Minimum Supply Voltage for Sequential Logic Circuits in a 22nm Technology

Chia-Hsiang Chen[†], Keith Bowman[‡*], Charles Augustine[‡], Zhengya Zhang[†], and Jim Tschanz[‡]
[†]Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI
[‡]Circuit Research Lab, Intel Corporation, Hillsboro, OR
[*]Now with the Processor Research Team, Qualcomm, Raleigh, NC

Abstract

The minimum supply voltage (V_{min}) is explored for sequential logic circuits by statistically simulating the impact of within-die process variations and gate-dielectric soft breakdown on data retention and hold time. As supply voltage (V_{cc}) scales, statistical circuit simulations demonstrate that hold time increases faster than circuit delay or cycle time, consequently the required number of min-delay buffers increases. For this reason, a new hold-time violation metric defines V_{min} as the V_{cc} in which the hold time exceeds a target percentage of the cycle time. Simulation results in a 22nm tri-gate CMOS technology indicate a data-retention V_{min} of $0.61V_{norm}$ and a hold-time V_{min} of $0.73V_{norm}$, where V_{norm} represents a normalized voltage for the process technology node. A key insight reveals that upsizing the first clock inverter in the sequential circuit reduces the hold-time V_{min} by 18% and the overall V_{min} by 16%.

Keywords: Logic V_{min}, data retention, hold time, gate-dielectric soft breakdown, within-die variation, sequential circuit

1. Introduction

Supply voltage (V_{cc}) scaling is the most effective technique for reducing the energy consumption of digital integrated circuits [1]. As V_{cc} reduces, however, the adverse effect of process parameter variations on performance and reliable operation increases [2, 3]. Furthermore, today's systems execute applications at a wide dynamic operating range to provide high-performance and low-power modes to trade-off performance and energy efficiency based on application requirements. The high-performance or turbo mode operates at the highest V_{cc} and induces the largest amount of stress on the transistor gate dielectric, consequently amplifying the transistor susceptibility to gate-dielectric soft breakdown [4]. This effect increases the transistor gate leakage current and is commonly modeled in circuits with an external gate-to-source resistor (R_g) [5]. Although the gate-dielectric stress is greater at the highest V_{cc}, the negative impact of the gate-dielectric soft breakdown on circuit performance and reliability is most pronounced in the low-power mode at the minimum supply voltage (V_{min}). Since lower V_{min} operation significantly enhances the system energy efficiency, scaling V_{min} within the presence of process variations and gate-dielectric soft breakdown is one of the primary goals and challenges in microprocessor and system-on-chip (SoC) product designs.

Traditionally, SRAM and register file circuits limit V_{min} scaling due to read, write, or data-retention failures [6, 7]. Recent circuit-assist techniques [8, 9] and multiple V_{cc} power domains [10] have improved V_{min} for both SRAM and register file designs. Looking forward, sequential circuits (i.e., flip-flops and latches), which contain feedback circuitry for storing data similar to SRAMs and register files, may start to limit V_{min} scaling for the logic portion of the processor design. Since sequential logic circuits do not have the regularity of array structures, the circuit-assist techniques for SRAM and register file designs may not be applicable or incur impracticable overheads for the sequential circuits.

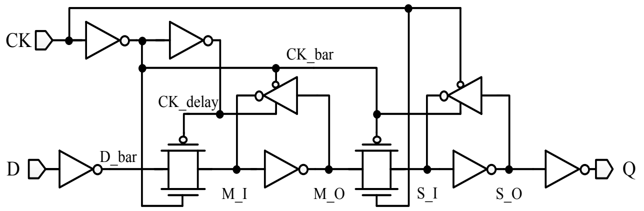


Figure 1: Master-slave flip-flop (MSFF) schematic.

Three fundamental metrics for a sequential circuit are the data retention, setup time, and hold time. As V_{cc} reduces within the presence of process variations and gate-dielectric soft breakdown, the data retention degrades, and the setup and hold times lengthen. Although setup-time violations at low V_{cc} can be avoided by reducing the clock frequency (F_{clk}) [2] in post-silicon testing, hold-time or data-retention failures cannot be resolved by changing F_{clk}. Rather, post-silicon data-retention or hold-time violations are only resolved by increasing V_{cc}. Furthermore, adding min-delay buffers and upsizing transistors in the sequential circuit are necessary to prevent potential data-retention and hold-time failures at low V_{cc}. These approaches, however, incur an expensive power overhead when worst-case variations are assumed in pre-silicon design. For this reason, the data retention and hold time dictate the V_{min} for sequential circuits.

This paper explores the V_{min} for sequential logic circuits in a 22nm tri-gate CMOS technology [11] by statistically simulating the impact of within-die (WID) process parameter variations and gate-dielectric soft breakdown on data retention and hold time for over 10^6 standard-cell master-slave flip-flops (MSFF), as illustrated in Fig. 1, to represent the sequential circuits in a high-performance microprocessor or SoC design. Section II describes the statistical circuit analysis. Sections III and IV explain the circuit simulation methodologies for the data-retention V_{min} and hold-time V_{min}, respectively. Section V compares the data-retention V_{min} and the hold-time V_{min} values while providing insight for reducing the overall V_{min} for the sequential logic circuits. Section VI summarizes the key results.

2. Statistical Circuit Simulation Analysis

The Monte Carlo (MC) simulation is the most common statistical methodology for capturing the effects of process variations in circuits. The MC simulation performs many MC samples. Based on the input device-level parameter distributions and spatial correlations, each MC sample assigns variations to the device parameters in the circuit. These

978-1-4799-1235-3/13/$31.00 ©2013 IEEE

device-level parameters include channel length, channel width, and threshold voltage. After simulating the circuit with the assigned parameter variations, the circuit output (e.g., delay) corresponds to one MC sample. The MC output distribution is generated after performing a sufficient number of samples.

In a high-performance microprocessor or SoC design, the number of flip-flops can reach or exceed 10^6 [12]. Quantifying the impact of WID variations on this number of flip-flops in a design requires a statistical analysis corresponding to a cumulative probability of ~5 standard deviations (σ) from the mean. To accurately capture the 5σ WID-variation probability in the tail of the MC output distribution, more than 10^7 samples are needed, resulting in excessive simulation time, and consequently, rendering the MC approach impracticable as a statistical analysis approach.

The most probable point (MPP) simulation [13] provides an exponentially faster alternative to an MC simulation for cumulative probabilities larger than 4σ. In contrast to the MC approach, the MPP only generates a single output value that corresponds to a specific cumulative probability (e.g., a 5σ probability in a normal distribution). The MPP first performs a sensitivity analysis to identify the most sensitive parameters in the circuit that are susceptible to variations. Then, the WID-variations are distributed to either maximize or minimize the circuit response, depending on the output function, for an input cumulative probability corresponding to a target number of σ values (e.g., 5σ). As an example of the MPP hold-time simulation, the WID parameter variations in channel length, channel width, and threshold voltage are distributed among the most sensitive transistors in the MSFF, as described in Fig. 1, to maximize the hold-time delay for a cumulative probability corresponding to a 5σ target in a normal distribution. In an MC simulation, the required number of samples increases exponentially as the target σ number increases. In contrast, MPP only requires a fixed number of samples for the sensitivity analysis and for calculating the maximum or minimum circuit output, which depends on the number of transistors in the circuit and is independent of the target σ number. For this reason, MPP is a highly practical statistical simulation methodology for evaluating a circuit response for a target of 4σ or higher. Furthermore, the MPP simulation provides key insight to the most vulnerable transistors in the circuit by specifying the assignment of the device-level parameter variations.

Recent statistical SRAM and register file circuit simulations employ the MPP methodology [14]. Table I and Fig. 2 provide a comparison of the MPP and MC simulations for validating the accuracy of the MPP approach. The MC simulations consist of 10^4 samples to enable a highly accurate analysis of the distribution tail for cumulative probabilities corresponding to a 3σ target and below. As described in Sections III and IV, separate statistical circuit simulations quantify the hold time and data-retention V_{min} for cumulative probabilities targeting a 2.5σ and a 3.0σ of WID variation. For the hold-time simulations, V_{cc} equals V_{norm} and 0.75V_{norm}, where V_{norm} represents a normalized voltage for the process technology node. From Table I, the MPP error as compared to MC is less than 5% for all four hold-time statistical simulations. Fig. 2(a) highlights one of the four comparisons by plotting the probability density from the MC simulation with V_{cc} at V_{norm}. From Fig. 2(a), the MPP normalized delay of 1.72 agrees closely (i.e., 1.8% error) with the MC normalized delay of 1.69 for a cumulative probability corresponding to a 3σ WID-variation target. For the data-retention V_{min}, simulations are performed with and without the R_g model that captures the gate-dielectric soft breakdown. In Table I, the MPP error in data-retention V_{min} is less than 4%. From Fig. 2(b), the MPP output error is 1.9% of the MC simulation value. In summary, the MPP methodology provides a highly accurate result as compared to an MC approach while exponentially reducing the simulation time for cumulative probabilities targeting 4σ and beyond.

Table I. Comparison of MPP and MC simulations. Table shows percentage difference between MPP and MC (10^4 samples) simulations for both hold time and data-retention V_{min}, at two WID variation targets (2.5σ and 3.0σ). Hold time is performed at two voltages and data-retention V_{min} is performed with and without R_g.

WID Variations	Hold Time		Data-Retention V_{min}	
	V_{norm}	0.75V_{norm}	w/ R_g	w/o R_g
2.5 σ	2.1%	0.2%	3.6%	1.6%
3.0 σ	1.8%	4.4%	1.9%	0.5%

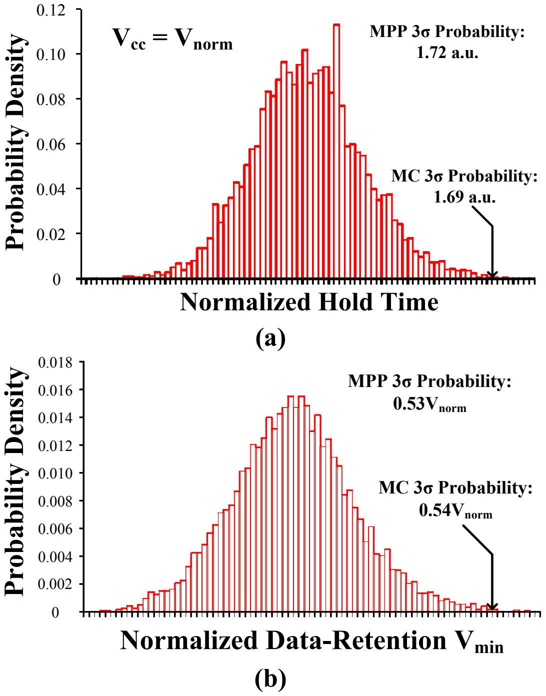


Figure 2: Probability density distributions from MC simulations with 10^4 samples for (a) normalized hold time at V_{norm} and (b) normalized data-retention V_{min} with R_g. The MC (a) hold time and (b) data-retention V_{min} corresponding to a 3σ WID-variation probability are compared to MPP results.

3. Data-Retention V_{min}

As described in Fig. 1, an MSFF consists of a master latch followed by a slave latch. An MSFF retains data in both the master and slave latches. Since the data retention for the master latch and the slave latch are similar, the data-retention simulation focuses on the slave latch to simplify the analysis. While the number of latches to consider for the data-retention analysis is twice the number of MSFFs, the change in the number of standard deviations for the WID variations is negligible. Fig. 3 zooms-in on the schematic of the slave latch for describing the data-retention analysis. The two primary sources of data-retention degradation are WID process variations and gate-dielectric soft breakdown [5]. Gate-dielectric soft breakdown is modeled by adding a resistor (R_g) between the gate and source of a transistor as illustrated in Fig. 3 at node S_I. The value of R_g is empirically extracted from device measurements. Although soft breakdown can occur in any transistor, the most probable node for soft breakdown in the slave latch is either S_I or S_O in Fig. 3. These two nodes receive the longest time of DC stress while the transistors on the clock path receive less DC stress because the clock nodes transition twice every cycle [15]. Thus, the data-retention analysis for the MSFF is similar to the SRAM [5]. The conventional SRAM data-retention analysis is based on a static DC simulation [16]. The conventional SRAM circuit analysis breaks the feedback inverter loop to simulate the DC response for the voltage transfer curve (VTC). Similarly for the MSFF, the simulation varies the input voltage from 0V to V_{cc} on S_I to generate one VTC and S_O to generate another VTC as illustrated in Fig. 4(a). From this butterfly curve which is formed by the two VTCs, the static noise margin (SNM) is calculated as the voltage corresponding to the smallest side of the two largest squares bounded inside the curve. The data-retention V_{min} is defined as the V_{cc} in which the SNM collapses to zero.

From MPP simulations, drive-current degradation in the top PMOS of the tri-state inverter, as circled in Fig. 3, with an R_g connected between S_I and ground limits the data-retention V_{min}. This occurs for two reasons. First, the tri-state inverter is designed with minimum width transistors to minimize the impact on the CK-to-Q delay and area. In comparison to the S_O node which is driven by an inverter, the S_I node is weakly driven by the tri-state or stacked inverter, resulting in greater susceptibility to the gate leakage from soft breakdown as modeled by R_g. Second, the PMOS drive current is slightly weaker than the NMOS drive current for iso-sized transistors [11], thus retaining a logic “1” on S_I is more difficult than holding a logic “0.” For these reasons, the worst-case simulations for soft breakdown occur while placing an R_g between S_I and ground while retaining a logic “1.” In addition to the location of R_g, the inverter drive current is more sensitive to the top PMOS of the tri-state inverter, as circled in Fig. 3, as compared to the bottom PMOS.

In contrast to an SRAM or register file design, the MSFF refreshes the data in both latches every cycle. Thus, the slave latch only needs to retain the data for half of the clock cycle (i.e., low phase of the clock). The retention time is inversely proportional to F_{clk}. The traditional static DC analysis assumes an infinite retention window, thus failing to capture the interaction between the data retention and the clock cycle time. To investigate the impact of cycle time (or retention window) on the data-retention V_{min}, a transient simulation is performed while varying the retention window as illustrated in Fig. 4(b). Table II lists the normalized data-retention V_{min} simulation results for retention windows ranging from 0.01µs to 10µs. From this data, the clock cycle time has a negligible influence on the data-retention V_{min} over the cycle time range of interest.

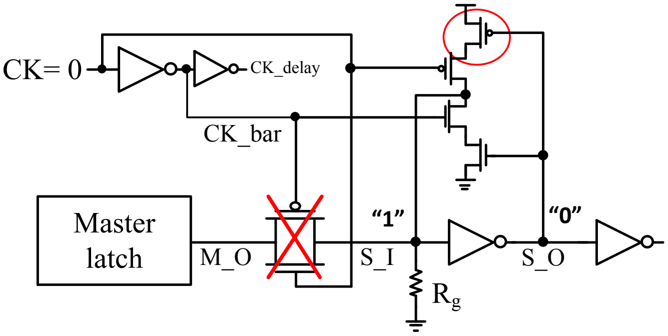
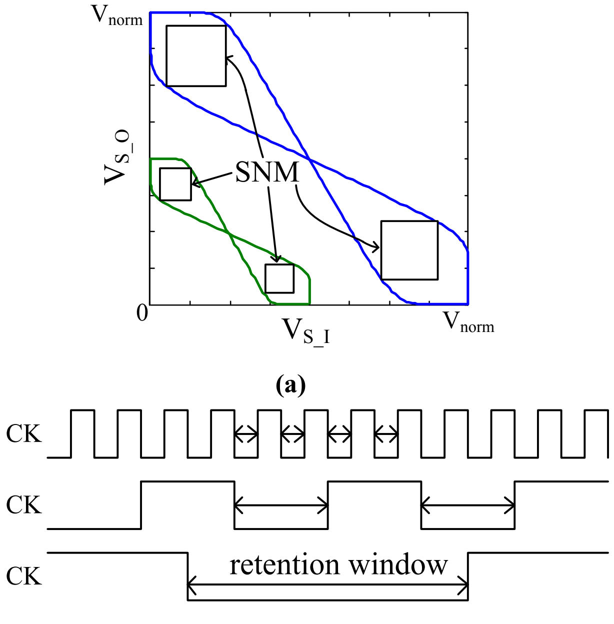


Figure 3: Data-retention analysis for the slave latch of the MSFF while holding a logic “1.” The PMOS process variation (circled) and gate-dielectric soft breakdown (R_g) on node S_I limit the data-retention V_{min}.



(b)

Figure 4: (a) Butterfly curves for the static DC analysis of data retention. (b) Description of the retention windows for the dynamic transient analysis of data retention.

Table II. Normalized data-retention V_{min} with R_g and a 5σ WID-variation target across various retention windows for the dynamic transient simulations.

Retention Window	0.01 µs	0.1 µs	1 µs	10 µs
Norm. Data-Retention V_{min}	0.607	0.608	0.609	0.609

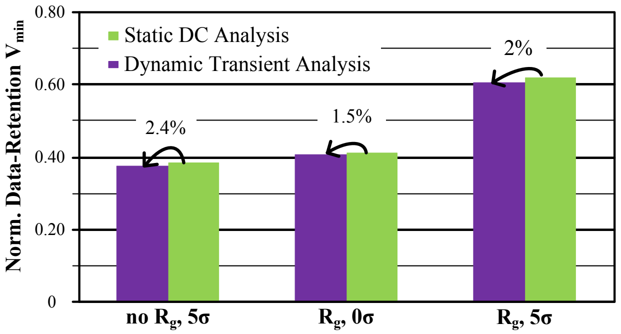


Figure 5: Normalized data-retention V_{min} with the individual and combined contributions of WID variation and R_g for the conventional static DC analysis and the dynamic transient analysis that captures the cycle time effect.

Fig. 5 quantifies the individual and combined impact of WID process variations and gate-dielectric soft breakdown on data-retention V_{min} for the static DC analysis and the dynamic transient analysis. First, the static DC analysis agrees closely (i.e., within ~2%) with the more rigorous and accurate dynamic transient analysis. Since the static DC analysis is significantly faster than the dynamic transient analysis, the conventional static DC simulation is the recommended approach for the data-retention V_{min} analysis in sequential logic circuits. Second, the results in Fig. 5 indicate that the WID variations at a 5σ target have a similar effect as the gate-dielectric soft breakdown on the data-retention V_{min}. The combination of both WID variations and gate-dielectric soft breakdown limits the data-retention V_{min} to $0.61V_{norm}$ for the 22nm technology.

4. Hold-Time V_{min}

Referring to Fig. 1, hold time is the minimum delay that the MSFF input (D) needs to be held after the rising edge of the clock (CK) to ensure the data is sampled correctly. The hold-time simulation sweeps the transition of D relative to the rising CK edge until the CK-to-D delay results in a 50% V_{cc} glitch at node M_O as described in Fig. 6. Hold time is influenced by the same factors considered in the data-retention analysis, including gate-dielectric soft breakdown and WID process variations. Hold time is also data dependent, as the hold time for a logic “1” at the input D is longer than the hold time for a logic “0” for a positive-edge-triggered MSFF as illustrated in Fig. 7. This phenomenon is attributed to the misalignment of clock signals to the transmission gate (i.e., transistors M1 and M2) and the tri-state inverter (i.e., transistors M3 and M4). The internally generated CK_delay and CK_bar nodes are separated by an inverter delay. As a result, the transmission-gate PMOS (M1) is always turned off after the transmission-gate NMOS (M2), thus creating a longer transparency window for a logic “1” on D_bar (i.e., “0” on D) as compared to a logic “0” on D_bar (i.e., “1” on D). The longer transparency window directly increases the hold time to prevent the high-to-low transition on D from entering the master latch and corrupting the desired state. In parallel, the M4 NMOS turns on after the M3 PMOS, thus the pull down to maintain “0” at node M_I (i.e., the “1” from D) is weakened. Thus, the hold time for a logic “1” at the input D is significantly longer than the hold time for a logic “0.”

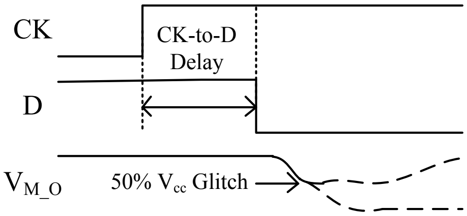


Figure 6: Hold time is defined as the CK-to-D delay where a 50% V_{cc} glitch occurs on node M_O.

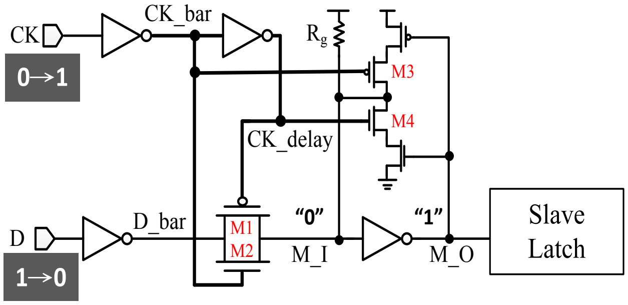


Figure 7: Dynamic transient simulation description for the worst-case data-dependent hold-time analysis.

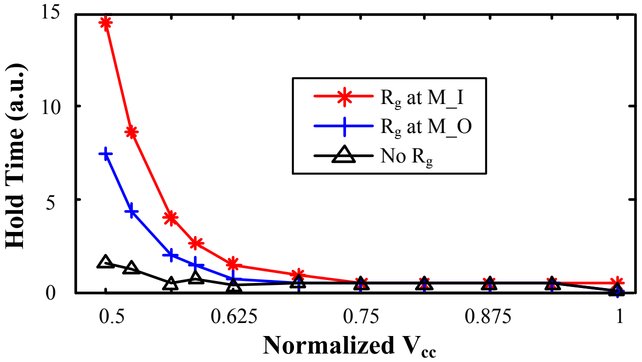


Figure 8: Hold time versus normalized V_{cc} with different placement of R_g while not including WID variations.

Initial hold-time simulations consider the effect of gate-dielectric soft breakdown by placing the R_g at different nodes in the MSFF. The hold time for a logic “1” degrades by either inserting R_g between V_{cc} and node M_I or placing R_g between node M_O and ground. Fig. 8 compares these two scenarios by simulating the impact of R_g on hold time without considering WID variations. Fig. 8 demonstrates that the worst-case hold time for a logic “1” occurs for an R_g between V_{cc} and M_I.

Fig. 9 plots the impact of WID process variations for a 5σ target on the hold time with and without inserting R_g between V_{cc} and M_I. From Fig. 9, the impact of WID variations dominates the hold time as V_{cc} scales while the gate-dielectric soft breakdown has a negligible effect for a 5σ WID-variation target. Fig. 9 also plots the normalized fan-out of 4 (FO4) inverter chain delay as a representative of logic path delay as V_{cc} reduces. From Fig. 9, the hold time increases at a much faster rate as compared to the FO4 inverter delay as V_{cc}

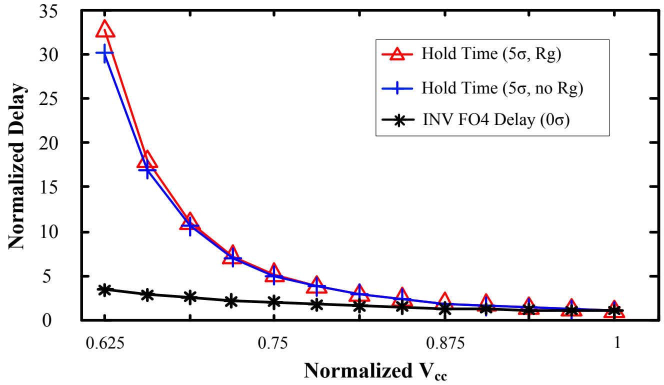


Figure 9: Normalized hold time with and without R_g for a 5σ WID-variation target and an FO4 inverter chain delay versus the normalized V_{cc}.

decreases. As V_{cc} reduces from V_{norm} to $0.625V_{norm}$, the normalized FO4 inverter chain delay increases by 3.3× while hold time increases by more than 30×. This large discrepancy between the hold time and the inverter chain delay amplifies the susceptibility of sequential logic circuits to min-delay race conditions, thus limiting V_{cc} scaling.

To avoid the hold-time violations, logic circuit designs must insert additional buffers to allow further V_{cc} scaling, which negatively affects the logic area and power at the high-performance mode. A critical step for evaluating the hold-time V_{min} is establishing the maximum hold-time delay for a given V_{cc}. In the simulations, hold time equals ~3% of the cycle time at V_{norm}. The normalized clock cycle time is assumed to scale as the inverter chain delay. A practical definition of hold-time V_{min} is determined by normalizing the hold time to the cycle time at each V_{cc} value as plotted in Fig. 10. This data demonstrates that the hold time increases as a larger fraction of the available cycle time as V_{cc} reduces. Consequently, the number of buffers for min-delay protection must become an increasing fraction of the total cycle time. The increasing cost of buffer insertion diminishes the energy benefits of reducing V_{cc}. From Fig. 10, a practical limit for hold time is ~10% of the cycle time. Beyond this point, the hold time increases exponentially, thus requiring an exponential increase in the number of min-delay buffers to avoid hold-time violations. By defining the maximum hold time as 10% of the cycle time, the hold-time V_{min} equals $0.73V_{norm}$ in the 22nm technology.

Fig. 11 describes the assignment of device-level parameter variations from the MPP simulation to maximize the hold time for a 5σ WID variation across four V_{cc} values. This data provides key insight to the most sensitive transistors in the MSFF for hold time. From Fig. 11, the MPP simulation places the vast majority of the WID variation on the NMOS of the first clock inverter of the MSFF (i.e., the inverter with input CK and output CK_bar in Fig. 1). As V_{cc} reduces, this NMOS transistor receives a larger portion of the WID variation. The variation in the NMOS of the first clock inverter changes the falling delay of CK_bar and the rising delay of CK_delay, which controls the NMOS and PMOS transistors in the master transmission gate, respectively. A longer channel length, shorter channel width, and/or higher threshold voltage on the NMOS of the first clock inverter weakens the drive strength of this inverter during a rising clock edge, thus increasing the delays for both clock inverters in the MSFF. The longer delays for the two clock inverters expand the transparency window, thus degrading the hold time for the MSFF. In summary, the hold time is most sensitive to the NMOS of the first clock inverter in the MSFF.

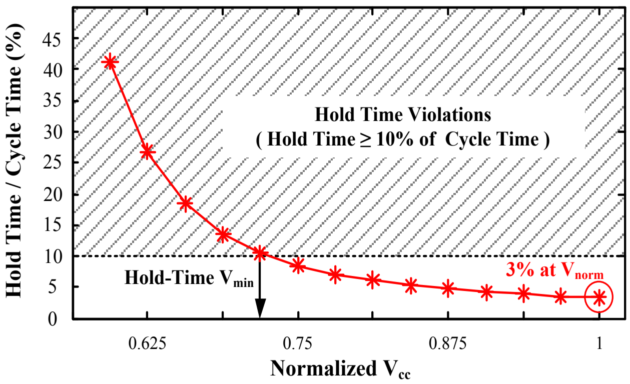


Figure 10: Hold time as a percentage of cycle time versus normalized V_{cc}. Hold-time V_{min} is defined as the V_{cc} in which the hold time exceeds 10% of the cycle time.

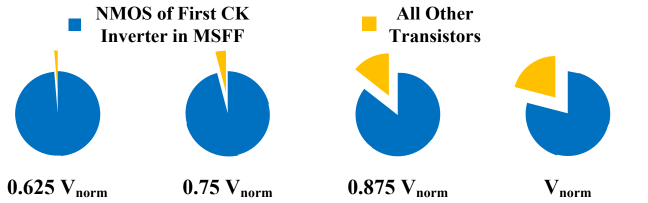


Figure 11: Breakdown of the 5σ WID variation across all the transistors in the MSFF from the MPP simulations.

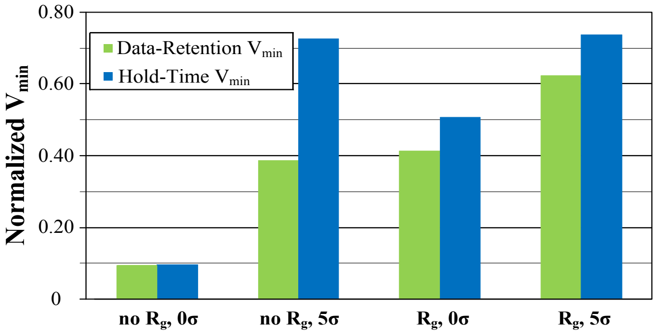


Figure 12: V_{min} for data retention and hold time versus different combinations of R_g and 5σ WID variation.

5. V_{min} of Sequential Logic Circuits

Fig. 12 compares the data-retention and hold-time V_{min} values while considering the individual and combined effects of WID variations and gate-dielectric soft breakdown. When neither WID variations nor gate-dielectric soft breakdown are considered, the data-retention and hold-time V_{min} equals the fundamental V_{cc} scaling limit for CMOS circuits [17, 18]. When only accounting for the 5σ WID variation, the data-retention and hold-time V_{min} values increase to $0.39V_{norm}$ and $0.72V_{norm}$, respectively. When only considering the gate-

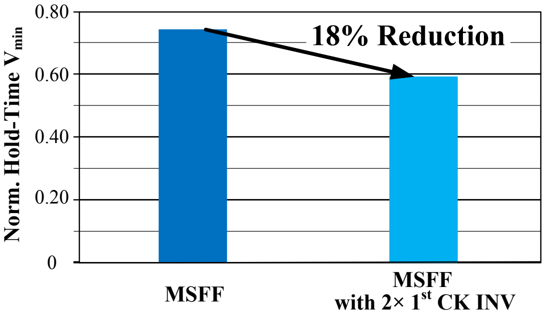


Figure 13: Hold-time V_{min} for the original MSFF in Fig. 1 and for the MSFF with a 2× larger first clock inverter.

dielectric soft breakdown, the data-retention and hold-time V_{min} values are 0.41V_{norm} and 0.5V_{norm}, respectively. As discussed previously in Sections III and IV, WID variation and gate-dielectric breakdown affect the data retention similarly while the WID variation dictates the hold time. When combining the effects of both WID variation and gate-dielectric soft breakdown, the data-retention and hold-time V_{min} values rise to 0.61V_{norm} and 0.73V_{norm}, respectively. From this analysis, the hold-time V_{min} limits the V_{cc} scaling for sequential circuits in a high-performance microprocessor or SoC in a 22nm technology.

Since the WID variation is the dominant contributor to the hold-time V_{min}, reducing the hold-time sensitivity to WID variations enables an overall lower V_{min} for the sequential circuits. As described in Section IV, the hold time is most sensitive to variations on the NMOS of the first clock inverter in the MSFF. Increasing the transistor width allows more averaging of the random uncorrelated WID variations, consequently reducing the drive current sensitivity to WID variations. Fig. 13 reveals an 18% reduction in hold-time V_{min} by doubling the size of the first clock inverter. This design change in the MSFF results in an overall 16% V_{min} reduction since the data-retention V_{min} of 0.61V_{norm} now limits the V_{cc} scaling. Although the larger clock inverter width increases the capacitive load on the clock network and the dynamic power at a given V_{cc} value, this analysis highlights the opportunity for optimizing the sequential circuit design for enhancing the energy efficiency of a high-performance microprocessor or SoC design.

6. Conclusions

Data-retention V_{min} and hold-time V_{min} are studied to avoid logic failures on sequential circuits while capturing the effect of WID process variations and gate-dielectric soft breakdown. Statistical circuit simulations demonstrate that the data-retention V_{min} depends on both WID variations for a 5σ target and gate-dielectric soft breakdown, which limit the data-retention V_{min} to 0.61V_{norm}. As hold time increases faster than the cycle time while lowering V_{cc}, a new hold-time violation metric is introduced to define V_{min} as the V_{cc} in which the hold time exceeds a target percentage (10%) of the cycle time. As a result, the hold-time V_{min} is found at 0.73V_{norm} and is primarily affected by WID variations.

Furthermore, a detailed circuit analysis reveals that the data-retention V_{min} is highly sensitive to the gate-dielectric soft breakdown and the variations on the top PMOS of the tri-state inverter. Hold-time V_{min} is most sensitive to the variations on the NMOS of the first clock inverter. Upsizing the first clock inverter in the MSFF by 2× reduces the hold-time V_{min} by 18% and the overall V_{min} by 16%.

ACKNOWLEDGMENT

The authors sincerely appreciate the support from Jaydeep Kulkarni, Vivek De, and Rick Forand.

REFERENCES

[1] A. Chandrakasan, *et al.,* "Low-power CMOS digital design," *JSSC*, pp. 473-484, Apr. 1992.

[2] K. A. Bowman, *et al.*, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *JSSC*, pp. 183-190, Feb. 2002.

[3] S. Borkar, *et al.,* "Parameter variations and impact on circuits and microarchitecture," in *DAC,* 2003, pp. 338-342.

[4] A. M. Yassine, *et al.,* "Time dependent breakdown of ultrathin gate oxide," *IEEE TED*, pp. 1416–1420, Jul. 2000.

[5] M. Agostinelli, *et al.*, "Erratic fluctuations of SRAM cache Vmin at the 90nm process technology node," in *IEDM,* 2005, pp.655-658.

[6] H. Qin, *et al.,* "SRAM leakage suppression by minimizing standby supply voltage," in *ISQED*, 2004, pp. 55-60.

[7] R. Heald *et al.*, "Variability in sub-100nm SRAM designs," in *ICCAD*, 2004, pp. 347-352.

[8] S. Ohbayashi, *et al.*, "A 65nm SoC embedded 6T-SRAM designed for manufacturability with read and write operation stabilizing circuits," *JSSC*, pp. 820-829, Apr. 2007.

[9] E. Karl, *et al.,* "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active V_{MIN}-enhancing assist circuitry," in *ISSCC*, 2012, pp.230-231.

[10] O. Hirabayashi, *et al.*, "A process-variation-tolerant dual-power-supply SRAM with 179 um^2 cell in 40nm CMOS using level-programmable wordline driver," in *ISSCC*, 2009, pp. 458-459.

[11] C. Auth, *et al.,* "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *Symp. VLSI Tech.*, 2012, pp. 131-132.

[12] Xilinx. *7 Series FPGAs Overview.* [Online]. Available: http://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf.

[13] X. Du, *et al.*, "A most probable point based method for uncertainty analysis," *J. Design and Manufacturing Automation*, pp. 47-65, Feb. 2001.

[14] D. Khalil, *et al.,* "SRAM dynamic stability estimation using MPFP and its applications," *J. Microelectronics*, pp. 1523-1530, Nov. 2009.

[15] S. Kumar, *et al.*, "Impact of NBTI on SRAM read stability and design for reliability," in *ISQED*, 2006, pp. 213-218.

[16] E. Seevinck, *et al.*, "Static-noise margin analysis of MOS SRAM cells," *JSSC*, pp.748-754, Oct. 1987.

[17] J. von Neumann, "Theory of self-reproducing automata," A. W. Burks, Ed., *University of Illinois Press*, Urbana, 1966.

[18] R. M. Swanson *et al.*, "Ion-implanted complementary MOS transistors in low-voltage circuits," *JSSC*, pp. 146-153, Apr. 1972.