WHITE PAPER

how big can you dream?™

PARASITIC EXTRACTION FOR DEEP SUBMICRON AND ULTRA-DEEP SUBMICRON DESIGNS

TABLE OF CONTENTS

Introduction 1

Design performance 1

Design methodology changes 2

  Interconnect verification 3

  Full-chip vs. critical net extraction 3

Extraction requirements 4

  Scalability 4

  Speed, capacity, and accuracy 5

  Flexibility 5

Extraction methods 6

  Field Solvers 6

  Geometric extractors 6

  Context-based extraction 7

Fire & Ice technology 8

Summary 8

TABLE OF FIGURES

Figure 1 Total delay is equal to device delay plus interconnect delay and slew rate 1

Figure 2 Interconnect verification provides tapeout confidence 2

Figure 3 The need for interconnect verification depends on clock speed and process size 3

Figure 4 All capacitance is extracted with gross accuracy, even for critically coupled nets 4

Figure 5 A tighter distribution of accuracy is superior to a wide distribution 5

Figure 6 Boundary errors caused by pattern matching technique 6

Figure 7 Calculating a 3-D dynamic "halo" surrounding the central conductor 7

Figure 8 Comparison of extraction methodologies for DSM and UDSM designs 8

INTRODUCTION

Shrinking process technologies and increasing design sizes continually challenge design methodologies and EDA tools to develop at an ever-increasing rate. Before the complexities of deep submicron (DSM), gate and transistor delays dominated interconnect delays, and enabled simplified design methodologies that could focus on device analysis. The advent of DSM processes is changing all of this, invalidating assumptions and approximations that existing design methodologies are based upon, and forcing design teams to re-tool. High-capacity parasitic extraction tools are now critical for successful design tape-outs.

As process technologies shrink, five different major effects start to happen:

- Device (gate) delays decrease, due to the thinning gate oxide
- Interconnect resistance increases, because of shrinking wire widths
- Vertical heights of interconnect layers increase, in an attempt to offset increasing interconnect resistance
- The area component of interconnect capacitance no longer dominates. Lateral (sidewall) and fringing components of capacitance start to dominate the total capacitance of the interconnect
- Interconnect capacitance dominates total gate loading

These changes directly affect the performance of a design and indirectly force a change in design methodology.

DESIGN PERFORMANCE

The total delay associated with a net or path is governed by a simple equation that includes device delays, device loads, and slew rates. The delays caused by device loading are known as interconnect delays. The equation could be represented as:

Total delay = device delay + interconnect delay + slew rate *(see Figure 1)*

When process geometries were greater than one micron, the performance of a design could be accurately predicted by analyzing device delays and approximating (or in some cases, ignoring) the interconnect delays and slew rates. Device loads were typically treated as a lumped capacitance, an approximation enabled by the fact that device delays dominated the equation. Slew rates were also typically ignored for the same reason.

As processes shrank below one micron, these approximations became more and more inaccurate. With the total delay decreasing, slew could no longer be ignored, since it affected a more significant percentage of the total delay. For the same reason, device loads could no longer be accurately represented by a simple capacitance, so the lumped RC model was introduced. Despite these relatively minor changes, device delays still dominated the total delay equation.

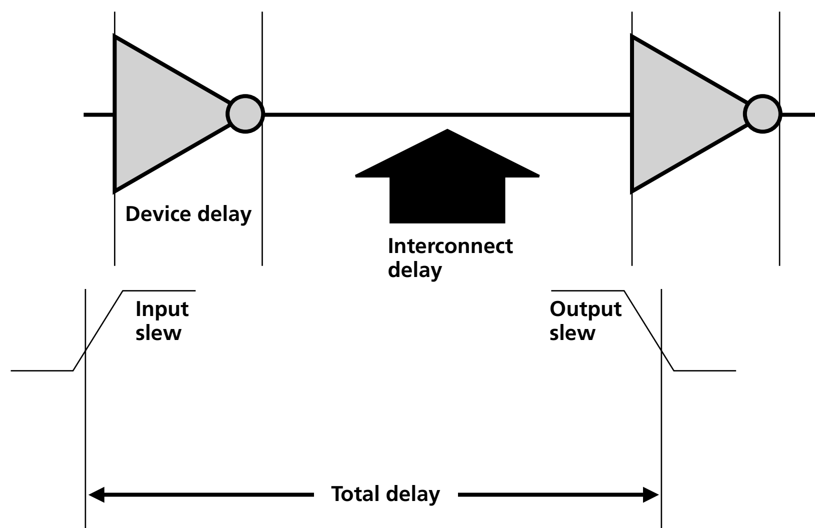

Figure 1: Total delay is equal to device delay plus interconnect delay and slew rate

An interesting phenomenon occurred as the process geometries shrank below 0.5 microns. Somewhere in the 0.5-0.35 micron process size, the interconnect delay caused by device loading became equal to the device delay. This event, while not perceived to be a dramatic change, emphasized the effects of DSM by changing the basic paradigm of design: gate delay no longer dominated interconnect delay. Original approximations based on this paradigm failed, and interconnect delays could no longer be treated as second order effects. Since interconnect delays began to play a major role in determining total delay, the distributed RC model was introduced to improve the accuracy of interconnect modeling.

Shrinking the process technology past DSM and down to ultra-deep submicron (less than 0.25 micron) will continue the delay trends and introduce some additional concerns. Total delay will continue to decrease, interconnect delay will continue to increase (eventually dominating device delay), and slew will continue to play an important role in determining total delay. Also, increased coupling capacitance between adjacent interconnect wires will increase delay times and cause hard failures due to noise injection. Voltage drop and ground bounce in poorly designed power rails will also impact delays due to weakened driver strengths. For high performance designs, inductance will start to play a role in interconnect delays, and the delay model will transition from a distributed RC model to a distributed RLC model.

DESIGN METHODOLOGY CHANGES

Cell-based and custom design methodologies were developed under the paradigm that device delay dominated the total delay equation. Accurate performance prediction was possible using analysis of the logic design and device delays only. Wire load models were sufficient to provide accurate loading information to synthesis, floorplanning, place & route and timing analysis design tools. Under this paradigm, DRC and LVS checks became the *de-facto* verification standard prior to tapeout under the assumption that the performance was guaranteed by an analysis of the logic and device delays.

With the advent of DSM and ultra-deep submicron (UDSM) process technologies, interconnect delays dominate device delays and therefore the basic paradigm is broken. Existing design methodologies must be re-evaluated based on this paradigm change, or design engineers will experience continuous tapeout failures due to unpredicted loading associated with critical nets and paths.

Both DSM and UDSM design methodologies must compensate for the physical effects of fabricating a design on silicon. It is no longer possible to predict a design's performance based on the logic design and device delays only. Instead, performance will be directly dependent on the physical implementation, cell placement, and signal routing that determines the interconnect delay. As performance prediction becomes an impossible task, post-layout interconnect verification becomes a necessary part of the design flow in providing tapeout confidence.

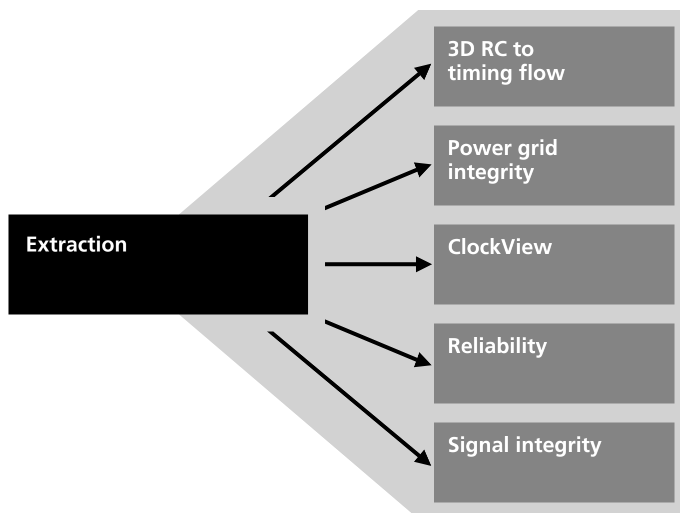

Figure 2: Interconnect verification provides tapeout confidence

INTERCONNECT VERIFICATION

With interconnect delay continuing to increase in importance, more effort must be spent in verifying interconnect, which involves the initial extraction of the interconnect parasitics followed by different types of interconnect analysis. This can be expressed in a simple equation:

Interconnect verification = extraction analysis

The amount of interconnect verification performed on a design is dependent on the target process technology, frequency, and the design itself. Designs targeted to a non-DSM process technology will not experience the same severity of interconnect-related problems as a design targeted to a UDSM technology. Similarly, high-performance designs such as today's microprocessors and microcontrollers will typically require significantly more interconnect verification than more conservative designs running at much slower clock frequencies. *Figure 3* provides guidelines on when interconnect verification may be required.

Interconnect verification is both frequency- and process-dependent; however, as process technologies continue to shrink, interconnect verification becomes more of a necessity.

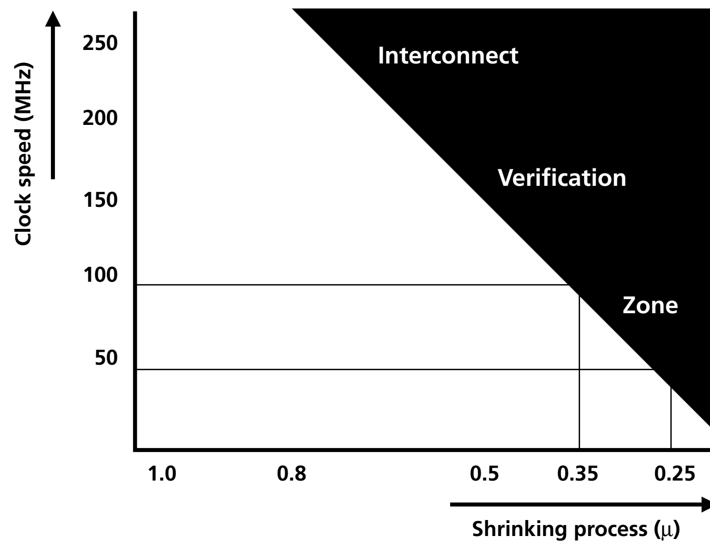

Figure 3: The need for interconnect verification depends on clock speed and process size

What is the risk of not performing interconnect verification? For a conservative design methodology, over-design could prevent most serious performance problems, but even over-design does not guarantee problem-free tape-outs. For example, one design team took a conservative design approach to avoid having to manage the complexities of additional clock verification. To ensure good clock performance, they over-sized the clock buffers. However, the increased buffer sizes caused voltage (IR) drops in the power grid that actually caused the clock performance to be worse.

For an aggressive design methodology, the chances of taping out a design that performs to its intended specification without performing interconnect verification are severely diminished. An example of this was a design team where, using a more aggressive design methodology, it eventually took 100 engineers hand modeling the layout to find a coupled capacitance that slowed down a critical net.

FULL-CHIP VS. CRITICAL NET EXTRACTION

Two design methodology approaches are used for parasitic extraction. The first assumes the full-chip netlist can be divided into two groups, critical nets and non-critical nets. Once divided, each group is treated differently, such that the critical nets are extracted with high accuracy and the non-critical nets are extracted with low accuracy. The reasoning behind this approach is that not all nets require accurate extraction, and so there is a runtime vs. accuracy tradeoff made for the non-critical nets.

The alternative approach assumes that the capacitance extracted from any net could be associated with a critical net, and therefore, all capacitance should be extracted with high accuracy. This is the most predictable approach for DSM and UDSM technology. The lateral and fringe components of capacitance clearly dominate below 0.25 micron, and coupling capacitance is dominated by these components. The ratio of coupled to non-coupled capacitance continues to increase as the process technology shrinks.

The danger with the first approach is that there is no guarantee that all critical nets are identified. The filtering mechanism used to divide the full-chip netlist must account for all factors that could determine criticality, including driver size, distributed RC load, coupling between nets, and signal timing. While it may be possible to identify critical nets based on the first two criteria, it is extremely difficult, if not impossible, to pre-determine criticality based on the last two.

Consider the following example: Three nets, pre-determined to be non-critical (based on driver size/loading criteria), are tightly coupled to each other because their routes are adjacent. The central net has a large driver and large capacitive load, which causes the adjacent nets to become critical due to high coupling capacitance *(see Figure 4)*. Since all of these nets were pre-determined to be non-critical, all of the associated capacitances are extracted with gross accuracy, despite the fact that high coupling capacitances A and B have caused two of the nets to become critical.

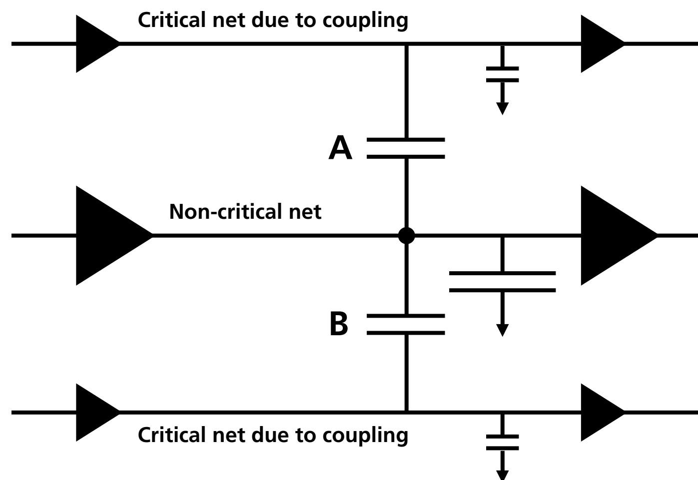

Figure 4: All capacitance is extracted with gross accuracy, even for critically coupled nets

EXTRACTION REQUIREMENTS

No design team can afford to wait multiple days to get an accurate set of parasitics for their design. Designers must be able to perform extraction and analysis on their entire chip within a reasonable amount of time and with acceptable accuracy. In addition, any solution must be able to maintain its speed, capacity and accuracy as the processes continue to shrink and as design size continues to increase. This is known as scalability.

SCALABILITY

Since a reasonable amount of time for one design team might be too long for another, extraction methodologies must be flexible to adapt to each set of design methodology requirements. While an extraction product that runs only on a single CPU sounds immediately attractive, the ramifications of being limited to that single CPU must be considered for future designs, where according to Moore's Law the design sizes will be doubled every one to two years. Today, an extraction methodology must be scalable with design size, where scalability implies that designers can apply additional CPU resources to the problem to accommodate increasing design sizes.

For example, a design team designing a three million transistor design today knows that the next design will be over seven million transistors. The design methodology requires that full-chip extraction be completed overnight. Their 3M-transistor design requires executing extraction distributed on five CPUs. For the 7M-transistor design they will need to execute on 10 CPUs if they want to continue to extract overnight. The architecture of the extraction tool must be able to accommodate this easily and without loss of accuracy.

While scalability with design size is one aspect of scalability to consider, another, maybe even more important aspect to consider is scalability with shrinking process technology. Process technologies continuously evolve, and an extraction solution must be able to maintain its accuracy as the process shrinks. If the extraction solution does not scale with process technology, the accuracy of the extracted results becomes unacceptable over a short period of time. Given this scenario, it becomes very difficult to justify the high price tag associated with extraction products that work for one or two years before becoming outdated.

SPEED, CAPACITY, AND ACCURACY

Three parameters that are normally measured to evaluate the performance of an extraction product are speed, capacity, and accuracy. Measuring these parameters can provide a basis to judge relative performance between competitive products. Including scalability and flexibility in the equation also provides a more comprehensive set of measurable parameters against which to judge.

Speed

Speed is the measure of turnaround time. The measurement of speed must include everything required to complete the extraction job, including loading the input data, processing time, and writing the output files. Simply measuring CPU time may not provide the complete story. Disk IO time on large designs is potentially a large factor in turnaround time. Measuring speed on extraction products that execute in a distributed manner across many CPUs must be done very carefully. Distributed processing uses a network for communication, and the transfer time through a slow network may dominate the actual extraction time, leading to misleading results. In addition, considering speed without considering accuracy may not solve your problem.

Capacity

Capacity is the measure of an extraction product's ability to handle large designs. When evaluating capacity, future design sizes (Moore's Law) and process technologies must always be considered since the extraction product must be scalable with both design size and process technology. Capacity will vary with different design styles.

Accuracy

Accurate extraction is essential when designing in DSM or UDSM process technologies. Trading off accuracy for speed typically does not solve your problem. Accuracy is not a single number or percentage, but a distribution over a range. A solution that produces accuracy within a tightly over a range. A solution that produces accuracy within a tightly controlled range with known bounds is to a solution that produces a smaller average, but has large error range *(see Figure 5)*.

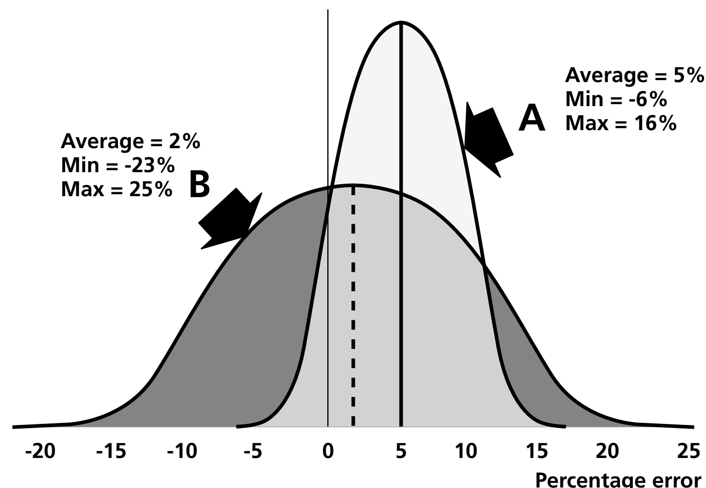

Figure 5: A tighter distribution of accuracy is superior to a wide distribution

FLEXIBILITY

Is your single extraction solution flexible enough to support all your analysis requirements? When you consider a new extraction product, one of the first questions to answer is what type of analysis is required. Is extraction to be used only in a timing verification flow? Is the extraction data used by more advanced analysis, such as power grid integrity, signal integrity, clock skew or electromigration risk?

If the extraction product is to drive more than a single analysis flow, then this should be added as a measurable to determine the most suitable extraction product for your needs. For example, if the extraction product works only at the gate level, it will not meet the transistor-level requirements of electromigration analysis.

While certain analysis flows require unreduced parasitic data, other flows can accept reduced data without any significant loss of accuracy. This should be taken into account when evaluating extraction products, since an extraction product that depends on reducing the data to obtain capacity will not be suitable to drive more advanced analysis, such as signal integrity or electromigration, without a significant loss of accuracy.

EXTRACTION METHODS

The goal of any extraction product is to provide the best combination of speed, capacity and accuracy together in one package. Different extraction methods may trade off one of these parameters for another. Resistance extraction is a much easier task than capacitance extraction since resistance can be extracted using 2-D modeling techniques. Capacitance is a 3-D effect, since objects surrounding a wire will influence its capacitance, and as such presents a much more difficult extraction problem. The remainder of this white paper will focus on the capacitance extraction problem.

Different types of extraction products on the market today serve to highlight speed/capacity/ accuracy tradeoffs:

FIELD SOLVERS

Field solvers discretize the layout structure and solve a Poisson equation for given boundary conditions using a variety of methods (finite element, finite difference, boundary element, boundary element with multiple acceleration, Monte Carlo, and mean equation of invariance). Results from each method will differ.

While they are indisputably the most accurate extraction tools, field solvers are slow. They look at the actual physical phenomena of electrical lines of force for given charges as well as the voltage differences on a set of conductors. Due to the millions of data points they have to process, field solvers do not scale easily to full-chip extraction, and runtimes increase sharply as higher accuracy and capacity are required.

The input to field solvers must be painstakingly prepared and re-prepared for any new process, which can take weeks of design time and require an expert in field solvers. They are compute-intensive and require large amounts of disk resources. As such, field solvers can be used only for small testcases, small cells, or single nets of a design. Usually field solvers are applied on critical nets such as a particular clock, or on a critical cell such as an SRAM cell. Critical net extraction will give you lumped RC values only (unless run in an even more time consuming matrix mode), limiting its usefulness for interconnect verification to timing analysis only.

GEOMETRIC EXTRACTORS

The most common types of extractors are geometric extractors. Initially developed for 1.0 micron process technologies, these extractors have limitations when applied to DSM and UDSM designs. The different techniques for identifying parasitics are outlined below:

Rules-based extractors

Rules-based extractors give you lumped RC values which are good for estimating timing and iterative place and route, but lumped RC values can’t be used for DSM and UDSM designs since they do not contain the critical coupling capacitance components. In general, these extractors use a rules extraction file to identify the type of interconnect (e.g. lateral, fringing, crossover, overhang) by its geometric features and then apply a look-up table, preset formula or oversimplified equation to calculate the capacitance of the interconnect. Rules-based methods are typically not comprehensive and they tend to be unpredictable from design to design.

Boolean extractors

Boolean extractors analyze transistors and interconnect by applying Boolean operations (A = B AND C) to the polygon shapes and form a Boolean pattern. They use a pattern matching technique to identify sub-patterns of previously characterized capacitance values, and calculate the capacitance of the layout by summing all of these patterns. All pattern-matching techniques are inaccurate for DSM and UDSM since they identify regions simplistically and ignore the boundary effects between adjacent patterns *(see Figure 6)*.

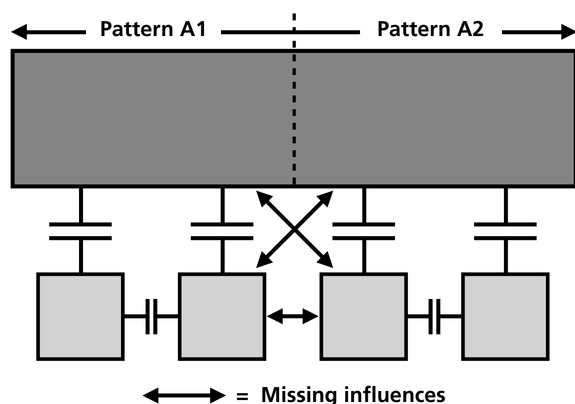

Figure 6: Boundary errors caused by pattern matching technique

Edge-based extractors

Edge-based extractors identify the edges of transistors and interconnect by tracing the outer portions of the polygon shapes, and then for each edge, apply some formula that matches with a previously characterized capacitance value, using a similar test structure. This is another variation of pattern matching, and suffers from the same boundary error problems as Boolean extractors. While edge-based methods can be "tuned" to produce accurate results for a single design/process combination, using the same set of extraction rules on a different design on the same process does not produce the same results.

Featured-based extractors

Featured-based extractors match pre-calibrated library structures to a design's geometry and then interpolate capacitance values from a table. These extractors become extremely slow when they detect a pattern or a particular structure (e.g. crossover) that has not been previously characterized, because they resort to field solving for the capacitance values. Since this extraction methodology is once again based on pattern matching, boundary conditions between adjacent patterns are not accounted for, which leads to inaccuracy. Since it is rare to find a design where all of the patterns match, the slowness caused by utilizing a field solver directly in the extraction process means that these tools are usually unable to perform full-chip extraction.

Context-based extractors

Context-based extractors utilize a suite of pre-characterized analytical (parameterized) models. During the extraction process, parameters are continuously generated (based on very specific 3-D regions or "contexts") and passed to the analytical models for capacitance calculation. Unlike the pattern matching techniques used in other methods, analytical models are parameterized such that all influencing objects are accurately modeled. Therefore context-based extraction does not suffer from boundary errors. Context-based extractors are scalable as design size increases and process geometries shrink.

CONTEXT-BASED EXTRACTION

A new breed of extraction products has recently been developed to specifically address extraction problems associated with shrinking process technologies and increasing design sizes. Context-based extraction is a new method, based on newer paradigms that include DSM and UDSM requirements. Context-based extraction starts with the creation of 3-D analytical models, then proceeds to full-chip extraction using these parameterized models.

Prior to the actual extraction process, 3-D analytical models are automatically generated from process technology information, including conductor and dielectric information. This generation process utilizes specialized Laplace solvers that have been tuned for accuracy and efficiency.

During the extraction process, each conductor of the design is traversed and geometrically analyzed in all three dimensions. This process is known as "contexting" and is achieved with the use of a special influence region, known as a "dynamic halo." The dynamic halo is centered on the conductor of interest and extends out in three dimensions to capture surrounding influences *(see Figure 7)*.

The halo accounts for all near-body and multi-level interconnect capacitive effects including the impact of crossover fringe, corners and capacitive shading. Objects within the halo determine what parameter values need to be applied to the 3-D models, and then the 3-D models are evaluated to provide the capacitance. Lumped and coupled capacitances are maintained separately, enabling advanced signal integrity analysis.

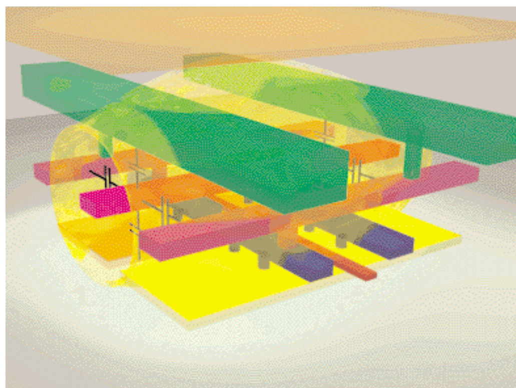

Figure 7: Calculating a 3-D dynamic "halo" surrounding the central conductor

In contrast to geometric methods, context-based extraction captures non-isoplanar effects (i.e. effects between conductors on different layers) because of its ability to identify in three dimensions all the interacting objects within the context of the conductor being evaluated. A context-based extractor works with one conductor in a design at a time and is capable of looking ahead and behind for the influence of any other conductors within the halo region. Because context-based extraction does not rely on pattern matching techniques, it does not suffer from the boundary errors associated with pattern matching.

The advantages of context-based extraction are listed below.

- It is the only 3-D extraction methodology that has the capacity for full-chip extraction
- It is the only 3-D extraction methodology that provides the accuracy required for DSM and UDSM designs
- Full 3-D accuracy is maintained in full-chip extraction
- 3-D analytical models are created once per process. There is no need to tune each design
- All capacitance components required for DSM and UDSM designs are accounted for
- Context-based extraction can produce RC data suitable for both timing and advanced analysis flows
- Coupling capacitance components are maintained separately

	Speed	Capacity	Accuracy	Scalability	Flexibility
Field solvers	x	x	✓	x	x
Geometric solvers	✓	✓	x	x	x
Context-based extraction	✓	✓	✓	✓	✓

Figure 8: Comparison of extraction methodologies for DSM and UDSM designs

FIRE & ICE TECHNOLOGY

Fire & Ice, extraction product suite, was created to provide the best combination of speed, capacity, accuracy, scalability and flexibility. As the first context-based extraction product on the market, Fire & Ice has proven that this methodology meets the requirements for DSM and UDSM designs.

- Speed and scalability are built into the product via efficient distributed processing wherever possible
- Capacity is achieved by initially “striping” a design into smaller partitions for distributed processing. This striping technique, enables accurate full-chip extraction for the largest of designs
- Accuracy is built into Fire & Ice. From the 3-D analytical models to the context-based extraction, the full-chip extraction produced by Fire & Ice results in accuracy within 10% of both reference field solvers and silicon
- Scalability is also obtained with the 3-D analytical models, which maintain their accuracy beyond UDSM
- Flexibility is defined by Fire & Ice. Architected to drive analysis products, Thunder & Lightning and ClockView, Fire & Ice has the flexibility to perform efficient extraction for timing verification flows as well as extraction for power grid, clock, signal integrity and electromigration risk analysis

SUMMARY

Parasitic extraction for DSM and UDSM has a much stricter set of requirements than ever before.

- Shrinking process technologies are forcing a retooling for parasitic extraction. Extraction tools must be scalable to 0.25 micron and below
- Unless a complete and accurate extraction is performed, design performance will suffer
- The capacity and speed to do full-chip extraction is required because it is no longer possible to identify all of the critical nets before extraction
- One extraction solution is needed for both timing and advanced analysis
- Extraction requirements should include speed, capacity, accuracy, scalability and flexibility
- A new extraction methodology, called context-based extraction, is setting the standard for DSM and UDSM extraction

Cadence Design Systems, Inc.

Corporate Headquarters
555 River Oaks Parkway
San Jose, CA 95134
800.746.6223
408.943.1234
www.cadence.com

© 2002 Cadence Design Systems, Inc. All rights reserved. In the U.S. and numerous other countries, Cadence and the Cadence logo are registered trademarks and “how big can you dream?” is a trademark of Cadence Design Systems, Inc. All others are properties of their holders.

Stock #3993 06/02