Defect or Variation? Characterizing Standard Cell Behavior at 90nm and Below
Robert Aitken
ARM R&D
Sunnyvale, CA USA
rob.aitken@arm.com

Abstract
Historically, design margin and defects have been viewed as different topics, one part of design and the other part of test. Shrinking process geometries are making the two part of a continuum. This paper discusses the leakage and delay behavior associated with classic resistive defects and compares it with transistor variation due to lithography.

1 Introduction
Test and design have evolved over the history of integrated circuits as separate topics, of mutual interest. Design for testability, DFT, is concerned about structures and methods to improve detection of defective parts. Defects cause faults, which are deterministic, discrete changes in circuit behavior. Design is not concerned with faults. Design worries instead about parametric variation, manifesting itself as design margin. This has led recently to renewed interest in design for manufacturability, or DFM, which is concerned with structures and methods to improve manufacturability and yield, mainly by reducing the susceptibility of design to lithographic variation. This state of affairs is shown in Figure 1. Test is concerned with distribution #2, while design is concerned with distribution #1.

Figure 1 Classic separation between design margin and defects

It has always been known that this separation is artificial, but it has remained because it is convenient. The first signs of a breakdown came in IDDQ test. Results such as those of Josephson [1] and Nigh [2] showed that there was no readily identifiable breakpoint between good and bad parts, leading to questions about the usefulness of IDDQ test [3].

Rather than being an anomaly, leakage variability, as shown by IDDQ test, has proven to be an indicator of the future. Circuit delay and operating voltage are now suffering from the same inability to distinguish parts with defects from inherent variability. While it may be argued that there is no need to discriminate, since both are “bad”, it is clear that choices made in these matters will directly affect yield and quality.

The remainder of this paper is organized as follows. Section 2 gives background on design margin and variability. Section 3 looks at process based leakage variability across a standard cell library and shows how it compares this to defect behavior. Section 4 expands the analysis to delay behavior. Section 5 examines the consequences of the overlap of defects and variability. Finally, section 6 concludes the paper.

2 Design Margin and Variability
Standard cell libraries for integrated circuits are specified to meet operating standards across a range of defined variations. These variations are derived from both the expected operating variations and manufacturing variations. Common variations include process (relative strength of N and P transistors), voltage, temperature, and threshold voltage. The design space in which circuit operation is guaranteed is called the operating range. In order to ensure that a circuit functions properly across the operating range, it is necessary to tolerate stress variations beyond the operating range. The difference between the design space defined by the stress variations and the operating region is called the design’s margin. The margin envelope is multidimensional, but for simplicity of discussion we will consider a single measurable property of a circuit at a time; e.g. leakage or delay.

2.1 Margin and Yield
Increased design margin generally translates into improved product yield, but the relationship is complex and difficult to quantify. Yield is composed of three components: random (defect-related) yield, systemic (design-dependent) yield, and parametric (process-related) yield. Electrical design margin translates mainly into improved systemic yield, with a small parametric component as well. Most layout margins (adjustments for improved manufacturability) translate into improved random yield, although some are systemic and some parametric.
2.2 Margin and Performance

In general, margin and performance are opposing requirements – an increase of one implies a decrease in the other. This can be seen in Figure 1. Unless the design is changed, the worst case margin point cannot be moved. Increasing the amount of margin can only come by reducing the worst case performance point. A new circuit design might allow for a new worst case margin point, presumably at the cost of extra design time and possibly some other design constraint (power, area), but margin and performance need to be traded off for this design as well. Of course, specific requirements exist for both margin and performance – without adequate margin, a design will not be robust enough for general use, and without adequate performance a design cannot meet market requirements.

2.3 Defects and Variation

Historically, test terminology distinguishes between “defects”, random events that cause silicon failure, and “design errors”, problems with a design that prevent it from functioning correctly. Process variation, sometimes referred to as parametric variation, occupies a middle ground between these two: design margin will account for some amount of variation, but beyond that a circuit will fail to operate. Had additional margin been provided, the design would have worked, so it is in some sense a design error, but the excess variation can equally well be thought of as a process defect.

2.4 Sources of variability

Variability occurs throughout a process, lot-to-lot, wafer-to-wafer, die-to-die and intra-die [4][5]. Three key sources of variation will be considered here: transistor length variation, oxide thickness variation, and ion implant variation.

Much length variation across a chip is continuous. A map of variation versus X,Y location will show a general trend. However, optical effects will lead to some variations that are more localized. Transistors with shorter channel lengths will be faster and leakier, while those with longer channels will be slower but less leaky. The leakage change is exponential for changes in L, so even small amounts of variability are significant.

Gate oxide thickness directly affects the electric field of a transistor, and so variation can significantly change its behavior. Wafer-to-wafer variation can be caused by changes in temperature and time for the oxidation process. Because oxide thickness is so vital, great care is taken to ensure uniformity, and localized variation is small, but the fact that modern oxides are less than 10 molecules thick means that oxide variation does contribute to leakage variation, especially at low temperatures where gate leakage is more dominant.

![Figure 2 Leakage and delay variation with threshold voltage](image)

Figure 2 Leakage and delay variation with threshold voltage

Ion implant variation is both localized and (mainly) random. The distribution of implanted ions is roughly Gaussian and depends on the ion’s energy. For lighter ions such as Boron, lower implant energy will result in higher variability [5]. Implants will also scatter beyond masks, resulting in changes to active areas, and off of nearby structures, resulting in higher variability near well boundaries. Implant variation between transistors leads to threshold voltage variation, leading to changes in both leakage and delay, as shown in Figure 2. As a result, leakage variation due to implant variation can be assumed to be independent on a per-cell basis.

Small variations in threshold voltage lead to substantial changes in leakage, but minor changes in performance. For example, an 80mV decrease in $V_T$ might lead to 10% improvement in drive current, at the cost of a 10X increase in leakage.

3 Experimental Results, Leakage

Changes in leakage come from many sources, for both defects and variability. For simplicity, two are considered here: a resistive short between a cell output and ground will be taken as a representative defect, and decreasing gate length will be taken as representative of a variability-based increase in leakage. Both are chosen because they represent common classes of behavior, and because they are relatively easy to model.

Experiments have been performed on a subset of commonly used cells from a commercial standard cell library in 90nm technology. Two variants of cells are considered: Standard $V_T$ (RVT) and high $V_T$ (HVT). A typical chip built in this technology would expect to use about 20% RVT cells in high performance paths and the remainder HVT cells. Since RVT cells tend to be about 10X leakier on average than HVT cells, this
means that about 70% of the total leakage comes from RVT cells.

For the first portion of the experiment, a set of standard cells was characterized with its nominal gate length increased by 8%, which was a value used by Kahng et al in their experiments [6]. The original characterization methodology was used to ensure good correspondence between results.

Figure 3 Leakage distribution for RVT cells

Figure 3 shows the leakage variation across a set of RVT cells (the leakage is listed in picowatts). On average, the 8% increase in gate length produced a 20% decrease in leakage. This value was very consistent across the cells. An 8% decrease in gate length similarly produces about a 25% increase in leakage.

Figure 4 Leakage distribution for HVT cells

Figure 4 shows the leakage variation for HVT cells. Again, approximately 20% reduction was observed, but the absolute value is significantly lower for these cells. Again, a 25% increase is observed for an 8% decrease in gate length.

In addition to these distribution results, Monte Carlo simulation was carried out for 5000 examples of a fixed cell. This showed the combined effects of all variations, not just gate variation. The results are shown in Figure 5. It can be seen that the distribution of leakage values is skewed right (towards higher values). This is likely due to the exponential dependence on transistor length [12]. The distribution has skewness of 4, and its standard deviation is equal to two thirds of the mean. The 5000 Monte Carlo samples show variation most of the way from the slow to fast corners. For commonly used cells, this suggests that variation outside the SPICE model is likely.

Figure 5 Monte Carlo distribution for one HVT cell

The average leakage for an RVT cell is 120nW, compared with 5.7nW for an HVT cell. Consider a design with 800K HVT cells and 200K RVT cells. Assuming that the cells are randomly placed across the die, the total expected leakage would be (120x200K) + (5.7*800K)nW or about 28.5mW. Assuming general variability similar to that shown in Figure 5, we would expect part to part variation of 16% in this total (assuming 1M independent cells with 67% individual variation, split 80-20 between HVT and RVT), or 4.7mW in this total. Now consider two scenarios:

Scenario 1, widespread length variation. In this scenario, all transistors become 8% shorter. In this case, the expected leakage changes to about 125% of 28.5mW, or about 36mW. This change will hold in all circuit states. This level of variation could be noticed at test time.

Scenario 2, local length variation. In this scenario a group of 50000 cells becomes 8% shorter. Again, assuming a consistent 80%-20% distribution of HVT and RVT in this group, the chip leakage will increase by about 300uW. There will be some state dependent behavior in this reduction, but it will be on the order of 10uW. This change will not be measurable amidst the noise of other states and is well below the 4.7mW of expected part to part variation.
Resistance values for defects can range arbitrarily. Studies have found that many defects concentrate in low resistance ranges [7], but there is speculation that this may have as much to do with detection techniques as it does with defect properties [8].

Consider a defect with similar behavior to the two scenarios above. For scenario 1, a defect that produces 7mW of additional leakage would be, by Ohm’s law, a short of about 140 Ohms. State dependent behavior for such a short connected to logic would be significant and readily detected via current signatures [9]. For scenario 2, a defect of 35kOhms will cause a 300uW increase in leakage. Current signatures might expose it, depending on equipment resolution, but this is about the limit of signature resolution [10].

Overall, it can be seen that with leakage measurement, low ohmic short defect behavior will be observable, as will large shifts in transistor parameters, and the two will be distinguishable based on current signatures. Higher resistance shorts and small or localized parameter shifts will not be observable at all due to high background leakage. Note that there may still be some measurement tricks that would help distinguish some defects from variability: changing voltage or temperature will change the behavior of defects and variable transistors differently, but care needs to be taken.

4 Experimental Results, Delay

While individual transistors do not affect overall leakage much, except in the case of low resistance defects, individual transistor variability is much more likely to be measurable in delay testing.

The concept of critical resistance has been defined to characterize delay defects. Shorts are active when the two nodes involved are driven to opposite values. A conducting path is formed from the power supply through active P transistors through the short (modeled as a resistance) through active N transistors to ground. The shorter (less resistive) the short, the more likely this conducting path will disrupt circuit operation. This basic idea can be extended to the concept of a “critical resistance” [7] below which the short “wins” and the circuit operates incorrectly, and above which the circuit “wins” and continues to operate correctly. There are actually multiple critical resistances for any short, as shown by Table 1, which lists delays associated with a bridge between an inverter output and ground. The waveforms associated with this table can be seen in Figure 6. There is a logical critical resistance, where the circuit will fail under all circumstances (below about 1700 ohms), a set of timing critical resistances (e.g. at 1800 ohms, a delay of about 150ps results), where the critical resistance depends on required timing and also on operating environment, and finally, a set of $I_{DDQ}$ critical resistance, where the defect will cause a significant enough increase in $I_{DDQ}$ to be observed.

It is clear that defects of relatively low resistance can cause significant changes in circuit delay. What about parametric variations? The same cell characterization data used in section 3 was examined for its effects on circuit delay. Interestingly, a length increase did not result in a uniform change in cell behavior, the way it did with leakage. Instead, delay effects varied widely, as shown in Figure 7.

![Figure 6 Delay waveforms, inverter output to GND short](image)

<table>
<thead>
<tr>
<th>Resistance (Ω)</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1700</td>
<td>$SA0$</td>
</tr>
<tr>
<td>1720</td>
<td>600ps</td>
</tr>
<tr>
<td>1730</td>
<td>400ps</td>
</tr>
<tr>
<td>1750</td>
<td>250ps</td>
</tr>
<tr>
<td>1800</td>
<td>150ps</td>
</tr>
<tr>
<td>2000</td>
<td>70ps</td>
</tr>
<tr>
<td>3000</td>
<td>&lt;10ps</td>
</tr>
</tbody>
</table>

Table 1 Delay versus resistance, inverter output to GND short

The X axis of Figure 7 shows the initial cell delay in ns. The Y axis shows the increased delay resulting from increased gate length. The data are clearer when summarized as a histogram in Figure 8. Typically, the 8% increase in gate length results in a 4% to 5% increase in delay.

Now consider the two scenarios from section 3 again, as they adapt to the delay case:

**Scenario 1a, widespread variation.** In this scenario, all transistors become 8% longer. In this case, the overall circuit delay increases by about 5%. For a circuit on the slow end of the process, this increase will likely mean a timing fail. On the fast end of the process, it may not be noticed.
Scenario 2a, local variation. In this scenario all cells on a path have 8% longer transistors than expected. Just as with scenario 1a, a single path that is 5% longer than expected can cause the circuit to fail, if it happens to be a near critical path.

Figure 7 Fractional delay increase versus initial cell delay

Figure 8 Fractional delay increase of Figure 6 as a histogram

In both cases above, a defect can produce identical behavior. A resistive short in a major branch of the clock tree can cause multiple clocks to be delayed, resulting in numerous failures of timing tests, similar to scenario 1a. For scenario 2a, a 5% increase in a nominal path delay of 3ns is equivalent to a 150ps delay increase, which could be caused by a 1800 ohm output to ground short in an inverter along the path. Thus, in this case at least, parametric variation and defect behavior are identical.

5 Consequences of Defect and Margin Overlap

Two results stand out from the preceding discussion: 1) there are classes of defects that behave externally in a very similar fashion to classes of design margin, and 2) extending design margins will result in untestable defects. The reliability implications of the latter are not clear. Also, in the absence of perfect path delay testing, some functional failures may still occur if these defects are not identified. The behaviors are summarized in Tables 2 and 3 (note that the extension to open defects in these tables is straightforward).

### Table 2 Leakage behavior and causes

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Possible Short Defect</th>
<th>Possible Margin Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher than expected leakage</td>
<td>Low resistance active short</td>
<td>Unlikely</td>
</tr>
<tr>
<td>(state dependent)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher than expected leakage</td>
<td>Low resistance inactive short</td>
<td>Widespread parametric variation</td>
</tr>
<tr>
<td>(not state dependent)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State dependent</td>
<td>Medium resistance active short</td>
<td>Substantial localized variation</td>
</tr>
<tr>
<td>leakage change around expect value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No change in</td>
<td>Medium resistance inactive short, high</td>
<td>Minor variation, either local or</td>
</tr>
<tr>
<td>leakage behavior</td>
<td>resistance active short</td>
<td>widespread</td>
</tr>
</tbody>
</table>

### Table 3 Delay behavior and causes

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Possible Short Defect</th>
<th>Possible Margin Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple timing test failures</td>
<td>Low to medium resistance short in clock</td>
<td>Widespread parametric variation</td>
</tr>
<tr>
<td></td>
<td>logic</td>
<td></td>
</tr>
<tr>
<td>Single timing test failure</td>
<td>Low to medium resistance short on near</td>
<td>Minor or substantial localized</td>
</tr>
<tr>
<td></td>
<td>critical path</td>
<td>variation on near critical path</td>
</tr>
<tr>
<td>No timing failures</td>
<td>Medium to high resistance short on non-</td>
<td>Variation on non-critical or untested</td>
</tr>
<tr>
<td></td>
<td>critical or untested near critical path</td>
<td>near critical path</td>
</tr>
</tbody>
</table>

As shown by Tables 2 and 3, significant numbers of shorts are not tested. For the short in Figure 5, the
critical resistance range between a hard fault and a delay fault is very small, less than a 5% change in magnitude. Such a fault could easily change over time and result in a system reliability failure. It may be possible to detect these using “faster than normal clock” testing [11], but the results in this area are very preliminary.

What is clear is that the ideal margin and defect distributions shown in Figure 1 actually overlap, and this leads to two possible approaches, shown in the Figure 9.

In the figure, the dashed line represents the point where a test is run and the solid line represents the point where the design margin guarantees it will work. In the conservative approach, shown in 9a, there is potential yield loss due to marginal devices that fail the test but would work in the system, but reliability issues are reduced since many defects are found. In the “bad guess” approach, where a test is set outside the margin bounds, a group of untested defects remains, along with a set of out-of-margin variations. It is unknown what the behavior of these would be in practice, so they are a potential quality issue.

![Figure 9 Relationship between test point and margin point](image)

6 Possible Solutions

Given that a single test may not be able to distinguish variation from defectivity, the simplest approach may be to apply multiple tests. For example, the same test may be run at different temperatures. The temperature behavior of defects tends to be linear, while the temperature behavior of devices is exponential, so marginal overlap at one temperature may be well separated at another. For cost reasons, it might be better to run high temperature test at package test and low temperature test at wafer sort. This approach also has the advantage of detecting poorly formed vias and other temperature dependent defects.

Similarly, earlier work has shown that lowering VDD can distinguish between defects and slow circuits [13], mainly due to the tendency of resistive wiring defects to cause circuits to slow down much more at low voltage than they would simply through weak transistors. Current technologies lack the voltage headroom assumed in earlier work, but the expected differences in behavior will continue to apply, albeit at a reduced magnitude.

7 Conclusions

This paper shows that in the presence of realistic design variation, there will be classes of shorts that are difficult to distinguish in delay and/or leakage behavior from parametric variation, and that in addition, a substantial class of defects may not be detected at all. The small range of resistances between logic critical and delay critical resistance show that the potential for these untested faults to be reliability problems is large. Testing at multiple voltages and/or temperatures is suggested as an option to mitigate these problems.

8 References


