An All-Digital, Highly Scalable Architecture for Measurement of Spatial Variation in Digital Circuits

Nigel Drego, Anantha Chandrakasan, and Duane Boning
{*ndrego, anantha, boning*}*@mtl.mit.edu*
Massachusetts Institute of Technology, USA

***Abstract*— Increased variation in CMOS processes due to scaling results in greater reliance on accurate variation models in developing circuit methods to mitigate variation. This paper investigates specific variation parameters and their measurement approach for use in such models, leading to critical considerations in aggressive voltage scaling systems. We describe a test-chip in 90nm CMOS containing all-digital measurement circuits capable of extracting accurate variation data. Specifically, we use replicated 64-bit Kogge-Stone adders, ring-oscillators (ROs) of varying gate type and stage length and an all-digital, sub-picosecond resolution delay measurement circuit to provide spatial variation data for digital circuits. Measurement data from the test-chips indicate that 1) relative variation is significantly larger in low-voltage domains, 2) within-die variation is spatially uncorrelated, and 3) die-to-die (or global) variation is strongly correlated, but degrades toward uncorrelated as the power-supply voltage is lowered.**

I. Introduction

Control and mitigation of process variation is increasingly a major factor in scaled designs. Reduced channel dimensions and sub-wavelength lithography result in increased relative significance of Random Dopant Fluctuation (RDF), Line-Edge Roughness (LER) and channel length variability. Consequently, a combination of improved process control, accurate variation modeling and circuit/architecture mitigation techniques are necessitated to achieve robust designs. Design and implementation of effective mitigation techniques relies fundamentally on an accurate modeling strategy, without which such techniques are sub-optimal.

Often, mitigation techniques rely on high correlation between actual critical paths and monitor circuits, particularly if the mitigation technique is not in-situ. The modeling community has begun to pursue statistical models which include spatial correlations [1]–[3]. However, many of the proposed models have not been validated using real manufacturing data to our knowledge - likely due to the criticality to circuit design. Correlation data at a device parameter level (such as V_T or channel length) has been extracted and evaluated in the literature [4]–[7], but with the exception of [8], [9], carrying this data through to correlation in circuit performance has not been performed.

In this work, we provide an all-digital variation test-chip architecture (Section II), from which we are able to extract statistical variation data for use in and validation of variability models. In particular we extract spatial correlation of delay in digital circuits, cross-correlation between critical paths and "monitor" circuits, as well as accurate sub-picosecond resolution delay measurements, not only at nominal process voltages $(1.0-1.2V)$ but also in low-voltage $(\sim 0.5V)$ domains where the effects of variation are exacerbated. The data presented in Section III show no within-die spatial correlation across all voltage domains but do show strong die-to-die correlation, decreasing with power-supply voltage. Conclusions and implications for design are presented in Section IV.

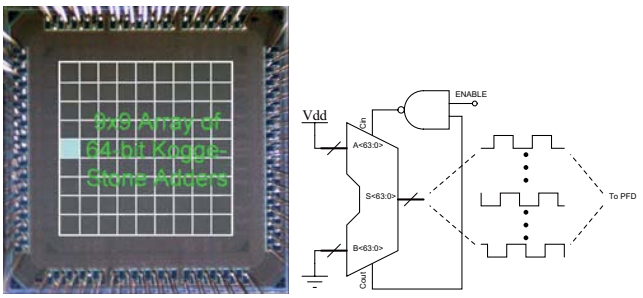
II. Variation Chip Architecture

The variation test-chip is composed of a large number of replicated blocks containing various digital circuits and all-digital measurement circuitry. The following subsections further detail the architecture of each component.

A. Extraction of Spatial Correlation

A correlation coefficient, ρ, quantifies the degree to which two devices (or circuits) vary together, and can take values $-1 \leq \rho \leq 1$, with -1 implying the circuits move oppositely in response to variation and 1 meaning they move in the same direction, to within a scale factor. Spatial correlation quantifies this as a function of separation distance, d. Understanding circuit performance correlation enables design of appropriate mitigation techniques ("replica" circuits if $|\rho|$ is high, or "in-situ" techniques if $|\rho|$ is low) and determination of spacing criteria between monitor circuits and potential critical paths to ensure high correlation.

To extract spatial correlation, we array 80 nominally identical "adder-blocks" in a $4mm^2$ area, each containing a 64-bit Kogge-Stone (KS) adder, implemented with standard cells, in a 9x9 matrix as shown in the die photo (Fig. 1a). The adder inputs are configured such that the outputs oscillate: $A\langle 63:0\rangle = 1, B\langle 63:0\rangle = 0, C_{in} = \overline{C_{out}}$, as shown in Fig. 1b. To capture correlation between different circuit



(a) Die Photo (b) Oscillating 64-bit Kogge-Stone adder

Fig. 1: Arrayed Kogge-Stone adders instrumented for internal delay measurement

structures, we include 16 ROs around each adder, since they are often used as monitor circuits due to their simplicity and small area overhead. The ROs used are: INV9,11,13,15, NAND9,11,13,15 and NOR9,11,13,15, denoting the type of gate and number of gate delays. The INV9, NAND11, NOR13 and INV15 are duplicated for a total of 16 ROs per adder. The NAND and NOR gates are two-input gates with inputs shorted to produce an inverting gate. These all differ from the AND-OR-INV gate used in the KS adder, allowing for quantification of correlation between disparate gate types and transistor stacks.

Asynchronous frequency counters made up of simple toggle flip-flops measure the oscillation frequency of the adder and ROs. Circuitry to measure the delay of each individual adder bit relative to the first bit in the adder is also included.

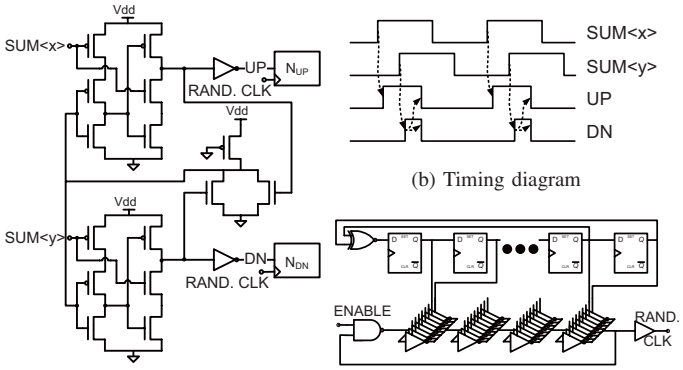
B. All-Digital Delay Measurement

Capturing variation data with higher spatial and temporal resolution than simple ROs requires alternative techniques. Specifically, we sought a highly-scalable, all-digital measurement technique capable of sub-picosecond delay resolution occupying small area. We describe the theory and operation of a random-sampling technique, in the context of measuring delays between bits of the adder, meeting these criteria.

The critical path in the adder is from C_{in} to C_{out}, resulting in all bits oscillating at an identical frequency but with delays determined by the logarithmic structure of the KS adder, giving 64 out-of-phase oscillators. Quantifying this phase-delay is equivalent to quantifying the difference in delay between each bit, and allows for variation analysis with improved spatial resolution. Delay measurement is done by randomly sampling the signals and counting the number of occurrences when one of the signals is logic high and the other simultaneously low, or vice-versa, dividing by the total number of samples taken and multiplying by the signals' period.

Use of random-sampling techniques for this purpose are not new [10]–[12]. However, in each of those implementations, an XOR gate is used to determine when one of the signals is logic high and the other logic low. XOR gates can limit the minimum detectable delay since some minimum delay is necessary to register a clear logic level at the output. In our implementation, we transform the measurement to be that of the difference in two pulse widths using a Phase-Frequency Detector (PFD) (Fig. 2a), which latches the rising edges of both input signals and simultaneously self resets both output signals following the rising edge of the delayed signal, as shown in Fig. 2b. The difference of these two pulse widths can be arbitrarily close to zero, eliminating any bound on minimum measureable delay. A random clock samples the UP/DN signals and stores the counts in the counters shown. A 5-stage RO, containing 32 tri-state drivers controlled by a Linear Feedback Shift Register (LFSR) producing pseudo-random bit sequences, generates the random clock (Fig. 2c).

For this technique to be accurate, the random samples must be uniformly distributed across all points in the sampled signal's cycle as non-uniformity in this distribution results in



(a) PFD [13] (b) Timing diagram (c) Random clock generation

Fig. 2: PFD and associated timing diagram

non-linearities in the measured delays. Mathematically, the operation shown in Fig. 3 can be modeled as:

$$S_{N+1} = S_N + X_N = \sum_{i=1}^{N} X_i \tag{1}$$

where S_N is the Nth sample, X_i is the time between samples and is a random variable, and t_{per} is the sampled signal's period. We also define $P_N = mod(S_N, t_{per})$ as the position of the Nth sample in the sampled signal's cycle . If μ_X and t_{per} are co-prime, it is clear that the distribution of P_N will be uniform even without random edges. However, if they are not co-prime, Eq. 1 describes a random walk, which we have simulated in Matlab using a distribution of random periods taken from Hspice simulations of the LFSR-controlled RO. The results of this simulation (shown in the inset plot of Fig. 3) empirically show convergence to a uniform distribution as $N \rightarrow \infty$. Clearly, with $N > 10^8$ samples the non-uniformity of the distribution, and thus non-linearity in the measured delay values, is significantly reduced. The theoretical framework in [12] is used in Section III-C to find standard errors and confidence intervals given $N \approx 1.7 \times 10^8$ samples. Furthermore, the time resolution of this technique can be arbitrarily increased with large enough N.

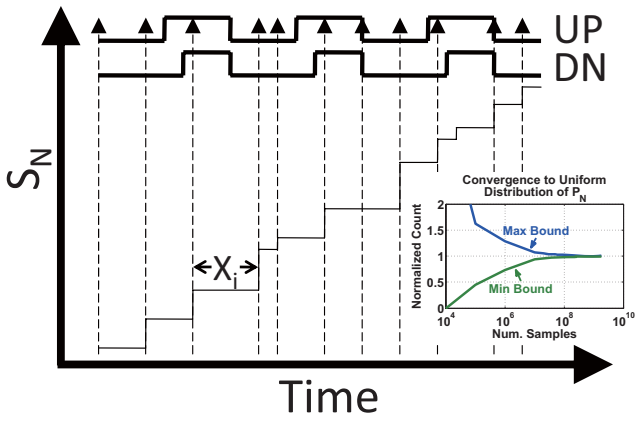


Fig. 3: Random sampling to determine pulse widths

III. Results and Data Analysis

A. Variation Measurements

The test-chip described above was fabricated in a 90nm process technology (Fig. 1a) and frequency measurements were carried out on 41 chips from two wafers. Variation measurements show that the effect of variation is larger at lower power-supply voltages (Fig. 4). As V_{DD} is decreased beyond 0.7V, both within-die and total variance increases dramatically with the within-die component becoming a larger and larger component of the overall variance. While noise is typically a concern at low-voltages, all of the measurements are taken by allowing the circuits to run freely for $> 10^8$ cycles, averaging out any (assumed white) noise.

As expected, ROs with smaller gates (e.g. INV) and fewer number of delay stages show greater within-die variation (due to less averaging, see below) than ROs with either larger gates (e.g. NOR) or more delay stages. At the extremes, the INV9 RO is most variable ($\sigma_{WID} = 4.9\%$ at 0.5V and $\sigma_{WID} = 1.13\%$ at 1.2V) and the adder, being the largest circuit, is least variable ($\sigma_{WID} = 2.8\%$ at 0.5V and $\sigma_{WID} = 0.82\%$ at 1.2V). Total variance shows the same general trends except that the NAND ROs are more variable than the INV ROs - we believe this may be due to larger NMOS than PMOS global variability as NAND gate delays are typically dominated by the series NMOS stack.

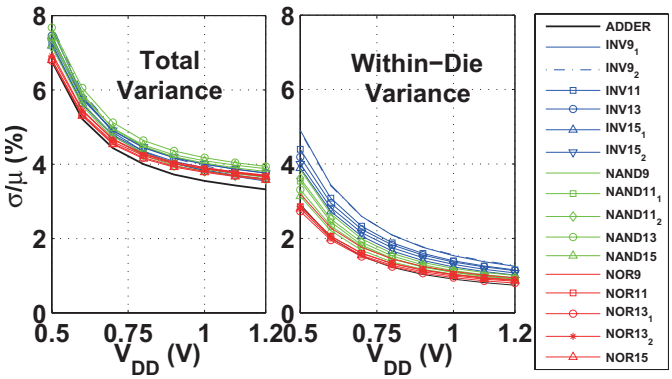


Fig. 4: Variation as a function of V_{DD}

B. Spatial Variation and Correlation

The large number of replicated circuits allows extraction of spatial correlation. We find that even at high V_{DD}, there is no discernible *within-die* spatial correlation (Fig. 5). The dip at the $V_{DD} = 1.2V$ and $2mm$ separation distance corner is likely due to limited data, as there are only 82 pairs of adders at this separation distance over 41 die. Furthermore, all points are within a 95% confidence interval, giving additional credence to this argument. Lack of within-die spatial correlation does not imply a lack of systematic within-die variation: rather, it simply means such variation is not a function of separation distance. Indeed, a systematic within-die pattern in adder frequency is noticed (Fig. 6). To elucidate this further, consider equidistant pairs of adders {(6,6), (6,7)} and {(7,6), (7,7)} that exhibit both unequal and directionally opposite frequency differences, $\triangle F(\{6,6\},\{6,7\})$ and $\triangle F(\{7,6\},\{7,7\})$, which implys lack of correlation.

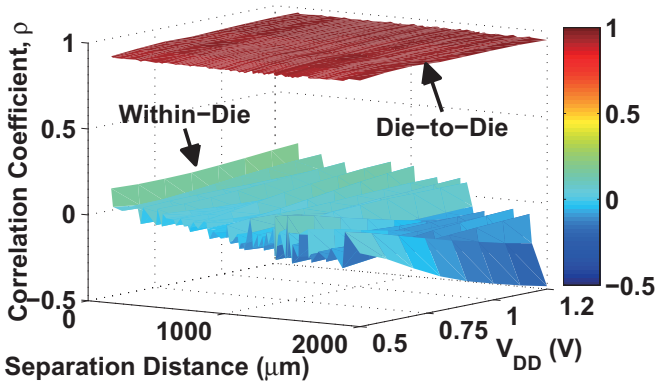


Fig. 5: Adder spatial correlation as a function of V_{DD}

Nevertheless, Fig. 5 also shows strong die-to-die correlation, at high V_{DD} but decreasing with V_{DD}, indicating that the effect of random variation increases at lower power-supply voltages. This is consistent with the effect of threshold voltage (V_T) variation, which is dominated by Random Dopant Fluctuation (RDF), increasing as gate overdrive decreases [14]. Furthermore, die-to-die correlation shows only weak dependence on separation distance. While Fig. 5 only shows adder correlations, we find similar spatial correlation results for all of our ring-oscillators as well, with the notable exception that die-to-die correlation decreases with decreasing V_{DD} more quickly with smaller circuit size. As an example, the INV9 based RO has a correlation coefficient, ρ, of 0.55 at $V_{DD} = 0.5V$ compared with $\rho = 0.65$ for the INV15 and $\rho = 0.75$ for the adder, consistent with smaller circuits being more susceptible to random variation sources as less averaging occurs.

C. Adder Bit Delays

Using the all-digital delay measurement circuits described in Section II-B, we measure the delay of each bit within each KS adder, relative to $Sum\langle 0\rangle$ of the same adder. All 64 bits,

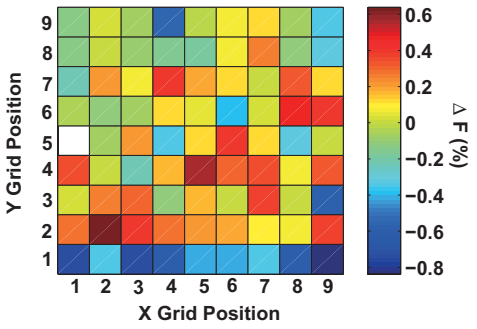


Fig. 6: Average within-die spatial variation pattern (Per location $\triangle F$ from die mean averaged over all die)

including $Sum\langle 0\rangle$, are muxed into a single PFD. By using the same PFD for all bits, offsets and other non-idealities due to mismatch in the PFD structure can be measured (by measuring the delay between the muxed and non-muxed versions of $Sum\langle 0\rangle$) and subtracted from subsequent measurements.

Measurements over all 80 adders on 40 chips are shown in Fig. 7 with a post-layout extracted simulation for comparison. There is good agreement between measured delays and simulation: the upward shift between measured data and simulation indicates a slower process than nominal simulation, and is consistent with 20% slower frequency measurements than nominal post-layout simulations. The measured delay pattern is consistent with simulation for all but three of the bits: $Sum\langle 16, 32, 48\rangle$, which are typically the fastest due to the logarithmic structure of a KS adder. However, in this layout, these three bits contain longer wires than the other bits, corresponding to the peaks in the post-layout simulation. While these three bits do have larger measured delays than nearly all of the other bits, the difference is not as large as in simulation, possibly due to slower transistors but "faster" wires, which would decrease the delay peaks formed by these three bits. Spatial correlation analysis of identical bits between different adders is consistent with the lack of correlation found previously. Each data point in Fig. 7 consists of no less than

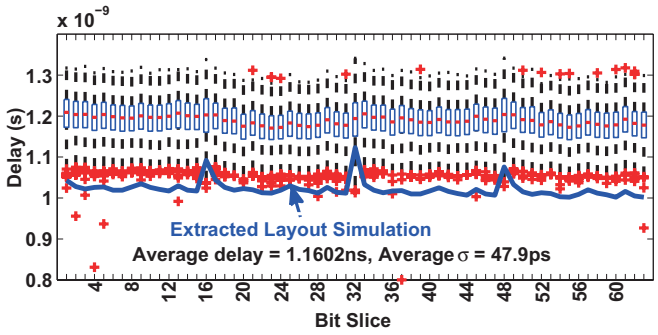


Fig. 7: Bit delay measurements relative to $Sum\langle 0\rangle$

1.7×10^8 random samples. Using Eq. 3 in [12], we compute a possible observed error of $\approx 200fs$, with a confidence level of 99.9999%. Combined with the general agreement between measured data and simulation, this computation gives high confidence in sub-picosecond accuracy of the measured data.

D. Implications for Variation Mitigation

The data suggests that variation mitigation strategies must be a function of the voltage/power domain in which circuits are operated. In high-performance domains where gate over-drive is sufficiently large ($> 2V_T$), although within-die variation is random and uncorrelated spatially, the absolute variance is also small. Mitigation schemes involving small "monitor" circuits, such as ring-oscillators or replica critical paths, require little margin - perhaps as little as 1%, as determined by the absolute variance and desired confidence level - to be effective.

In low-performance, low-power domains in which $V_{DD} \approx V_T$, variation as a percentage of the mean is significantly increased, as much as $5\times$ relative to high-performance voltage domains. Furthermore, within-die variation is a significant component of the overall variation seen. In combination, these two factors necessitate in-situ circuits capable of accurate measurement of timing data as the basis of a robust mitigation strategy. All-digital random sampling techniques with low area-overhead (especially if clock generation and PFDs are multiplexed over multiple critical circuits) provide accurate measurements but at the cost of delayed feedback due to the large number of samples required.

IV. Conclusion

We have implemented a 90nm CMOS test-chip capable of characterizing spatial variation in digital circuits. This test-chip also included a sub-picosecond resolution, all-digital delay measurement circuit using random sampling techniques for use in accurate variation characterization. While similar to prior published random-sampling methods, our method has no lower-bound on the minimum measurable delay. The measured data from these circuits shows that variation is a function of power-supply voltage and is random and spatially uncorrelated within-die. Finally, effective variation mitigation techniques must consider all possible circuit operating ranges.

Acknowledgment

The authors acknowledge the support of the Focus Center for Circuit & System Solutions (C2S2), one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation Program. The authors also thank D. Lim, K. Balakrishnan, N. Verma, and Y. Ramadass and DARPA/TAPO for chip fabrication.

References

[1] A. Agarwal, *et al.*, "Statistical Delay Computation Considering Spatial Correlations," *ASP-DAC*, Jan. 2003.
[2] A. Agarwal, *et al.*, "Statistical Timing Analysis for Intra-Die Process Variations with Spatial Correlations," *ICCAD*, Nov. 2003.
[3] H. Chang *et al.*, "Statistical Timing Analysis Under Spatial Correlations," *IEEE TCAD*, Sept. 2005.
[4] B. Cline, *et al.*, "Analysis and Modeling of CD Variation for Statistical Static Timing," *ICCAD*, 2006.
[5] K. Agarwal, *et al.*, "A Test Structure for Characterizing Local Device Mismatches," *VLSI Symp.*, 2006.
[6] S. Mukhopadhyay, *et al.*, "Statistical Characterization and On-Chip Measurement Methods for Local Random Variability of a Process Using Sense-Amplifier-Based Test Structure," *ISSCC*, Feb. 2007.
[7] R. Rao, *et al.*, "A Completely Digital On-Chip Circuit for Local-Random-Variability Measurement," *ISSCC*, Feb. 2008.
[8] L.-T. Pang *et al.*, "Impact of Layout on 90nm CMOS Process Parameter Fluctuations," *VLSI Symp.*, 2006.
[9] M. Kotani, *et al.*, "A 90nm 8x16 LUT-based FPGA Enhancing Speed and Yield Utilizing Within-Die Variations," *ESSCIRC*, Sept. 2006.
[10] S. Maggioni, *et al.*, "Random Sampling for On-Chip Characterization of Standard-Cell Propagation Delay," *ISQED*, Mar. 2003.
[11] R. Bhatti, *et al.*, "Duty Cycle Measurement and Correction Using a Random Sampling Technique," *IEEE MWSCAS*, vol. 2, Aug. 2005.
[12] R. Bhatti, *et al.*, "Phase Measurement and Adjustment of Digital Signals Using Random Sampling Technique," *ISCAS*, May 2006.
[13] W.-H. Lee, *et al.*, "A High Speed and Low Power Phase-Frequency Detector and Charge-Pump," *ASP-DAC*, Jan. 1999.
[14] N. Drego, *et al.*, "Lack of Spatial Correlation in MOSFET Threshold-Voltage Variation and Implications for Voltage Scaling," *Submitted to TSM*, 2008.