# 20nm Design

How this Advanced Technology Node Will Transform SoCs and EDA

A number of manufacturing issues specific to 20nm pose a challenge to developing high-quality silicon and SoCs on time and on budget. Silicon Realization at such an advanced node requires a holistic approach consisting of three critical and interrelated components: unified design intent, higher levels of design abstraction, and design convergence at each stage of the flow. This approach, coupled with new lithography techniques and in-design DFM technologies, provides the most practical and predictable path to 20nm silicon.

## Contents

Introduction1	
20nm Design Solutions2	
A Quick Look at 20nm Challenges2	
Unified Design Intent Ensures Manufacturable Designs	
The Challenges of Double Patterning5	
Design Abstraction Supports Greater Complexity and Capacity7	
Convergence Throughout the Flow Eases Final Signoff9	
Conclusion9	

# Introduction

Manufacturing issues take center stage for design flows at the 20nm technology node, and no wonder: at 20nm, chips have feature sizes 10 times smaller than the wavelength of the laser light typically used in lithography. At these wavelengths, lithography is clearly pushing its luck—and design teams are not in the habit of relying on luck to get working silicon.

Designers need automated solutions that deal with the manufacturing requirements of 20nm chip designs and, specifically, with new lithography techniques. One technique in particular—<u>double-patterning technology</u>—has such a pervasive impact on the silicon that nearly every step in the design flow must anticipate and prevent double-patterning problems.

Each new technology node brings design and integration complexity to a new level. Server signals and power electromigration create challenges. Decreasing metal pitch leads to coupling effects and signal integrity issues. Increasing wire and via resistance requires more advanced and variable wire sizing and tapering techniques. Extraction, timing, signal integrity analysis, and modeling must take a multitude of variation issues into account to achieve accuracy without compromising performance. Lithography limitations at 20nm often require a great deal of fixing to achieve signoff. Moreover, there are multiple chip and IP integration challenges, packaging issues, and the complexity of all these issues interacting.

It is said that smart people can fix anything, and this assumption may continue to be true at 20nm. But how many iterations will it take to reach design closure on all counts? How many re-spins afterwards? How about a predictable yield? In short: what are the costs, both in time and money, of the "fix-it-at-signoff" approach? The specific costs are not yet known, but solutions for minimizing design risk and costs are available.

## 20nm Design Solutions

Because of manufacturing challenges and many other issues, the big idea to grasp at 20nm is interdependency: each stage of the design flow affects and is affected by every other stage (see Figure 1). This idea is not really new, since a chip's performance became dependent on physical layout way back when CMOS geometries got smaller than a micron. What is unique to 20nm is the deep and complex dependency of manufacturing and variability issues on top of the timing, area, and power issues designers were already facing.



Figure 1: With the added complexity of manufacturing issues at 20nm, each stage of the flow is interdependent

A new solution is available today to deal with this complexity. This solution integrates tools, methodologies, and models in a unified approach to <u>Silicon Realization</u>, and it hinges on three critical components:

- 1. Unify the design intent throughout the design flow Ensure convergence by including mixed-signal interoperability and using the same timing/area/power/manufacturing constraints for each tool, from chip planning to signoff.
- 2. Abstract portions of the design to improve productivity Abstracting a highly complex design into smaller data sets enables faster design exploration and prototyping, as well as verification and layout. Managed correctly, the design variables of the abstracted sections will correlate in the overall implementation. Abstraction capabilities also need to support power-aware IP reuse as well as the ability to modify block-level parameters during top-level optimizations.
- 3. Begin design convergence from the beginning of the flow To avoid a crush of problems at signoff, run signoff checks throughout the flow. By using fast, parallelized tools to run these checks, they take only a little additional time.

All three components are necessary to meet 20nm challenges in an efficient, predictable way and the three components are interrelated.

Which brings us back to the big idea: everything about 20nm design and implementation is interrelated and interdependent in complex ways. So even though good point tools are crucial in a 20nm design flow, focusing on point tools alone misses the point. The focus needs to expand beyond traditional tool boundaries to include coherent integration so that tools work together in intelligent ways—not just obeying lists of rules, but interactively determining design features that are manufacturable.

## A Quick Look at 20nm Challenges

Before getting to the details of the design solutions, it is useful to clarify why the 20nm technology node needs special care. The areas of concern include:

- Chip complexity and capacity
- Meeting performance goals

- Power management
- Lithography limitations
- Process complexity and variability
- Packaging complexity
- Schedule and project management challenges (engineers/project cost)

## **Design Challenges**

Most of these issues have been growing in importance over several technology nodes. In the case of chip layouts, for example, a smaller die means less I/O space. Additionally, the higher levels of integration available from advanced process technologies lead to greater use of memory blocks and mixed-signal blocks that are implemented as large hard macros. With these blocks no longer confined to the periphery of the die, most 20nm SoCs will have an irregular layout that can cause severe routing congestion. At the same time, many blocks in these chips need to talk to each other, often using wide buses, which further increase congestion. A custom design methodology is usually needed to handle such designs.

## Manufacturing Challenges

When the designs get to manufacturing, they encounter a new world of advanced-node lithography, process, and packaging issues. This world includes an ever-growing number of foundry rules, with relatively new complications due to the use of high-K metal gates (HKMGs) and silicon-on-insulator (SoI) technology that can involve as many as 50 layers. The use of stress-and-strain engineering causes varying electrical effects that depend on layout features (described later). Rule-based metal fill does not take into account multi-layer effects and so may cause varying results. Double-patterning technology has its own ecosystem of issues (also described later).

In addition, due to the increased complexity in lithography, conventional model-based optical proximity correction (OPC) and resolution enhancement techniques (RETs) are not sufficient to deliver the required silicon pattern fidelity. Source mask optimization (SMO) enables more accurate computational lithography assessments and tradeoffs that improve pattern fidelity and increase product yield. Packaging issues may include thermal and stress challenges due to 3D stacked dies (3D-ICs) and through-silicon vias (TSVs).

## The Interdependency Challenge

The critical take-away from this overview is not that 20nm is burdened with a lot of design rules. If that were the only challenge, EDA tools would carry on as usual, automatically applying and verifying all the rules. But that approach is unlikely to prevent re-spins at 20nm.

The important idea at 20nm is that just following rules is not enough to deal with the interrelated nature of the challenges. A chip's electrical properties vary with lithography and process and packaging details that vary depending on the chip's layout. From floorplanning to signoff, tools must interact to make countless adjustments and tradeoffs.

At the floorplanning stage, these adjustments may be simple: do not place cells with certain internal characteristics too close together, for example. If these adjustments are not made during floorplanning, the router may find it impossible to meet all design objectives even after weeks of tweaking to pass signoff.

While the complexity and interdependency of design and manufacturing challenges at 20nm has multiplied, the time to revenue remains the same. That leaves design teams in dire need of more deterministic, automated solutions.

## Unified Design Intent Ensures Manufacturable Designs

As mentioned earlier, one critical component of the solution is to unify design intent throughout the flow. This can best be understood by looking at power optimization and DFM examples. The importance of these examples is easy enough to see: the main reason for re-spins at 20nm is the need to modify designs to minimize leakage and increase yields (source: IBS).

#### Unified intent for low-power design

One low-power design approach coarse-grain power gating demands automated support to make the necessary tradeoffs among gate size, placement, routing, simultaneous switching analysis, and the slew rate of the gate-control signal. (Note that signoff-quality signal integrity (SI) is vital for this control network because a spurious signal caused by an SI aggressor could shut down an entire module.) By making tradeoffs far more rapidly and accurately than is possible with manual choices, automated power-switch prototyping reduces the number of design iterations and makes more efficient use of die area.

This automation depends on several tools sharing the same design intent. For example, even if the synthesis tool automatically optimizes power-related structures (such as multiple power domains) and even if analysis tools simplify verification, the physical implementation methodologies must follow through on the necessary strategies. In fact, physical implementation—ranging from floorplanning to routing—might be considered the most critical aspect of most power-management techniques.

Early in the flow, floorplanning and design partitioning play important roles in reducing power consumption. Since interconnect capacitance is a big component of dynamic (switching) power, for example, wires with high switching probabilities must be kept as short as possible. That usually means keeping such wires within a partition.

Cell placement needs to account for this requirement without taking a lot of iterations to determine which wires contribute the most capacitance to the power equation. In particular, multi-power-domain designs need partitioning and floorplanning strategies that support each domain's requirements. Doing so requires integration with analysis tools that can accurately predict power relationships even at the floorplanning stage—a nontrivial task, but an essential capability for meeting power goals.

Similar requirements apply to routing, and better information is available at this stage of the flow to support accurate power analysis. The physical implementation tools must use this information in a number of ways to meet power goals. For multiple power domains, the tools must handle signal routing and power connections with level shifters in suitable ways. The router must "understand" the power-down-switch placement scheme to implement any power-down approach. For distributed switch layouts, it is important to place the switch and then make sure it provides adequate current-carrying capacity. The switch-enable distribution is also critical because this signal utilizes a high-fanout net that is always on. Thus, buffers on this net must be kept out of partitions that will be powered-down.

Use of the Common Power Format (CPF, managed by the Si2) simplifies the task of specifying design intent. For example, CPF 1.1 macro modeling capabilities help in mapping the power domain of an IP block to the top-level design. This mapping recognizes and honors IP-level rules and modes, allowing easier integration of hard and soft IP into a design. The arrival of CPF 2.0 brings more complete coverage of power parameters and provides an expanded set of low-power constructs.

## Unified intent for DFM

Two DFM examples show how unifying design intent throughout the flow is vital for 20nm SoCs and IP. The first example is about layout-dependent effects (LDE) variability due to the use of stress liners, which is mechanical stress intentionally applied to improve CMOS transistor performance. The stress itself is useful; variability is the problem. At 20nm, stress-induced variability can result in unexpected timing variations of 15 percent or more.

Stress is also unintentionally and non-uniformly applied due to the constraints of a chip's layout, and is unintentionally induced through various technologies such as shallow trench isolation (STI). Thus, design teams still need to expect variability even if a silicon process does not intentionally induce stress.

Evaluating LDE and its variability is not straightforward, because the evaluation must consider proximity effects. The location and dimensions of a transistor's neighboring layout features change the surrounding stress, and therefore the timing performance.

Most IC design teams handle stress-induced variability by applying guard bands to critical paths—ironically robbing some of the performance most often needed in critical paths. A better approach is for tools to model stress-induced variability and intelligently adjust timing parameters only as needed. This approach requires a specialized analysis tool to evaluate standard cells for LDE variability in various contexts. This capability becomes even more critical with double patterning as the cells need to be analyzed for multiple placement possibilities.

This stress variability analysis can be used two different ways. Library designers can use the information to optimize their libraries, and the information can become part of the "design intent" for use by the tools in an SoC flow. For example, LDE variability information can drive double patterning-aware placement optimization by enabling the placement engine to swap cells in the layout to mitigate both double patterning- and LDE-induced variability.

The variability information can also be used for de-rating standard cells during static timing analysis (STA) to reduce excessive design margins. Unlike the use of one-size-fits-all guard bands, an intelligent modeling approach adjusts margins based on a known level of sensitivity for a specific cell. This approach retains the performance of the 20nm process while preventing failures due to timing variations.

# The Challenges of Double Patterning

One wrinkle, new with 20nm design, is the need for double patterning. To be fair, double patterning is a useful lithography technique. In fact, it is an essential technique at 20nm. On the other hand, poor color resolution, mask misalignment, and pattern interference problems can easily defeat an SoC whose layout is not double patterning–friendly.

Double patterning uses two or three masks to image one layer of a chip on silicon. The exposures from multiple masks overlap to create features that are half the pitch that would otherwise be possible using these wavelengths of light. The patterns of the two masks can be thought of as printing two or three different colors (see Figure 2) that combine to form a single layer.



Figure 2: Double patterning

The semiconductor industry has developed several versions of the double patterning technique. Triple and quadruple patterning techniques are being investigated for 14nm and beyond. The success of all these techniques depends on accurate decomposition of the design layout into the multiple masks, precise mask alignment during lithography imaging, and control of variables such as dosage, focus, etch, and overlay.

Chip design teams do not need to know much about the specific lithography variables. But teams do need to know this: you cannot print just any pattern you please using double patterning. Using a method of layout decomposition, multiple masks have to be created and combined in specific ways, and some combinations will not work.

As a result, managing double patterning effects cannot be left to the physical signoff tool—as was possible at previous technology nodes—but needs a holistic approach. The entire design flow must take double patterning into account to have optimal layout for manufacturing. Specifically, the implementation tool needs more manufacturing cause-and-effect knowledge, so this tool needs to work closely with the physical design and analysis tools. Simply integrating the signoff tools is not enough. Accurate abstraction technologies must be built into placement and routing to handle early convergence of double patterning issues. Throughout the design flow, this integration for double patterning needs to be tighter than existing integration for dealing with physical design issues.

#### **Placement Challenges**

The placement engine needs to understand that certain cells cannot be next to one another because their internal patterns will interact in undesirable ways—because of either double patterning problems (see Figure 3) or LDE proximity variability. Analyzing these problems during placement maintains the designer's timing intent and minimizes the need to fix problems at signoff, thereby creating loops to fix them in implementation. (By the same token, standard cells and other IP should be designed to avoid double patterning and LDE variability problems.)



Double pattterning-aware placement



Figure 3: Double patterning and LDE variability in digital implementation

#### **Routing Challenges**

The router needs to understand double patterning issues, just as the router understands other physical design and DFM rules. However, the double patterning issues are more complex because the router must model the way a layer will be separated into two layers or "colors" and recombined lithographically. Simple design rules cannot specify the decompositions that work versus those that do not. The manufacturing and lithography effects must be abstracted intelligently into models and heuristics for fast analysis that are integrated into the place-and-route tools, and the tools must support multiple double pattering technologies and mixed optical processes.

#### Verification and Signoff Challenges

Finally, the implementation cannot work without knowledge of the decomposition done prior to tapeout. For example, the parasitic extraction tools must account for the overlay errors while calculating parasitics so as to ensure accurate static timing and SI analysis. Physical verification and DFM signoff tools also need to ensure that the final decomposition is accurate before hand-over to the foundry.

To avoid an overwhelming number of color conflicts and LDE variability issues at signoff, 20nm flows need a correct-by-construction approach that begins where the design begins: cell and block placement and routing as well as signoff-quality verification within the flow. After double patterning-aware routing and an in-design signoff double patterning and DFM verification step, all the physical and LDE variability issues are analyzed and resolved before proceeding to final signoff. This last signoff should thus be a final check only, with minimal iterations—or none at all—to fix problems.

#### **Ecosystem Challenges**

At 20nm, it is no longer possible for design tools to simply implement foundry rules. DFM requires tighter collaboration with the foundry early on, using design process co-optimization in the tools. In fact, collaboration is crucial throughout the entire design and manufacturing ecosystem, from understanding process requirements to validating results with the design community and library providers.

The 20nm design process also demands accurate technologies and flexibility in the design tools to adapt to the multiple mix-and-match use of double patterning techniques that different foundries are implementing for individual layers as well as for whole designs. This approach provides the best cost of ownership and return on investment to chip designers for their tool investment as they assess and decide on their 20nm foundry partners.

At the very least, 20nm EDA tools and design flows must support the goal of delivering a complete end-to-end Silicon Realization product line that enables unified intent, abstraction, and convergence. Achieving this goal requires accurate abstraction of complex models for DFM convergence and signoff using models that are foundry qualified.

## Design Abstraction Supports Greater Complexity and Capacity

As mentioned earlier, part of the overall solution for 20nm design is to create a new level of design abstraction dividing a highly complex design into multiple modules and dealing with the modules at a higher level of abstraction for fast prototyping. Of course, this prototyping is useful only if the abstractions correlate back to the actual details of the design. One way to ensure correlation is to use intelligent placeholder cells in each section that carry physical placement and congestion knowledge. Done correctly, the placeholders retain enough accuracy to get realistic analysis for floorplanning and partitioning.

Tools can create models for use in abstraction based on the netlist, timing constraints, placement constraints, and other variables. Table 1 compares analysis times and the resulting worst negative slack, with and without the use of abstraction.

	Full Netlist	Abstraction (2 CPUs)	
Placement (hours)	15:00	1:13	
Timing analysis (hours)	2:50	0:55	
Slack (ns)	-0.406	-0.092	

Table 1: Analysis times with and without abstraction

Intelligent abstraction also proves useful in DFM, where design tools must anticipate manufacturing problems that cannot be found using traditional rule-based checks. Figure 4 shows an example of proximity effects that DRC does not detect. A final signoff lithography check catches the proximity effects, but fixing the hotspots this late could lead to delayed design cycle time.



DRC Clean



Litho Check Passes





Still DRC Clean

Litho Check Fails

Figure 4: Proximity effects not detected by DRC (source: GLOBALFOUNDRIES)

The best approach is to catch these problems during layout. Model-based simulation is the most accurate and should be the final step in the design signoff. But, since model-based printability simulation takes far too long during implementation, abstraction offers a quick solution. In this approach, failures on silicon are abstracted into yield detractor patterns of lithography hotspots (see Figure 5). Using these patterns for 2-dimensional shape-based pattern matching, a tool can find layout features that may create proximity problems—an analysis method that can find problems up to 10,000 times faster than printability simulation.



Figure 5: Abstraction of yield detractor patterns finds problems 10,000 times faster than printability simulation (source: GLOBALFOUNDRIES)

The cost of this analysis increases is the runtime for routing, yet the analysis minimizes the risk of yield-limiting litho hotspots and reduces the electrical variability impact early in the digital implementation flow. Reducing the variability issues early in the flow also offers enormous time savings at signoff, as shown in Table 2.

Design	Size (mm2)	Model-based runtime (sec)	In-design pattern- based runtime (sec)	Speed-up
А	1.21	40,100	175	229x
В	2.16	56,900	343	166x
С	8.64	1,41,350	1,027	138x

Table 2: In-design pattern matching brings early DFM convergence and speed-up from use of abstraction

At signoff, the accurate abstraction helps to reduce the DFM check analysis turnaround time during digital implementation—by more than 100 times compared to traditional signoff methods. More importantly, tests have shown a 10x reduction in the number of hotspots after lithography analysis and correction, without any severe hotspots. The use of intelligent, accurate abstraction can therefore cut several days of DFM error iteration cycle time from the signoff stage. Model-based DFM signoff should thus be a final check only with minimal iterations—or none at all to fix problems.

## Convergence Throughout the Flow Eases Final Signoff

In the 20nm design flow, tools must optimize designs across multiple objectives, from timing and area to power, double patterning, and DFM. Traditionally, timing/area optimizations have achieved a fairly high level of accuracy but exactly what that level of accuracy is remains a mystery until signoff. Too often, the story has a surprise ending. As more and bigger process variabilities come into the picture for smaller geometries, more surprises are in store. Using conventional methods at 20nm, no one should be surprised to see hundreds of problems at signoff that require many iterations before reaching decent yield.

Since running full signoff checks are the only way to guarantee that a design meets objectives, design teams need to run these checks at multiple points in the flow. Floorplanning, synthesis, and physical design tools can still use approximations and abstractions to minimize runtimes, but then a signoff check must ensure that the design really does meet timing, power, and yield targets. Similarly, parasitics must be extracted using the signoff engine. With this accuracy, subsequent steps in the flow can build upon an accurate foundation.

Running multiple signoff checks is practical today thanks to parallelization and inexpensive hardware. Even modest multi-core server farms can run checks in a matter of hours, so long as the tools are designed to take full advantage of task parallelization. The checks do add extra time, but at a cost of about 5 percent more time at any given point, they enable 95 percent faster convergence at the end of the flow.

A signoff check at a point in the middle of the flow is meaningful if design intent is accurately captured at that point. The final component of the 20nm solution convergence ensures that accurate design intent is maintained at every point in the flow.

## Conclusion

Today, a separate CAD team handles much of the DFM signoff for a typical SoC. By the 20nm technology node, the whole design team needs access to DFM technology. This access can only happen through integration that spans the design flow, using techniques such as in-design DFM and signoff checks. Facilitating seamless DFM closure at 20nm also requires a tightly coupled feedback mechanism between the design implementation and signoff databases, where intelligent abstraction, smart computing, and comprehensive analysis will help design teams meet the prescribed design cycle times.

Design approaches for complex projects at the technology nodes preceding 20nm (45/40nm and 32/28nm) show an increased emphasis on custom design methodologies rather than digital design flows with added mixed-signal blocks. Since the majority of designs at 20nm will include mixed-signal functionality, the custom design approach provides the highest probability of first-time success. The custom design approach is engineering-intensive, however, and design implementation costs are high because of the need for a large number of highly skilled engineers—another reason to leverage computing power and intelligent analysis as much as possible.

Keeping design intent in mind and leveraging smarter abstraction methods and technologies, advanced node design teams will be able to achieve convergence and deliver timely silicon for the EDA industry. With double patterning technology leading the way to enable 20nm through 10nm lithography, the industry requires a change in thinking to leverage in-design DFM technologies that are built into the implementation environments and that enable rapid identification, analysis, and fixing of many manufacturing problems during the implementation phrase, rather than via long iterative feedback loops after the designs are completed.

CAD organizations and design engineers tend to focus on the performance of specific EDA tools, while upper management of semiconductor companies worry about market positioning and financial performance. The 20nm node holds the promise of uniting all these people through the recognition that holistic design flows offer the only practical, predictable path to working silicon.



Cadence is transforming the global electronics industry through a vision called EDA360. With an application-driven approach to design, our software, hardware, IP, and services help customers realize silicon, SoCs, and complete systems efficiently and profitably. www.cadence.com

© 2011 Cadence Design Systems, Inc. All rights reserved. Cadence and the Cadence logoare registered trademarks of Cadence Design Systems, Inc. 22020 07/11 MK/DM/PDF