In-Field Aging Measurement and Calibration for Power-Performance Optimization

Shuo Wang[1], Mohammad Tehranipoor[1], and LeRoy Winemberg[2]
[1]Dept of ECE, University of Connecticut, {shuo.wang, tehrani}@engr.uconn.edu
[2]Freescale Semiconductor, leroyw@freescale.com

ABSTRACT

Aging of transistors has become a major reliability concern especially when the VLSI circuits are in the nanometer regime. In this paper, we propose a novel methodology to address circuit aging in the field. On-chip aging sensor is designed to monitor transitions on functional paths capturing functional mode workload. Path delay is then accurately measured and converted to a digital value. Diagnosis and calibration are performed in the field, thereby achieving power-performance optimization throughout the entire lifetime. Simulation results demonstrate the efficiency of the proposed structure.

Categories and Subject Descriptors

B.8.0 [**PERFORMANCE AND RELIABILITY**]: General

General Terms

Design, Performance, Reliability, Measurement

Keywords

Aging, On-chip measurement, Performance calibration, Path delay measurement, Power-performance optimization

1. INTRODUCTION

As technology scaling continues, VLSI circuits are facing challenge of reliability degradation. This gets exacerbated when various physical factors have become increasingly significant in the nanometer regime [1, 2]. In particular, aging mechanisms, such as negative/positive bias temperature instability (NBTI/PBTI) [12, 13], hot carrier injection (HCI) [14] and time-dependent dielectric breakdown (TDDB), cause parametric shifts and eventually device failure.

Traditionally, one-time worst-case guardbands, e.g., clock frequency reduction, supply voltage increase, and gate sizing, are added to address the aging issues. For example, in [10], characteristics of each PMOS transistor, and hence each gate, are modeled considering NBTI aging. Then, synthesis is executed using NBTI-aware standard cells. However, worst-case guardbands are over-pessimistic and inefficient [3]. It is necessary to develop efficient solutions that can monitor the aging process and react dynamically.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2011, June 5-10, 2011, San Diego, California, USA.
Copyright 2011 ACM ACM 978-1-4503-0636-2/11/06 ...$10.00.

To monitor/measure aging, many existing solutions are based on online aging sensors that either monitor aging of stand-alone circuit or monitor that of functional circuit. In [7, 8], a ring oscillator is stressed and its aging degradation is monitored to predict that of the functional circuit. In contrast, [17] uses tunable replica circuits to account for variations caused by aging. These methods have small area overhead and do not introduce performance penalty to the functional circuit. However, a major limitation of monitoring a stand-alone circuit is that aging depends on workload, while the stand-alone circuit does not share the same workload with the functional circuit, thereby leading to inaccurate aging measurement. In [5, 6], online aging sensors are designed and integrated into flip-flops on the functional paths. A guardband interval is generated to catch late transitions that indicate serious aging. However, only being able to predict failure when the degradation is already serious may be too late for optimal solutions.

From a different approach in [4], offline self-test is scheduled periodically on one or more selected cores in a multi-core system using test patterns pre-stored in off-chip nonvolatile memory. However, normal execution of the core under test has to be stalled during the test. This inevitably causes performance penalty.

Once aging information is obtained by monitors, a dynamic reliability management (DRM) such as [11] can utilize it and make reliability-control decisions accordingly. Adaptive body bias (ABB) [18] and adaptive supply voltage (ASV) [19] traditionally used for compensating process variations can be applied for the purpose of aging recovery [9].

In this paper, we propose a novel methodology to address circuit aging in the field. On-chip aging sensors are integrated to monitor transitions on functional paths capturing functional mode workload. Actual timing margin, hence path delay, is then accurately measured and converted to a digital value. Diagnosis and calibration are performed in the field, thereby achieving power-performance optimization throughout the entire lifetime.

Our work makes the following major contributions:

- Our technique can accurately measure and convert delay degradation into digital values. In contrast to existing work, the quantization of circuit aging enables the designers to exploit opportunities for power-performance optimization throughout the lifetime using actual measurement results, rather than only being capable of predicting failure at the end of lifetime. This feature also largely relaxes the requirement on guardbanding.
- The proposed on-chip aging sensor can capture any transitions *on the fly* and has negligible performance penalty to the system. This is a major advantage over existing work, most of which either has to stall the normal execution of the circuit or only monitors aging of a stand-alone circuit that do not accurately represent ag-

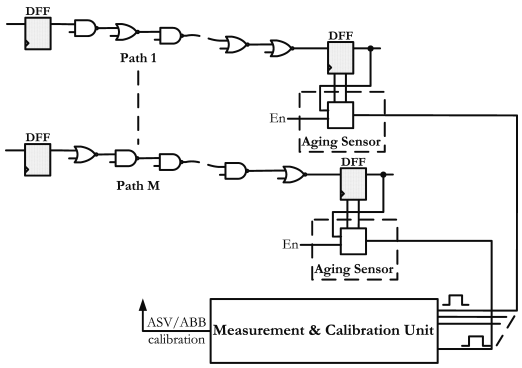

Figure 1: In-the-field aging measurement and calibration architecture.

ing of the functional circuit. The proposed aging sensor can also be used in offline measurement for diagnosis if needed.

- The proposed method is robust to glitches at the output of paths. This has not been paid much attention in existing aging sensors.
- The proposed aging measurement has high-resolution measurement at largely reduced area overhead compared to traditional delay line-based methods.
- The proposed method takes into account variations of the clocks. Not only clock skews, but also clock cycle variations, are covered in the measurement. This is especially effective considering clock aging and frequency scaling.

The rest of the paper is organized as follows. Section 2 presents the architectural overview of the proposed method. In Sections 3 and 4, we describe the implementation details of aging measurement and calibration, respectively. Section 5 discusses the simulation results. Section 6 concludes this paper.

2. PROPOSED ARCHITECTURE

Figure 1 shows the proposed architecture. The architecture consists of two major parts: aging sensors integrated into M selected paths and one measurement and calibration (MC) unit.

M functional paths on the chip, e.g., critical paths and aging-sensitive paths that might potentially become new critical paths under aging in the field, are selected (path selection algorithm is not the focus of this work; however, the limitations of path selection are taken into account in our calibration policy, see discussion in Section 4). Aging sensors are then inserted in the flip-flops at the end of each path. One sensor can actually monitor multiple paths ending at the same flip-flop. The sensors capture transitions at runtime or during offline diagnostic tests, then generate pulses whose widths represent the actual timing margins of the paths. Note that glitches are handled by the aging sensors and will not mislead the decision making.

Next, pulses are delivered to the MC unit. During the delivery, the pulse widths are well-preserved using buffers, so are the timing margin information. The MC unit then accurately measures timing margin from the pulses. Note that all sensors can share the same MC unit for measurement and calibration, thereby reducing area overhead. The MC unit also serves as the central control for reliability management. It can enable (or disable) the sensors for periodical aging measurement (or for saving power). In each round (i.e., time point) of aging measurement, timing margin information is collected for all the paths. As timing margin of

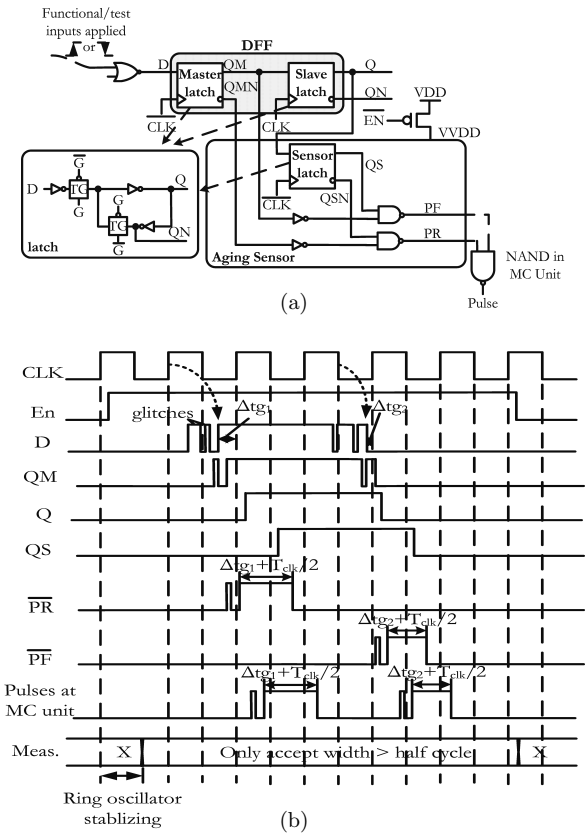

Figure 2: On-chip aging sensor: (a) implementation, and (b) a sample waveform.

a path varies on different transition types (0-to-1 and 1-to-0 transitions), off-path inputs, and voltage drops, etc., the sensor needs to record the minimum timing margin observed from multiple transitions on each path, as only the minimum timing margin is critical to reliability. To account for clock aging and scaling, clock cycle is also measured and stored at the beginning of each round. Based on the aging measurement results, ASV/ABB calibration decisions can be made to achieve power-performance optimization.

We organize the above operations into two phases: aging measurement and ASV/ABB calibration. They will be elaborated in Sections 3 and 4, respectively.

3. AGING MEASUREMENT

In this section, we will discuss the implementation details of aging measurement. We start from the design of aging sensor. Then, the procedure of aging measurement in the MC unit will be discussed.

3.1 On-Chip Aging Sensor

Figure 2(a) shows the implementation of the on-chip aging sensor and Fig. 2(b) illustrates its mechanism using a sample waveform. The on-chip aging sensor consists of the following components: a latch (sensor-latch as shown in Fig. 2(a)), two small-size inverters, and two NAND gates. The area overhead of the proposed aging sensor is only 22 transistors. As the sensor is embedded into D flip-flop, the performance impact is isolated by the master-latch and slave-latch from affecting the path. Moreover, sleep transistors can be used to manage the power overhead and aging of the sensors.

To save power, sensors are only enabled ("En=1") period-

ically when new measurement is needed. When the sensor is on, as shown in Fig. 2(a) and (b), a transition on signal "D" when "CLK=0" will pass through the master-latch and leads to a transition on signal "QM". This change will eventually cause a transition on signal "QS" at the output of the sensor-latch in the following clock cycle when "CLK=0". The two NAND-gates and a third one in the MC unit essentially form an XOR-gate and will capture this transition and generate a pulse for 0-to-1 or 1-to-0 transition.

Assume that the timing margin for this transition is Δtg_1 (or Δtg_2), which is defined as the time difference between the transition arrival time at "D" and the clock rising edge. The width of the pulse generated by the sensor is $\Delta tg_1 + T_{clk}/2$ ($\Delta tg_2 + T_{clk}/2$). This is because the pulse width at the NAND-gate is determined by the time difference between the "QM" transition and the "QS" transition. Thus, by subtracting a latch delay from both, pulse width is determined by the time difference between the "D" transition and the clock falling edge, hence half clock cycle is added to Δtg_1 (Δtg_2). The constant $T_{clk}/2$ is intentionally added to differentiate pulses generated upon actual transitions and those from glitches. As shown in Fig. 2(b), when $CLK = 1$, glitches are blocked by the master-latch and do not affect the output of the sensor; on the other hand, when $CLK = 0$, glitches can pass through but the widths of which do not get extended thus are always less than half cycle. In contrast, pulses generated upon actual transition are extended by half cycle and hence stand out. Thus, measurement results less than half cycle will be discarded. Eventually, after propagating through the interconnect between sensor and MC unit, the pulse arrives at MC unit and can be measured. Note that the shape of pulse can be well preserved during the propagation. When the timing margin reduces because of aging, the amount of degradation can be accurately measured.

It is noteworthy that if the sensor-latch and slave-latch have different delay and if the logic gates in the sensor have imbalanced delay for 0-to-1 and 1-to-0 transitions, which are common in practice, there will be a slight difference between the pulse width observed and the ideal value of $\Delta tg + T_{clk}/2$. However, by careful design, this systematic error can become negative, i.e., the observed pulse width is always smaller than $\Delta tg + T_{clk}/2$ (see column 7 of Table 2 in Section 5). This is effectively having a small guardband from being over-optimistic about the timing margin of the paths.

We acknowledge that when the timing margin is larger than half cycle, which means the path is very short and transitions happen early when "CLK=1", the pulse width would be confined to roughly one clock cycle hence is not accurately affecting the actual timing margin value. However, the inaccurate pulse width in this case would not affect the judgement that this path is not timing-critical.

3.2 Measurement and Calibration Unit

Figure 3 shows the implementation of the MC unit and how it performs pulse width measurement. The MC unit consists of the following components: MUXes, an NAND gate, a ring-oscillator (RO), an N-bit counter, a delay line, a lookup table, a write buffer, and calibration control. Note that only one MC unit is used for all paths.

The MUXs are used to select either a pulse signal or a reference clock for the measurement. Recall that we add half clock cycle to the pulse width for distinguishing actual transitions and glitches. Thus, the clock cycle measurement is performed at the beginning of each round of aging measurement, so that timing margin can be calculated by subtracting half clock cycle from the pulse width. This is very useful when clock aging or frequency scaling occurs. Note that the reference clock is only provided for clock cycle measurement,

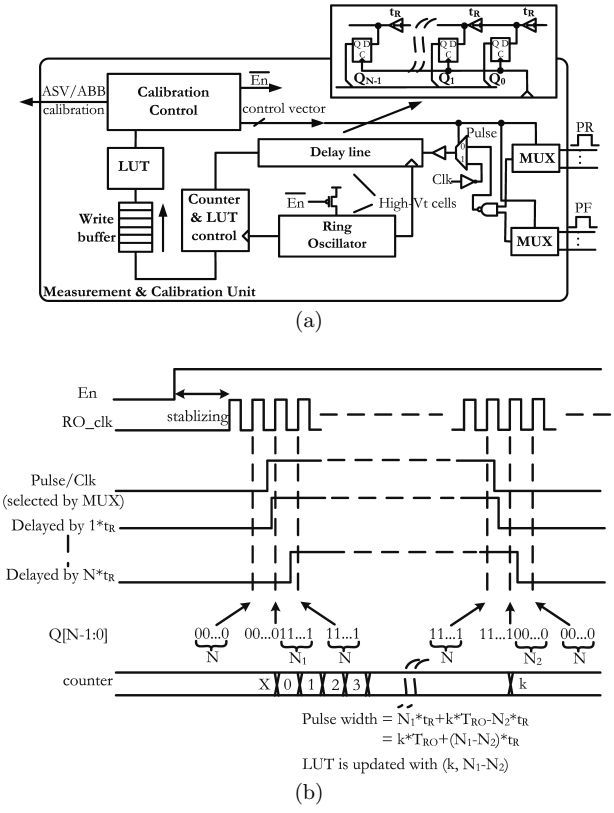

Figure 3: Aging measurement: (a) MC unit implementation, and (b) aging measurement mechanism.

skew at local clock is already taken into account when timing margin is converted to pulse width at each aging sensor. After clock cycle is measured, measurement is performed for each path individually by setting the control bits for MUXs. The NAND gate shown in Fig. 3(a) (also shown in Fig. 2 (b)) is used to combine the "PR" and "PF" and complete the pulse generation.

To collect timing margin information for each path in a timely manner, efficient path measurement scheduling is needed. Programmable timer should be set for the measurement on individual paths, so that each path is initially allocated a small amount of time to quickly collect measurement results for those paths that have relatively more switching activities during measurement and then more time is spent waiting for those relatively silent. This way, performance calibration can start as soon as needed. Note that it is preferred to perform multiple measurements for each path to account for variations due to multiple input switching, voltage drop, temperature, etc. Only the minimum timing margin detected for a path matters and should be stored in LUT.

A delay line alone can measure pulse width at high resolution. However, it would require a very long delay line to perform this task, especially when clock frequency scaling technique is used, which results in a large range of pulse width. On the other hand, measuring pulse width using counter alone is more flexible and area-efficient, but at the cost of low resolution (determined by the fastest RO clock cycle that the counter can handle). We develop a hybrid design using both an N-bit counter and a delay line to make high-resolution measurement at low cost. This is in principle similar to [15] but achieved in a different approach. To further improve resolution of the delay line, two rows of delay

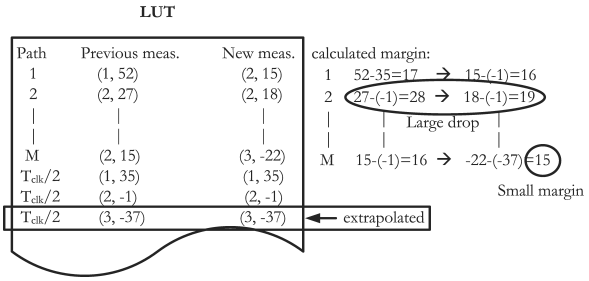

Figure 4: An example of the lookup table entries and timing margin calculation, where (k, m) represents $k * T_{RO} + m * t_R$. The numerical values are just an example for the purpose of illustration.

cells can be used similar to [16] to improve the resolution from t_R to $t_R/2$.

To manage the power overhead and aging of the MC unit, RO and delay line are both implemented with high-V_t devices and can be put in power-saving and long-recovery mode by setting "En" to 0. This procedure is controlled by calibration control in software.

In Fig. 3(b), the mechanism of aging measurement is illustrated. The MC unit measures both clock and the pulses from aging sensors in the same way as follows. As long as the delay line is longer than one RO clock cycle T_{RO} (i.e., $N * t_R > T_{RO}$), it can guarantee that when the pulse/clock edges arrive the delay line readings sampled by the RO clock will change from all 0 to 00...011...1 or from all 1 to 11...100...0 as shown in Fig. 3(b). The number of 1's (0's) in the least significant bits (LSBs), i.e., N_1 (N_2) when the pulse/clock rising (falling) edge arrives are determined by how many delay cells the pulse edge has traversed before a RO clock edge arrives. Thus, delay line provides a high-accuracy measurement for the beginning and the end of the pulse. In between, the counter counts how many RO cycles that the pulse has been observed. Combining the results from delay line (N_1 and N_2) and counter (k), the pulse width can then be calculated at $Width = k * T_{RO} + (N_1 - N_2) * t_R$. When the measured pulse width is less than half clock cycle on record, it is considered a glitch and will be discarded. Otherwise, by subtracting half clock cycle, we can obtain the timing margin. Note that a rough value of t_R is enough for aging recovery decision, because as long as (k, m)'s are greater than a certain value defined at design stage, the path under monitoring can be considered safe. In addition, the RO clock cycle T_{RO} does not affect the measurement accuracy, hence the exact value of which is also not important.

After the measurement is done, the MC unit will update the LUT with new results. A write buffer (can be shift-registers) is implemented to account for the speed gap between fast measurement and relatively slower LUT write. To diagnose aging degradation, both results of the current round of measurement and the results from previous round are kept in the LUT, as shown in Fig. 4. One practical issue is that pulse width and half clock cycle can be measured at different counter value k (e.g., $k = 3$ is not observed in clock measurement but is observed in pulse measurement), especially when timing margin is greater than the length of delay line. In this case, we cannot use a direct subtraction to obtain timing margin. A solution to this is to extrapolate different half cycle values at various k's as shown in Fig. 4. This is performed by calibration control implemented in software. Extrapolation is safe as pulse width equation is linear in terms of (k, m).

Finally, calibration control will analyze the aging measurement results and make decisions to achieve power-performance optimization. This will be elaborated in Section 4.

4. LOOKUP TABLE-GUIDED CALIBRATION

In this section, we present the calibration policy using ASV/ABB based on the aging measurement results stored in lookup table.

First, available options $\{(V_{dd}, V_{bb})\}_i|_{i=1}^{l}$ are ranked from low performance to high performance (usually also in the order from high power to low power), where the slowest option is determined during the first measurement after fabrication using the proposed aging measurement method. Second, as time goes, periodical aging measurements in the field are used to calculate timing margin for each path. Changes of timing margin are also calculated based on the most recent measurements and previous measurements. Meanwhile, after calculation for each path is done, the minimum margin and the largest drop of margin are both recorded.

Next, ASV/ABB calibration decisions are made based on the above results. Specifically, three scenarios are considered in the calibration policy: (i) if the minimum timing margin is smaller than a certain threshold, ASV/ABB calibration should be directed towards using (V_{dd}, V_{bb}) that improves performance in order to recover from aging degradation; (ii) if the minimum timing margin is larger than a threshold, ASV/ABB calibration setting (V_{dd}, V_{bb}) should target at saving power; (iii) if a large drop of timing margin is observed, indicating fast aging and the path might become new critical path under certain workload, thus the next time point t for aging measurement should be scheduled sooner than originally planned. An example of ASV calibration is shown in our simulation in Section 5.3.

Although path selection is not the focus of this work, we can see that the paths selected for aging monitoring should include not only the critical paths at time 0 but also the paths that are more sensitive to aging and might potentially become new critical paths under certain workload. When selecting such paths, the proposed ASV/ABB calibration policy can monitor closely the changes and make proper decisions in a timely manner.

5. SIMULATION RESULTS

In this section, we evaluate the effectiveness and efficiency of the proposed method using Synopsys HSpice. For the purpose of demonstration, four functional paths are implemented using 90-nm technology. Among them, three paths are critical paths selected from ISCAS $s9234$ benchmark circuit and one path is built from random logic of various gate types. Aging sensors and an MC unit are implemented. Clock frequency in the experiments is 500 MHz. First, we evaluate the accuracy of the proposed method on aging measurement. Second, area and power overhead is examined. Third, we demonstrate the efficiency of using adaptive supply voltage (ASV) based on the aging measurement results. The temperature throughout all the simulations is 75^oC. Note that process variation is not included in the simulation in order to simplify the experiments.

5.1 Aging Measurement Accuracy

Tables 1 and 2 show the procedure and results for aging measurement. At the beginning of a new round of aging measurement, the MC unit first selects the clock for measurement on the value of a half cycle. This is shown in Table 1. The largest readings from 10 measurements on counter value k and delay line value $m = N_1 - N_2$ are stored in LUT. By doing so, the margin calculation tends to be conservative. Based on the (k, m) stored in LUT, the measurement results

Table 1: Clock measurement on $T_{clk}/2$ (where $T_{RO} \approx 785ps$, $t_R \approx 20ps$).

LUT	k	m	Meas. (ps)	Error (ps)
clk	1	9	965	-35
clk	2	-30	970	-30
clk (extrapolated)	3	-69	975	-25

can be calculated. This is shown in column 4 in Table 1 as "Meas.". Compared to the nominal value of half cycle $T_{clk} = 1000ps$, the error (listed in column 5) is less than $35ps$. The extrapolation for $k = 3$ is also shown in Table 2 in case pulse measurements obtained $k = 3$. This procedure is discussed earlier in Section 3.2 and Fig. 4.

After clock cycle is measured, the MC unit can then select and measure pulses generated from paths by the aging sensors. Due to the space limit, Table 2 shows the aging measurement results for the 4 different paths only at time 0 and 10 years. Subtracted by $T_{clk}/2$ from clock measurement results stored earlier in Table 1, timing margin for each path at any time point of measurement can be obtained. Note that to simplify the simulation, we keep the clock from aging. Thus, we can use the same clock measurement results at time 0 to calculate timing margin at time 10 years as well.

Column 1 shows the type of transition observed for each path, e.g., $R_1P_1|_{t=0}$ represents a rise transition observed at the output of path P_1 at time 0 and $F_1P_1|_{t=10y}$ represents a fall transition observed at the output of path P_1 at time 10 years. We measure two rise transitions and two fall transitions for each path at time 0 and 10 years. Column 4 "Meas. margin" represents the margin measured using MC unit while "Actual margin" in column 5 represents that measured using Hspice directly. Their difference is shown as "Overall error" in column 6. The overall error is contributed by two sources: (i) aging sensor and (ii) measurement circuitry in MCU. The error contributed by aging sensor is due to the fact that gates in aging sensors have imbalanced delay for 0-to-1 and 1-to-0 transitions, thereby slightly changing the pulse width that represents the timing margin, as discussed in Section 3.1. It is extracted and labeled as "Sensor error".

From the results, we can see that the overall errors are small (from $-62ps$ to $+21ps$). We also notice that the error contributed by aging sensor is constantly negative, effectively giving a guardband in the range of $25ps$ to $42ps$ at time 0 and $12ps$ to $32ps$ at time 10 years. As a result, measurement on margin tends to be smaller than the actual margin (shown as negative overall error in most cases). This makes the aging measurement conservative and triggers the calibration procedure at a timely manner, which offers a favorable effect on our measurement. It can be seen that aging sensors themselves also age, contributing different sensor errors at time 0 and 10 years. However, the magnitude of sensor error is decreasing over time, making the measurement more accurate at the end of lifetime.

5.2 Area and Power Overhead

The MC unit does not introduce performance overhead to the chip and the sensor does not affect the path length. Therefore, in this section we only study area and power overhead of the proposed method. We use three benchmark circuits s9234, s35932, and b19 to estimate area and power overhead of the proposed method, assuming aging sensors are inserted into 5% of the DFFs (covering much more than 5% of all the paths in the circuit) and only one MC unit is used.

As shown in Table 3, for a small circuit such as s9234, the area overhead and switching power overhead are relatively large (shown in columns 4 and 5, respectively). However, for large designs, such as b19, area overhead drops to 1.1% and switching power overhead is less than 0.5% of the entire circuit. In addition, aging measurement only needs to be taken at a frequency on the order of weeks to months depending on the application. When aging measurement is not needed, both sensors and MC unit can be shut down by setting $En = 0$. Thus, power overhead of the proposed method over lifetime of the circuit then averages down to nearly zero (shown in column 6).

Furthermore, the proposed aging measurement can benefit the system-level power consumption by trading performance for power when the timing margin of the paths is large (e.g., during early lifetime). This will be demonstrated in next subsection.

Table 2: Accuracy analysis for aging measurement at time 0 and time 10 years.

Transitions	k	m	Margin (ps)		Error (ps)	
			Meas.	Actual	Overall	Sensor
$R_1P_1\|_{t=0}$	2	-10	400	413	-13	-34
$R_2P_1\|_{t=0}$	1	29	400	417	-17	-35
$R_1P_1\|_{t=10y}$	1	17	160	166	-6	-18
$R_2P_1\|_{t=10y}$	1	18	180	175	$+5$	-19
$F_1P_1\|_{t=0}$	1	27	360	399	-39	-39
$F_2P_1\|_{t=0}$	1	27	360	399	-39	-39
$F_1P_1\|_{t=10y}$	1	16	140	157	-17	-30
$F_2P_1\|_{t=10y}$	1	15	120	157	-37	-30
$R_1P_2\|_{t=0}$	2	0	600	625	-25	-33
$R_2P_2\|_{t=0}$	2	0	600	623	-23	-33
$R_1P_2\|_{t=10y}$	2	-9	420	399	$+21$	-18
$R_2P_2\|_{t=10y}$	1	30	420	425	-5	-18
$F_1P_2\|_{t=0}$	2	-6	480	508	-28	-42
$F_2P_2\|_{t=0}$	1	33	480	508	-28	-42
$F_1P_2\|_{t=10y}$	1	21	240	262	-22	-32
$F_2P_2\|_{t=10y}$	1	19	200	262	-62	-32
$R_1P_3\|_{t=0}$	2	-4	520	552	-32	-38
$R_2P_3\|_{t=0}$	2	-3	540	552	-12	-38
$R_1P_3\|_{t=10y}$	1	22	260	274	-14	-23
$R_2P_3\|_{t=10y}$	1	22	260	275	-15	-24
$F_1P_3\|_{t=0}$	2	2	640	671	-31	-34
$F_2P_3\|_{t=0}$	2	3	660	671	-11	-34
$F_1P_3\|_{t=10y}$	2	-5	500	531	-31	-27
$F_2P_3\|_{t=10y}$	1	33	480	531	-51	-27
$R_1P_4\|_{t=0}$	2	12	840	858	-18	-25
$R_2P_4\|_{t=0}$	2	12	840	860	-20	-26
$R_1P_4\|_{t=10y}$	2	2	640	668	-28	-17
$R_2P_4\|_{t=10y}$	2	4	680	671	$+9$	-17
$F_1P_4\|_{t=0}$	2	16	920	944	-24	-28
$F_2P_4\|_{t=0}$	2	16	920	944	-24	-28
$F_1P_4\|_{t=10y}$	2	9	780	794	-14	-12
$F_2P_4\|_{t=10y}$	3	-30	780	794	-14	-12

Table 3: Area and power overhead.

Circuit	Gates	DFFs	Area Overhead	Power Overhead	
				Switching	Lifetime
s9234	5597	211	6.3%	5.2%	0.2%
s35932	$16k$	$1.7k$	3.8%	2.1%	0.1%
b19	$231k$	$7k$	1.1%	0.3%	0.0%

5.3 Power-Performance Optimization

In this subsection, we demonstrate power-performance optimization under aging enabled by the proposed method. Only ASV is used in this experiment to reduce the possible combinations of available options for simplicity, but it is expected to see more flexibility hence better chance of finding an optimal setting, using a combination of ASV and ABB [9]. In this experiment, all the available options considered for Vdd are ranging from $1.08v$ to $1.23v$ in increments of $0.30v$, where $1.20v$ is the nominal Vdd. Aging measurement

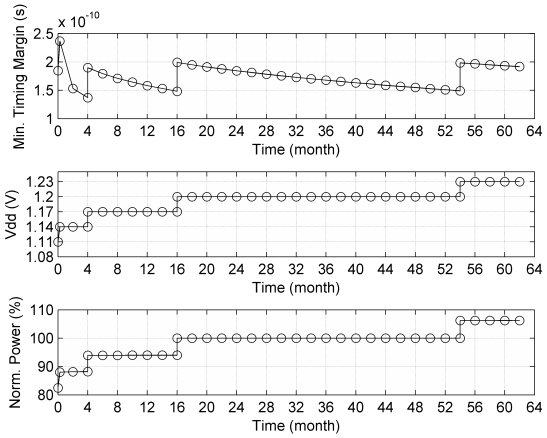

Figure 5: Adaptive supply voltage (ASV) for power-performance optimization under aging.

is executed for 62 months (over 5 years) and the measured timing margin is used to guide ASV decision. We use the calibration policy explained in Section 4. Specifically, if the minimum timing margin found during aging measurement is more than $250ps$, Vdd is reduced to save power; whereas if the minimum timing margin found is less than $150ps$, Vdd is increased to compensate for performance degradation due to aging. The ASV operation and its benefit are shown in Fig. 5. The first subfigure shows the minimum timing margin found in the measurement. The second subfigure shows the changes of Vdd at different time points based on aging measurement results and the proposed calibration policy. The third subfigure shows the power consumption normalized as a percentage compared with the power if $Vdd = 1.20v$ would have been used all the time.

At time 0, the first aging measurement is performed to find that the nominal $Vdd = 1.2v$ gives too much timing margin (greater than $250ps$). Therefore, Vdd is reduced and $Vdd = 1.11v$ is applied. The normalized power is 82.4%. Very soon (after one week), aging measurement detects that the timing margin drops below $150ps$, thus, Vdd is immediately increased to $1.14v$ in the first week. Compared to using the nominal Vdd, normalized power consumption is only 88.0% and the absolute value of power consumption is slowly reducing because of the increasing V_{th} under aging. Next, in month 4, timing margin is found to be too small. Thus, Vdd is again increased, to $1.17v$. Normalized power consumption at this point is 94.0%. Till month 16 when $Vdd = 1.2v$ has to be applied, the circuit has been enjoying low power enabled by the proposed method.

In month 54, the aging measurement results show that the minimum timing margin detected is again smaller than the safe margin $150ps$. Thus, ASV calibration can trade power for performance by increasing Vdd to $1.23v$. The normalized power then increases to 106.2%. Thus, the lifetime of the system gets extended. It is also important to notice that without the proposed aging measurement and calibration, a design targeting lifetime longer than 54 months in this case would have to introduce a larger guardband of extra power consumption and/or slower performance. This clearly shows the benefits of our proposed method.

One observation from this example is that when Vdd increases as ASV calibration is taken to recover aging, performance degradation slows down. This happens because although Vth degrades faster at higher Vdd, the degradation of performance has a non-monotonic behavior with increasing Vdd [12].

6. CONCLUSIONS

The proposed in-field aging measurement and calibration method not only provides accurate measurement on aging degradation but also enables opportunity for exploiting power-performance optimization throughout the entire lifetime. Simulation results demonstrate the effectiveness and efficiency of this method. In future work, we plan to further investigate the impact of using both ASV and ABB on power-performance optimization.

7. ACKNOWLEDGEMENTS

This work was supported in part by Cisco Systems, Inc. and Semiconductor Research Corporation (SRC) under grants 2053 and 2094.

8. REFERENCES

[1] S. Borkar, "Designing reliable systems from unreliable components," *IEEE Micro*, vol. 25, no. 6, pp. 10–16, 2005.

[2] J. W. McPherson, "Reliability challenges for 45nm and beyond," *Design Automation Conference*, pp. 176–181, 2006.

[3] E. Mintarno, J. Skaf, R. Zheng, J. Velamala, Y. Cao, S. Boyd, R. W. Dutton, and S. Mitra, "Optimized self-tuning for circuit aging," *Design, Automation and Test in Europe Conference*, pp. 586–591, 2010.

[4] Y. Li, S. Makar, and S. Mitra, "CASP: concurrent autonomous chip self-test using stored test patterns," *Design, Automation and Test in Europe*, pp. 885–890, 2008.

[5] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit failure prediction and its application to transistor aging," *IEEE VLSI Test Symposium*, pp. 277–286, 2007.

[6] J. C. Vazquez, V. Champac, A. M. Ziesemer, R. Reis, I. C. Teixeira, M. B. Santos, and J. P. Teixeira, "Low-sensitivity to process variations aging sensor for automotive safety-critical applications," *IEEE VLSI Test Symposium*, pp. 238–243, 2010.

[7] T.-H. Kim, R. Persaud, and C. H. Kim, "Silicon odometer: an on-chip reliability monitor for measuring frequency degradation of digital circuits," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 874–880, 2008.

[8] J. Keane, D. Persaud, and C. H. Kim, "An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB," *Symposium on VLSI Circuits*, pp. 108–109, 2009.

[9] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in digital circuits," *Asia and South Pacific Design Automation Conference*, pp. 284–289, 2009.

[10] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "NBTI-aware synthesis of digital circuits," *Design Automation Conference*, pp. 370–375, 2007.

[11] C. Zhuo, D. Sylvester, and D. Blaauw, "Process variation and temperature-aware reliability management," *Design, Automation and Test in Europe Conference*, pp. 580–585, 2010.

[12] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and minimization of PMOS NBTI effect for robust nanometer design," *Design Automation Conference*, pp. 1047–1052, 2006.

[13] R. Zheng, J. Velamala, V. Reddy, V. Balakrishnan, E. Mintarno, S. Mitra, S. Krishnan, and Y. Cao, "Circuit aging prediction for low-power operation," *IEEE Custom Integrated Circuits Conference*, pp. 427–430, 2009.

[14] E. Takeda, C. Y. Yang, and A. Miura-Hamada, "Hot-carrier effects in MOS devices," *Academic Press*, 1995.

[15] E. Saneyoshi, K. Nose, and M. Mizuno, "A precise-tracking NBTI-degradation monitor independent of NBTI recovery effect," *IEEE International Solid-State Circuits Conference*, pp. 192–193, 2010.

[16] C. Metra, M. Omana, T. M. Mak, A. Rahman, and S. Tam, "Novel on-chip clock jitter measurement scheme for high performance microprocessors," *IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems*, pp. 465–473, 2008.

[17] J. Tschanz, K. Bowman, S. Walstra, M. Agostinelli, T. Karnik, and V. De, "Tunable replica circuits and adaptive voltage-frequency techniques for dynamic voltage, temperature, and aging variation tolerance," *Symposium on VLSI Circuits*, pp. 112–113, 2009.

[18] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. 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 International Solid-State Circuits Conference*, pp. 422–478, 2002.

[19] J. Tschanz, S. Narendra, R. Nair, and V. De, "Effectiveness of adaptive supply voltage and body bias for reducing the impact of parameter variations in low power and high performance microprocessors," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 826–829, 2003.