Challenges and Opportunities for Design Innovations in Nanometer Technologies

Jason Cong

Computer Science Department

University of California, Los Angeles, CA 90095

(E.mail: cong@cs.ucla.edu, Tel.: 310-206-2775)

1   Introduction

The driving force behind the spectacular advancement of the integrated circuit technology in the past thirty years has been the *exponential scaling* of the feature size, i.e., the minimum dimension of a transistor. It has been following the Moore's Law [1] at the rate of a factor of 0.7 reduction every three years.   It is expected that such exponential scaling will continue for at least another 10-12 years as projected in the recently published 1997 National Technology Roadmap for Semiconductors (NTRS'97) [2] shown in Table 1. This will lead to over half a billion transistors integrated on a single chip with an operating frequency of 2-3 GHz in the 70nm technology by Year 2009. The challenges to sustain such an exponential growth to achieve gigascale integration have shifted in a large degree, however, from the process and manufacturing technologies to the design technology. A great deal of design innovation, in terms of both significant extension of the existing design capability and the development of new design paradigm and methodology, is needed to achieve the projected targets in the NTRS'97.   This paper discusses the challenges and opportunities for design innovations in the future IC designs, especially in the areas of interconnect design and high-degree of on-chip integration. Section 2 presents a number of design challenges in these areas, with *quantitative measurements* derived based on the technology projection in NTRS'97.   Section 3 discusses opportunities and possible directions for design technology innovation to meet various design challenges in the road ahead.

Technology (nm)	250	180	150	130	100	70
Year	1997	1999	2001	2003	2006	2009
# transistors	11M	21M	40M	76M	200M	520M
Across chip clock (MHz)	750	1200	1400	1600	2000	2500
Area (mm^2)	300	340	385	430	520	620
Wiring Levels	6	6-7	7	7	7-8	8-9

Table 1. Overall technology roadmap from NTRS'97 [2].

2   Challenges in NTRS Implementation

The rapid scaling of IC technology has two profound impacts. First, it leads to much smaller and faster devices, but more resistive interconnects with larger coupling capacitance. This results in *interconnect limited designs*. Second, it enables much *higher degree of on-chip integration*, which leads to significant increase in the design complexity. This section discusses the implications and

challenges in these two areas based on the projections from NTRS'97. Possible solutions and innovations in the design technology to address these challenges will be discussed in Section 3.

2.1 Interconnect-Limited Designs

With the rapid feature size scaling, the circuit performance is increasingly determined by the interconnects instead of devices. Rapid scaling has introduced many challenges on interconnect performance, modeling, and reliability. It also drives the development for new interconnect materials. In order to better understand the significance of interconnects in future technology generations, we collected basic interconnect parameters provided in NTRS'97 (shown in boldface in Table 2), and set up a proper 3D interconnect model to extract various components of interconnect capacitance (shown in the remaining rows in Table 2) using the 3D field solver FastCap [3]. We also collected the basic device parameters provided in NTRS'97 (shown in boldface in Table 3) and derived the driver/buffer input capacitance, effective resistance, and intrinsic delay in each technology generation (shown in the remaining rows in Table 3) using HSPICE simulation. These data are used for quantitative analysis of device and interconnect performance in each technology generation listed in NTRS'97. The following subsections discuss the challenges on interconnect performance, modeling, and reliability as implied by NTRS'97.

Technology (nm)		250	180	150	130	100	70
Metal resistivity ρ (uΩ-cm)		3.3	2.2	2.2	2.2	2.2	1.8
Dielectric constant		3.55	2.75	2.25	1.75	1.75	1.5
Min. wire width (nm)		250	180	150	130	100	70
Min. wire spacing (nm)		340	240	210	170	140	100
Metal aspect ratio		1.8:1	1.8:1	2.0:1	2.1:1	2.4:1	2.7:1
Via aspect ratio		2.2:1	2.2:1	2.4:1	2.5:1	2.7:1	2.9:1
2X min. width & spacing	Ca (aF/um)	29.0	21.2	16.2	12.0	14.4	8.56
	Cf (aF/um)	41.8	30.2	24.8	18.3	14.1	14.8
	Cx(aF/um)	71.0	58.3	49.4	42.8	45.3	41.6
5X min. width & spacing	Ca (aF/um)	73.5	53.6	40.6	30.0	26.6	19.5
	Cf (aF/um)	63.5	47.3	38.4	28.5	28.2	23.6
	Cx(aF/um)	18.3	16.9	15.4	14.8	16.5	16.7

Table 2. Interconnect parameters used in this paper. The basic parameters in the first six rows (in boldface) are taken from NTRS'97. The breakdown capacitance values in remaining rows under different width and spacing assumptions are obtained using a 3D-field solver (FastCap). Ca, Cf, and Cx are the unit-length area, fringing, and same-layer line-to-line coupling capacitance for the given width and spacing under the assumption that the wires are located between two ground planes.

Technology (nm)	250	180	150	130	100	70
Vdd (V)	2.15	1.65	1.35	1.35	1.05	0.75
Ion[NMOS/PMOS] (uA/um)	600/280	600/280	600/280	600/280	600/280	600/280
Buffer input cap. (fF)	0.17	0.12	0.11	0.085	0.070	0.042
Buffer Rd (x10KΩ)	1.71	1.86	2.26	2.25	2.39	2.42
Buffer intrin. delay (ps)	70.5	51.1	48.7	45.8	39.2	21.9

Table 3.   Device parameters used in this paper. The values of voltage and transistor on-current are taken from NTRS'97. The remaining values are obtained using HSPICE simulation. A buffer is a pair of cascaded inverters with the size of the second one being five times that of the first one.

2.1.1 Interconnect Performance Limitation:

As we move from the 250nm technology with the clock frequency of 700-800 MHz to the 70nm technology with the clock frequency of 2-3 GHz in the next 10 years, the interconnect delay will far exceed the device delay and become the dominating factor in determining the system performance. The majority of the clock period will no longer be spent on computing or generating the data but on transmitting and communicating the data between various parts of the chip. The simulation results in Figure 1 and Table 4 show that although the intrinsic device delay of a minimum size transistor will decrease from 70ps in the 250nm generation to about 20ps in the 70nm technology, the delay of an average interconnect (1mm metal line) will decrease only from about 60ps to 40ps, while the delays of a 2cm un-optimized global interconnect will actually increase from about 2ns to 3.5ns. Recent advances on interconnect optimization techniques, such as interconnect topology optimization, optimal buffer insertion and sizing, optimal wire-sizing, etc, can help to reduce interconnect delays significantly (see [4] for details). But they are not able to reverse the trend of growing gap between device and interconnect performance. After we apply the simultaneous driver sizing, buffer insertion, buffer sizing, and wire-sizing, the 2cm global interconnect delay can be reduced by a factor of 2x to 5x across different technology generations, but it still cannot meet the required performance to support the across-chip clock rates projected in NTRS'97 as shown in Figure 1 and Table 4. The gap between the optimized performance and required performance grows to over a factor of 2 in the 70nm generation. This study indicates clearly that if we simply use the process geometry provided in NTRS'97 and follow the existing design style, we are not able to achieve the projected performance targets in NTRS'97 due to the interconnect limitation! Therefore, innovations are needed at every level of the design process, from system architecture design to careful process geometry optimization, to eliminate the interconnect performance bottleneck.
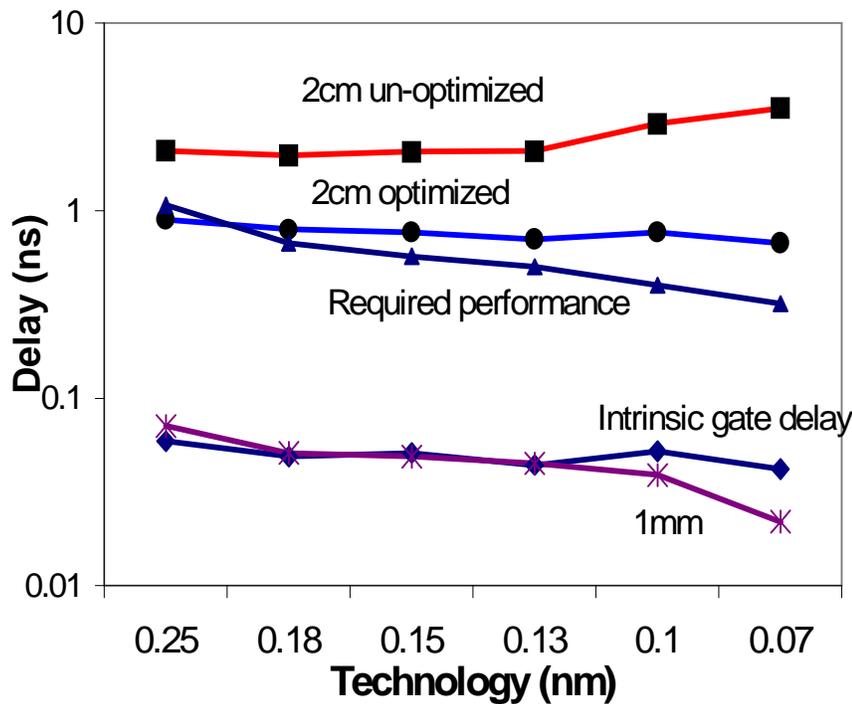
Figure 1. Trends of intrinsic gate delays and delay values (in log scale) of an average interconnect (1mm), a 2cm un-optimized global interconnect, a 2cm optimized global interconnect, and the required interconnect performance to support the across-chip clock rates projected in NTRS'97 in different technology generations. Detailed delay values are shown in Table 4.

Technology (nm)	250	180	150	130	100	70
Device intrinsic delay(ps)	70.5	51.1	48.7	45.8	39.2	21.9
1mm (ns)	0.059	0.049	0.051	0.044	0.052	0.042
2cm un-optimized (ns)	2.08	1.97	2.06	2.07	2.89	3.52
2cm optimized (ns)	0.89	0.79	0.77	0.70	0.77	0.67
Required performance for global interconnects (ns)	1.07	0.67	0.57	0.50	0.40	0.32

Table 4. Intrinsic gate delays[1] and delay values for an average interconnect (1mm), a global interconnect (2cm) [2] (both with 2x minimum width and 2x minimum spacing), and a 2cm optimized global interconnect after simultaneous driver sizing, buffer insertion, buffer sizing, and wire sizing, using the TRIO package developed at UCLA [5] in different technology generations[3]. The last row shows the required performance for global interconnects to support across-chip clock rates projected in NTRS'97 (computed as 80% of the required clock periods).

---

[1] Same as intrinsic buffer delays shown in Table 3.

[2] For both the 1mm average interconnect and 2cm un-optimized global interconnect, the drivers are optimally sized to match the interconnect loads.

[3] The capacitance values used in the optimization are based on the set f 5x minimum width and spacing as shown in Table 2. Other details of the assumption and results of the optimization procedure are shown in Table 6. Optimization results may vary under different assumptions.

2.1.2 Interconnect Modeling Complexity:

Not only has the interconnect become more important, it has also become much more difficult to model, analyze, and predict, as we move into nanometer designs with high clock frequencies. Due to aggressive scaling of interconnects, even an average-length metal line may have significant resistance comparable to its driver resistance. Thus, the *distributive nature* of the interconnect has to be modeled. Moreover, in order to limit the increase of interconnect resistance, the wire aspect ratio (height over width) will increase considerably, from its current value of 1.8:1 in the 250nm technology to 2.7:1 in the 70nm technology. The increase of wire aspect ratio together with the decrease of line-to-line spacing results in rapid increase of *coupling capacitance*. Figure 2 shows that the coupling capacitance contributes to over 70% of total capacitance at the minimum spacing and over 50% at 2x the minimum spacing in *all* technology generations[4]! The value of such coupling capacitance, on the other hand, is very difficult to be computed accurately, as it depends on both the *spatial locations* of all neighboring wires in the 3D structure and the *temporal relations* between the signals on these wires. The significance of wire resistance and coupling capacitance makes it nearly impossible to have any accurate RC delay estimation during logic or high level synthesis in absence of the layout information. *This invalids almost all timing models currently used in the high level designs.*

Furthermore, as the IC operating frequency approaches multi-gigahertz in nanometer designs, the signal rise-time will be comparable to the time-of-flight delay of a global interconnect across the chip, and the current change rate (dI/dt) will also increase significantly. As a result, *interconnect inductance*, especially that from global interconnects, such as power and ground nets and the clock nets, also needs to be properly modeled, and global interconnects need to be considered as lossy transmission lines. Efficient extraction of on-chip interconnect inductance has been difficult, however, due to the computation of return current.
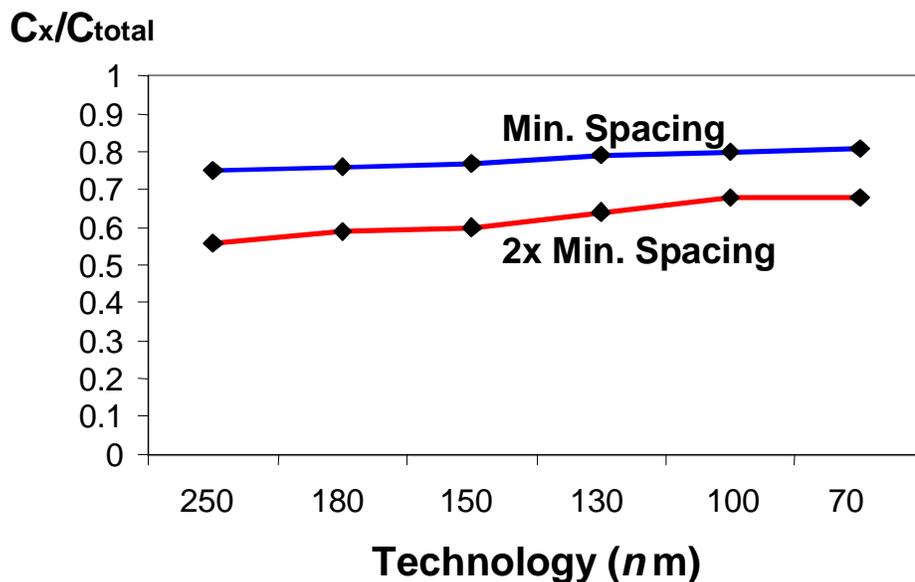

Figure 2. The percentages of coupling capacitance over the total capacitance under the minimum and 2x the minimum spacings in each technology generation.

---

[4] Comparing to those in NTRS'94, the wire aspect ratios in NTRS'97 are substantially smaller and wire spacings are larger. This results in much slower increase of coupling capacitance as shown in Figure 2.

In short, efficient and accurate extraction of interconnect parasitics in a complex 3D structure will be very important but difficult. They are needed by design tools to build adequate interconnect models at each level of the design process.

2.1.3 Interconnect Reliability:

As the IC technology further scales, the interconnect reliability becomes another challenge to the design technology. The interconnect reliability includes both the signal reliability and manufacturing reliability. The *signal reliability* requires that the signal carried by the interconnect always stabilizes at its intended value within its specified delay bounds. The *manufacturing reliability* requires that the interconnect structures meet the design rules and the connectivity specification throughout the manufacturing process and the life span of the ICs.
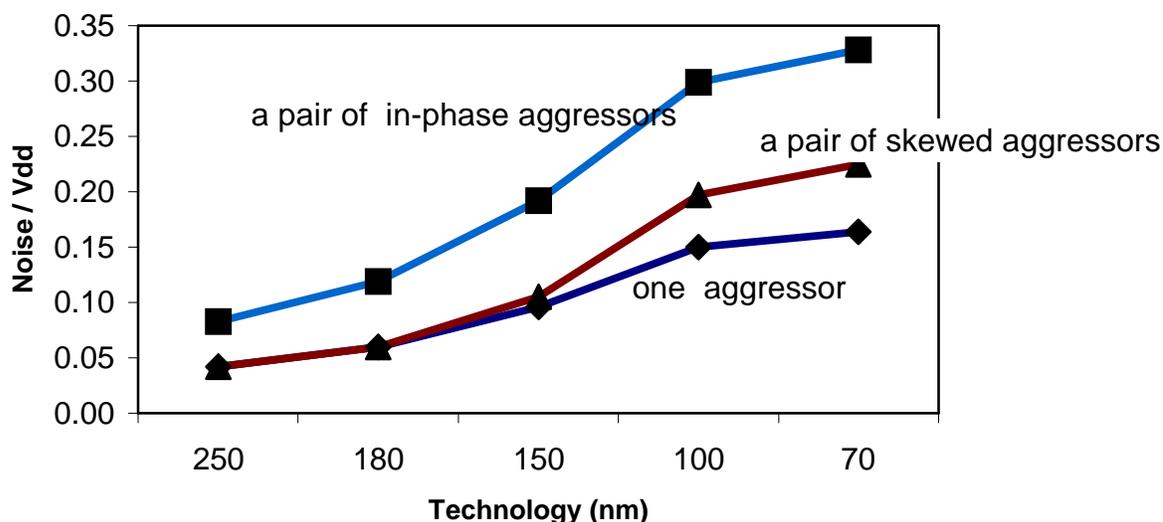

Figure 3. The ratios of peak capacitive crosstalk noise to Vdd for a 1 mm line with 2x the minimum width and spacing to its neighbors under three different temporal relations: only one neighbor is switching, both neighbors are switching simultaneously (in-phase), both neighbors are switching but one after the other. The rise time of the switching signal is 10% of projected clock period in NTRS'97.

The signal reliability is affected mainly by noise and process variation. One major source of noise is *crosstalk*, especially the capacitive coupling crosstalk, which becomes very significant due to the rapid increase of coupling capacitance with the technology scaling. Figure 3 shows the peak values of capacitive crosstalk noise in each technology generation for a 1 mm line with 2x the minimum width and spacing to its neighbors. It reaches over 30% of Vdd in the 70nm generation even for an average length wire. The value of crosstalk noise depends on not only the coupling capacitance of adjacent wires, but also a number of other factors, such as the driver and receiver sizes, the patterns and relative timing of the signals on neighboring wires, etc. For example, under different switching patterns of neighboring wires, the noise value may differ by a factor of 2x to 3x as shown in Figure 3. Another source of noise is the *power and ground bounce*, caused by simultaneous switching of a large number of devices on the chip. As a result, the voltage levels in power supply or ground planes may fluctuate considerably. The noise in the power and ground planes also depend on the

temporal correlation of the signals on the chip as well as their distribution. Both types of noise are are extremely difficult to predict and calculate, especially in the early phase of design process, as they depend on the detailed layout and timing information.

The process variation in deep submicron designs also adds a large degree of uncertainty to the signal delay and skew values in various portions of the chip. Experts indicate that the across the chip length variation (ACLV) of wire widths can be as large as 20% in the 0.1um technology and below[5]. In this case, layout parameters have to be treated as *statistical variables* instead assuming fixed values. Models and tools are needed to handle and optimize a large number of statistical variables to assure the signal reliability.

The interconnect manufacturing reliability of interconnects is affected by both the *defect density* and *electromigration*. The defect density is largely determined by the manufacturing technology. Electromigration, which forms open or short to neighboring lines due to the transport of the metal atoms when an electric current flows through the wire, may limit the lengths and widths of interconnects in order to control the current density. It is predicted that the electromigration current density limitation of Cu will become an issue at the 70nm technology [2]. The electromigration constraint and other constraints related to design for manufacturing needs to be properly considered by future design tools.

2.1.4 Impact of New Interconnect Materials:

We expect to see much progress in the IC manufacturing technology in using new interconnect materials to improve the interconnect performance in the near future. For example, copper will be used to replace aluminum to reduce interconnect resistance, and new dielectric materials with lower dielectric constants, such as polyimide (permitivity = 2.6), will be used to reduce the interconnect capacitance. Although these improvements can be substantial, they do not provide the ultimate solution to the increasing performance mismatch between devices and interconnects. At best, they can improve the interconnect performance by one or two technology generations. The interconnect parameters shown in Table 2 from NTRS'97 have already considered the advances in the new interconnect materials, with the use of copper at the 180nm generation and the decrease of dielectric constant from 3.55 in the 250nm generation to 1.5 in the 70nm generation. However, the global interconnects remain to be the performance bottleneck as shown in Figure 1 and Table 4. Therefore, on one hand, the design technology needs to be prepared for the advent of new interconnect materials, studies its impact, and makes appropriate adjustments of the design tools and methodology. On the other hand, one needs to realize that the ultimate solution to overcome the interconnect performance and reliability bottlenecks has to come from *design innovations*.

2.2 Higher Degree of On-Chip Integration

The exponential shrinking of the feature size, together with the consistent growth of the chip size, has another profound impact -- it enables rapid increase of the chip capacity and the degree of on-chip integration. However, this potential will not be realized unless a number of key challenges faced by the design and test community are successfully addressed, such as the mismatch of design complexity and productivity, limitations of current design abstraction and hierarchy, the complexity of system-on-chip, and the power barrier. These issues will be discussed in this subsection. Again,

---

[5] Private communication with David LaPotin at IBM Research.

we postpone the discussion about possible solutions and innovations to overcome these issues to Section 3, after we complete the discussion of all aspects of design challenges.

2.2.1 Complexity and Productivity:

Based on the projection of the NTRS'97, the number of transistors on a single chip will reach over 500 millions in the 70nm technology, which presents a 50 fold increase in the device count as compared to the current 250nm generation. At the meantime, the amount of interconnect to be dealt with by the design tools will also increase drastically as shown in Table 5, with 8-9 routing metal layers, an estimated total wire length of 10 kilometers, and close to 800,000 buffers inserted for performance optimization in the 70nm generation. Moreover, the interconnects, especially global interconnects, become much more complex due to aggressive interconnect optimization for performance and reliability. A number of optimization techniques have been developed in recent years for improving the performance and reliability of the interconnects, including interconnect topology optimization, optimal wire sizing, optimal buffer insertion, simultaneous device and buffer sizing, etc. Table 6 shows the optimization results of a 2cm global interconnect in each technology generation obtained using simultaneous driver sizing, buffer insertion, buffer sizing, and wire sizing using the TRIO package [5]. For the 70nm technology, an optimized 2cm global interconnect contains 8 buffers (16 inverters), with over 99% wire segments being sized using 10 different wire widths (within the same 2cm line!). Clearly, a global interconnect is no longer a simple metal line. It becomes a *complex circuitry* with optimized devices and wires in nanometer designs! The sheer volume of devices and interconnects on a single chip, combined with the complex device and interconnect models and tight design constraints on performance and reliability, presents a great challenge to every design team and design tool, and seriously damper the growth of design productivity. The recent study by SEMATECH showed that although the level of on-chip integration, expressed in transistors per chip, increases at an approximately 58% per year compound growth rate, the design productivity, measured in transistors per person-month, grows only at a 21% per year compound rate. Such a *mismatch of silicon capacity and design productivity*, if not resolved timely, will seriously limit the potential of achieving high-degree on-chip integration, and significantly increase time-to-market.

Technology(nm)	250	180	130	100	70
Length (m)	820	1,480	2,840	5,140	10,000
Wire Levels	6	6-7	7	7-8	8-9
#buffers per chip	5K	25K	54K	230K	797K

Table 5. Complexity of interconnects in each technology generation as projected in NTRS'97. The number of buffers per chip is estimated using the wire length distribution model developed in [6] and the TRIO interconnect optimization package [5].

Technology (nm)	250	180	150	130	100	70
Delay (ns)	0.895	0.793	0.77	0.70	0.77	0.672
Number of buffers	3	4	4	4	5	8
Avg. wire width	6.52	6.61	6.93	7.31	7.70	8.00
% wire segment sized	98.6%	98.9%	99.1%	99.3%	99.5%	99.8%

Table 6: Optimization results of a 2cm global interconnect in different technology generations after simultaneous driver sizing, buffer insertion, buffer sizing, and wire sizing using the TRIO package [5]. The maximum wire width is limited to 10x the minimum and the maximum buffer size is limited to 200x the minimum. A long wire is divided into a sequence of unit-length segments, each of 100um. The wire width is uniform within each unit-length segment.

2.2.2 Limitations of Current Design Abstraction and Hierarchy:

One natural approach to designing a complex system is to provide various *levels of abstraction*. The current flow goes through behavior level design, RTL level design, logic design, and physical design. The success of such an approach depends heavily on the good correlation between the abstract model at each level and its physical implementation in the final design. Such a correlation, however, becomes very difficult to maintain, as the existing abstractions are incapable of modeling the performance, reliability, and complexity of the interconnects, which in fact have become the dominating factors to be considered and optimized in nanometer designs.

An orthogonal approach to designing highly complex systems is to apply the 'divide-and-conquer' methodology to decompose the large system into a set of smaller subsystems recursively and carry out the design *hierarchically*. Nowadays, it is common to decompose a complex IC into a number of functional blocks, each of them designed by one or a team of engineers with manageable complexity, and then go through a 'full-chip assembly' phase to interconnect these blocks together. Such hierarchical design methodology is also facing serious difficulty in nanometer designs. Although the functionality and structure of a circuit can usually be decomposed hierarchically, many performance and reliability related issues in nanometer designs, such as interconnect delay, crosstalk, power and ground bounces due to simultaneous switching, *do not fit naturally into the function hierarchy*. As a result, we are facing a *methodology crisis* with no adequate abstraction model and scalable methodology to handle the rapid growing design complexity.

Efficient reuse of existing designs and intellectual property will play an important role in reducing the mismatch between silicon capacity and design productivity. *Efficient design reuse*, however, also introduces new challenges to design abstraction and the hierarchical design methodology, as it requires proper representation, abstraction, and characterization of existing designs in terms of its functionality, performance, reliability, and possible interactions with the environment. It also requires a systematic approach to update the characterization and implementation of an existing design when we migrate from one technology generation to another, from one foundry to another, or even from one design environment to another. Such capabilities have yet to be developed and validated by the industry.

2.2.3 Systems on a Chip

Rapid increase of IC capacity enables the possibility of integrating a complex system on a single chip. It will be feasible to integrate microprocessors, embedded memories, application-specific integrated circuits (ASICs), and field-programmable logic arrays (FPGAs) into a single IC. As a result, future IC design will involve development of embedded operating systems, code-generation of embedded application software, synthesis and layout of application-specific circuits, mapping and configuration of programmable logic, and many more. Analog components, such as A/D or D/A converters and radio frequency (RF) transceiver circuits, may co-exist with digital components on the same chip. In this case, the interference between the digital and analog circuits, such as the noise from high-speed digital signals to low-level analog signals and electromagnetic interference (EMI) generated by the high-frequency circuits, needs to be properly modeled, analyzed, and considered by the design and test tools. The needs of integrating various design techniques and design tools developed for completely different design styles, automatically synthesizing of their interfaces, optimizing the overall system-on-chip, simulating, testing, and verifying the correctness of the entire system far exceed capability of today's design technology.

2.2.4 Power Barrier

Power consumption in CMOS circuits includes both static and dynamic power dissipation. The static power dissipation, caused by leakage currents and sub-threshold currents usually contributes to a small percentage of total power consumption, while the dynamic power dissipation, resulted from charging and discharging of capacitive loads of interconnects and devices dominates the overall power dissipation. Although rapid shrinking of device dimensions and reduction of the supply voltage reduce the power dissipation of individual device significantly, the exponential increase of degree and operating frequencies still results in a *steady increase* of total power consumption. For example, in the 70nm technology, with over 500 million transistors and 10 kilometer metal wires integrated in a single chip operating at multi-gigahertz, the overall power consumption is estimated to be 170W for high-performance microprocessors [2] even after drastic power supply scaling. Therefore, if power dissipation is not controlled and optimized carefully, it soon will become a *limiting factor* for system integration and performance improvement.

3   Opportunities for Design Innovations

Given the growing challenges discussed in the preceding section, an increasing amount of research and development is needed in the design technology in order to support the growth rate as predicted in the NTRS'97. This section identifies a set of key areas that are of critical needs to improve the design productivity, and are likely to lead to design innovations. These areas cut across many dimensions of the design technology, with emphasis on interconnect-driven design methodology, coupling of synthesis and layout, complexity management, and new system and interconnect architectures. We hope that our suggestions can also stimulate the exploration of many other fundamentally new concepts and techniques to cope with the challenges faced in future IC designs.

3.1   Interconnect-Driven Design Methodology

The conventional IC design process focuses mainly on *logic functions and devices*. Interconnects are largely ignored until the final step physical design in which they are implemented by either layout designers or automatic Place-&-Route tools as an *after-thought*. Given the growing

importance of interconnects on system performance, reliability, power dissipation, and overall cost, an *interconnect-driven* or *interconnect-centric* design methodology is clearly needed so that interconnect design can be considered and optimized *throughout the design process*. Such a design methodology should support interconnect planning, estimation, synthesis, and verification at every level of the design process.

Interconnect planning and estimation will allow the designer to explore various interconnect design alternatives, study the trade-off between devices and interconnects, plan the overall interconnect structures, and quantitatively evaluate the impact of interconnects on performance, power, reliability, and cost at each level of design process. For example, during design exploration, interconnect planning and estimation tool should provide models and metrics to evaluate the options of changing the number of routing layers, varying the design and manufacturing rules (such as width, spacing, wire and via aspect ratios in each layer), using different packaging alternatives, etc, and quantify their impact on various design objectives. During design synthesis, it should interact with high-level and logic-level synthesis tools to analyze the interconnect requirement of different design alternatives, perform interconnect planning with floorplanning at each level, and provide accurate estimation of interconnect performance to enable *forward constraint propagation*, so that the interconnect design can *drive* the entire synthesis process. Such an interconnect planning capability is critical, given the importance and complexity in nanometer designs. For example, it will not be feasible to insert over half a million buffers (as predicted for the 70nm technology shown in Table 5) in post-layout optimization. Such interconnect planning capability, however, is almost entirely missing in the current tools.

Interconnect synthesis will be a key part of the interconnect planning system. Given the constraints on performance, reliability, and power dissipation, it determines the optimal or near-optimal interconnect topologies, layer assignment, wire widths and spacings, buffer locations and sizes, etc, for overall cost optimization. In the early stage of the design process, the results of interconnect synthesis can be flexible and imprecise, depending on the level of abstraction and accuracy of interconnect modeling. They will be gradually refined to a complete design-rule correct detailed routing solution satisfying all design constraints as the overall design process progresses.

Interconnect verification needs to be carried out at every level of the design process in junction with the existing capabilities of simulation and verification of logic functionality at corresponding levels. It will extract and model the interconnect parasitic parameters, neighborhood structures, temporal correlation among different signals, and interaction between devices and interconnects, so that it can simulate and verify the performance and reliability of the interconnects at each level of design process with sufficient accuracy and efficiency. Since the performance and reliability of interconnects depend heavily on the *interaction of signals* in nanometer designs, simulation along will not be sufficient, as enumeration of all correlation patterns for simulation is not computationally feasible. *Formal verification techniques for interconnect performance and reliability* need to be developed, in addition to the current effort on formal verification of circuit functionality. The statistical variations of the interconnect designs should also be taken into consideration in the interconnect verification.

3.2   *Coupling between Synthesis and Layout*

The growing significance of interconnects requires careful consideration of layout design in the higher levels of design process, as layout design finally determines the interconnects. Effective coupling between synthesis and layout, however, has been very difficult due to the high complexity

of the resulting problem, lack of support in the traditional design flow, and the historical division of the EDA industry. New design *infrastructure, methodology, and techniques* are needed to overcome such difficulties. We classify the possible approaches to synthesis and layout coupling broadly into two categories.

One approach is to employ a highly *iterative design flow*. It follows the design steps in the traditional design flow, but feeds the layout result in the current iteration back to synthesis tools at higher levels to improve the synthesis results in the next iteration to better meet the design constraints. This approach is currently being adapted by the industry. But many questions need to be answered before accepting such a *'construct-by-correction'* approach in nanometer designs. In particular, we need to understand if this approach *guarantees the convergence* to a feasible and close-to-optimal solution satisfying all design constraints, and whether the number of iterations as well as the design compilation time within each iteration can be well controlled to lead to an acceptable overall design cycle.

Another approach uses a *concurrent design flow*, which perform architecture/behavior-level synthesis, RTL/logic synthesis, and layout synthesis concurrently. It constructs an initial design in high-level abstraction and gradually refines the logic and interconnect designs to a complete solution. The design constraints are satisfied in constructing the initial solution, and maintained after each refinement through careful *constraint propagation*. The feasibility of such a *'correct-by-construction'* approach shows a greater promise, but has yet to be demonstrated successfully to handle the complex constraints in nanometer designs. Many critical issues, such as design abstraction, constraint propagation, and efficient techniques for design refinement, need much more research.

We believe that an *combination of iterative and concurrent design approaches* is likely to be successful in practice, which performs concurrent synthesis, layout planning, and solution refinement in all levels of the design process, with possibly limited number of iterations within the same or adjacent levels to correct unacceptable estimation errors. For all three approaches, the interconnect-driven design methodology discussed in the preceding subsection is needed to guide the design process. The existing high-level and logic-level synthesis, simulation, and verification capabilities need to be integrated with the interconnect planning, estimation, synthesis, verification capabilities to support a complete interconnect-driven design flow.

3.3   Innovations in Complexity Management

Complexity management is critical for improving the design productivity to meet the fast-growing silicon capacity. All areas of CAD are facing the complexity crisis. We believe that innovations in the areas of *design abstraction*, *efficient re-use* of existing designs and intellectual properties, and *large-scale constrained global optimization* are most needed to enable successful complexity management in all problem domains.

Design abstraction is critical to the hierarchical design methodology, which has been the most effect way to control the design complexity. The current approach provides good abstraction of the functionality and structure of the design. But much research is needed to develop adequate models and metrics for performance, power, area, reliability, and cost at each level of design hierarchy with consideration of interconnects to enable high-level design decisions. Such models should consider not only the functionality of the circuit, but also model the underlying *design process* and *design tools* to be used, as they have substantial impact on the final design as well. They should provide

sufficient information concerning issues in nanometer designs, such as the RLC loads seen by other blocks, the delays under different input combinations, and possible coupling with other blocks through crosstalk and simultaneous switching. These models should provide simulation and emulation capabilities, and can be gradually refined into final implementation.

Design reuse is another important mean of improving the design productivity. Effective reuse of existing designs or intellectual properties (IP) requires proper *representation, abstraction, and characterization* of an IP in terms of its functionality, performance, reliability, and possible interactions with the environment, so that it can be seamlessly integrated with the rest of design for synthesis, simulation and verification. Such representation and abstraction should also support *efficient update* of an existing IP when we migrate from one technology generation to another, from one foundry to another, and from one design environment to another. Design reuse should also include the development *of reusable design process, methodology, and tools* so that they can be re-targeted for different technology generations and easily shared among different design projects. For example, an high-quality re-usable cell generator will significantly reduce the development cost and design time compared to manual design of cell libraries for every technology generation.

Development of efficient, highly scalable optimization algorithms for global performance, power, area, and reliability optimization will also play a fundamental role in complexity management. Many optimization techniques used in today's design tools were developed in early 1980s to handle a few thousands of circuit elements, and many of them have serious limitations to handle the complexity of future designs. One example is the widely used simulated annealing algorithm, which produced satisfactory solutions to many VLSI CAD optimization problems in the past, but is difficult to scale to today's design complexity in terms of both runtime and solution quality. Innovations in *highly scalable optimization algorithms* which can handle complex design constraints, multiple design objectives, and rapid increasing design sizes will significantly improve the capability, efficiency, and quality of the design tools for future ICs. Investment in this area will have very high potential of returns. The VLSI CAD community enjoyed a period of close interaction with the theoretical computer science community in early 1980s, which led to a number of significant progresses, as evident by a number of significant results on VLSI routing, logic optimization, and circuit retiming developed jointly by researchers from both communities during that period of time. Many of them are still used widely in today's design tools. For various reasons, such collaboration did not continue into this decade. Given the challenges we are facing today, we believe that it is important to revive such *inter-discipline collaborations* again, so that the design technology can benefit from the advances and breakthroughs in many other related areas, such as operation research, mathematical programming, combinatorial optimization, computational geometry, in order to cope with the fast growing silicon capacity and design complexity.

3.4   New System and Interconnect Architectures for Predictable Performance

The interconnect bottleneck problem cannot be eliminated by the improvement of design methodology and design tools alone. The results in Table 4 show that even after the best possible interconnect optimization based on the process geometry given in NTRS'97, the performance of global interconnect still cannot meet the target clock rate. Innovations in the system architecture and the interconnect architecture are needed to overcome the interconnect bottleneck problem.

The development of *new system architectures* may take many different directions. *The trade-off between devices and interconnects* should be explored to utilize the abundant silicon capacity to improve the interconnect performance (buffer insertion can be viewed as a simple case of it*)*.

Regular structures and architectures with predominately local interconnects should be strongly favored (the systolic array is an extreme example). Use of *different communication protocols* over the global interconnects, such as asynchronous or differential signaling, should also be considered.

The exploration of *new interconnect architectures* also has many dimensions. *Design and optimization of process geometry and interconnect physical architecture* may substantially improve the interconnect performance by optimally determining the width, spacing, wire and via aspect ratio, metal and dielectric materials of each layer to be used in fabrication. An aggressive reverse scaling scheme was suggested in [7], where an exponentially increasing reverse scaling factor is applied to all conductor and dielectric cross sectional dimensions when each pair of wiring layers are added. This scheme effectively controls the growth of global interconnect delay at the expense of design density. Other schemes should be explored as well to consider the performance and density trade-off. *Development of interconnect structures with predictable performance* may greatly simplify the design flow and enables efficiently interconnect planning. One example is the design of today's FPGA routing architecture, where different interconnect resources, such as direct interconnects, average interconnects, and global interconnects, are carefully planned at the chip architecture level in anticipation of the interconnect needs. This will give much predictable interconnect performance. *Fundamentally new interconnect technologies*, such as optical interconnect schemes and on-chip wireless radio frequency connections, should also be explored jointly with the scientists in other field, as they may completely change the nature of interconnect problems and provide revolutionary advancement.

The search for system and interconnect architecture innovations is likely to involve the system architects, circuit designers, and fabrication process experts. It is very important, however, for the design technology community to come up with appropriate models, design methodology, and tools to *drive* the exploration of new design alternatives, *evaluate and validate* their potentials.

4   Concluding Remarks

This paper identifies a set of key challenges to the design technology for future nanometer IC designs, especially in the areas of interconnect designs and large-scale on-chip integration. It provides quantitative measurements of the scale and severity of the problems faced by the design community based on the detailed modeling and simulation using the data provided in the newly published NTRS'97. This study clearly indicates the that the design technology may well be the bottleneck in sustaining the exponential growth of the IC technology, unless we can successfully handle the increasing challenges on interconnect performance and signal reliability, overcome the power barrier, manage the rapid growing design complexity, and significantly improve the design productivity. The paper also identifies a set of areas where design innovations are most needed, including development of interconnect-driven design methodology, coupling synthesis with layout, design complexity management, and development of new system and interconnect architectures for predictable performance. Possible innovations and research opportunities in these areas are discussed. Specific directions and recommendations include (i) development of an interconnect-driven design methodology with interconnect planning, estimation, synthesis, and verification capabilities, (ii) coupling synthesis with layout designs through a hybrid of concurrent and iterative design flows, (iii) innovations in complexity management, especially in the areas of better design abstraction, efficient design re-use, and development of efficient, highly scalable optimization algorithms, and (iv) development of new system and interconnect architectures, including the exploration of trade-off between device and interconnects, use of regular structures and different

communication protocols, design and optimization of process geometry and interconnect physical architecture, development of interconnect structures with predictable performance, and search for fundamentally new interconnect technologies.

5 Acknowledgements

6 References

1. G. E. Moore, "Cramming More Components onto Integrated Circuits", Electronics Magazine, Vol. 38, April 1965, pp. 114-117.
2. Semiconductor Industry Association, National Technology Roadmap for Semiconductors, 1997.
3. K. Nabors and J. White, "FastCap: A Multiple Accelerated 3-D Capacitance Extraction Program", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 10, Nov. 1991, pp. 1447-1459.
4. J. Cong, L. He, C. K. Koh, and P. H. Madden, "Performance Optimization of VLSI Interconnect Layout", Integrations, the VLSI Journal, Vol. 21, 1996, pp. 1-94.
5. J. Cong, L. He, K. Y. Khoo, C. K. Koh, and Z. Pan, "Interconnect Design for Deep Submicron ICs", Proceedings of International Conference on Computer-Aided Designs, Nov. 1997, pp. 478-585.
6. J. A. Davis, V. K. De, and J. D. Meindl, "A Stochastic Wire Length Distribution for Gigascale Integration (GSI)", Proceedings of Custom Integrated Circuits Conferences, 1997, p. 145.
7. G.A. Sai-Halasz, "Performance Trends in High-End Processors", Proceedings of the IEEE, vol. 83, no. 1, Jan. 1995, pp. 20-36