

The DFM Pandemic: How Many Chips Have to Die?

By P.T.Patel, Pyxis Technology

Introduction

In the context of digital integrated circuits (ICs), the phrase *design for manufacture (DFM)* refers to a variety of techniques used during the process of creating the design so as to facilitate its being manufactured. Meanwhile, the term “yield” refers to the number of die that work as a percentage of the total number of die on the silicon wafer; hence, the phrase *design for yield (DFY)* refers to any techniques used to improve the yield of a particular device. In reality, these concepts and techniques are so intertwined that it is becoming common to consider them as being a single DFM/DFY entity.

Until recently, design engineers had relatively little concern with regard to manufacturability or yield. In the case of the current 90 nanometer technology node, however, manufacturing and yield issues are being pushed upstream into the design process, because these factors are strongly design dependent and are affected by how the design is laid out. If these problems are not addressed, it may not be possible to achieve economically viable yields at the forthcoming 65 and 45 nanometer technology nodes.

There have been some efforts with regard to DFM/DFY random (statistical) yield; to date, however, almost all of the focus on systematic yield issues has been in the area known as *front end-of-line (FEOL)*. (The concepts of random and systematic yield are introduced later in this paper.) FEOL encompasses the diffusion, poly, and contact layers forming the transistors; that is, everything below metal layer 1. At the 90 nanometer node, however, the routing in the metal layers is becoming a significant systematic yield issue which no one has addressed thus far. As we move to the 65 nanometer node and below, routing will become a major yield limiter for designs.

This paper first introduces the DFM/DFY problems associated with technology nodes of 90 nanometers and below. Next, the paper considers the traditional split between design and manufacturing coupled with the inadequacies of current DFM/DFY techniques. Finally, the paper discusses the way in which these problems can be addressed by means of next-generation routing technologies.

DFM/DFY Problems at 90 Nanometers and Below

As is discussed below, there are several different aspects to manufacturability and yield. The reasons why yield and manufacturability are important may be summarized as follows:

- The chips (and associated products) may completely miss the market window.
- The chips (and associated products) may hit the market window, but the chips may cost too much to make the products economically viable.
- The chips may not perform at required level; that is, they still may function, but not at the required speed.

With regard to yield itself (where yield is a function of the device's manufacturability), there are three main "buckets" into which problems may be categorized; these are commonly referred to as *Random Yield* (sometimes called *Statistical Yield*), *Systematic Yield*, and *Parametric Yield*.

Random (Statistical) Yield

As its name suggests, this form of yield is a function of random "happenings" that may occur during the manufacturing process flow. For example, no matter how clean the wafer manufacturing environment, there are always some number of small particles in the atmosphere that may land on the surface of the chip (Figure 1).

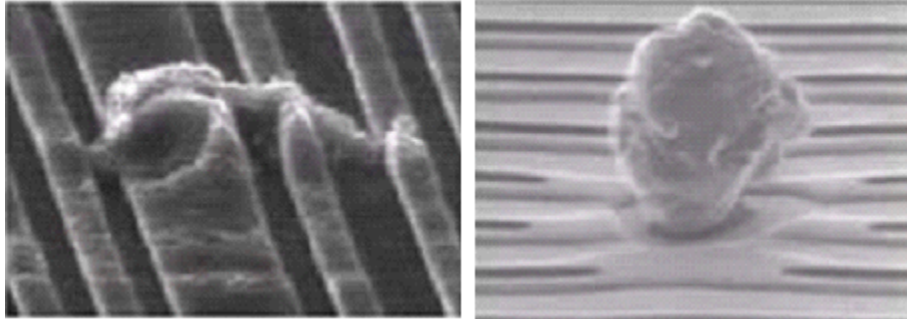


Figure 1. Examples of random defects affecting statistical yield

Such particles may cause catastrophic faults in the form of open or short circuits. Alternatively, in some cases they may cause parametric variations. For example, a particle may land on a non-critical area of a particular layer and may cause a non-planar feature (bump) in subsequent layers. In turn, this bump may end up varying the width or thickness of a wire on a higher layer, thereby changing the electrical characteristics of that wire resulting in a parametric yield failure as discussed below.

By their very nature, random happenings are difficult to control. However, it is possible to create the design in such a way as to minimize their effects on final yield.

Systematic Yield (Including Printability Issues)

The term "systematic" encompasses the concepts of "logical," "methodical," and "ordered." Thus, systematic yield refers to a class of manufacturability issues that can be identified and addressed in a systematic way. Systematic yield issues relate to the way in which the various process steps are performed. For example, the intensity of the light source and the focus of the lens used in the lithographic processes may vary; similarly for the temperature, duration, and concentration of chemicals used in the various development and etching related process steps.

Many systematic yield issues are design-dependent. For example, some designs may have high densities (concentrations) of wires in certain areas and low densities in others. Such density variations can affect the amount of etching that takes place in the various regions. Similarly, in the case of process steps like chemical mechanical polishing (CMP), variations in wire density can cause differences in the effectiveness of the polishing process, which can result in areas where some wires are thinner than others. In turn, this affects the resistance and capacitance values associated with these wires, which can modify the power and performance (timing) of the design.

Another facet of systematic yield that has become extremely significant at the 90 nanometer technology node may be referred to as *printability* issues. In order to appreciate what this means, it's necessary to understand that the output from the design portion of the process is a set of GDSII (or similar) files. These files are used to generate the photo-masks, which are – in turn – used to create the silicon chip.

In earlier technology nodes, the wavelength of the light used to create the structures and wires on a silicon chip was smaller than the size of the structures themselves. To a large extent, this meant that the patterns in the GDSII files and the ensuing photo-masks could be directly replicated on the surface of the silicon chip. By comparison, in the case of the 90 nanometer node the wavelength of light used to create the patterns on the chip is larger than the structures and

wires themselves. The resulting physics creates substantial interference effects, which means that – unless treated – the patterns in the GDSII files and photo-masks will be distorted when printed.

Furthermore, these interference effects are now so significant at the 90 nanometer node that placing one wire in close proximity to another wire, for example, can result in one or both wires printing badly. This is similar to the way in which electrical signals can experience crosstalk effects, but in this case the crosstalk occurs at the lithographic level causing the patterns to be printed to interfere with each other.

Some of these issues can result in catastrophic failure in the form of a dead chip. Others can leave the chip functioning, but can modify the power and performance (timing) of the design as shown in Figure 2 (the white outlines reflect the ideal wire patterns.).

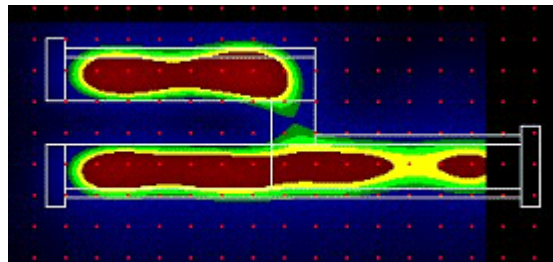


Figure 2. Examples of printability issues affecting systematic yield

In this example, the two horizontal portions of the wires to the left of the image reinforce each other and print correctly. With regard to the vertical wire segment in the middle of the image, however, the two horizontal wires interfere with each other resulting in the vertical segment not printing at all. Furthermore, the upper horizontal wire is not present to reinforce the lower horizontal wire in the right-hand side of the image; this results in the lower wire printing badly and being "pinched," which will affect its resistance and associated timing.

In many cases, detailed analysis of the failed device can trace the problem back to some root cause. Once a particular problem is understood, it may be possible to define a design rule to prevent such an occurrence happening in the future (the concept of design rules is introduced later in this paper). In many cases, however, it may not be possible to even formulate such a rule.

Parametric Yield (Including Variability Issues)

The concept of parametric yield refers to the fact that a chip may perform its logical function correctly (*"stimulus X returns response Y"*), but variations in the device's parameters may mean that it does not achieve its specified performance goals. If transistor channels aren't formed quite as expected, for example, the result may be lower drive capabilities, increased RC time constants, and slower chips. Alternatively, issues in the etching and CMP processes may cause non-planarity in the surface of the chip; in turn, this can cause wires to have higher resistances and/or capacitances than expected, which will result in the device's speed falling and its power consumption rising (Figure 3).

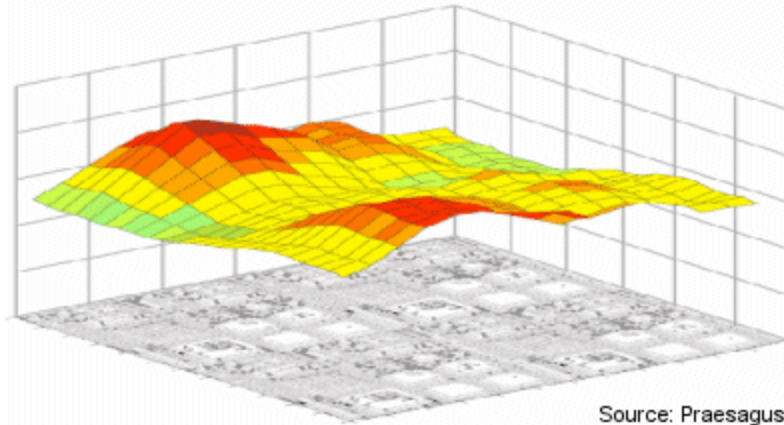


Figure 3. Copper "dishing" affecting parametric yield

In this figure, the colored region, which reflects the topology of the surface of the device, should ideally be flat. However, non-uniform wire densities have affected the etching and CMP processes resulting in a non-planar surface.

In some cases, such as a microprocessor, the chip may still be useful – although less profitable – in that it can be sold at a lower speed grade (1.5 GHz versus 2.0 GHz, for example). In other cases, such as network switches, failing to meet specified performance goals means that the chips are essentially useless.

In the case of the 90 nanometer technology node, one aspect of parametric yield that has become extremely significant is that of *variation* or *variability*. There has always been an issue with regard to inter-wafer variation, which refers to slight differences – such as variations in the sheet resistance – between wafers in a lot. More recently, intra-wafer effects started to become apparent; this refers to variations between different die on the same wafer and how the die track each other. Now, in the case of the 90 nanometer node, we are starting to see substantial intra-die (on-chip) effects, which refers to variations between different areas on the same die. The reason this is such a problem is that the layers forming the chip are much thinner than before – and the structures being created are much smaller – so small changes are becoming a significant proportion of the total value.

Complex Interrelationships

Although there is a tendency to wish to categorize different problems into specific "buckets," in reality many of the issues presented above are interrelated. For example, a random fault may cause a catastrophic failure in the form of an open or short-circuit, or it may result in a parametric modification which – if significant enough – may affect the parametric yield. Similarly, a systematic effect such as a printability issue may result in a parametric yield problem.

The real problem is that, with the introduction of the 90 nanometer node, effects that used to be third or fourth order – and could therefore be largely ignored – are now assuming first order status. Existing design tools and flows are finding it difficult to meet the demands of this technology node, and will be totally inadequate when it comes to the forthcoming 65 and 45 nanometer nodes.

The Split between Design and Manufacturing

Until recently, design engineers had relatively little concern with regard to manufacturability. As long as the design met a few simple rules – such as wires meeting minimum width and spacing values – it was assumed that the device could be manufactured.

Similarly, with the exception of specialist teams working on extremely high volume products such as SRAM devices, design engineers simply did not concern themselves with yield issues, which were considered to fall wholly in the fab's domain. Once the device was in production, it was the fab's responsibility to analyze and modify the process so as to bring up the yield.

Furthermore, yield issues were not significantly design-dependent. With the introduction of a new technology node, assuming that Design #1 had been brought up and the process flow had been tuned for maximum yield, the fab's engineers were reasonably confident that subsequent designs could be fabricated with minimal problems.

With the advent of the 90 nanometer technology node, however, such assumptions no longer hold true. For example, it is now possible to bring Design #1 up and tune the process flow. When Design #2 is introduced to the fab, however, the yield may fall dramatically or the device may fail in its entirety. In many cases, this is because the way in which the second design was laid out interferes with the manufacturing process in a non-friendly way.

The end result is that the complex relationships and dependencies between the various factors as discussed earlier in this paper are pushing manufacturing and yield issues upstream into the design process, because these factors are strongly design dependent and are affected by how the design is laid out. This is a very difficult problem to solve, but it *must* be solved in order to produce yieldable products at the forthcoming 65 and 45 nanometer technology nodes.

Inadequacies with Current DFM/DFY Techniques

There are a variety of techniques that are currently employed to increase manufacturability and yield. These approaches are generally considered to be *rule-based* or *model-based* as discussed below.

Rule-Based Techniques

The term *design rules* refers to a collection of rules that must be met by the physical design engineers and their tools. Examples of these rules would be the minimum width of wires and the minimum spacing between wires. One problem is that the number of such rules is increasing dramatically with each new technology node. In the case of the 180 nanometer node, for example, there were typically only a few dozen such rules; the 130 nanometer node saw several hundred design rules; current 90 nanometer node processes have design rules in the high-hundreds to low-thousands; the forthcoming 65 nanometer node may have several thousand design rules; and it's difficult to conceive the state of play with regard to the forthcoming 45 nanometer node.

In addition to the fact that we now potentially have thousands of such rules, the situation is further complicated by the fact that there are often not only a set of parameters associated with each rule, but each parameter may have gradations in the form of a range of possible values. This is akin to telling the tool: "Try *this*, but if you can't do this try *this*, and if you can't do this try *this*," and so forth.

Applying thousands of rules to millions of patterns takes a substantial amount of computational power. In many cases the rules are so restrictive that the result is to guard-band the design so as to leave a significant amount of performance on the table. In some cases, the design ends up being so guard-banded that it is impossible to achieve its original performance goals. Even worse, the complex relationships between different manufacturability and yield mechanisms means that in many cases it is simply not possible to actually formulate an appropriate rule in such a way as to be meaningful to the design tool.

Model-Based Techniques

There are a whole class of tools that are used to modify the GDSII (or similar) files generated by the physical design process. These tools, such as optical proximity correction (OPC), are collectively known as resolution enhancement techniques (RET). In the case of OPC, for example, the tool may be used to increase the widths and spacing of wires.

Similarly, some tools may augment the GDSII file with special constructs known as sub-resolution assist features (SRAFs). In the case of printing a single wire in isolation, for example, it may be advantageous to space two much thinner traces on either side of the original wire. These features are small enough on the ensuing photo-mask that they won't actually print onto the silicon, but they will reinforce the desired pattern for the original wire. Similarly, multiple wires may act as assist features for each other. The problem is that moving wires around and widening them and changing the spacing between them can cause other wires to not print as well as they should.

Early versions of these GDSII-modification tools were rule-based, but modern versions use a different approach known as model-based. The idea here is that, before making a modification, the tool uses appropriated mathematical models to simulate the effects of that change to see if there's going to be any collateral damage. This approach allows the number of rules used by the engine to be reduced and the remaining rules to be simplified, while still returning a much higher quality of results.

Problems with Post-Processing

A key consideration with regard to today's extremely complex, high-performance designs, is that any tradeoffs need to be evaluated during the design portion of the flow so as to fully understand the impact of any decisions. The problem with post-processing the GDSII files as discussed above is that they occur "open-loop" outside the main design flow.

For example, in the case of a DRC/DFM post processing tool widening tracks and increasing the spacing between tracks in order to minimize random yield effects, these changes may negatively impact the timing of the design. Furthermore, this post-processing tool may introduce lithographically "unfriendly" jogs into the design, which may interfere with the printability of surrounding patterns.

Yet another consideration is that these post-processing engines typically work in isolation with little understanding of the implications of their modifications with regard to other aspects of the design. In addition to the design rules discussed earlier in this section, for example, there are also a variety of *recommended rules*, such as adding redundant vias wherever possible. This type of recommended rule is often applied by means of a special engine that post-processes the GDSII file. But adding redundant vias at this stage may create lithographically "unfriendly" patterns; these patterns may need to be addressed by re-running the OPC tool, which can – in turn – cause further problems.

Summary: The Solution is in the Routing

A very important point that is often overlooked is that the "D" in both "DFM" and "DFY" stands for "Design." On this basis, post-processing the GDSII files to correct problems introduced by the upstream design tools cannot truly be considered to be DFM/DFY.

The solution is to bring DFM/DFY upstream into the design process; to create a design that is correct by construction; and to hand-off a design that is as manufacturing and yield-friendly as possible. The obvious candidate to subsume DFM/DFY analysis and implementation is the routing engine, because routing currently accounts for approximately 80% of the design's delays, and the routing engine is the "front door" into the manufacturing process.

Having decided that the routing engine needs to become truly DFM/DFY-aware, another big consideration is that today's EDA tools in general – and routing engines in particular – are choking on the sheer number of design and recommended rules. The number of these rules is increasing dramatically with each new technology node; the rules themselves are becoming extremely complex; but, at the same time, it is becoming increasingly difficult to adequately address manufacturability and yield issues using a rule-based approach.

The answer is to bring model-based techniques into the routing domain. In this case, as an integral part of making any decisions or modifications, the router would use appropriate mathematical models to simulate the effects of such decisions or modifications (Figure 4).

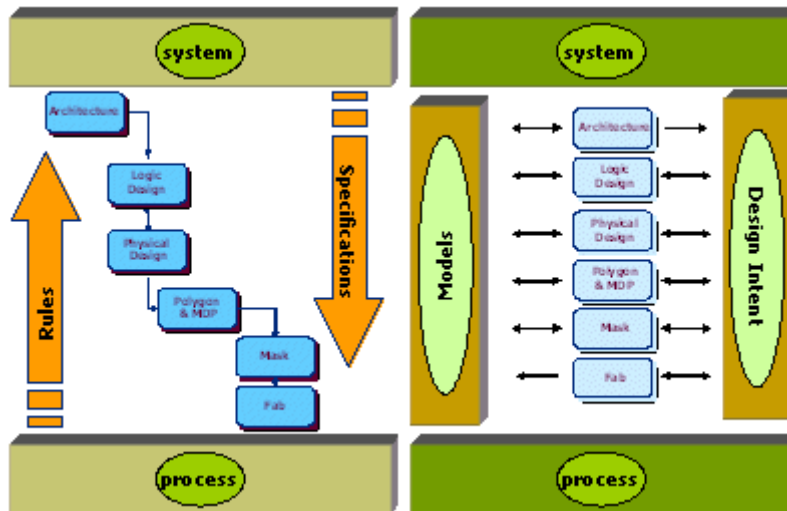


Figure 4. Old rule-based flow (left) compared to new model-based flow (right)

One final consideration is that existing routers are simply not capable of addressing these issues. What is required is a completely new routing architecture. A routing engine based on this new underlying architecture should be capable of performing as many DFM/DFY-related actions as possible. These actions include, but are not limited to, wire widening, wire spreading, redundant via insertion, and minimizing jogs. While performing these actions, the router must take full account of timing and signal integrity effects like noise and crosstalk. Such a router must be capable of performing multi-variable, multi-value optimizations, and it must also be capable of making decisions such as using minimum width wires or not introducing redundant vias in certain cases, where all such decisions are made in the context of predicted yield.

About the author

P.T.Patel is Founder and CEO of Pyxis Technology and brings over 30 years of microprocessor physical design expertise to the company. Prior to founding Pyxis, PT spent 3 years at Sun Microsystems where he was Senior Distinguished Engineer, responsible for setting the vision for microprocessor tools and methodologies for 65nm and below technologies.

Prior to Sun, PT was with IBM for 27 years where he served as Distinguished Engineer and led the physical design effort for the Power PC family of microprocessors. He developed structured design methodologies targeted towards reducing total time to market and provided technical leadership for physical design tools development at IBM.

PT is a recipient of Corporate and Outstanding Technical Achievement awards from IBM for designs, tools and methodology innovations for Power and PowerPC microprocessors. He received the Chairman's Award at Sun for CAD tools and methodology innovations. He has been granted 16 U.S. patents and has several publications on chip design, tools and methodology. PT has an MS degree in electrical engineering from University of Connecticut.