Long-Tail Behavior of Process Variation with Application to Domino Keeper Sizing

Ben Keller and David Money Harris
Harvey Mudd College
Claremont, California 91711
Email: {ben_keller, david_harris}@hmc.edu

***Abstract*—Designers use Monte Carlo simulations to evaluate the impact of variability on circuits, but such simulations require prohibitive amounts of computation to characterize rare events. In this paper, we propose a method by which the long tail behavior of circuits can be modeled with a reasonable number of simulations. This technique is then applied to the problem of domino keeper sizing to determine the sizing necessary to ensure a reliable circuit. We find that to ensure reliability for a commercial 45 nm process, the width of the keeper must be 0.17 times the effective width of the pull-down stack. Such a wide keeper results in a delay penalty of 9.9% compared to a circuit with no keeper.**

I. Introduction

Variability increasingly affects the reliability, yield, and performance of modern CMOS processes [1] [2]. Accordingly, statistical methods have become a necessary technique in circuit design, and are widely employed to measure performance characteristics [3]. Monte Carlo simulation is a straightforward method of capturing such probabilistic behavior.

Modeling the extremes of device variation is important in systems with many devices because a single low-probability deviation can result in failure of the entire system. Unfortunately, Monte Carlo simulation cannot accurately capture this long-tail behavior unless a very large number of simulations are computed, a process which can take prohibitively long.

Subthreshold leakage currents are particularly sensitive to process variation because they vary exponentially with threshold voltage [4]. In general, a 100 mV threshold variation causes an order of magnitude increase in leakage. This problem impacts the sizing of domino keepers. Even though leakage through OFF transistors is nominally insignificant, extreme variations in leakage mandate a strong keeper.

A recent probabilistic approach to quantifying domino circuit leakage [5] avoids extensive Monte Carlo simulation but requires detailed knowledge of the distributions and correlations of every physical parameter contributing to leakage. This paper avoids the requirement of parameter extraction by combining a reasonable number of Monte Carlo simulations with curve-fitting techniques to describe the long-tail behavior. Another approach, known as "importance sampling," oversamples the long tail of the distribution and then weights these samples to match the true probabilistic behavior [6]. This approach does not require a complete characterization of the underlying distribution, but still requires an assumption of the overall form of the distribution. The technique presented in this paper does not require such assumptions about the nature of the underlying probability distribution.

Domino circuits require keepers to improve resilience to leakage and noise. A strong keeper degrades domino circuit performance, reducing its advantage over static circuits. Various circuit techniques to alleviate this problem have been proposed and surveyed in [7], but the conventional static keeper remains standard industry practice [8]. One common modification of the canonical design is the addition of a separate inverter on the feedback line [9]. The separate feedback inverter improves the noise resilience of the circuit by isolating the keeper from noise on the output node. Since this inverter is only driving the keeper, it can be sized minimally, and so has a negligible effect on energy and delay.

II. Yield Analysis

Consider an chip with M circuits, where the failure of a single circuit results in the failure of the entire chip. We wish to determine the value of a circuit parameter r (such as the ratio of the keeper size to the size of the pulldown stack) to achieve acceptable yield. Let $\epsilon_i(r)$ be the probability that a particular circuit on chip i fails; this probability is typically very low. Then the yield $Y_i(r)$ for chip i is the probability that any circuit on that chip fails:

$$Y_i(r) = (1 - \epsilon_i(r))^M \approx 1 - M\epsilon_i(r). \tag{1}$$

The average overall yield $Y(r)$ across many such chips is then given by

$$Y(r) = \frac{1}{N}\sum_{i=1}^{N} Y_i(r) \approx \frac{1}{N}\sum_{i=1}^{N}(1 - M\epsilon_i(r)) = 1 - M\bar{\epsilon}(r), \tag{2}$$

where $\bar{\epsilon}(r)$ is the average probability of failure of a single circuit. Even though circuits on the same chip may display spatial correlation, this result proves that analyzing a single circuit is sufficient to predict yield.

Circuit designers generally desire that special-purpose circuits degrade yield negligibly. For example, 3σ reliability is a common design goal [10]. Therefore, we wish to find r such that the chip yield $Y(r)$ = .9987, the 3σ probability of the normal cumulative distribution function. If we let $M = 10^4$, this yields a very small $\bar{\epsilon}(r) = 1.3 \cdot 10^{-7}$. Approximately 10^8 Monte Carlo simulations would be required to generate an

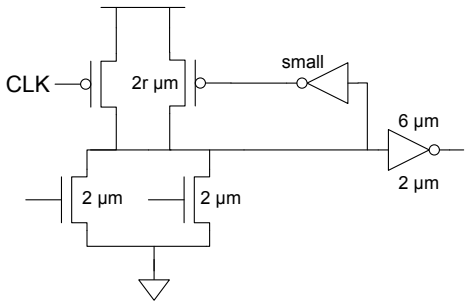

Fig. 1. Wide-NOR domino gate.

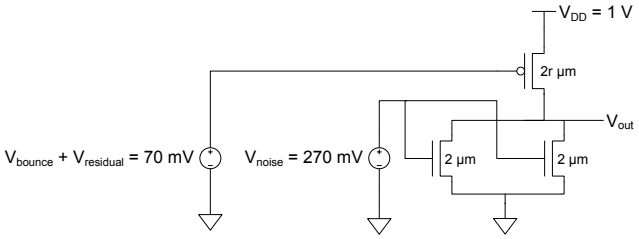

Fig. 2. Simulated circuit.

empirical distribution that could be used to accurately estimate failure probabilities of this magnitude.

III. Proposed Simulation Method

A more computationally efficient method of estimating low-probability events involves fitting a shifted exponential distribution to the tail of an empirical distribution. While the exact nature of the empirical distribution may be unknown, the use of this so-called "quasiempirical" distribution is justified because the tail of a wide class of distributions behaves as an exponential function [11]. After finding the empirical cumulative distribution function (CDF) of the simulated data, the last k observations are replaced with a shifted exponential function, where k is typically small relative to the number of samples. This exponential curve is chosen so that overall expected value of the distribution remains the same.

The overall process for constructing this quasiempirical distribution is as follows [11]: first, order the data points X_1 through X_n and construct the empirical CDF. Then, replace the last k points with an exponential function, such that the cumulative density $F(x)$ at any value of the random variable x in this exponential range is given by

$$F(x) = 1 - (k/n)\exp\left(-(x - X_{n-k})/\theta\right), \tag{3}$$

where

$$\theta = \left(X_{n-k}/2 + \sum_{i=n-k+1}^{n} (X_i - X_{n-k})\right)/k. \tag{4}$$

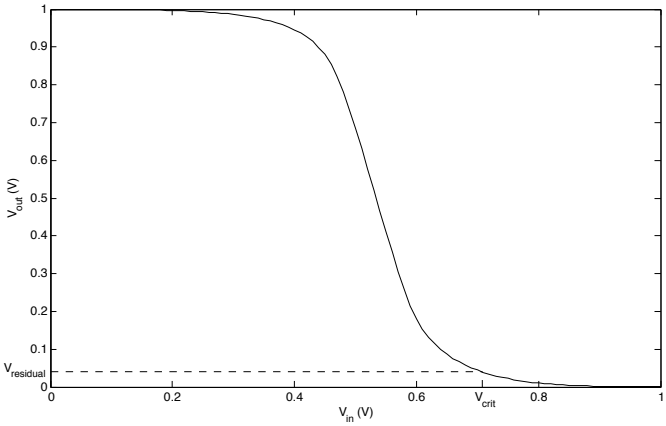

Fig. 3. DC transfer characteristics of high-skew inverter.

IV. Application to Domino Keepers

We will now apply this yield model to domino circuits. In general, we are interested in the leakage behavior of arbitrarily complex domino gates. In the worst case, only a single transistor in each parallel stack is OFF . The OFF transistors subject to crosstalk noise dominate the overall leakage through the pull-down network. In domino circuits with 1-of-n encoding and bundled routing [8], at most two victim inputs will see crosstalk noise caused by the single aggressor. Therefore, a two-input domino OR gate is representative of the worst-case leakage behavior of an arbitrary domino gate. Such a gate is shown in Fig. 1. The P/N ratio of the keeper relative to a pull-down stack is r. The high-skew inverter has a P/N ratio of 3:1.

Domino circuits were simulated in HSPICE on both a low power and a high speed 45 nm process at a temperature of 125°C and a supply voltage $V_{DD} = 1$ V. We used the test setup in Fig. 2. Normal V_t transistors were used in all cases. Leakage is strongly affected by noise on the dynamic inputs. We employ a noise condition with three contributing components:

$$V_{\text{noise}} = V_{\text{bounce}} + V_{\text{residual}} + V_{\text{coupling}},$$

where $V_{\text{bounce}} = 30$ mV represents the ground bounce, $V_{\text{residual}} = 40$ mV is the residual noise caused by noise at the input to the previous stage, and $V_{\text{coupling}} = 200$ mV is the coupling noise. Our worse-case "noisy corner" therefore applies a total of $V_{\text{noise}} = 270$ mV on each coupled input. Since the keeper is isolated by the separate feedback inverter, it is not affected by coupling on the output node. The noise on the keeper input is therefore limited to 70 mV. While the transient characteristics of the noise spike will likely have some impact on the necessary keeper size [12], we assume for the sake of simplicity that V_{noise} is effectively a DC level.

Circuit failure occurs when leakage pulls V_{out} below V_{crit}, the level that results in V_{residual} at the output of the subsequent high-skew inverter. DC simulation of a high-skew inverter finds the nominal value of V_{crit} to be 629 mV for the low power process and 708 mV for the high speed process. The

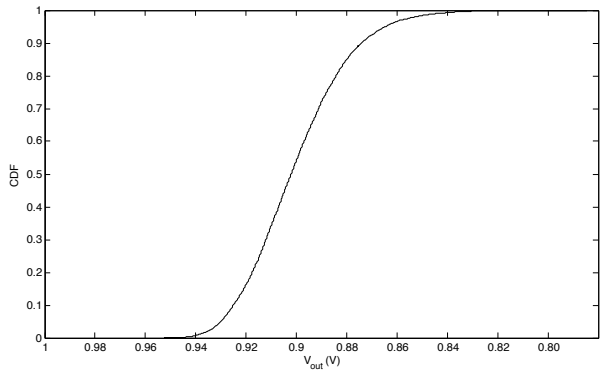

Fig. 4. Empirical CDF constructed from Monte Carlo simulations.

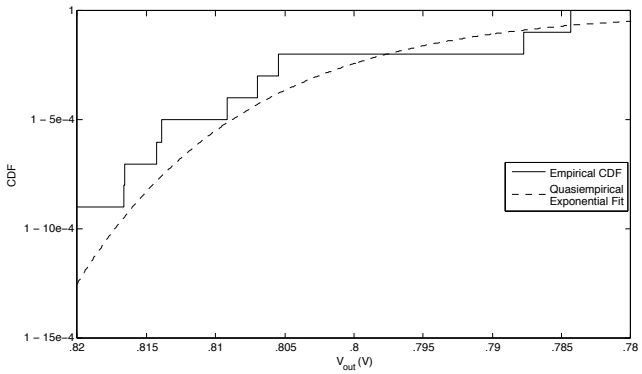

Fig. 5. Empirical CDF with exponential fit.

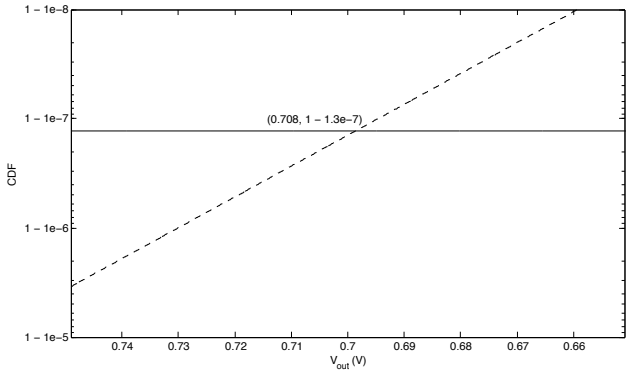

Fig. 6. Exponential fit on a semi-logarithmic scale.

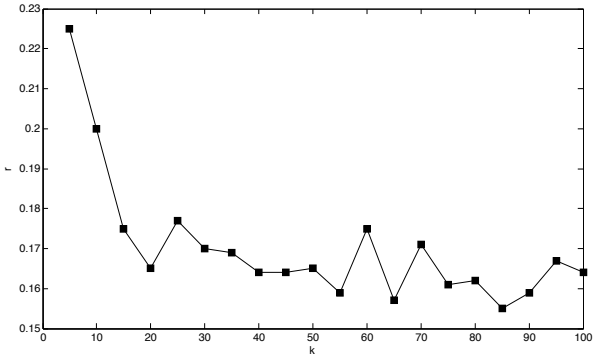

Fig. 7. Dependence of the fitting technique on k.

DC transfer characteristics of the high-skew inverter for the high speed process are shown in Fig. 3. The input voltage that yields and output voltage of V_{residual} is marked. We ignore the impact of variation on the transfer characteristics of this inverter.

V. Results

We applied the proposed simulation method to a two-input domino OR gate under the conditions described above. We performed 10^4 Monte Carlo simulations to obtain empirical CDFs.

The proposed simulation method allows an accurate determination of V_{out} for a given keeper size. However, we wish to determine an appropriate P/N ratio r such that the probability that $V_{out} < V_{crit}$ is less than ϵ. To accomplish this, iterative simulations were performed, adjusting the size of the keeper until $V_{\text{out}} = V_{\text{crit}} = 708$ mV for the high speed process. This condition was satisfied at $r = 0.17$.

Fig. 4 shows the CDF of the high speed process with a standard keeper using $r = 0.17$. Even with 10^4 simulations, the output level never approached the failure point V_{crit}. Fig. 5 shows the exponential fitting method applied to the empirical CDF with $k = 50$. The exponential tail allows the quasiempirical CDF to predict lower probability events. Fig. 6 shows the quasiempirical CDF on a semi-logarithmic scale. The probability of failure $\bar{\epsilon}(0.17)$ is achieved at $V_{\text{out}} = V_{\text{crit}} = 708$ mV.

Evaluating the curve-fitting technique with different k can yield different results for the expected V_{out} at $\bar{\epsilon}(r)$. Fig. 7 shows the r necessary to satisfy $V_{\text{out}} = V_{\text{crit}} = 708$ mV using various k from 5 to 100. These r were determined by iterative simulation as described above. If $k \geq 15$ is chosen, the results of the simulation remain relatively stable, varying by no more than 10% from $r = 0.17$. Thus, the specific choice of k is unimportant as long as it satisfies this condition and is still small relative to the total number of points in the empirical distribution.

We also performed iterative simulation using this technique with a larger number Monte Carlo simulations ($5 \cdot 10^4$). The iteratively determined best keeper size differed by less than 10%, demonstrating the stability of our curve fit. Additional validation was performed by generating a huge number of Monte Carlo simulations (10^8), sizing the keeper such that $r = 0.17$. Fig. 8 compares the empirical distribution generated by this computation to the quasiempirical distribution shown in Figs. 5 and 6. The generated long tail tracks plausibly with the simulated empirical distribution in the range of interest.

As it becomes clear that a large keeper may be required for a reliable domino circuit, it is instructive to consider the impact of keeper sizing on delay, as shown in Fig. 9. The keeper delay penalty is the increase in delay for a reliable gate as compared to a circuit with no keeper. Each stage drives a fanout of four, for a total fanout of sixteen across the gate. The delay of FO16 static buffers (back-to-back inverters) using normal and low V_t

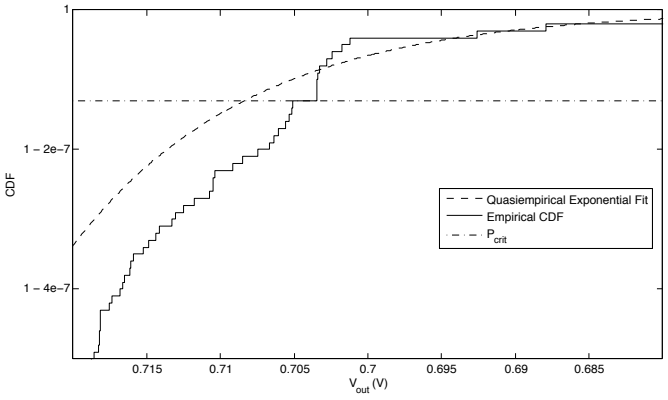

Fig. 8. Comparison of exponential fit to empirical validation.

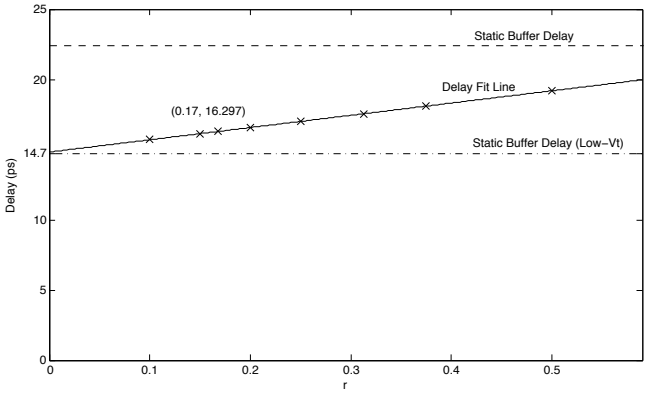

Fig. 9. Delay for 45 nm high speed process.

TABLE I
SIMULATION RESULTS

Process	P/N Ratio	Delay Penalty
45 nm High Speed	0.17	9.9%
45 nm Low Power	0.038	2.2%

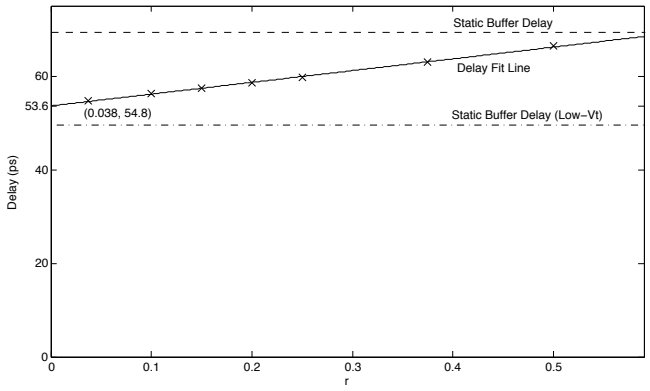

Fig. 10. Delay for 45 nm low power process.

transistors are shown for reference. Extrapolating the data to $r = 0$ estimates the performance of a domino gate with no keeper. The advantage of domino circuits over static logic for simple gates is substantially reduced for this process, although high-fanin structures would still see benefit. Fig. 10 shows the same information for the low power process.

Table I gives the iteratively determined P/N ratio and keeper delay penalty for the high speed and low power processes. The low power process has much lower leakage, so the keeper size and delay penalty are negligible.

VI. CONCLUSIONS

We have presented a technique by which the long-tail behavior of circuits can be evaluated with a reasonable number of Monte Carlo simulations without making any assumptions about the underlying probability distribution. We then applied this technique to the problem of domino keepers to determine the keeper sizing that ensures a reliable domino circuit. The keeper in a modern high speed process must be undesirably wide to cope with leakage, especially under severe process variation. This significantly impacts the performance of the circuit, causing in a 9.9% delay penalty compared to a domino circuit with no keeper. The same technique could be applied to other yield problems such as voltage scaling in SRAM cells.

ACKNOWLEDGMENTS

We would like to thank Intrinsity Inc. and the Clay Family Foundation for their financial support, and Michael Runas of Intrinsity for his advice and contributions.

REFERENCES

[1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Design Automation Conference*, Anaheim, CA, June 2–6, 2003, pp. 338–342.

[2] S. Nassif *et al.*, "High performance cmos variability in the 65nm regime and beyond," in *Electron Devices Meeting*, Washington, DC, Dec. 10–12, 2007, pp. 569–571.

[3] A. Srivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*. New York: Springer, 2005.

[4] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester, "Statistical analysis of subthreshold leakage current for vlsi circuits," *IEEE Transactions on VLSI Systems*, vol. 12, no. 2, pp. 131–139, Feb. 2004.

[5] L. Brusamarello, R. da Silva, G. I. Wirth, and R. A. L. Reis, "Probabilistic approach for yield analysis of dynamic logic circuits," *IEEE Transactions on Circuits and Systems*, vol. 55, no. 8, pp. 2238–2248, Sept. 2008.

[6] G. K. Chen, D. Blaauw, T. Mudge, D. Sylvester, and N. S. Kim, "Yield-driven near-threshold sram design," in *International Conference on Computer-Aided Design*, San Jose, CA, Nov. 4–8, 2007, pp. 660–666.

[7] P. Zhao, J. McNeely, M. Bayoumi, G. Pradeep, and K. Weidong, "A low power domino with differential-controlled-keeper," in *International Symposium on Circuits and Systems*, New Orleans, LA, May 27–30, 2007, pp. 1625–1628.

[8] M. Runas and M. Seningen. (2006, Oct.) Noise mitigation of domino logic in nanoscale designs. [Online]. Available: http://www.intrinsity.com/images/stories/Papers/noise _mitigation.pdf

[9] M. Anders, R. Krishnamurthy, R. Spotten, and K. Soumyanath, "Robustness of sub-70nm dynamic circuits: Analytical techniques and scaling trends," in *Symposium on VLSI Circuits*, Kyoto, Japan, June 14–16, 2001, pp. 23–24.

[10] N. H. E. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. Boston: Pearson, 2010.

[11] P. Bratley, B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, 2nd ed. New York: Springer-Verlag, 1987.

[12] S. H. Choi and K. Roy, "Domino noise model: A new crosstalk noise model for dynamic logic circuits," in *SOC Conference*, Portland, OR, Sept. 17–20, 2003, pp. 207–210.