

# Process Variation Immunity of Alternative 16nm HK/MG-based FPGA Logic Blocks

Ahmad Alzahrani and Ronald F. DeMara

Department of Electrical Engineering and Computer Science  
University of Central Florida, Orlando, FL 32816-2362

Email: {alzahrani@knights, demara@mail}.ucf.edu, http://cal.ucf.edu\*

**Abstract**—Continued miniaturization of semiconductor technology to nanoscale dimensions has elevated reliability challenges of high density Field-Programmable Gate Arrays (FPGA) devices due to increasing impacts of Process Variation (PV). The issue is addressed herein using a systematic bottom-up analysis by determining the relative influence of PV on alternate design realizations of FPGA logic blocks. Results for conventional design structures are obtained through detailed SPICE simulations and related to structural risk features. Namely, Transmission Gate (TG) and Pass Transistor (PT) based MUX architectures for realizing Look-Up-Tables (LUTs) are compared. At threshold voltage variation  $\sigma_{V_{th}} = 14\%$ , PT-based designs that meet the 95% yield objective can exhibit as high delay variation as 23.3%. PV impact can be reduced to 4.9% if TG-based LUT is considered. Finally, the impact of transistor sizing is investigated as a method of mitigating PV susceptibility in FPGA structures.

## I. INTRODUCTION

Advancement of CMOS manufacturing technology to reduce device dimensions has ushered in significant challenges resulting from *Process Variation (PV)* [1]. Significant sources of variation in sub-45nm manufacturing processes include imprecise lithography, etching, deposition, and dopant implantation [2]. These can lead to Random Dopant Fluctuation, Line-Edge Roughness, and structure dimension variance, e.g. channel length and oxide thickness. Variation in these physical parameters translates into deviation in device electric characteristics, such as  $V_{th}$  and drive current  $I_{dsat}$ , from the intended specifications. Therefore, PV can lead to slow, weak, or defective transistors, thus affecting yield, final product performance, efficiency, and reliability. The International Technology Roadmap for Semiconductors (ITRS) has estimated that  $V_{th}$  variation, given by three-sigma ( $3\sigma_{V_{th}}$ ), has already reached 42% ( $\sigma_{V_{th}} = 14\%$ ) and can reach up to 79% ( $\sigma_{V_{th}} = 26\%$ ) for near-future process technology, according to table DESN10 in [3]. Fortunately, PV exhibits a statistical nature which makes it feasible to study at various levels of design abstraction, which are compared in this paper for alternate functional realizations. Traditionally, Statistical Static Timing Analysis (SSTA) technique is used to predict design behavior at design-time and accordingly devise the appropriate mitigation strategies to minimize PV effects and increase yield.

SRAM-based *Field Programmable Gate Arrays (FPGAs)* have been at the frontier of technology scaling owing to the increased demand for high performance and low-power reconfigurable systems. Thus, the design of FPGA logic blocks have an increasing need to cope with PV issues emerging at each new process node. Contemporary SRAM-based FPGAs

designs are composed of array of tiles, which contain Logic Clusters (LCs), Connection Boxes (CBs), and Switching Box (SBs). A logic cluster contains Look-Up-Tables (LUTs) and flip-flops which implement logic functionality. Connection and switching boxes provide the required connectivity among LCs and routing channels. Commercial FPGAs have utilized multiplexers (MUXes) to implement LUTs, CBs, and SBs due to the lower required number of control inputs and favorable area-delay product [4][5]. Because of the uniform fabric of modern SRAM-based FPGAs, multiplexers can be viewed as a dominating fundamental logic structure in these devices besides SRAM cells. To reduce cost and area, FPGA vendors have relied on NMOS Pass-Transistors (PTs) as the preferred fundamental switching elements for realizing multiplexers. NMOS PTs are known for conveying a weak high logic signal level at a saturated output of  $V_{DD} - V_{th}$ . As aggressive scaling of devices continues, the voltage difference between  $V_{DD}$  and  $V_{th}$  decreases; thus half-latch restoration logic has been used to mitigate resultant reliability and performance issues. Recently, the work in [4] investigated the use of Transmission Gates (TGs) as an alternative design option to implement FPGAs blocks while achieving lower area-delay product. In [6], PV-induced failure rate in a PT-based multiplexer without the restoration logic is studied. However, the work did not consider that transistor sizing can substantially reduce defect rate as demonstrated later in this work. To the best of our knowledge, there has not been a study on how PT and TG-based structures compare as design options for FPGA structures under the effect of variations. In this paper, we study the impact of variation on these two design alternatives and report the *Defective Rate*, *Delay*, and *Energy Delay Product (EDP)*.

The remainder of the paper is organized as follows. Section II provides a background on PV and the FPGA structures to be considered for the evaluation. Section III describes the evaluation framework and toolset adopted for simulation. Results and Conclusions are discussed in Section IV.

## II. EFFECTS OF PROCESS VARIATION

Process variation is a key challenge for continued technology scaling. It can affect functional, leakage, or timing yield and power efficiency of final design due to the need for wider voltage margins. Variation can include any nanoscopic imprecision in manufacturing processes during physical realization of design layout. PV can be manifested as Die-to-Die (D2D) or Within-Die (WID) variations. WID variation becomes a significant factor in the impact of variation and thus is the scope of this work. At the individual transistor-level, the prominent negative effect of WID variation is observed as a variation in device threshold voltage and the amount of current flow during transistor ON-OFF states. Threshold

\* This work is supported by the Ministry of Higher Education of Saudi Arabia under scholarship grant SACM64923.

voltage variation  $\sigma_{V_{th}}$  is essentially a function of device dimensions and dopant density in the channel as expressed below [7].

$$\sigma_{V_{th}} \propto \frac{t_{ox}}{\epsilon_{ox}} \sqrt{\frac{n_{ch}}{3 \cdot w \cdot l}} \quad (1)$$

where  $w$  and  $l$  are the channel width and length respectively,  $t_{ox}$  is the gate oxide thickness,  $\epsilon_{ox}$  is the permittivity of oxide layer, and  $n_{ch}$  is the concentration of channel doping. Since device delay  $t_g$  is tightly dependent on  $V_{th}$  as given by the well-cited *alpha-power law* in (2), high variation can severely impact transistor speed and cause timing yield loss.

$$t_g \propto \frac{l_{eff} \cdot V_{DD}}{(V_{DD} - V_{th})^\alpha} \quad (2)$$

where  $l_{eff}$  is the effective channel length,  $\alpha$  is a constant depending on process technology. Similarly, threshold voltage affects the transistor ON saturation current  $I_{D_{sat}}$  as given in (3) which results in a weak or slow driving transistor.

$$I_{D_{sat}} = \frac{w \cdot \nu_{sat} \cdot \epsilon_{ox}}{t_{ox}} \cdot (V_{gs} - V_{th} - V_{d_{sat}}) \quad (3)$$

where  $V_{gs}$  is the gate voltage, and  $V_{d_{sat}}$  is the saturation drain voltage. This effect leads to a situation where a transistor is either not strong enough to trigger downstream gates, or does so at a slow pace causing higher sub-threshold current to flow in fanout gates. Thus, PV can cause a functional yield loss and power constraint violation even if timing requirements determined by the proportion of gates in the critical path are met. In this paper, we consider the case where PV can cause FPGA structures to functionally fail. Previous work for the effects of PV in FPGAs have considered the timing and leakage yield [8] [9], whereas in [6] functional failure is studied for a single MUX design without considering other design alternatives or device-level mitigation strategies such as transistor sizing to combat variation. FPGA group testing studied for hard faults can offer an emerging alternative [10].

The key logic structures for realizing SRAM-based FPGAs are SRAM cells and MUXes. Due to their ubiquitous applications, variation in SRAM cells has been extensively studied [11]. In the case of FPGAs, the effect of PV on SRAM cells is less limited since SRAM cells are not packed in an array structure in the same organization used in other custom VLSI designs, e.g. cache memory. In addition, SRAM cells in FPGAs are not often written; thus, the overhead imposed by any deployed technique to avoid PV-induced write failures can be negligible. To that end, we focus on MUX-based structures and consider two commonly accepted design options for implementing them in FPGAs [12]. Namely, the effect of variation on PT-based MUXes with half-latches and TG-based counterparts are compared.

#### A. Pass Transistor-based Multiplexers with Half-latch

Fig. 1(a) shows an example of 2:1 PT-based multiplexer. Two NMOS pass transistors  $t_0$  and  $t_1$  with complementary control inputs are used to select which input signal to pass to the multiplexer output. Due to their higher mobility, NMOS transistors are favored for relative driving strength over PMOS transistors. Since NMOS transistor passes a weak high logic level with a voltage swing ranging from 0 to  $V_{DD} - V_{th}$ , some restoration logic, or *half-latch*, is required to pull-up the

weak-1 output to recover a strong-1 level. The half-latch is an inverter with a pull-up transistor  $t_r$  controlled by the inverter output. When a weak-1 is propagated to the inverter input, inverter output transition to a low voltage will activate the pull-up transistor  $t_r$  to boost the inverter input to a strong-1 for a stable operation. The restoration logic can be placed after every two or more cascaded PTs in a large multiplexer, as depicted in Fig. 1(b), to reduce area and latency overhead at the expense of less reliable signal propagation. High variation in a PT-based MUX can lead to a low  $V_{DD} - V_{th}$  voltage difference insufficient to trigger the inverter, precipitating functional failure as demonstrated in Section IV.

#### B. Transmission Gate-based Multiplexers

A transmission gate is composed of NMOS and PMOS transistors connected in parallel to avoid the issue of passing a weak-1 or weak-0 at the expense of the added area of a PMOS transistor. The two transistors are controlled by two complementary signals to activate both during TG transparent state. Fig. 2(a) depicts the structure of a 2:1 TG-based MUX. Two TGs  $t_{g0}$  and  $t_{g1}$  are connected to complementary signals to select one of the two inputs. Restoration logic is not needed as TG can pass both strong-0 and strong-1.

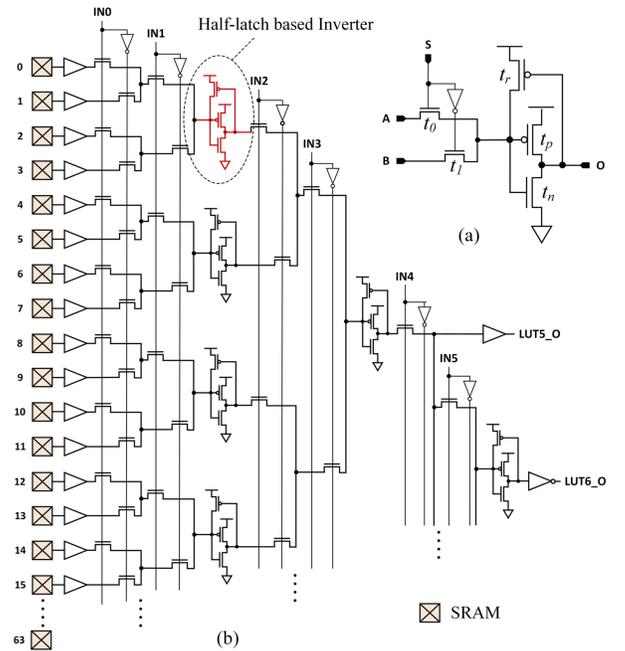


Fig. 1. Netlist for (a) 2:1 PT-based MUX (b) PT-based 6-input LUT (partial view)

### III. EVALUATION FRAMEWORK

To facilitate valid comparisons, a fracturable 6-input LUT that can be utilized as 6-input or 5-input LUT is adopted as a case study. 6-input LUTs are considered the optimal size in terms of area-delay product [13] and are also used in latest commercial FPGAs, e.g. Xilinx Virtex 7 and Altera Stratix V. Design and simulation for this evaluation are based on the High-K Metal-Gate (HK/MG) 16nm Predictive Technology Model (PTM) from Arizona State University. Moreover, the insight gathered from this case study can be generalized to other MUX-based FPGA structures. The 6-input LUT is

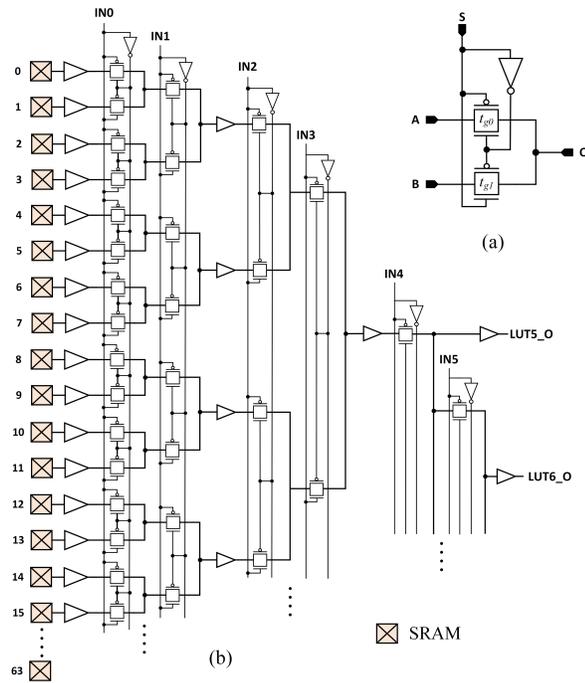


Fig. 2. Design Diagram for (a) 2:1 TG-based MUX (b) TG-based 6-input LUT (partial view)

implemented as a fully-encoded MUX tree using  $2^6 - 1 = 63$  multiplexers. Fig. 1(a) and Fig. 2(a) depict a partial view of PT-based and TG-based implementation. Internal re-buffering with proper sizes are also used to maintain optimal latency. The baseline sizes for NMOS and PMOS transistors are determined based on the optimal DC Voltage Transfer Characteristic (VTC) curve.

Initially, Cadence Virtuoso platform with Spectre simulator was used to determine optimal transistor sizes and extract corresponding threshold voltages for each sized device. The Gaussian random variable [14] was used to study the effect of WID variation on delay, efficiency, and output correctness. For each LUT implementation, 1,000 Monte Carlo samples are generated by assigning a random deviation in threshold voltage using a Gaussian distribution. Monte Carlo simulations are carried out using Synopsys HSPICE. To check transition faults at all MUX nodes, the commercial Synopsys TetraMAX APTG tool was used to generate the minimum number of test patterns that can be applied as LUT inputs and configuration values to check all possible transition at MUXes ports. The generated input sequence was used to test each Monte Carlo sample during SPICE simulation. Delays and power consumption data are collected for each sample. Due to the limited drive of pass transistors, different PT-based designs with increasing transistor sizes were included. The unit size parameter  $w$  signifies multiplication factor for transistor size whereby  $w = 2$  indicates twice its original  $w/l$  ratio.

#### IV. RESULTS AND CONCLUSIONS

Fig. 3 (a) and (b) display the obtained frequency distribution of delay values for the 1,000 Monte Carlo sample designs of PT-based and TG-based LUTs, respectively, at  $\sigma_{V_{th}} = 10\%$ . The red vertical line in each figure indicates delay value for baseline design without variation  $d_{baseline}$ . It is evident that the

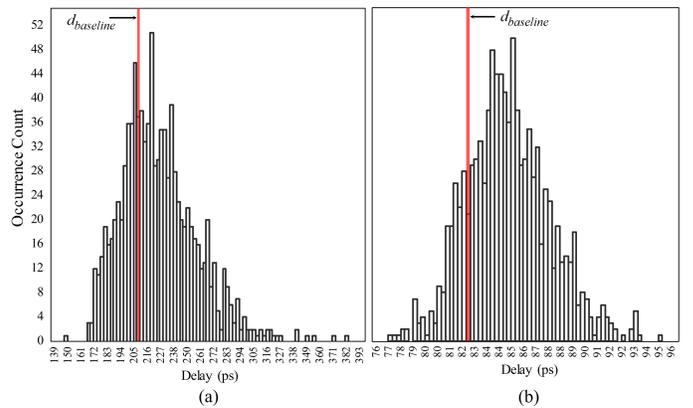


Fig. 3. Delay Distribution for (a) PT-based 6-Input LUT ( $w = 2$ ) (b) TG-based 6-Input LUT ( $w = 1$ ) at  $\sigma_{V_{th}} = 10\%$

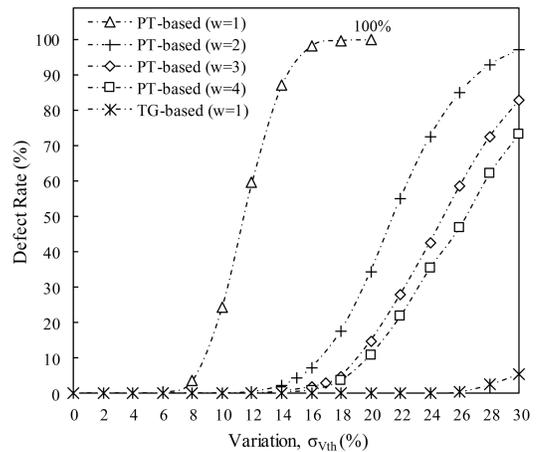


Fig. 4. LUT Defect Rate vs. Variation  $\sigma_{V_{th}}$  using 16nm PTM Model 1  $\leq w \leq 4$

effect of PV on performance follows a Gaussian distribution. It is also observed that the mean of delay distribution  $\mu_{V_{th}}$  is higher than delay of baseline design.

The test patterns used during simulation provide high coverage to check against transition failures caused by any failed multiplexer. The defect rate defined by the proportion of 1,000 LUT designs that fail at least one test pattern for different design implementations is given in Fig. 4. These simulation results reveal interesting observations. As expected, designing a structure using PT-based multiplexers without proper transistor sizing results in a substantially high failure rate that exponentially increases as variation increases beyond 6%. Results also show that sizing pass transistors has considerably decreased defect rate as seen in the design cases where  $w = 2, 3$ , and 4. A diminishing improvement in variation tolerance is also observed as  $w$  increases. On the contrary, TG-based structure offers much less sensitivity to variation than any PT-based design in this evaluation. This is achieved while using the minimum optimal  $w/l$  ratio. The results here are restricted to a maximum variation of 30% which covers the ITRS expected variation range for current and future technology.

Fig. 5 shows variation impact on mean delay for Monte Carlo simulations across variation range where functional yield  $\geq 95\%$ , i.e. defect rate is less than 5%. The TG-

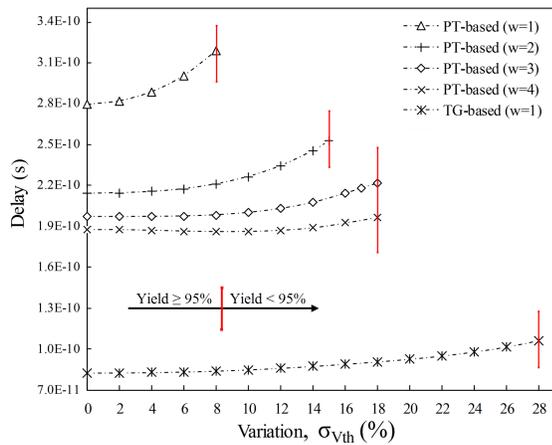


Fig. 5. Effect of Variation on Mean Delay using PT and TG MUXes.

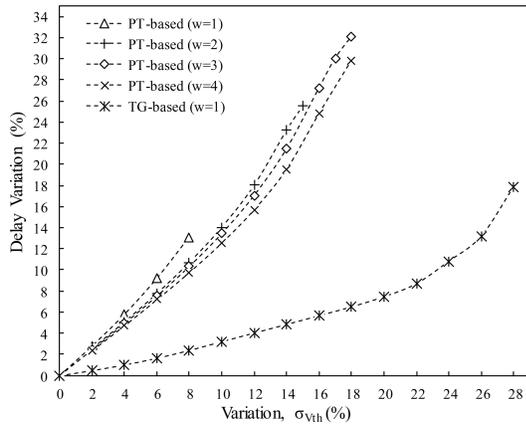


Fig. 6. Design Delay Variation vs.  $\sigma_{V_{th}}$  of PT and TG MUXes.

based implementation maintains another significant advantage in terms of latency. For instance, at  $\sigma_{V_{th}} = 14\%$ , TG-based LUT enables 64.4% (57.9%) reduction in latency compared to PT-based alternatives for  $w = 2$  ( $w = 3$ ). Transistor sizing for PT-based designs allows expected reduction in delay at a diminishing rate as  $w$  increases.

The effect of threshold voltage variation  $\sigma_{V_{th}}$  on delay variation  $\sigma_{delay}$  is shown in Fig. 6. The results reveal a pseudo-linear relation with exponential amplification under high variation. PT-based designs are indeed much more susceptible to variation than TG-based structures. Results also show that transistor sizing for pass transistors has a minor effect on mitigating design delay variation. At  $\sigma_{V_{th}} = 14\%$ , PT-based designs that meet the 95% yield objective can exhibit as high delay variation as 23.3%. This variation impact can be reduced to 4.9% if TG-based LUT is considered. Variation impact on design efficiency given by EDP is shown in Fig. 7. We observe that the EDP increases with variation while TG-based design provides substantially lower EDP than any PT-based design. Results show that TG-based designs offer a substantially superior resilience to WID variation compared to other PT-based alternatives. This is achieved while using the optimal minimal transistor sizing.

#### REFERENCES

[1] K. Kuhn, M. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. Ma, A. Maheshwari, and S. Mudanai, "Process technology variation," *IEEE*

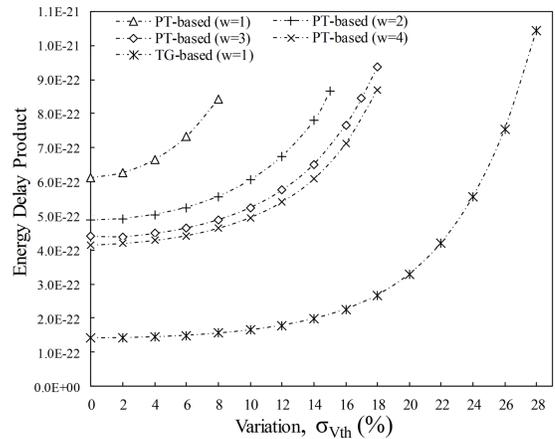


Fig. 7. Energy-Delay Product vs.  $\sigma_{V_{th}}$  for PT and TG MUXes.

*Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug 2011.

- [2] K. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in *IEEE International Electron Devices Mtg*, Wash., DC, Dec. 2007, pp. 471–474.
- [3] "Design for manufacturability technology requirements table," International Technology Roadmap for Semiconductors ITRS, Tech. Rep., 2011. [Online]. Available: [http://www.itrs.net/ITRS\\_1999-2014\\_Mtgs\\_Presentations\\_&Links/2011ITRS/2011Tables/Design\\_2011Tables.xlsx](http://www.itrs.net/ITRS_1999-2014_Mtgs_Presentations_&Links/2011ITRS/2011Tables/Design_2011Tables.xlsx)
- [4] C. Chiasson and V. Betz, "Should FPGAs abandon the pass-gate?" in *Proc. Intl. Conf. on Field Programmable Logic and Appl.*, Porto, Portugal, Sep. 2013, pp. 1–8.
- [5] C. Chen, R. Parsa, N. Patil, S. Chong, K. Akarvardar, J. Provine, D. Lewis, J. Watt, R. T. Howe, H.-S. P. Wong, and S. Mitra, "Efficient fpgas using nanoelectromechanical relays," in *Proc. 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays (FPGA'10)*, Monterey, CA, Feb. 2010, pp. 273–282.
- [6] A. DeHon and N. Mehta, "Exploiting partially defective LUTs: Why you don't need perfect fabrication," in *Proc. International Conference on Field-Programmable Technology (FPT'13)*, Kyoto, Japan, Dec. 2013, pp. 12–19.
- [7] Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S.-H. Lo, G. Sai-Halasz, R. Viswanathan, H.-J. Wann, S. Wind, and H.-S. Wong, "CMOS scaling into the nanometer regime," *Proceedings of the IEEE*, vol. 85, no. 4, pp. 486–504, Apr. 1997.
- [8] A. Kumar and M. Anis, "FPGA design for timing yield under process variations," *IEEE Trans. VLSI Syst.*, vol. 18, no. 3, pp. 423–435, Mar. 2010.
- [9] H.-Y. Wong, L. Cheng, Y. Lin, and L. He, "FPGA device and architecture evaluation considering process variations," in *Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD'05)*, San Jose, CA, Nov. 2005, pp. 19–24.
- [10] S. A. Sharma, Carthik A., A. Alzahrani, and R. F. Demara, "Self-healing reconfigurable logic using autonomous group testing," *Microprocessors and Microsystems*, vol. 37, no. 2, pp. 174–184, Mar. 2013.
- [11] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscale CMOS," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 12, pp. 1859–1880, Dec. 2005.
- [12] T. Pi and P. J. Crotty, "FPGA lookup table with transmission gate structure for reliable low-voltage operation," Dec. 23 2003, US Patent 6,667,635.
- [13] E. Ahmed and J. Rose, "The effect of LUT and cluster size on deep-submicron FPGA performance and density," *IEEE Trans. VLSI Syst.*, vol. 12, no. 3, pp. 288–298, Mar. 2004.
- [14] S. Nassif, K. Bernstein, D. Frank, A. Gattiker, W. Haensch, B. Ji, E. Nowak, D. Pearson, and N. Rohrer, "High performance CMOS variability in the 65nm regime and beyond," in *IEEE International Electron Devices Meeting*, Wash., DC, Dec. 2007, pp. 569–571.