Understanding Vmin Failures for Improved Testing of Timing Marginalities

Adit D. Singh
Department of Electrical and Computer Engineering
Auburn University
Auburn, AL 36849 USA
adsingh@auburn.edu

***Abstract*— There has been speculation that the source of many of the unpredictable and hard to diagnose intermittent errors being increasingly observed in operation are timing marginalities, accentuated in low voltage operation, that escape detection during test. To investigate this possibility, we present a comprehensive study, combining analytical modeling with simulation, of the impact of random process variations on the timing of CMOS gates and circuit paths when operating at significantly reduced voltages. Our analysis is validated with the help of production test data recently published by a large industrial team for an advanced FinFET technology [1]. A key, somewhat unexpected, observation from this study is that virtually all variability paths that are statistical outliers, slow enough to cause timing failure, contain a single extremely weak transistor which contributes a large share of the increased delay. This suggests that TDF timing tests, that target localized "lumped" delay defects, may also detect many timing failures caused by distributed delays from process variations. The perceived need for path delay testing to target such failures is at least partially mitigated. We further show how the results of the analysis in this paper can be leveraged for conditioning the voltage and timing of the applied TDF scan tests to help enhance detection of such marginal timing parts, thereby reducing test escapes and minimizing system level tests.**

Keywords—Timing failures, process variations, Vmin fails

I. INTRODUCTION

Traditional scan structural tests frequently fail to achieve DPPM (defective parts per million) targets for SOCs fabricated in state-of-the art technologies, despite the introduction of advanced new fault models for test generation in recent years. Most notably, these new test generation methods include defect-oriented Cell Aware Tests (CAT) [2], and timing aware delay tests [3]. Consequently, functional system level tests (SLTs) [4] are increasingly needed as a final test screen to eliminate scan test escapes. However, there is growing evidence of malfunctioning ICs evading even the SLT screens and finding their way into parts deployed in operation. At a recent presentation at the Hot Topics in Operating Systems (HotOS) 2021 conference, Google engineers reported that some ICs do not always perform their calculations as expected [5]. "These machines were credibly accused of corrupting multiple different stable well-debugged large-scale applications. Each machine was accused repeatedly by independent teams, but conventional diagnostics found nothing wrong with them." The Google researchers analyzing these silent execution errors concluded that "mercurial cores" were to blame – CPUs that miscalculated occasionally, under different circumstances, in a way that defied prediction. Other companies operating large server farms are also reporting similar failures [6].

This research is an attempt to better understand failures that escape both advanced scan methodologies and functional system level test, so that test effort can be better directed towards the largest contributors of scan test escapes and field failures. This is challenging because of the lack of detailed test, failure, and diagnosis data from advanced test processes in the public domain [7]. Some of the silicon failures most difficult to detect during test are those that occur infrequently and only under very specific circuit operating conditions, such as the mercurial failures discussed above. These often result from timing marginalities caused not by hard defects, but by manufacturing processes variations, that are accentuated by some combination of switching noise and adverse voltage and temperature conditions. Failures resulting from an interaction of distributed delays from process variability are, by their very nature, virtually impossible to locate accurately in silicon through diagnostic tests, and consequently hard to confirm using physical failure analysis. Since their occurrence is often random and unpredictable, it is critical to better understand such failures so that their activation conditions can be properly characterized for test development. This is essential for ensuring that marginal and unstable circuit behavior can be efficiently screened out using carefully calibrated test conditions and testing margins, without aggressive overtesting which can result in excessive yield loss.

Given the paucity of industrial data from advanced nodes in the public domain, in this work we exploit some recently published test data from Intel's 14nm FinFET technology [1], in a manner quite unrelated to its use in the original paper, which was to study the effectiveness of the Cell Aware Test methodology and its "defect-oriented" extensions. Our aim here is to understand and characterize an important hard-to-detect failure mode that is being observed in significant numbers at advanced technology nodes. These "Vmin only" failures are observed only at low voltages, very close to the specified Vmin for a given operating frequency, while passing all tests at higher VDD. Due to the inevitable presence of power supply noise, both during test and in operation [8], such parts can display marginal and unpredictable behavior. There is speculation that many such failures escape scan tests and contribute significantly

to the fall-out at system level tests; some parts likely remain undetected even during SLT and are the cause or unexplained intermittent errors in the field. Failures in low voltage operation are a particular concern because many processor SOCs today have multiple power saving operating modes that reduce the supply voltage to maximize battery life in mobile devices during periods of light computational loads [9]. High-performance server processors employ extensive voltage frequency scaling for power and thermal management.

The main contributions of this paper are as follows: We study the low voltage timing behavior of CMOS gates, and circuit paths, in the presence of random process variations using both analytical modeling and simulation. While in earlier work we have presented an initial analysis using a very simplistic analytical model applied to inverters to argue for an adaptive approach to minimize system level tests [4], in this paper we work with the accurate and well validated gate delay model based on the Sakurai-Newton alpha-power law [10]. We have also performed over 25 million SPICE simulations for a circuit path comprising 20 NAND gates with randomly assigned transistor parameters drawn from realistic process variability distributions to obtain meaningful delay statistics for circuit path delay in the presence of manufacturing variations. A key observation from this simulation data is that the slowest, and statistically rarest, outlier paths in the tail of the distribution almost always contain a single extremely slow transistor (gate) which contributes most of the increased delay in the path. This can be observed for outlier paths that occur less than about once in a million path delay simulations with random transistor parameters that realistically model process variations and rapidly becomes more pronounced as the occurrence probability drops by additional orders of magnitude. We offer an analytical explanation for this observation. Such extreme outlier paths can be the cause of many timing marginalities that occur in manufactured circuits; larger delays are likely to be more reliably detected and screened out during test. Note that because complex SOCs today have many billion transistors, and often path counts that can run in the hundreds of millions, even extreme low probability variability paths can be expected to be commonly observed, and can therefore potentially contribute tens, even hundreds of DPPM in timing failures.

To re-state our key finding, the excessively delay in extreme outlier circuit paths resulting from random process variations is largely due to a single excessively slow transistor in some gate, and less from the accumulation of distributed delays from multiple gates along the failing path as is sometimes assumed. This suggests that TDF timing tests, that have traditionally targeted localized defects causing lumped delay faults, can remain at least partially effective for testing timing failures caused by variability, mitigating the need for path delay testing that has so far remained elusive in practice. In the absence of this knowledge of a single transistor/gate causing the majority of the excessive delay in failing circuits, test experts have correctly assumed that path delay testing would be essential for effective delay screening of distributed delays from random process variations. We further show in this paper how our analysis here can be leveraged to condition the voltage and timing of the applied TDF test to help enhance the detection of parts with timing marginalities.

The insight provided by this paper can be quite useful, given the uncertainty regarding the root cause of SLT fails. Our observations also explain the success of CAT-delay tests in detecting many Vmin failures that appear to be caused by process variations in the experiments reported by Intel [1]. The authors of the Intel study expressed surprise that their lumped delay CAT scan tests detected all the 146 voltage sensitive Vmin only failures that had been identified by a final functional SLT screen after being missed by traditional SA and TDF tests. Quoting from [1]: "It was also expected that at least some units would fail CAT patterns as these units were essentially "known" SLT failures. What was not expected was that ***all*** (emphasis in the original paper) units failed with CAT patterns." Indeed, chance detection of all of these statistically large number (146) parts appears extremely unlikely, suggesting that the two-pattern CAT timing tests generated to targeting localized defects in cell layouts have an inherent ability to detect Vmin timing failures. While left unanswered by [1], this observation is explained by the results presented in this paper. We also use this published production data for Intel's advanced 14-nm FinFET technology to validate our analysis in this work in other ways.

While timing errors in low voltage operation from process variations have been a suspected cause of SLT fails in some recent work [4,7,20], to our knowledge this is the first paper to attempt to validate this conjecture using manufacturing test data. Such failures, caused by timing marginalities, may also explain the unpredictable "mercurial" errors recently reported from field operation [4,5]. The majority of the SLT fails analyzed in [1] only failed the CAT-delay very close to the specified Vmin, where gate delays are maximized. If some similar parts evade the final SLT test screen, they may display occasional errors in low voltage operation from power supply noise, perhaps accentuated by additional conditions such as temperature.

The rest of the paper is organized as follows. Section II provides background on state-of-the-art defect-oriented cell aware and timing aware test methodologies and discusses their effectiveness at advanced nodes. Section III introduces test data from the 14nm FinFET process, recently presented by Intel, to highlight the importance of the voltage sensitive "Vmin only" failures that are the subject of this paper. Section IV analyzes the impact of reducing VDD on gate delay changes from random process variations with the help of analytical modeling and simulations. This analysis is extended to path delays in Section V. The results from the previous two sections are then used to analyze the supply voltage (VDD) versus failure detection data in the Intel paper in depth in Section VI to better understand VDD sensitive failures. Section VII presents our new approach to optimize the screening of circuit timing marginalities. Section VIII concludes the paper.

II. Background and Review of Related Methodologies

A. Scan Tests and System Level Manufacturing Tests

Scan testing of ICs and SOCs to screen out manufacturing defects is generally performed at least twice during the manufacturing test flow, as shown in Fig. 1. Wafer probe tests are applied when the dies are still part of the wafer. These tests, which detect most of the faulty dies, help avoid wasteful packaging of defective parts. Then, after the wafer is diced and the dies individually packaged, the finished IC is tested again.

Both the wafer probe and post packaging tests employ scan DFT architecture for test application [11]. Test generation, traditionally based on the classical stuck-at (SA) and transition delay fault (TDF) models [11], is today commonly supplemented with new Cell Aware Tests (CAT) [2] to improve defect coverage. These new tests have been observed to significantly reduce defect levels DPPM in many applications [12]. However, as shown in Fig.1, SOCs fabricated in advanced technology nodes are increasingly requiring an additional functional system level test (SLT) screen to reach acceptable defect levels. Here the packaged part is temporarily mounted on a test board that mimics the intended application hardware [4]. It is then extensively tested anywhere from 10 to 120 minutes at the rated speeds over a range of user applications under varying operating conditions. SLTs require an additional test insertion step in the manufacturing flow, and also new test equipment that can support highly parallel functional testing of the test boards. More recently, a scan test capability has also been incorporated into SLT testers, which offers the potential for moving post packaging scan tests to SLT, combining the second and third test insertion steps in Fig.1. This can save the extra test insertion step for the post packaging scan tests which are still required to ensure low DPPM.

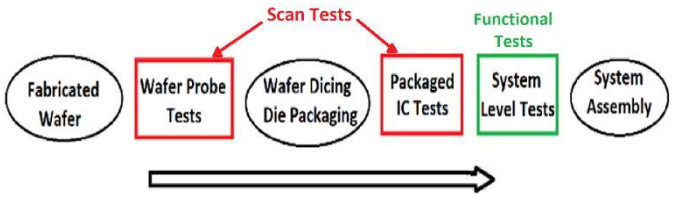

Fig. 1: Simplified manufacturing flow showing test insertion stages.

B. *Cell Aware and Defect Oriented Scan Tests*

Traditional SA and TDF test generation targets the circuit nodes that are the inputs and outputs of the gates (standard cells) used as building blocks in a design but does not explicitly target faults/defects inside the cells. While most internal defects in simple logic gates are also detected when faults at the inputs are tested, this is not always true for the large complex CMOS standard cells commonly used in modern processors. The defect coverage improvement from Cell Aware Tests results from the explicit targeting of potential short and open defects within the standard cells in this new approach. Both single pattern (CAT static) and two-pattern (CAT delay) tests are generated by this methodology. Encouraged by the significant DPPM improvements from CAT [12], this "defect-oriented" test generation approach has recently been extended to shorts and opens outside individual standard cells: shorts between adjacent cells in the layout, and also to shorts and opens in the interconnect [1]. However, the DPPM gains from these extensions of the CAT methodology appear to be more modest. It is becoming increasingly apparent that attempts to eliminate, or even reduce, SLT fallout by improving scan tests must also look at additional techniques to improve coverage.

C. *Timing Aware Tests*

Beyond logic correctness, circuits must also meet timing requirements that ensure that all switching transitions reach the correct stable signal value within a clock period. Unfortunately, path delays in CMOS logic are highly dependent on off-path signals and can display very varied timing under different inputs [13]. For example, a simple 3-input NAND gate displays 3 very different rise time delays. The largest delay, when only one pull-up PMOS turns on to charge the output capacitance (e.g. for the 011 input), is roughly 3 times the smallest delay when all three parallel PMOS transistors turn on (for the 000 input). More complex gates can experience an even larger range of input dependent delays at their outputs, spanning a factor of 5 or more. Ensuring that timing tests set up worst case delay conditions for each test is extremely difficult. Random process variations further complicate the picture; the worst-case timing test for the same delay fault in two manufactured copies of the same design may not always be the same.

Observe that for critical paths that are 20-30 gates long, a 15% timing margin translates to 3-5 gate delays. A delay fault must exceed this delay, plus any additional timing slacks in the faulty path, to cause an error. In older technologies, these slacks and timing margins are large enough to absorb the impact of the modest process variations plus any circuit noise, and even many resistive delay defects. Consequently, relatively few circuits fail timing unless the increased delay due to a defect is quite large. Such "gross" delays are generally detected by TDF tests that target localized lumped delays at circuit nodes. The concern today is that this may no longer hold for advanced technologies which display much larger random process variations. Here multiple small delays from process variability can potentially accumulate and build up along some path to cause timing failure, even in the absence of any defect or manufacturing abnormality. Note that if a single excessively slow statistical outlier path out of the many millions in a large chip fails timing under any condition encountered in operation, the part is defective, although such failures may be rare and extremely difficult to detect and screen out during test. As was first shown in [4] and more accurately analyzed in Section IV of this paper, gate delay variations are greatly amplified in today's aggressive low voltage operating modes commonly used to achieve power savings. This can potentially result in many more failures caused by "small delays" which can escape TDF tests aimed at gross delay faults.

Timing aware TDF and CAT-delay test generation [3] attempts to partially mitigate this problem. Here delays, which are still assumed to be lumped at the target node, are tested by activating a transition along the longest path containing the node. The aim is to minimize the timing slack that adds to the timing margin in the tested path, allowing detection of the smallest possible delay fault in the path. However, timing aware TDF (or CAT-delay) tests still explicitly target only localized delays. They are not path delay tests that aim to detect failure from the accumulation of excessive delays in multiple gates along any path in the circuit. The number of paths tested by timing aware TDF (and CAT) is typically orders of magnitude fewer than all the paths in a circuit. Consequently, timing aware tests can be effective in detecting scan test escapes from traditional timing tests that are caused by somewhat smaller

localized defects. However, such tests are generally not believed to be effective in detecting faulty parts if multiple random distributed gate delays in the circuit caused by process variations can add up to cause timing failure in some path -in the absence of any single large delay fault in the path. Unfortunately, effective scan-based path delay testing [13] to detect such failures has not yet proven practical. Consequently, this has limited confidence in the effectiveness of scan timing testing in screening process variability induced failures.

A mentioned earlier, a key contribution of this paper is to show that the rare circuit paths that fail due to greatly increased delays from random process variations, overcoming the conservative timing margins that are common, always contain a single extremely slow transistor (and corresponding gate) which contributes most of the increased delay in the path. This suggests that traditional lumped delay scan timing tests may remain effective in screening out variability failures.

III. THE ITC 2018 [1] DATA ON LOW VDD FAILURES

To get a better understanding of the kinds of failures in modern technologies that escape scan testing and contribute to system level test fall out (Fig. 1), we revisit recent experimental data from volume production in [1]. Our motivation here is to exploit the very limited industrial test data for an advanced process available in the public domain in the context of the new understanding developed by our research, and to check its consistency with our analysis to validate our work. This data presented by Intel at ITC 2018[1], was from a number of test experiments conducted on processors fabricated in a 14nm FinFET process. The tests investigated include classical stuck-at and TDF tests, traditional Cell Aware Tests (with CAT-Delay generated without any attempt to maximize delay along the fault propagation path), Timing Aware Cell Aware Tests (TA-CAT), and System Level Tests (SLT).

TABLE I. FALL-OUT FROM THE DIFFERENT CAT TESTS IN [1]

CAT-Static (Applied Before Delay Tests)	CAT-Delay (Fails at Vmax & Vmin)	CAT-Delay (Vmin Only Fails)
400 DPPM	2500 DPPM	1400 DPPM

The first experiment in [1] studied the additional fallout from traditional CAT tests following initial testing using classical SA and TDF tests run at nominal VDD. Table 1 reproduces the data from [1] that summarized the main results from this experiment. Observe that in total CAT detects an additional 4300 DPPM over the classical tests, the large majority of which were detected by the two-pattern CAT-delay (not TA-CAT) tests. These CAT-delay fails were further divided into two bins based on further tests: (1) "hard" fails that are observed over the entire range of supply voltages that the circuits were expected to work at the specified clock frequency, from Vmax to Vmin; (2) power supply voltage sensitive fails that only fail close to the minimum specified operating voltage Vmin, but pass at higher VDD values. Recall that CMOS circuits slow down significantly as the supply voltage is reduced, so these low voltage fails are likely "soft" timing failures, where one or more paths exceeded the clock period at Vmin.

Vmin only timing failures are of particular interest since, as seen in Table I, the other scan test escapes are mostly hard defects (e.g. shorts and opens) that can be reliably screened out by CAT patterns applied at nominal VDD. In modern processors, the on-chip power management system is provided with a Vmin specification, corresponding to each operational frequency, that must be always met to ensure error free operation as VDD is lowered during voltage/frequency scaling. To maximize power savings, these Vmin values are not conservatively picked to work across multiple instances of the same IC, but are instead customized and hard programmed into individual parts. In theory, each individual IC can be expected to pass an ideal (exhaustive) timing test at a unique Vmin because systematic and random manufacturing variations result in part specific device parameters. However, precisely determining this Vmin for each part and frequency is prohibitively expensive because it requires repeatedly rerunning high coverage test content as VDD is lowered in very small steps until failure. Therefore, part specific Vmin estimates are used instead. These are typically computed using parametric test results from the IC, combined with statistical models that add additional guard bands for process variability and circuit noise. However, using such estimated Vmin values leaves open the possibility that some ICs, containing statistically extreme slow outlier paths, may actually fail at a VDD above (although in practice generally close to) the specified Vmin. These are the Vmin only failures in Table I that must be screened out during test to ensure error free operation. Unfortunately, scan tests are often unable to catch such small delay failures that are only active very close to Vmin because scan timing tests do not accurately capture actual functional timing. Scan tests are well known to be impacted by several "out of normal functional mode" factors [16] such as switching noise [17], power supply droop, clock-stretching [18], temperature variations etc. Consequently, even if a part passes a scan timing test, there is no certainty that it will be timing error free in functional operation. Several studies, e.g. [19], have suggested that scan tests are optimistic and tend to overestimate functional Fmax.

It should be noted that a key focus of the Intel paper [1] was highlighting the success of CAT in detecting SLT fallout that would otherwise escape if only SA and TDF scan tests are used. To determine if the failures uniquely detected by the CAT scan tests cause functional failures, in a second experiment in [1] using a different set of parts, 156 SLT failing ICs that escaped SA and TDF testing, were selected to be retested with traditional CAT patterns (not TA-CAT). However, only 10 failing parts were detected by CAT at nominal VDD. The remaining 146 out of the 156 required VDD to be significantly lowered for detection, on average to within 55 millivolts of Vmin. This again suggests that there are clearly many Vmin only timing failures that escape traditional scan tests and are detected by functional tests. As discussed earlier, the authors were surprised to find that all the 156 SLT fails were also detected by the CAT-delay tests.

Our interest in this work is to better understand the characteristics of the power supply voltage sensitive Vmin only timing failures that pose a challenge to scan testing. These appear better detected by (at speed) function testing that is performed during SLT, although some such test escapes may give rise to intermittent errors in the field.

IV. Impact of Reduced VDD on Gate Delays

In this Section we study the impact VDD reduction on gate and path delays in the presence of process variations using both analytical modeling, as well as SPICE simulations. We begin by modeling gate delay using an analytical gate delay model based on the well validated Sakurai-Newton alpha-power law [10]. This approximates switching delay to be proportional to $1/(VDD-Vth)^{\alpha}$. The literature suggests that for advanced FinFET technologies, the best fit α appears to be 1.25 [14], although our analysis below is quite robust over any reasonable range around this value. We further assume that process variations only influence delay through variations in the threshold voltages of individual transistors, and that Vth is normally distributed around some nominal Vth0 with standard deviation σ. In [15], Intel has reported measured values of σ to be 19 and 24 millivolts for NMOS and PMOS 14nm FinFET transistors respectively. Importantly, this threshold voltage distribution was further also found to be Gaussian at least out to +/-5σ [15], which allows for easily tractable analytical analysis. Consistent with the above, we take $\alpha = 1.25$ and $\sigma = 25$ mV for all transistors to compute the illustrative changes is switching delay due to process variations shown in Table II.

TABLE II. Gate delay vs. ΔVth and occurrence probability

ΔVth		Percentage Delay Change due to ±ΔVth				Probability of Vth > Vth0 + ΔVth
σ	Volts	VDD = 1.0V		VDD = 0.6V		
0	0.000	0%	0%	0%	0%	1 in 2.0E0
1	0.025	-4.9%	+ 5.4%	-13.7%	+18.1%	1 in 6.0E0
2	0.050	-9.6%	+ 11.4%	-24.4%	+43.2%	1 in 4.4E1
3	0.075	-13.6%	+18.1%	-32.7%	+79.9%	1 in 7.4E2
4	0.100	-17.5%	+25.6%	-39.7%	+137.8%	1 in 3.2E4
5	0.125	-21.0%	+33.9%	-45.5%	+240.7%	1 in 3.5E6
6	0.150	-24.3%	+43.2%	-50.3%	+465.6%	1 in 1.0E9
7	0.175	-27.3%	+53.8%	-54.4%	+1018%	1 in 7.8E11

Table II shows the percentage change in gate delays, both speed-up and slow-down, at VDD = 1.0V and 0.6V, for transistors at σ = 0,1,2,3,4,5,6 and 7 from the Vth distribution. Observe that if Vth increases due to variability, the transistor slows down, and gate delay increases. However, based on a normal variability distribution for Vth, there is an equal chance of Vth decreasing, thereby speeding up gate delay. These are shown as negative delay increases in Table 1. Also shown in the table is the cumulative probability of a transistor having the given or higher Vth shift from the nominal. For example, the chance of a transistor with Vth at 6σ or higher is approximately 1 in a billion. While this probability appears small in absolute terms, given that many modern SOCs contain hundreds of billion transistors, one can expect many such extreme outlier transistors in *every* chip. And since DPPM is typically evaluated over a million such ICs, cumulatively containing many trillion transistors in total, even transistors at 7σ and beyond in the Vth distribution can contribute to the measurable DPPM numbers observed in production. This informs the range of σ values, through 7σ, considered in Table II. We assume that the Vth distribution remains Gaussian over this window. Recall that it has been shown to behave so at least until 5σ [15].

Note from Table II that the magnitudes of speed-up and slow-down for the same Vth deviation from the nominal are not the same. This is particularly accentuated in low voltage operation, suggesting that, on average, process variations slow down circuits. *Delays from random variations do not average out along long paths as is sometimes assumed.* Two additional important observations from Table 1. (1) Delay variability from random process variations increases greatly in low voltage operation. Consequently, timing margins, as a percentage of the clock period, need to be increased significantly to minimize yield loss caused by timing failures in circuits even when they are free of hard defects. This can greatly degrade performance beyond the large (2-3X) slowdown already observed even in nominal transistors operated at low voltages to save power. (2) If sufficiently large timing margins, and consequent loss in performance, cannot be afforded to avoid nearly all the variability related timing failures, as has been traditionally achieved at nominal VDD values where the impact of process variations is much less, then the statistical outlier slow parts that experience timing failures must be detected and screened out during test. The general perception is that current TDF scan tests developed for localized lumped delay faults are unable to reliably screen out timing failures from an accumulation of these distributed variability delays. And because of the uncertain coverage of the applied functional tests, even the SLTs currently employed as an additional test screen after scan structural tests appear to be an imperfect solution to this problem.

V. Characteristics of Slow Statistical Outlier Paths

In this section we show that the common understanding that TDF(and CAT) "lumped delay fault" timing tests are ineffective against timing failures caused by random process variations may be misplaced. We begin by establishing a key result of this paper that extremely slow paths, in the tail of the statistical path delay distribution caused by process variations, nearly always contain a single outlier transistor that contributes a large share, even a majority, of the increased delay. Furthermore, the probability of such an extreme transistor in the slow paths increases dramatically for paths further out in the tail of the path delay distribution and is a virtually certainty once the path probability falls below approximately one in a billion. Consequently, the output of the gate containing such a transistor displays a large, lumped delay that is frequently detectable by two-pattern scan tests that are generated to target slow transitions at gate outputs.

To get some intuition on why virtually every path that can potentially cause timing failure contains a statistical outlier transistor from the tail of the variability distribution, consider a design with a 30-gate critical path. Assume that that in some copy of the manufactured circuit, extreme process variations increase the delay of this path by 10 additional gate delays, to a total of 40 nominal gate delays. Assume further that this 33% increase in path delay can potentially cause timing failure even in view of the somewhat larger timing margins employed in low voltage operation. Now this large delay increase due to process variations can be distributed over the path in a number of different ways. For example, it can mostly come from a single extremely slow gate with ~11X the nominal delay, or the

combination of two gates that are each approximately half as slow, or it can be more broadly distributed over multiple gates.

We first focus on just the first two cases, with only one or two gates contributing a large delay increase with the potential to cause timing failure. We investigate which of these two cases is more likely given realistic manufacturing statistics. It is reasonable to assume that because of the large number of remaining gates (29 or 28 respectively) in the 30-gate path, the delay probability distribution for the rest of the path will be very similar, i.e. statistically, the rest of the path will contribute nearly equally to the added delays in the two scenarios.

Observe from Table II that there is a 1 in 7.8E11 chance of a random transistor having a Vth of 7σ or beyond, thereby providing the extra 10-gate delay needed for timing failure all by itself. Also, from Table II the random probability of encountering a 6σ or greater transistor is 1 in 1.0E9. Note that two such transistors can together contribute roughly comparable delay to a single 7σ transistor. The multiplicative joint probability of randomly picking two 6+ σ devices compounds to 1 in 1.0E18. This number informally suggest that it is orders of magnitude more likely that a path contains a single transistor with Vth of 7σ or beyond, compared to two transistors having a Vth of 6σ or higher resulting in similar increased delay. Furthermore, the monotonic relationship between Vth, gate delay, and the probability of observing a specific Vth value in the right half of the probability distribution allow the same reasoning to be more generally extended to make an informal case for a single extremely slow device over multiple moderately slow ones contributing the same total delay.

The above discussion is presented only to offer insight on why a single extreme transistor contributes most of the delay in a very slow outlier path. We do not yet have a formal proof to support this conjecture; that appears quite complex to develop. Therefore, we have conducted SPICE simulations of path delays in a 20-gate chain of two-input NAND gates to further investigate this problem. Several million distinct copies of these circuits were constructed with individual transistor threshold voltages randomly drawn from normal distributions centered at the nominal PMOS and NMOS threshold voltages, with σ = 25mV to account for process variations. These circuits were then all simulated to obtain path delays for a rising transition at the output, corresponding to a single bit change at the input. Note that while this simple logic chain does not represent the wide range of circuit paths encountered in practical designs, performing millions of SPICE simulations on practical designs for Monte Carlo experiments is computationally prohibitive. Also, while we would have liked to have performed billions, even trillions of path delay simulations to accurately characterize outlier statistics leading to 50-100 DPPM timing failures in manufactured parts, even the modest 25 million simulations performed took weeks of computation time.

Table III presents early results from only the first 175K simulations performed, while Table IV shows the same data after all 25 million simulation runs were completed. The second row in Table III shows the delay for each of the 10 slowest paths observed in the 20-gate NAND chain simulations with random process variations at the end of the first 175K simulations. The path ranked #1 is the slowest path observed with a 3.27ns delay. Additionally, below each path delay, the columns of Table III include the Vth values for the 10 slowest (weakest) transistors, out of the 30 active transistors (20 NMOS and 10 PMOS), in each of these 10 NAND-gate paths. These thresholds are indicated as deviations from the nominal Vth using a standard deviation σ measure, and rank ordered with the slowest (highest σ) transistor on top. Table IV shows the same data for the much larger set of 25 million path delay simulations. Observe from comparing the two tables that, as expected, the 10 worst-case path delays are significantly longer (slower) in the larger simulation because more extreme Vth values are encountered in the much bigger population of random transistors simulated. For example, in Table IV the three most extreme transistors have Vth of 5.2σ, 5.1σ and 4.9σ among the 10 slowest paths. This is to be expected for the 25 million simulated paths because each path has 30 active transistors for a total of 750 million transistors. Approximately 225 transistors with Vth of 5σ or higher (5+σ) can be expected in this population since Table II shows that on average 3 such transistors are found among 10 million devices. However, because of the random selection of transistors in the NAND chains, not all the slowest 10 paths out of the 25 million simulated are guaranteed to include a 5+σ transistor. Meanwhile other paths, not among the slowest 10 may include such a transitor. With Table III reporting results for only 175K path delay simulations, a 4.5σ transistor is the most extreme observed. Far more rarely occurring 5σ transistors were not encountered in this much smaller population of transistors.

TABLE III. THE 10 SLOWEST PATHS AFTER 175K SIMULATIONS. VTH FOR THE 10 WEAKEST TRANSISTORS IN THE PATH IS ALSO SHOWN IN EACH COLUMN

Path Rank	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**
Delay(ns)	3.27	3.23	3.2	3.2	3.2	3.2	3.19	3.19	3.18	3.18
Tr #1(σ)	**3.7**	**4.1**	**4.0**	**3.7**	**3.6**	**4.5**	**4.1**	**4.2**	**3.8**	**3.8**
Tr #2(σ)	**3.5**	**3.3**	**3.5**	**3.6**	**3.6**	**2.5**	**3.0**	**3.0**	**3.0**	**3.1**
Tr #3(σ)	3.4	2.1	3.0	3.2	3.2	2.5	2.9	2.9	2.0	2.7
Tr #4(σ)	2.7	2.1	2.2	2.9	2.9	2.4	2.0	2.9	2.0	2.5
Tr #5(σ)	2.7	2.0	2.1	2.4	2.5	1.7	1.9	2.0	1.8	2.4
Tr #6(σ)	2.3	2.0	1.9	2.0	1.9	1.6	1.8	2.0	1.7	2.3
Tr #7(σ)	1.8	1.8	1.9	1.6	1.8	1.5	1.6	1.9	1.4	2.0
Tr #8(σ)	1.4	1.7	1.8	1.4	1.3	1.3	1.6	1.8	1.4	1.9
Tr #9(s)	1.2	1.4	0.8	1.3	1.2	1.2	1.5	1.6	1.4	1.3
Tr #10(σ)	0.8	1.2	0.8	1.3	1.2	1.1	1.4	1.2	1.3	1.2

TABLE IV. THE 10 SLOWEST PATHS AFTER 25 MILLION SIMULATIONS. MOST OF THE DELAY INCREASE COMES FROM THE SLOWEST TRANSISTOR

Path Rank	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**
Delay (ns)	3.45	3.41	3.4	3.4	3.4	3.39	3.39	3.38	3.38	3.38
Tr #1(σ)	**4.6**	**3.5**	**4.5**	**4.9**	**5.2**	**4.4**	**4.6**	**4.1**	**4.3**	**5.1**
Tr #2(σ)	**3.0**	**3.2**	**3.6**	**2.8**	**2.5**	**2.9**	**3.0**	**3.0**	**2.8**	**3.0**
Tr #3(σ)	2.9	3.0	3.4	2.2	2.0	2.1	2.3	2.7	2.5	2.2
Tr #4(σ)	2.9	2.5	2.1	2.0	1.9	2.0	2.0	2.4	2.3	2.1
Tr #5(σ)	2.5	2.5	1.9	2.0	1.9	1.9	1.8	2.3	2.2	1.8
Tr #6(σ)	2.0	2.1	1.5	1.8	1.8	1.9	1.7	2.2	2.1	1.8
Tr #7(σ)	1.8	1.7	1.5	1.7	1.7	1.9	1.7	1.8	1.8	1.7
Tr #8(σ)	1.8	1.6	1.3	1.6	1.6	1.8	1.6	1.7	1.7	1.6
Tr #9(σ)	1.7	1.5	1.3	1.6	1.6	1.8	1.5	1.5	1.7	1.6
Tr#10(σ)	1.5	1.4	1.2	1.5	1.4	1.7	1.5	1.5	1.7	1.5

Of particular significance in making our case for the existence of single extreme transistor in every slow path is the difference between the σ values in each of the columns of the tables above that represent individual slow paths. The σ

difference between the two top transistors in each column, shown in bold, reflects the difference in strength between the weakest and next weakest transistor. Observe that this difference increases markedly between Tables III and IV, as statistically the paths become more rarer, from the slowest 10 paths in 175K circuits to the slowest 10 in 25 million. Correspondingly, the delays of the slowest paths also increase from the 3.18 to 3.27ns range in Table III to 3.38 to 3.45ns in Table IV. The average difference between the highest σ value, and the next highest is 0.74σ in Table III, but more than doubles to 1.54σ in Table IV for the larger delays. This suggests that in the orders of magnitude rarer extreme paths containing transistors with Vth of 7σ and beyond, which are likely to be the source of timing marginalities in real ICs, the second slowest transistor will generally have at least 2-3 lower σ value on average. From Table II this translates to 4-8X difference in delay. This highly nonlinear relation between delay and σ difference implies, that much of the excess delay in the failing path will be therefore localized in a single slow transistor.

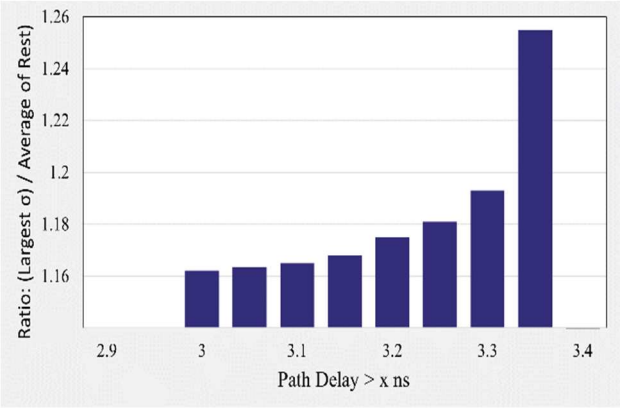

Fig. 2. Average ratio of σ of the weakest transistor in a path to the average σ for rest of the transistors for different path delays.

Fig. 2 presents an alternate way of visualizing the above result. The bar chart shows the ratio of σ for the slowest transistor and the average σ of the rest of transistors in a simulated path. This σ ratio is computed over all the paths that satisfy the target minimum delay; their number decreases quickly as delay increases. Note the rapid and non-linear increase in this σ ratio as paths become slower and their occurrence statistically much rarer, indicating the concentration of delay in a single transistor. Again, Fig. 2 is based on the same 25 million simulations as Table IV, in which the most extreme transistor observed had a Vth of 5.2σ. The σ ratio can be expected to be much more skewed for the transistors with 7σ Vth and beyond that are expected to contribute to timing marginalities in large SOCs.

To summarize, we have shown in this section that extreme statistical outlier slow paths that are found in circuits because of manufacturing process variations nearly always contain a single excessively weak transistor that contributes to a large share of the path delay. This skewed contribution to path delay from a single transistor (and corresponding gate) becomes more acute as path delay increases and the occurrence probability of such a path decreases to just a few paths among billions or even trillions. Nevertheless, such rare slow paths can still contribute to many DPPM faulty and marginal ICs, and are a primary target of new test methods for complex state-of-the-art ICs and SOCs.

VI. Interpreting the Vmin Test Data

In this Section we analyze some additional timing test data from the Intel paper [1] that was originally presented to evaluate the effectiveness of Timing Aware Cell Aware Tests (TA-CAT). We use this data entirely differently, to study the impact of process variations on device behavior and path delays as VDD is reduced in Intel 14nm FinFET processors. We show that the observed data is entirely consistent with our delay modeling. The specific data from [1] that we focus on in the discussion below plots the actual measured Vmin using a TA-CAT test set against the same for a TDF test set, for each of a large collection of processor circuits operated at five different frequencies.

Recall that TA-CAT targets the same defects as traditional CAT-delay tests, but additionally ensures that the target defect is activated and propagated along the longest (slowest) possible path. For these timing aware tests, ATPG must work with circuit timing information in the Standard Delay Format (SDF), which greatly increases test generation time and complexity. Also, TA-CAT significantly increases pattern count over timing unaware CAT. In the Intel experiment, TA-CAT patterns were generated only for a select 25% of all CAT-delay faults, with the aim to detect "small delay defects". Nevertheless, as shown in Fig. 3, these relatively low coverage TA-CAT tests still needed 2.5X the number of patterns required by TDF, while 100% of the timing unaware tests for CAT-delay faults only needed only about 2X as many patterns as the TDF tests.

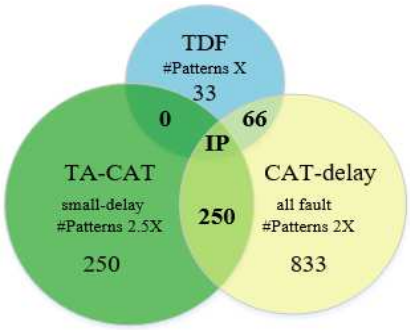

Fig. 3. Results reproduced from [1] showing TDF, CAT-delay and TDF fault detections. The circles reflect the approximate size of the test set.

Fault detection results from [1] aimed at detecting voltage sensitive failures using the three tests are shown in the Venn diagram in Fig. 3. The focus here is on what are commonly called "interesting" parts in test research: those ICs that fail at least one of the applied test types but not all. Those that fail all three tests fall into the IP classification in Fig. 3. This number is not shared because it is protected yield related information. The effectiveness of the CAT methodology, including TA-CAT, to detect a significant number of unique failures is clear from this figure. However, our interest is in analyzing this industrial data for the change in critical path delay as VDD is lowered using additional data from the same experiment plotted as Fig. 10 in [1]. We have directly copied the figure here as Fig. 4 because the numerical data for this plot is not available in the public domain, and unfortunately cannot be extracted due to the limited resolution of the figure available in the published paper.

For each of the interesting parts (contributing the total of 1083 DPPM) in Fig. 3, Fig. 4 shows the lowest passing VDD for the TA-CAT test set, shown as TA-CAT (Vmin) plotted against the passing Vmin for the TDF tests TDF(Vmin) for the same part. Note that these parts are all "defective", i.e. known to exhibit voltage sensitive failures. Each IC was tested at each of its 5 different operational frequencies, i.e. scan tested for the 5 different launch-to-capture time periods corresponding to the different frequencies. To generate this data, both the tests, TDF and TA-CAT, were run at multiple supply voltages and clock frequencies. In Fig. 4, the red markers are for tests run at the lowest frequency (longest period) F1, and the light blue markers (of a different shape) for the highest F5.

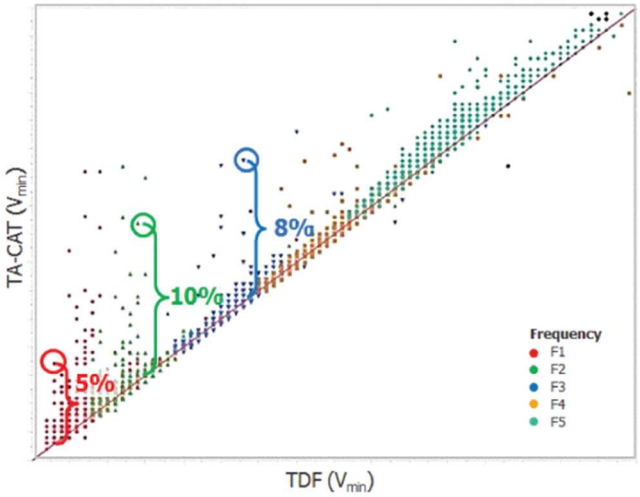

Fig. 4. VDDmin TA-CAT versus TDF for timing fails. Plot copied from [1].

The idealized timing diagram in Fig. 5 shows the 5 different clock periods (not to scale) and helps illustrate the impact on path delays as VDD is reduced. The arrow in the timing diagram (shown for only one signal) indicates how path delays increase as VDD is lowered. The 5 capture clock edges represent each color-coded frequency. The measured Vmin for any delay test set (TA-CAT or TDF) at a given frequency is the VDD value when, for some test pattern in the test set, the delay at any output equals the clock period, indicating timing failure. Observe that the Vmin voltage is largest for the highest frequency F5 (which has the smallest timing slack and therefore requires the least reduction in VDD before failure) and smallest for the lowest F1, which allows the greatest reduction in VDD before failure.

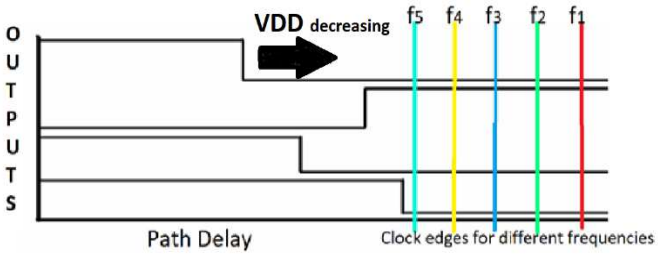

Fig. 5. Illustrative timing diagram showing clock edges for test frequencies. Path delays increase with lower VDD; Vmin is the VDD value when a detectable path delay equals the clock period.

Note also in Fig. 4 that if for any IC the TA-CAT tests fail at a higher Vmin than the TDF tests, then that IC is plotted in the upper triangle in the figure. This indicates that the TA-CAT test set contains some pattern that activates a path that is slower than all the paths activated by the TDF test set. Several such ICs are observed in the data plotted in Fig. 4. A point below the diagonal suggests that the TDF test has a higher Vmin, i.e. all the TA-CAT patterns passed at some VDD below the Vmin required to pass the TDF patterns. This is extremely unlikely (at least in the absence of process variations) because, in addition to cell internal defects, the TA-CAT tests also target all transition faults targeted by the TDF patterns while ensuring error propagation along the longest possible path. Theoretically then, the two Vmin values can at best be the same. In practice, Vmin for TA-CAT can, on occasion, be very marginally lower if the timing aware ATPG is inaccurate in modeling the actual delays experienced in silicon. Consequently, virtually no points are observed significantly below the diagonal in Fig. 4.

An IC exhibiting the same Vmin for both test sets ends up on the diagonal line, indicating that the slowest paths in both the test sets are about equal. TA-CAT generates tests for targeted defects along the slowest possible path, and typically activates somewhat longer paths. However, a large timing closed circuit has many near critical paths that have very nearly the same delay. It is therefore quite likely that the TDF test set also activates one or more paths with close to the same delay as the slowest TA-CAT path. Therefore, it is not surprising that most of the parts end up on or close to the diagonal in Fig. 4.

This band of points close to the diagonal indicates small delay differences between the slowest paths in the two test sets. Observe that the width of this band of ICs around the diagonal narrows at lower Vmin until around the dark blue test frequency F3. This is an indication that the transistors in most of the failing long paths in the band around the diagonal have close to nominal device parameters. Suppose, for example, that the longest TA-CAT path in some part is 5% longer than the longest TDF path for the part. Then the difference in Vmin for the two tests, i.e. the deviation of the point plotted for the part from the diagonal in Fig 3, is the additional reduction in VDD required to cause a further 5% delay in the TDF path after the TA-CAT path fails. Our modeling has established that small changes in VDD result in larger percentage changes in delays as operating voltages are lowered because of the non-linear relationship between VDD and delay as per the alpha power law. Thus this 5% difference is made up by a smaller difference in Vmin when timing failure occurs at lower VDD. i.e. when the same part is operating at the lower frequencies towards the bottom of the plot.For the highest frequencies F5, F4 and F3 (light blue, yellow and blue markers) this band is fairly well defined with relatively few outliers. However, for the lowest frequencies at the left of the plot, F1 and F0, which require very low VDD to force timing failure, there are many more points scattered above the diagonal. This is caused by much more significant delay increases at the low voltages from process variations even in near nominal transistors. Recall from Table II that the percentage gate delay increase over the nominal gate delay increases as VDD is reduced for all weak transistors in the variability distribution. Most of this delay increase is picked up by the TA-CAT patterns

because they are more than double in number compared to TDF patterns and they target long paths containing more transistors.

However, in general, we should also expect some cases from increased variability plotted below the diagonal line. Consider a case where at high frequencies, a part was plotted on the diagonal. This implies that when delays from variability are minimum, the longest detected paths in the TA-CAT and TDF test sets have approximately the same. In this scenario, it is certainly possible that at lower frequencies, where Vmin is lower and gate variability delays accentuated, that random process variation increases the critical TDF path more than the critical TA-CAT path. However, this is not observed at all in the data. There are virtually no variability points below the diagonal, even in the high variability region at the bottom of the plot. The only explanation for this is that any significant delay increase in a path due to variability is dominated by a single slow transistor or gate. While such a transistor could well be in the critical TDF path, and not on what was the TA-CAT critical path at the higher frequency, a TA-CAT test set will also contain a different test that tests the affected transistor through some other long path, which is now critical. Moreover, unless there are multiple transistors with significant variability delays, this new TA-CAT critical path will equal or exceed the TDF critical path. Thus Vmin data for the part will again be plotted on or above the diagonal. Therefore, the absence of points plotted below the diagonal in the high variability region at the bottom left of Fig. 4 validates our conjecture that timing in failing paths is dominated by a single extremely weak transistor that is detected by lumped delay TA-CAT tests. With a more even distribution of delays among the gates, the Vmin data in Fig. 4 would display many more points below the diagonal in this region.

VII. OVERTESTING TO SCREEN TIMING MARGINALITIES

In this Section we outline an optimized "overtesting" approach to more effectively screen out parts with timing marginalities at post-packaging scan tests. The goal is to minimize test escapes and fallout at functional system level tests (SLT) that often follow scan testing. The proposed approach has the potential to minimize SLT testing, and perhaps even eliminate it completely in some applications. Furthermore, a similar overtesting methodology can be applied during SLT to minimize test escapes from timing marginalities, and thereby reduce the unpredictable intermittent timing errors being reported in operation.

Recognizing that path delay increases at reduced voltages, the proposed overtesting approach tests the part more aggressively at a supply voltage below the specified Vmin, so as to reliably activate marginal timing failures that may just escape being detected by tests applied at and above Vmin. Many such failures can still be triggered during operation by less favorable circuit conditions and electrical noise, or by degradation due to aging. However, such overtesting risks yield loss because of the possibility of good parts failing when operated outside of functional specifications. To avoid this, we take advantage of the new result from this research that circuits that exhibit timing marginality likely contain a single extreme outlier transistor in the slow path. Consequently, the increase in the delay of this critical marginal path, when VDD is further reduced below Vmin, can be expected to be much larger compared to more reliable paths containing less extreme transistors. To understand this, notice from the (VDD-Vth) term in the Sakurai-Newton alpha-power law equation used to derive the increased delays for different threshold voltages in Table II, that a decrease in VDD has the same impact on delay as the same increase in Vth. Thus, given that the standard deviation σ in Table II is 25mV, a 25 mV reduction in VDD will increase the delay for a 5σ transistor by the same amount as a 1σ increase in Vth, which is shown to be approximately 225% (2.25 nominal gate delays) in the table. The same VDD decrease will increase delay of a 6σ transistor by a larger 550%, or 5.5 gate delays. The difference in the increased delay of will grow even larger between 6σ and 7σ transistors. This allows the possibility of overtesting with VDD below Vmin, but at a reduced clock rate that places the clock edge within this increased delay separation, to maximize the detection of extreme paths while minimizing yield loss. Moreover, some additional fallout from such overtesting, perhaps even as much as 500 DPPM (0.05% of the tested parts), may be acceptable if most of the hard-to-detect timing marginality failures in production are reliably screened out.

In the following we outline a two-step test optimization methodology for the final scan test that works with a relatively large (statistically significant) sample of production parts. Initial test results for each part are first generated and recorded for scan tests applied using the specified Vmin and corresponding test frequency, and also for the subsequent SLT tests performed on all the parts that pass this scan tests. Note that the additional failures observed only at SLT are functional failures that escape the scan test. Our goal is to find an optimum VDD, below Vmin, along with a reduced test frequency, that improves scan test effectiveness by detecting most, or at least many, of the parts that were earlier only detected by SLT. At the same time, scan failure of any additional parts that do not fail SLT (at the rated Vmin and frequency) and would therefore contribute to yield loss) should be minimized. Thus, scan test optimization in the proposed "overtesting" approach aims to use a more aggressive (lower) VDD, but a less aggressive (lower) clock frequency to apply the two-pattern timing tests to reduce the test escapes that are only detect at SLT by tests applied at Vmin and the rated clock frequency

Our test optimization approach first experiments with lowering VDD below Vmin in relatively small steps, e.g. 5mV, and re-running the scan test at each lower voltage, with the test frequency kept unchanged. Lowering VDD increases circuit delays and causes marginal scan test escapes that were earlier only detected as SLT failures to now also fail the scan tests. As VDD is further reduced, the detection of additional SLT failures ultimately drops off once most of the low voltage timing failures earlier only detected by SLT are now also detected by the reduced VDD scan tests. However, generally, not all the SLT failures will be detected. The SLT fallout may also include some timing independent hard defects that are not covered by the scan patterns. These will remain undetected. Note that the most effective voltage for applying the sub-Vmin scan tests is the highest VDD value that detects all or nearly all the voltage dependent SLT fallout. Meanwhile, at this lower voltage, the scan tests will likely also fail some number of good parts, which are not in the set of SLT failing parts. Clearly his yield loss needs to be minimized. For this, we can next optimally increase the

launch-to-capture time (reduce test frequency) of the scan test. This will prevent some of the scan timing fails, reducing yield loss. At the sweet spot in this trade-off, most of the voltage sensitive SLT fails will be detected by the scan test, with minimal yield loss. Our discussion earlier in this section describing how lowering VDD increases the spread between marginal and stable timing paths suggests the existence of a relatively wide sweet spot that can be exploited for this purpose. In practice, this optimization can be performed on the test floor in much the same way as test engineering currently sets VDD levels and test frequencies for scan timing tests.

VIII. CONCLUSION

Some recent studies have suggested that many of the hard to detect failures that escape traditional scan testing and increasingly require functional system level test (SLTs) for detection are supply voltage sensitive timing failures that are accentuated in low voltage operation. Where such timing marginalities escape all testing, they may be responsible for many of the unpredictable and difficult to diagnose intermittent errors being increasingly observed in field operation. To investigate this problem, we have presented a comprehensive study, combining analytical modeling with simulation, of the impact of random process variations on the timing of CMOS gates, and circuit paths, operating at significantly reduced voltages. Our analysis is validated with the help of production test data recently published by a large industrial team for an advanced FinFET technology. A key, somewhat unexpected, observation from our study is that virtually all variability paths that are statistical outliers slow enough to cause timing failure contain a single extremely weak transistor which contributes a large share of the increased delay. This suggests that TDF timing tests, that have traditionally targeted localized "lumped" delay defects, may remain effective for testing timing failures due to distributed process variations. Thus, the need for path delay testing to target such failures can perhaps be mitigated. We have further shown how the results of our analysis in this paper can be leveraged for conditioning the voltage and timing of the applied TDF and CAT scan tests to help enhance detection of such marginal parts. This can potentially reduce the need for expensive system level testing, and also minimize test escapes that can cause failures in the field. Future work will be focused on validating the results of this work in volume production in collaboration with industry.

ACKNOWLEDGEMENT

This paper is based upon research supported by the National Science Foundation under Grant CCF- 1910964. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

The author would also like to acknowledge Wendong Wang for assisting with the simulations.

REFERENCES

[1] W. Howell et al., "DPPM Reduction Methods and New Defect Oriented Test Method Applied to Advanced FinFET Technologies," Proceedings of the IEEE International Test Conference 2018, Phoenix, AZ.

[2] F. Hapke et al., "Cell-Aware Test," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 33, no. 9, pp. 1396-1409, Sept. 2014.

[3] X. Lin et al., "Timing-Aware ATPG for High Quality At-speed Testing of Small Delay Defects," Proceedings of the 15th Asian Test Symposium, Fukuoka, 2006, pp. 139-146.

[4] Adit D. Singh, "An Adaptive Approach to Minimize System Level Tests Targeting Low Voltage DVFS Failures", Proceedings of the International Test Conference, 2019, 1-10

[5] Hochschild, Peter H., et al. "Cores that don't count." Proceedings of the Workshop on Hot Topics in Operating Systems. 2021 https://dl.acm.org/doi/10.1145/3458336.3465297

[6] Dixit, Harish Dattatraya, et al. "Silent data corruptions at scale." arXiv preprint (2021) https://arxiv.org/abs/2102.11245

[7] Ilia Polian et al, "Exploring the Mysteries of System-Level Test. Proceedings of the IEEE Asian Test Symposium 2020: 1-6

[8] J. Wang, D. M. H. Walker, A. Majhi, B. Kruseman, G. Gronthoud, L. Elvira Villagra, P. van de Wiel and S. Eichenberger, "Power Supply Noise in Delay Testing", Proceedings of the International Test Conference, 2006, pp. 1-10.

[9] Dejan Markovic, Cheng C. Wang, Louis P. Alarcon, Tsung-Te Liu, Jan M. Rabaey, "Ultralow-Power Design in Near-Threshold Region" Proceedings of the IEEE, Vol. 98 , No. 2 , Feb. 2010.

[10] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas" IEEE Journal of Solid-State Circuits Volume: 25, Issue: 2, Apr 1990, pp. 584-594.

[11] M. L. Bushnell and V. D. Agrawal, "Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI Circuits", Springer, 2000.

[12] F. Hapke et al., "Cell-aware Production test results from a 32-nm notebook processor," Proceedings of the IEEE International Test Conference 2012, Anaheim, CA, pp. 1-9.

[13] Lin, C.J. and Reddy, S.M., 1987. On delay fault testing in logic circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 6(5), pp.694-703.

[14] Anees, Mohammad, et al. "Behaviour Shockley and Sakurai Models in 7nm FinFet." Proceedings of the IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) 2020.

[15] M.D. Giles et al "High sigma measurement of random threshold voltage variation in 14nm Logic FinFET technology" Proceedings of the IEEE Symposium on VLSI Technology 2015.

[16] T. M. Mak, A. Krstic, K. T. Cheng and L. C. Wang, "New challenges in delay testing of nanometer, multigigahertz designs", IEEE Design & Test of Computers, vol. 21, 2004, pp. 241-248.

[17] J. Saxena, K. M. Butler, V. B. Jayaram, S. Kundu, N. V. Arvind, P. Sreeprakash and M. Hachinger, "A case study of ir-drop in structured at-speed testing", Proceedings of the International Test Conference, 2003, pp. 1098-1104.

[18] Jeff Rearick, Richard Rodgers, "Calibrating clock stretch during AC scan testing". Proceedings of the International Test Conference 2005.

[19] D. Belete, A. Razdan, W. Schwarz, R. Raina, C. Hawkins and J. Morehead, " Use of DFT Techniques in Speed Grading a 1+GHz Microprocessor" Proceedings of the International Test Conference, 2002.

[20] Adit D. Singh, "Are Timing Marginalities Due to Process Variations the Cause of Silent Data Corruption?", Keynote, IEEE VLSI Test Symposium, 2022.