of ASA and TEP-FF insertion methodology.

5.2.1 Power Saving Effects

Figure 12 shows the trade-off curves between the minimum average power and the clock cycle under the MTTF constraint of 10^{17} cycles, where (a) in OpenRISC and (b) in
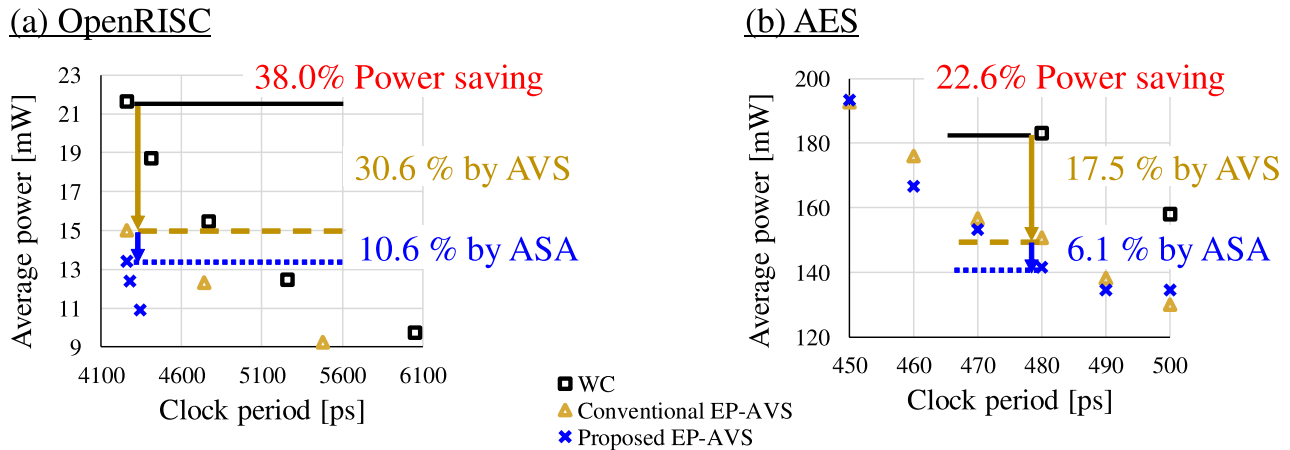
Fig. 12 Trade-off curves between clock period and average power. (a) OpenRISC, (b) AES.

AES, respectively. The black square plots represent the conventional WC design with guard-banding for PVTA variation. The yellow triangular and blue cross plots correspond to the conventional EP-AVS which optimizes only the sensing circuit, and the proposed EP-AVS which optimizes both the main logic under AVS and sensing circuit, respectively. Here, the TEP-FFs in the conventional EP-AVS were inserted by the method in Sect. 4. In this section, we examine our evaluation results from the following two aspects; (1) overall power saving effect thanks to the proposed EP-AVS, and (2) difference of the power dissipation between the proposed and conventional EP-AVS.

First, we compare the black square and blue cross plots for clarifying the overall performance improvement thanks to the proposed EP-AVS. Figure 12 shows that the proposed EP-AVS saves average power while keeping the target MTTF. For example, in Fig. 12(a), at a clock period of 4,260 ps, the proposed EP-AVS achieved the target MTTF with an average power of 13.4 mW, whereas the conventional WC design required 21.6 mW. In other words, EP-AVS achieved 38.0% power savings from 21.6 mW to 13.4 mW. Similarly, in Fig. 12(b), at a clock period of 480 ps, the proposed EP-AVS achieved 22.6% power savings from 183.0 mW to 141.5 mW. We experimentally confirmed that the proposed EP-AVS made the significant power savings both in AES and OpenRISC. The proposed EP-AVS increases the circuit area by 1.5% in OpenRISC and decreases the area by 5.4% in AES.

Next, we compared the conventional EP-AVS and proposed EP-AVS, i.e., yellow triangular and blue cross plots. Figure 12 shows that the proposed EP-AVS further improves power dissipation from the conventional EP-AVS. For example, the proposed EP-AVS achieved 10.6% power savings from 15.0 mW to 13.4 mW at a clock period of 4,260 ps in OpenRISC and 6.1% power savings from 150.8 mW to 141.5 mW at a clock period of 480 ps in AES. These power savings reveal that the ASA for the main logic works well and the simultaneous optimization of the main logic under AVS and the sensing circuit enhance the efficacy of EP-AVS. We also observe that the performance improvement thanks to
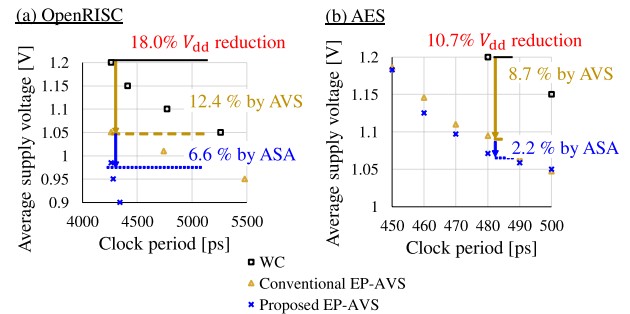

Fig. 13 V_{dd} reduction by the proposed EP-AVS. (a) OpenRISC, (b) AES.

ASA is the largest around the target clock periods of 4,260 ps in OpenRISC and 480 ps in AES and it becomes smaller as the period goes away from the target one since ASA optimized the circuit at the target clock period under MTTF constraint, which was also reported in [10]. There could be a room for improvement at different clock periods.

5.2.2 Effectiveness of ASA

The performance evaluation results in Sect. 5.2.1 showed that the proposed design saved power significantly. Let us investigate the results in detail.

Firstly, we examine the power saving effects by the main logic optimization with ASA in terms of V_{dd} and area. Figure 13 shows the trade-off curves between the average supply voltage and the clock period under the MTTF constraints of 10 years in OpenRISC and 1.6 years in AES. We can see that the proposed design, which corresponds to the blue cross plots, achieves the target MTTF at a lower supply voltage compared with the conventional EP-AVS, i.e., yellow triangle plots. For example, in Fig. 13(a), at a clock period of 4,260 ps, the proposed design achieves the target MTTF at an average supply voltage of 0.98 V, whereas the conventional EP-AVS design requires 1.05 V operation, which means the proposed design achieves 6.6% V_{dd} reduction from 1.05 V to 0.98 V. As for AES, the proposed EP-AVS reduces the supply voltage from 1.09 V to 1.07 V and achieves 2.2% V_{dd}
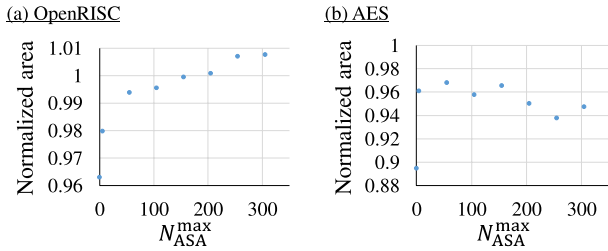
Fig. 14 Area of ASA circuit. (a) OpenRISC, (b) AES. Y-axis is normalized by the area of pre-ASA circuit laid out at FMAX.

Fig. 15 MTTF comparison in OpenRISC.

reduction as shown in Fig. 13(b). Thanks to these V_{dd} reductions, the circuit power dissipation is dramatically reduced as shown in Fig. 12.

Figure 14 shows the area of the ASA circuits. In this figure, the area is normalized by that of the pre-ASA circuit laid out at FMAX. Figure 14 shows that the ASA saves the area overhead or, in some cases, even reduces the area from the pre-ASA circuit thanks to the pre-ASA circuit selection and Δsetup$_{i,j}$ adjustment. For example, in Fig. 14(b), at N_{ASA}^{max} = 255, the ASA reduces the area by 6.2%. This reduction directly decreases the dynamic and static power dissipation. Thus, we experimentally confirmed that the ASA for the main logic is highly compatible with EP-AVS, and they mutually enhance the performance with margin elimination.

5.2.3 Effectiveness of the Proposed TEP-FF Insertion

Next, we evaluate the effectiveness of the proposed TEP-FF insertion methodology that takes into account the gate-wise failure probability. We compare the proposed methodology with the following two approaches.

C1: Choose the insertion locations according to the ascending order of FF setup slacks.

C2: Choose the insertion locations according to the descending order of the FF failure probability.

The first approach of C1 is the conventional TEP-FF insertion method, e.g., [7]–[9], that insert the sensors to critical paths. This approach needs only STA (or SSTA) timing reports and hence this approach is easier to adopt. The second approach of C2 places importance on not only the timing violation probability but also the activation probability. Even if some FFs are timing critical, timing errors never occur as long as they are not activated. To calculate the activation probability, we need to perform the logic simulation with prospective workloads or to calculate signal transition rates mathematically (e.g. [24]).

Figure 15 shows the comparison results of MTTF in OpenRISC. We can see that the proposed method and the C2 achieved much longer MTTF than the conventional C1. More seriously, the conventional C1 cannot satisfy the given MTTF constraint at all. Note that in AES, all the proposed, the C1, and the C2 satisfied the target MTTF since the AES is the highly activated circuit as previously explained with Fig. 9.
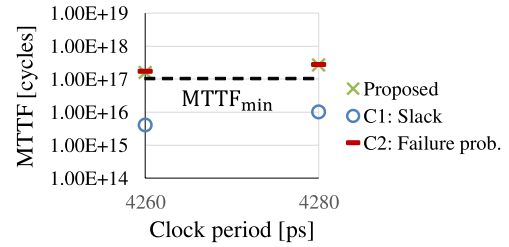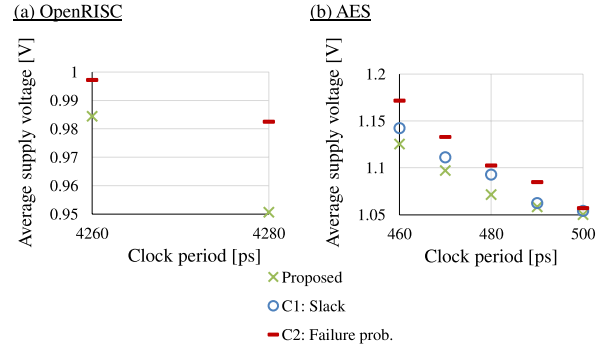

Fig. 16 Comparison of the average supply voltage. (a) OpenRISC, (b) AES.

Figure 16 shows the comparison results of the average supply voltage in (a) OpenRISC and (b) AES. Note that in OpenRISC, the average supply voltages of the proposed and the C2 are compared since the conventional C1 could not satisfy the MTTF constraint. We can see that the proposed method achieves the lower average supply voltage with satisfying the MTTF constraint. For example, in Fig. 16(a), at the clock period of 4,260 ps, the proposed methodology achieves the average supply voltage of 0.984 V, whereas C2 requires 0.997 V. Similarly, in Figure 16(b), at the clock period of 480 ps, the proposed methodology achieves 1.071 V, whereas C1 and C2 require 1.093 V and 1.102 V, respectively. From the above, we experimentally confirm that the maximizing the sum of the gate-wise failure probability better exploits the timing margins to the supply voltage reduction.

Lastly, we compared the performance between the proposed design and the ASA circuit without implementing AVS. Here, it should be noted that the ASA in [10] assumes the chip-wise voltage assignment can be performed, i.e. the supply voltage can be set for each chip individually for power minimization, which is a highly preferable situation for [10]. However, the assumed chip-wise voltage assignment is expensive for most of the products since the post-silicon delay test with LSI tester is necessary for individual chips. On the other hand, in our AVS, the supply voltage is automatically adjusted with the sensing circuit without LSI tester, which is a distinct difference with [10]. Figure 17 shows the trade-off comparison results between the clock period and the average supply voltage (a) OpenRISC and (b) AES. From Fig. 17, we observed that the proposed design achieves the similar performance with the ASA in [10], which means that the
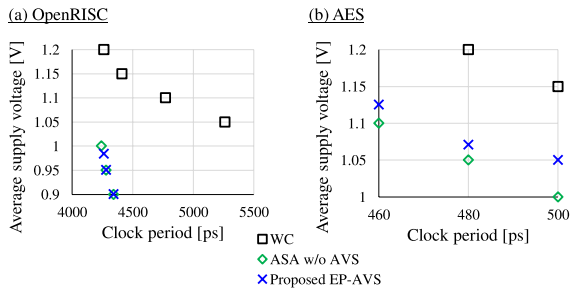
Fig. 17 Trade-off comparison between the proposed design and the ASA without AVS. (a) OpenRISC, (b) AES.

supply voltage control with the sensing circuit works well.

6. Conclusion

This paper focused on EP-AVS and proposed a design methodology for EP-AVS circuits. The proposed design methodology optimizes both the main logic under AVS and sensing circuits taking into account the timing failure probabilities of FFs. The quantitative MTTF and power evaluation results showed that the proposed EP-AVS design methodology achieved 38.0% power saving while satisfying target MTTF thanks to the ASA and failure probability based TEP-FF insertion.

Acknowledgments

References

[1] B. Zhang and M. Orshansky, "Modeling of NBTI-induced PMOS degradation under arbitrary dynamic temperature variation," Proc. International Symposium on Quality Electronic Design, pp.774–779, March 2008.

[2] T. Wang and Q. Xu, "On the simulation of NBTI-induced performance degradation considering arbitrary temperature and voltage variations," Proc. Design Automation Conference, pp.1–6, June 2014.

[3] 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, April 2006.

[4] K.A. Bowman, J.W. Tschanz, S.L. Lu, P.A. Aseron, M.M. Khellah, A. Raychowdhury, B.M. Geuskens, C. Tokunaga, C.B. Wilkerson, T. Karnik, and V.K. De, "A 45 nm resilient microprocessor core for dynamic variation tolerance," IEEE J Solid-State Circuits, vol.46, no.1, pp.194–208, Jan. 2011.

[5] M. Fojtik, D. Fick, Y. Kim, N. Pinckney, D.M. Harris, D. Blaauw, and D. Sylvester, "Bubble razor: Eliminating timing margins in an ARM Cortex-M3 processor in 45 nm CMOS using architecturally independent error detection and correction," IEEE J. Solid-State Circuits, vol.48, no.1, pp.66–81, Jan. 2013.

[6] S. Kim and M. Seok, "Variation-tolerant, ultra-low-voltage microprocessor with a low-overhead, within-a-cycle in-situ timing-error detection and correction technique," IEEE J. Solid-State Circuits, vol.50, no.6, pp.1478–1490, June 2015.

[7] A. Benhassain, F. Cacho, V. Huard, M. Saliva, L. Anghel, C. Parthasarathy, A. Jain, and F. Giner, "Timing in-situ monitors: Implementation strategy and applications results," Proc. Custom Integrated Circuits Conference, pp.1–4, 2015.

[8] T. Sato and Y. Kunitake, "A simple flip-flop circuit for typical-case designs for DFM," Proc. International Symposium on Quality Electronic Design, pp.539–544, 2007.

[9] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error predictive sensors for subthreshold circuits," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.20, no.2, pp.333–343, Feb. 2012.

[10] Y. Masuda, T. Onoye, and M. Hashimoto, "Activation-aware slack assignment for time-to-failure extension and power saving," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol.26, no.11, pp.2217–2229, Nov. 2018.

[11] S. Iizuka, Y. Masuda, M. Hashimoto, and T. Onoye, "Stochastic timing error rate estimation under process and temporal variations," Proc. International Test Conference, pp.1–10, Oct. 2015.

[12] Y. Masuda and M. Hashimoto, "MTTF-aware design methodology of error prediction based adaptively voltage-scaled circuits," Proc. Asia and South Pacific Design Automation Conference, pp.159–165, Jan. 2018.

[13] Y. Masuda, M. Hashimoto, and T. Onoye, "Critical path isolation for time-to-failure extension and lower voltage operation," Proc. International Conference on Computer-Aided Design, Nov. 2016.

[14] P. Cappello and K. Steiglitz, "Some complexity issues in digital signal processing," IEEE Trans. Acoust., Speech, Signal Process., vol.ASSP-32, no.5, pp.1037–1041, Oct. 1984.

[15] A. Teman, D. Rossi, P. Meinerzhagen, L. Benini, and A. Burg, "Controlled placement of standard cell memory arrays for high density and low power in 28 nm FD-SOI," Proc. Asia and South Pacific Design Automation Conference, pp.81–86, Jan. 2015.

[16] J.B. Velamala, K.B. Sutaria, H. Shimizu, H. Awano, T. Sato, G. Wirth, and Y. Cao, "Compact modeling of statistical BTI under trapping/detrapping," IEEE Trans. Electron Devices, vol.60, no.11, pp.3645–3654, Nov. 2013.

[17] H. Awano, M. Hiromoto, and T. Sato, "Variability in device degradations: Statistical observation of NBTI for 3996 transistors," Proc. European Solid State Device Research Conference, pp.218–221, Sept. 2014.

[18] H. Chang and S.S. Sapatnekar, "Statistical timing analysis under spatial correlations," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol.24, no.9, pp.1467–1482, Sept. 2005.

[19] C. Visweswariah, K. Ravindran, K. Kalafala, S.G. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," Proc. Design Automation Conference, pp.331–336, July 2004.

[20] V. Garg, "Common path pessimism removal: An industry perspective: Special session: Common path pessimism removal," Proc. International Conference on Computer-Aided Design, pp.592–595, Nov. 2014.

[21] J. Zejda and P. Frain, "General framework for removal of clock network pessimism," Proc. International Conference on Computer-Aided Design, pp.632–639, Nov. 2002.

[22] M.R. Guthaus, J.S. Ringenberg, D. Ernst, T M. Austin, T. Mudge, and R.B. Brown, "MiBench: A free, commercially representative embedded benchmark suite," Proc. International Workshop on Workload Characterization, pp.3–14, Dec. 2001.

[23] Y. Li, X. Zhang, Z. Zhang, and Y. Lian, "A 0.45-to-1.2-V fully digital low-dropout voltage regulator with fast-transient controller for near/subthreshold circuits," IEEE Trans. Power Electron., vol.31, no.9, pp.6341–6350, Sept. 2016.

[24] F.N. Najm, "Transition density: A new measure of activity in digital circuits," IEEE Trans. CAD, vol.12, no.2, pp.310–323, Feb. 1993.

Yutaka Masuda received the B.E., M.E., and Ph.D. degrees in Information Systems Engineering from the Osaka University, Osaka, Japan, in 2014, 2016, and 2019, respectively. He is currently an Assistant Professor in Center for Embedded Computing Systems, Graduate School of Informatics, Nagoya University. His research interests include low-power circuit design. He is a member of IEEE, IEICE, and IPSJ.

Masanori Hashimoto 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 2016, he has been a Professor in Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University. His research interest includes computer-aided design for digital integrated circuits, and high speed circuit design. Dr. Hashimoto served on the technical program committees for international conferences including DAC, ICCAD, ITC, Symposium on VLSI Circuits, ASP-DAC, DATE, ISPD and ICCD. He is a member of IEEE, ACM, IEICE and IPSJ.