

Getting to the Bottom of Deep Submicron

Dennis Sylvester

University of California, Berkeley
Electrical Engineering and Computer Sciences
dennis@eecs.berkeley.edu

Kurt Keutzer

University of California, Berkeley
Electrical Engineering and Computer Sciences
keutzer@eecs.berkeley.edu

ABSTRACT

In this tutorial, we take a fresh look at the problems posed by deep submicron (DSM) geometries and re-open the investigation into how DSM effects are most likely going to affect future design methodologies. We describe a comprehensive approach to accurately characterizing the device and interconnect characteristics of present and future process generations. This approach results in the generation of a representative strawman technology that is used in conjunction with analytical models, simulation tools, and empirical design data to obtain a realistic picture of the future of circuit design. We then proceed to quantify the precise impact of interconnect, including delay degradation due to noise, on high-performance ASIC designs. Having determined the role of interconnect in performance, we then reconsider the impact of future processes on ASIC design methodology.

Keywords

Interconnect modeling, gate delay, CMOS scaling, signal integrity, power dissipation, ASIC, wirelength

1. Introduction

The magnitude of the difficulties associated with the plunge of circuit design into deep submicron (DSM) process geometries has been the source of much speculation. One particularly significant impact has been the rise in the portion of critical path delay attributable to interconnect. A number of independent sources have made forecasts that in DSM geometries 80% or more of the delay of critical paths will be directly linked to interconnect [1]. This forecast has been further supported by the broad industrial experience of significant problems in obtaining timing closure in current high-performance integrated circuit (IC) designs. Based on the present difficulties in IC design methodologies and predictions of increasing migration of delay to interconnect, there is a sense in both industry and academia that current synthesis and physical design methodologies require a significant overhaul. In this scenario, traditional design flows will no longer be viable for any size of module, or block of gates.

Deep submicron effects, particularly interconnect, have thus been billed as potential showstoppers to the continuation of Moore's law. Among the effects that are commonly mentioned are rising RC delay of on-chip wiring, noise considerations such as crosstalk and delay unpredictability, reliability concerns due to rising current densities and oxide electric fields, and increasing power dissipation. Each of these issues has underlying physical explanations that shed insight into its potential impact as CMOS processes continue to scale. We address these issues and others in section 4.

Our goal in this paper is to get to the bottom of deep submicron effects and make an objective judgement of how they are likely to

impact future design methodologies. We will present a strawman view of future technologies that give what we believe is a realistic and substantiated projection of the impact of DSM on IC design. To achieve this we use simulation tools as well as analytical models, in combination with actual design data, to obtain a comprehensive view of the future of circuit design. We then aim to draw conclusions that assess the ability of current CAD flows to survive in the light of DSM effects.

2. Current Methodology

2.1 Traditional ASIC flow

Figure 1 gives the implementation portion of a design flow used in traditional application-specific integrated circuit (ASIC) designs. In this flow, the design is described in a hardware-description language (HDL) such as VHDL or Verilog. The *technology library* is the database containing the data that models the pre-designed cells in the underlying process technology for the logic synthesis and physical design tools. *User constraints* are the mechanism by which the user conveys constraints regarding the speed, area, and power of the design. Logic synthesis transforms the HDL description into a graph in which each vertex represents a cell in the technology library, and each edge represents a wired connection between the cells. This graph is called a *netlist*. Logic synthesis optimizes the circuit according to user constraints and ensures that design rules are met. A survey of logic synthesis can be found in [2].

Physical design is the process by which the synthesized *netlist* is transformed into a mask, which is used to fabricate a three-dimensional integrated circuit. Information contained in the *technology library* and *user constraints* ensures that the output of physical design can be fabricated in the designated semiconductor process. User constraints restrict the location of pads or signals, the area resources available for implementation and the timing behavior. Typical steps in this category are cell placement, global and detailed routing, sizing and clock/power distribution. An excellent reference for these physical design operations is [3]. In the ASIC flow described in figure 1, the first half of the flow is the responsibility of the design center while the second half is the responsibility of the ASIC or semiconductor vendor.

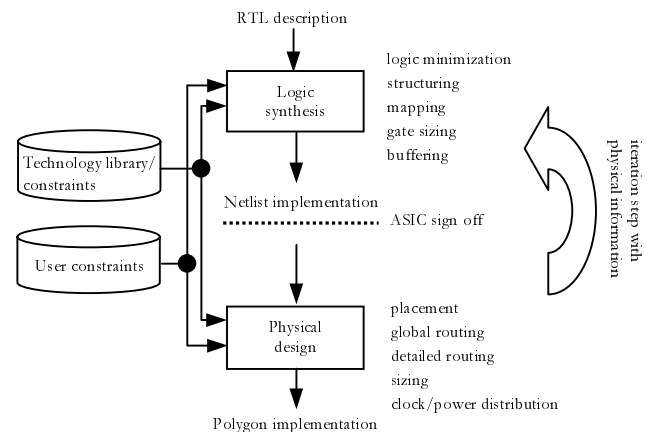


Figure 1. Traditional ASIC design flow

2.2 Problems with current flow

Let us now anticipate the effect of deep submicron technology on the tools and flows. The design flow in figure 1 contains a separation between the logic synthesis step and the physical design step. For designs with aggressive performance goals, we find that several iterations between synthesis and physical design are required to converge to a desired implementation. As a result, design teams have begun to bring more of the backend design flow in-house and the handoff to the semiconductor-vendor occurs only at the end. This approach is shown in figure 2 and is known as the *customer-owned tooling* (COT) approach. Note that the differences between the flow in figure 1 and that of figure 2 are more organizational than technical.

2.2.1 Problems with logic synthesis

During synthesis the capacitances and delays associated with final wiring are unknown. Models of interconnect, known as *wire-load models*, attempt to predict the amount of capacitance in a wire by reducing it to a function of fanout and block size. In this approach a single capacitance value is used for *all* nets in a block with the same fanout. Obviously, the capacitance values for nets in a block will vary and so the wire-load model is necessarily only an approximation of the actual future capacitance. When wire delay constitutes a small percentage of the total critical path delay, errors in capacitance estimation have little impact. However, if the capacitance values in wires increase then wire-load models become increasingly inaccurate. For this reason, precisely calibrating the increase in wire capacitance in future processes is one of the primary aims of this study.

2.2.2 Problems with physical design

Placement and routing tools tend to use primitive models of circuit delay. Often the timing constraints inherent in the synthesized netlist are translated into a single static net prioritization. Place and route tools attempt to honor this prioritization scheme, but because they have only a primitive internal model of critical path delay, they cannot be sure that they have in fact obeyed the initial timing constraints. As delay migrates to interconnect, the possibility that excess capacitance in routing violates a timing constraint increases. If capacitance is increased due to cross-coupled capacitances in routing, this is only likely to be discovered after the final layout is extracted.

2.2.3 Problems with the design flow

Problems with the design flows in figures 1 and 2 will be understood by stepping through the flow. Initially the netlist from synthesis is optimized by logic optimization using a wire-load model as described above. The resulting netlist is passed on to placement and routing tools. Place and route will attempt to realize the delay

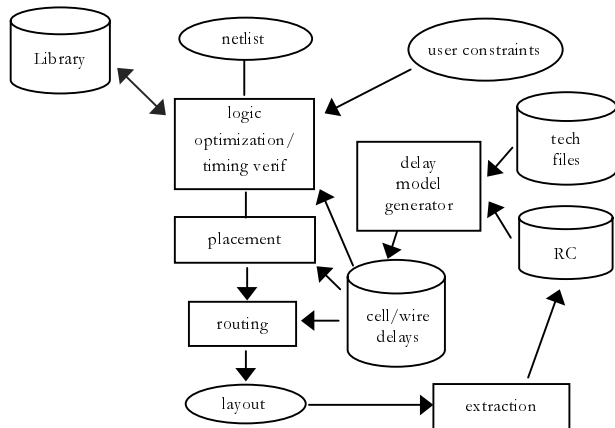


Figure 2. Today's high performance logical/physical flow

constraints in the netlist, but if the timing constraints are violated during place and route it is unlikely to be discovered until extraction is performed, new RC's are extracted, delay models are generated, and an accurate timing analysis is performed. Then the synthesized netlist can be back-annotated with more accurate capacitances and returned for another round of logic optimization. Unfortunately, if the original capacitance estimates of logic optimization were very inaccurate, then logic optimization, lacking good incremental optimization capability, is likely to produce a netlist that bears little resemblance to the original netlist. Placement, also operating with limited incremental capability, will then produce a new placement while routing produces a new set of routes and capacitances. Therefore the chances that the design produced by the second iteration through the flow realizes the timing constraints may not be any better than the first.

Generally speaking, logic optimization, placement, and routing are each inherently sensitive to changes in the input parameters. Modest changes in input parameters (*e.g.* wire capacitances or gates in the netlist) can cause significant changes in the output. Two especially important issues to calibrate in future processes is the amount of wiring capacitance and the prevalence of delay degradation due to noise. Therefore, the focus of this paper is to first accurately model future processes, then evaluate the impact of future processes on design characteristics such as noise and interconnect delay, and finally evaluate the impact of these characteristics on future design methodologies.

3. Our approach

A particular goal of this paper is to determine the size of a module which can still be reliably designed in the methodology of figures 1 and 2 without significant attention to DSM effects. Figure 3 shows the approach employed in this paper. To begin, we develop a strawman technology file that completely describes future process generations for both device and interconnect characteristics. To supplement this process data, we have obtained design data from an ASIC vendor for current 0.35 μm designs. From this data, we extract important design characteristics such as average wirelengths, average fan-outs, etc. Combining this empirical data with analytical models, we propose a "shrink" process down to 0.05 μm .

To facilitate our analysis, we develop models of future designs, at varying degrees of accuracy, that allow us to determine the performance of future ASIC's. These models emphasize the important issues of delay, noise, and power. Before introducing our strawman technology we will first give a high-level overview of device and interconnect issues in deep submicron.

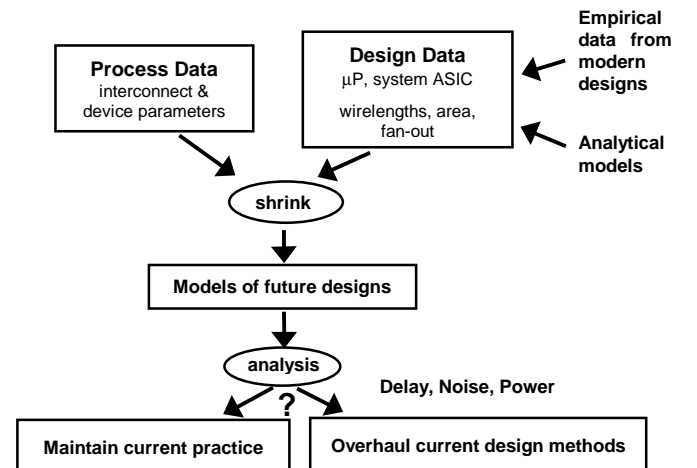


Figure 3. Overview of our analysis methodology

4. Device and Interconnect in DSM

In this section, we take an extensive look at the most significant circuit-level issues arising from the scaling of CMOS beyond the 0.25 μm feature size.

4.1 MOSFET's

4.1.1 Voltage scaling

While 5-volt power supplies were prevalent at feature sizes $> 0.5 \mu\text{m}$, DSM processes are seeing a continual drop in V_{DD} . One key reason for reducing voltages in scaled CMOS devices is that of reliability. For instance, gate oxides tend to break down if exposed to electric fields in excess of 5 to 6 MV/cm [4]. Since oxide thickness (T_{ox}) is being scaled to increase current drive, the maximum voltage on the gate (V_{DD}) must also be scaled to keep electric fields within reason. Another form of device reliability that must be considered is the effect of hot electrons. When the electric field in the channel is too high near the drain, electrons may gain enough energy to inject themselves into the gate oxide. This accumulation of charge in the oxide results in shifts in V_t and hence, changes in the device's I-V characteristics. By reducing the voltage supply, both the channel and gate electric fields will be proportionately reduced for improved reliability.

The impact of reduced supply voltages on power is another important reason why V_{DD} values are decreasing. Dynamic power consumption is defined by $P_{dyn} = \alpha C V_{DD}^2 f$, where α is a switching activity ratio, C is the switched capacitance, and f is the operating frequency. The quadratic dependency on supply voltage makes the reduction of V_{DD} a primary goal in power minimization. This issue will be discussed more in the analysis section.

4.1.2 Drive current

Classical long-channel MOS theory states that device current in the saturation mode of operation is proportional to the square of gate drive ($V_{DD} - V_t$) and inversely proportional to channel length. At ultra-short channel lengths most carriers travel at a maximum saturated velocity, v_{sat} , throughout the channel, which nearly eliminates the impact of channel length on current. A new and more accurate expression for this limiting case is as follows [4]: $I_{dsat} = W v_{sat} C_{ox} (V_{DD} - V_t)$. From this expression we can see the channel length independence as well as the move from quadratic to linear dependence on gate drive. Since v_{sat} is a material constant (approximately 8×10^6 cm/s for electrons, 6.5×10^6 cm/s for holes), we find that I_{dsat} (normalized to device width) will vary with $(V_{DD} - V_t) / T_{ox}$. With V_t fixed at $V_{DD}/4$ (necessary to maintain sufficient gate drive), we obtain $I_{dsat} \propto V_{DD} / T_{ox}$. Since both of these parameters are decreasing, DSM MOSFET's are not expected to provide enhanced drive current per unit width. This marks a new regime for scaled CMOS devices, since up to the 0.35 μm process generation additional current drive was obtained with each process shrink. Data taken from published reports confirms this analysis of saturation current in DSM CMOS devices [5-8]. Our own survey of 16 reported technologies shows both n- and p-channel devices to achieve negligible increases in drive current from 0.25 to 0.09 μm L_{drawn} .

4.2 Interconnect

Deep submicron interconnect effects present a variety of problems to process engineers, circuit designers, and CAD tool developers. This section presents an overview of these effects.

4.2.1 RC delay

The most commonly cited DSM interconnect problem is that of rising RC wire delays. For instance, the RC delay of a 1 mm metal 1 line in 0.5 μm technologies was 15 ps while in 0.1 μm technology it is 340 ps (without new materials). It can be clearly seen that wiring

delay is capable of consuming the majority of the shrinking clock cycle time in DSM designs. We now look at the reasons behind the rapid increase in RC delay and possible methods of slowing this trend.

Increasing line resistance is the main reason behind the increased wiring delay in DSM. Resistance is inversely proportional to the cross-sectional area of the wire. Due to the rising need for higher densities on-chip, wiring pitches are dropping rapidly at about the same rate as gate length. In an effort to keep resistance from increasing too quickly, many processes are scaling line thickness (or height) at a slower rate, which results in taller, thinner wires. For instance, [1] predicts an increase in wiring aspect ratio ($AR = \text{height}/\text{width}$) from 1.8 at 0.25 μm to 2.7 at 0.07 μm .

Besides the use of high AR lines, the other approach to reducing resistance is the use of better conductors for on-chip interconnect. Until recently, aluminum wires were used exclusively in back-end processes. Recent literature [5,6] has demonstrated the use of copper in sub-0.25 μm processes. The resistivity of copper interconnect is approximately 30% smaller than that obtained using aluminum wiring (2.2 $\mu\Omega\text{-cm}$ vs. 3.2 $\mu\Omega\text{-cm}$). Another key advantage gained through the use of copper wiring is increased resistance to electromigration (EM) effects. Electromigration occurs in metals when a large current density is being driven through the line. Metal ions are physically moved down the wire by the large current, resulting in opens in the line or shorts to neighboring wires. It has been shown that scaling aluminum wires to the 0.18 μm generation and beyond will cause severe restrictions in routing due to EM [9]. Copper, however, has a much lower susceptibility to ion transport from EM since it is a heavier metal. Results have shown copper to have an EM lifetime that is 100 times longer than aluminum wiring at the same current density [5].

Wiring capacitance is also increasing in scaled processes due to the higher densities needed to route modern chips. For instance, line-to-line spacing and insulator thickness are both shrinking, resulting in an overall increase in line capacitance (this is true despite smaller linewidths). Since the reduction of packing density is not an option in DSM, the only way of reducing wiring capacitance is by using a low-k dielectric instead of SiO_2 . SiO_2 has many natural advantages that have made it the dielectric of choice in microelectronics. However, its relative dielectric constant of 3.9-4.1 leaves room for improvement by using advanced materials such as polyimides. Significant research is ongoing in the area of low-k process integration as preliminary work has suggested that gate delay and power can be reduced by 39% and 47% respectively in a 0.25 μm process with $\epsilon=3.1$ [10]. The ultimate goal in low-k dielectric process integration is the use of xerogels, which are highly porous materials with dielectric constants approaching that of air, $\epsilon=1$ [11].

4.2.2 Noise

As mentioned above, one of the methods to reduce resistance has been to slowly scale line thickness, resulting in taller, thinner wires. These high aspect ratio lines have a detrimental side effect in that they result in a large amount of coupling capacitance. With an $AR > 1$, lines tend to have more parallel plate capacitance to neighboring wires than to upper and lower wiring layers, which effectively serve as ground planes. In addition, spacing between wires is shrinking quickly in an attempt to maintain high packing densities, further increasing coupling capacitance. As evidence, line-to-line capacitance between wires on the same level can be seen to make up over 70% of the total wiring capacitance at lower levels even at 0.25 μm processes [12].

The impact of this rise in coupling capacitance can be seen in the form of noise. In this paper, we will discuss two distinct forms of noise in regard to DSM interconnect. The first is delay deterioration, which refers to the fact that total capacitance seen by a

gate is no longer a constant value [13,14]. Due to the rising contribution of coupling capacitance to total load capacitance, the Miller effect can have a large impact on actual delay times on a chip. The Miller effect states that when both terminals of a capacitor are switched simultaneously, the effective capacitance between the terminals is modified. For instance, if a wire A switches from 0 to V_{DD} while an adjacent wire B switches from V_{DD} to 0, the effective voltage swing between the 2 terminals is actually $2V_{DD}$. Since $Q=CV$, the charge needed to switch wire A is now doubled with respect to the case where wire B is static. Alternatively, this is seen as a doubling of the “effective” capacitance. Clearly the increase in coupling capacitance is a potential timing hazard in that delay becomes a function of neighboring signal activity, making static timing analysis difficult.

The second form of noise we discuss in this paper is that of crosstalk or signal integrity [12]. In this scenario, a static wire (called the victim) is perturbed by switching activity on neighboring wires (aggressors). In the worst-case, 2 aggressors switch in the same direction simultaneously, leading to an undesirable voltage spike on the victim line due to capacitive coupling. This sort of noise can cause false switching (especially in dynamic circuits) or voltage overshoot effects, which may lead to enhanced device stress or forward-biasing of p-n junctions. Crosstalk is highly sensitive to the ratio of coupling capacitance to total capacitance, implying that signal integrity will become a larger issue as interconnect dimensions continue to scale.

5. Strawman Technology

Aiming to quantify the concerns raised in the previous section, we now present substantiated projections of DSM technology parameters at both the device and interconnect level. Care is taken to explain and justify these choices, and comparisons are drawn to predictions in [1].

5.1 Device roadmap

Process (μm)	T_{ox} [1] (\AA)	V_{dd} (V)	V_t (V)	L_{eff} (μm)
0.25	50 [40-50]	2.5	0.625	0.16
0.18	40 [30-40]	1.8	0.450	0.10
0.13	30 [20-30]	1.5	0.375	0.07
0.10	25 [15-20]	1.2	0.3	0.05
0.07	20 [<15]	0.9	0.225	0.035
0.05	15 [<10]	0.7	0.175	0.025

Table 1. Projected MOSFET characteristics in DSM

Table 1 presents the most important features of our strawman technology for DSM CMOS devices. Our gate oxide thickness projections are generally higher than those found in [1]. Due to oxide reliability requirements, one should keep the electric field in the oxide below 5-6 MV/cm [4]. The aggressive numbers found in the roadmap exceed this benchmark, especially at 0.13 μm and beyond. In addition, oxides thinner than 20 \AA raise serious concerns about leakage due to direct tunneling through such an extremely thin layer [15]. These concerns are not likely to be resolved by the 0.13/0.1 μm generations as anticipated in [1]. Therefore, we predict oxides that are free from tunneling problems until the 0.07 μm technology node. Finally, T_{ox} values below 10 \AA are extremely optimistic; this thickness represents only a few atomic layers of SiO_2 . We predict oxides to “bottom out” around 15-17 \AA due to leakage and fabrication issues.

Regarding voltage scaling, we select the high-performance V_{DD} values from [1], rather than the low-power scenario. Also, V_t is set at $V_{DD}/4$ to provide sufficient current drive to keep performance climbing [4]. Beyond 0.1 μm , such small V_t 's may yield high amounts of leakage current. In this case, several approaches may be taken. A dual- V_t process utilizes two different threshold voltages within the same design. Low V_t devices are used where speed is essential while the bulk of devices (e.g. 90%) have a higher V_t to keep overall leakage current small [16]. A second approach would be to use circuit techniques to raise the V_t in idle circuit blocks. This approach is similar to standby or sleep modes in current microprocessors where the clock is disabled in low-activity regions in order to reduce dynamic power consumption [17]. However, sleep modes do not eliminate static power and further work will be needed in this area.

Finally, effective gate length is extrapolated from current trends and also based on published reports of DSM devices. As can be seen from the table, channel overlap length ΔL ($= L_{drawn} - L_{eff}$) and T_{ox} are scaling similarly so that overlap capacitance will be roughly constant with scaling. From baseline 0.35 μm processes, we extrapolate the device capacitance parameters based on physical relationships, such as $C_{junction} \propto (N_{sub})^{1/2}$. Combined with the characteristics described in table 1, BSIM3 models have been developed to model DSM devices [18]. Default parameters are used except for T_{ox} , V_t , mobility, and capacitances. Resulting I-V curves yield good fit (within 20% in linear region, 10% in saturation) with measured results from 0.15 μm devices ($T_{ox} = 40 \text{ \AA}$). Simulated values of I_{dsat} for 0.25 to 0.1 μm processes show excellent correlation with published data.

5.2 Interconnect roadmap

Process (μm)	0.25	0.18	0.13	0.1	0.07	0.05
Thickness (μm)	0.5	0.46	0.34	0.26	0.2	0.14
Width / Space (μm)	0.3	0.23	0.17	0.13	0.1	0.07
Sheet Resistance (Ω/\square)	0.044	0.048	0.065	0.085	0.11	0.16
T_{ins} (μm)	0.65	0.5	0.36	0.32	0.27	0.21
Dielectric Constant	3.3	2.7	2.3	2.0	1.8	1.5

Table 2. Interconnect characteristics for metals 1 and 2

Thickness (μm)	Width/Space (μm)	Sheet Resistance (Ω/\square)	T_{ins} (μm)
2.5	2.0	0.009	1.4

Table 3. Top metal layer parameters for all generations

Tables 2 and 3 highlight key parameters from our interconnect roadmap. Table 2 presents dimensions for the lower 2 levels of metal for each process. Our interconnect hierarchy consists of several pairs of identical metal layers that fall under the categories of local, intermediate, and global wiring. The number of metal layers increases from 6 at 0.25 μm to 9 at 0.05 μm for enhanced connectivity. The first two levels of metal are used exclusively for routing local signals between gates within a larger block of gates (e.g. 50K). In these instances, the wavelength is typically short and the first concern is that of wiring density. Therefore, we predict a

continuing drop in lower-level wiring pitch. This will not only provide additional local routing capability but will also allow for smaller standard cell sizes as [19,20] have shown cell size to be set by contacted wiring pitch.

The second concern at this level is noise. Due to the shrinking pitches, larger coupling capacitances lead to enhanced noise. In order to limit noise, we recommend the use of “flat” wiring where the aspect ratio is capped at 2 [12]. Compared to predictions in [1], an AR of 2 will yield 30% smaller coupling capacitance at $0.07\ \mu\text{m}$ than $\text{AR} = 2.7$. The use of thinner wires in the “flat” approach can be seen as a tradeoff between noise and resistance, where the lower resistivity of copper is taken advantage of in order to limit capacitances. This approach has been used in early copper designs to limit capacitance and is especially beneficial at lower levels where device resistance tends to be much larger than wire resistance (due to short wirelengths) [16]. An additional point in noise reduction is the scaling of insulator thickness, T_{ins} . In order to prevent coupling capacitance from becoming an even larger portion of total wiring capacitance T_{ins} needs to be scaled appropriately. Also, by reducing T_{ins} vias with reasonable AR’s can be used, allowing for easier fabrication and lower resistance.

Copper is used for all sheet resistance calculations, with a resistance of $2.2\ \mu\Omega\text{-cm}$. The reduction of dielectric constant with scaling reflects the significant work being done on low-k materials to replace SiO_2 as the insulator of choice in ULSI [21]. The implementation of low-k dielectrics into processes represents a significant step in realizing very high performance designs. Smaller interconnect capacitances result in lower power dissipation as well as smaller delay times, making low-k dielectrics a more beneficial process advance than copper wiring. Our projections for low-k dielectrics are fairly conservative in comparison to [1], especially beyond $0.13\ \mu\text{m}$.

Table 3 shows projections for global interconnect throughout scaling. In contrast to lower level metals, where average wirelengths scale down due to shrinking gate sizes, global wires must actually become longer. For this reason, we select a large cross-section global wire that maintains a constant resistance of $44\ \Omega/\text{cm}$. This low resistance value will allow for unattenuated distribution of power grids, clocks, global busses, and other important signals on the top layers of metal. The approach follows the concept of “fat” wires suggested in [22]. Transmission line characteristics will need to be well-modeled and controlled in this scheme, as time-of-flight delay will be an important issue [23].

Intermediate metal layers not included in tables 2 and 3 provide inter-modular routing and offer minimum pitches and thicknesses between those previously presented. For instance, metals 3 and 4 have a minimum pitch that is approximately double that of metals 1 and 2.

6. Design Data

To supplement the analytical models and simulation tools to be used in the analysis section, we obtained detailed design data from modern ASIC’s. Empirical design data was compiled for thirteen $0.35\ \mu\text{m}$ ASIC designs from Symbios Corporation. Average wirelengths in the designs vary somewhat but tend to be in the range of 200 to $300\ \mu\text{m}$. Average fan-out is consistently between 2.2 and 3. Designs regularly incorporated large macro blocks in addition to standard cells. The percentage of standard-cell logic area to total chip area ranged from 30-95% but logic area typically encompassed about 70% of the chip. Data was also obtained showing average wirelength for each fan-out in each design. This data is used in section 7 to create a critical path model for future ASIC’s.

7. Analysis

In this section, we develop models for various performance metrics (delay, noise, power) and apply them to DSM ASIC’s to determine the size of a module which can be designed using the flow in figures 1 and 2.

7.1 Delay

7.1.1 Defining gate/interconnect delay

One of the primary goals in our analysis of DSM trends is to determine the impact of interconnect in $0.1\ \mu\text{m}$ technology. To assist in doing this, we propose a well-defined method of assessing gate delay vs. interconnect delay. Typically for a given process, gate delay (t_{gate}) is determined using an unloaded ($\text{FO} = 1$, no significant wiring load) ring oscillator made up of inverters. This is an elegant way to determine t_{gate} as it is independent of device sizing since increasing device width contributes equally to larger drive current and load capacitance.

We propose a modified version of the ring oscillator concept to determine gate and interconnect delay. First, 2-input NAND gates replace inverters since they better represent on-chip logic gates. One of the inputs in these gates is tied to V_{DD} , resulting in worst-case low-to-high delays. Next, the fan-out of the gates is varied from 1 to 4 (fan-outs greater than 4 are not of practical interest). At this point, gate delay is found for each fan-out individually. Finally, minimum-pitch interconnect of length L_{avg} is added between each stage. Average wirelength is a function of fan-out and is varied accordingly, using empirical design data as a reference. At this point, a stage delay (t_{stage}) is found for a given technology with fan-out ranging from 1 to 4. Interconnect delay, t_{wires} is defined as the difference between the stage delay and intrinsic gate delay. The basic approach is shown schematically in figure 4.

Sizing of gates with interconnect loading becomes non-trivial as extremely large devices could be used to make t_{wire} negligible. Likewise, the use of minimum-sized gates would yield a misleading depiction of interconnect delay. Since this is not practical due to area and power considerations, we define an optimal driver size. The optimal size is defined by a W/L ratio such that an increase in W/L of 1 does not yield a 2% drop in stage delay. This criterion was chosen to most closely approximate the knee of the delay vs. device sizing plot obtained when sweeping W/L.

7.1.2 Analytical approach

In this section, we discuss the first of two different approaches to the delay analysis. A first-order analytical model is presented to observe general delay trends in future ASIC’s. A full-scale simulation approach that employs accurate device models and distributed interconnect effects is described in the following section.

To the first-order, delay of a MOSFET is governed by a simple relationship between capacitance, voltage, and current: $T_d = CV/2I$. To study the scaling of delay through process generations, we employ this expression by approximating the trends of each of the variables involved. Several approximations are made and will be discussed in turn.

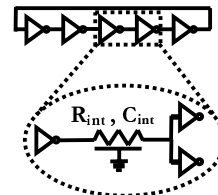


Figure 4. Determination of gate/interconnect delay involves 2-input NAND ring oscillators and average wirelengths

Our goal is to determine the maximum size of a block that could be reliably designed using current design flows without encountering significant DSM effects. After experimentation we arrive at a block size of approximately 50,000 gates although other block sizes will be discussed. In a block of this size, wirelengths are typically small and line resistance is much less than the effective resistance of the MOSFET drivers. Therefore, we ignore the impact of wire resistance in these calculations. We consider a single gate with a fixed fan-out in 2 processes, with a scaling factor of S . By scaling both channel length and width by S , we get a quadratic reduction in module area assuming that wiring pitch is also reduced by S [19,20]. We assume that average wirelength is a fixed fraction of the module side length so that L_{avg} scales by S . Since I_{dsat} is relatively constant in DSM processes, a reduction in channel width of S results in a proportional reduction in I_{dsat} .

Voltage swing, as discussed above, is also shrinking due to power and reliability concerns. From table 1, we estimate the scaling factor for V_{DD} as 0.75. Finally, we assume that the load capacitance is made up of two components, interconnect and fan-out gate capacitance. Interconnect capacitance, C_{wire} , is found by multiplying the average wirelength by the capacitance per unit length. From 2-D simulations that incorporate low-k dielectrics (table 2), we find the capacitance per unit length to be dropping by about 15% per generation, or a scaling factor of 0.85 [24]. Gate capacitance can be evaluated by estimating that T_{ox} is shrinking by 0.8X per generation (from table 1). We set interconnect capacitance to be twice as large as the fan-out capacitance based on 0.25 μm calculations ($FO=2$). Note that we are neglecting device junction and overlap capacitances in this analysis.

Normalizing $CV/2I$ to 1 for the original process, the shrunken process yields a $CV/2I$ value of 0.65. This represents a stage delay improvement of 35% from one generation to the next despite the smaller current drive. Translating to clock frequencies, these results predict a 54% increase for a generic process shrink. This analysis brings up several important points. First, it seems possible to maintain rising performance levels while scaling channel widths even in the era of velocity saturation. Second, the voltage and current components of $CV/2I$ cancel each other to some extent, leaving the bulk of the delay improvement to reduction of load capacitance. This is important because it highlights the most vital issues in DSM. In order to keep on pace with performance projections, wirelengths need to be reduced, low-k dielectrics need to be introduced, and device capacitances need to drop. These are among the most important factors contributing to faster designs.

Table 4 shows detailed results from this first-order analysis using actual scaling numbers taken from our strawman technology. Gate capacitance is determined using $W/L = 20$ and a fan-out of 2. As mentioned in the previous discussion, the sharp decrease in load capacitance ($C_{gate} + C_{wire}$) compensates for a slow rise in V_{DD}/I_{dsat} to yield overall speed improvements. A comparison is made to gate delay predictions in [25], which uses 3-input NAND's and several

Process (μm)	C_{gate} FO=2	C_{wire}	V_{DD}/I_{dsat}	Frequency	[25]
0.25	1	1	1	1	1
0.18	0.56	0.626	1.03	1.61	1.61
0.13	0.38	0.373	1.22	2.18	2.23
0.1	0.25	0.225	1.4	3.07	3
0.07	0.153	0.141	1.5	4.61	4.31
0.05	0.104	0.082	1.67	6.73 (5.77)	5.75

Table 4. First-order analysis results for delay scaling. 0.05 μm value in () is adjusted to match V_{DD} with [25].

empirical factors to account for device resistance and the use of dynamic logic. Our simple analysis gives very similar results regarding the scaling trends of gate delay in DSM.

7.1.3 Simulation approach

The previous approach is useful in determining trends and important parameters in delay scaling. However, it makes several assumptions that we will now remove. For example, the analysis ignored the impact of global wires in clock cycle determination. Since designs are getting larger, global wires become longer, which means that they will necessarily take up an increasing portion of the clock cycle. In addition, wiring resistance was ignored due to the assumption that device resistance dominated. Also, only the gate oxide component of device capacitance was treated. In this section, we will look more closely at the factors involved in determining ASIC performance using simulation tools.

We begin by characterizing a small ASIC library for each process (0.25 through 0.1 μm) consisting of 2-input NAND's with varying fan-outs. As mentioned earlier, BSIM3 device models are used in simulations. According to the procedure outlined in 7.1.1 we determine t_{gate} , optimal device sizing ($W_n=W_p$ in 2-input NAND's), and t_{wire} . After finding optimal device sizes at the 0.25 μm generation, we use this W/L ratio for all subsequent processes to allow for better comparisons.

Figure 5 shows plots resulting from device sizing optimization for 0.25 and 0.1 μm with a fan-out of 2. It can be seen that while the gate delay is constant throughout, the total stage delay decreases appreciably when increasing device sizes. In the limit, infinitely large devices will not even notice the presence of a relatively short wire, yielding a stage delay essentially equal to t_{gate} . We find that the optimal device size is around $W/L = 20$. At this size, interconnect represents 39 and 26% of the total delay at 0.25 and 0.1 μm respectively. The fact that interconnect delay is actually decreasing is somewhat surprising. The primary reasons behind this conclusion are the presence of shorter average wires and new materials.

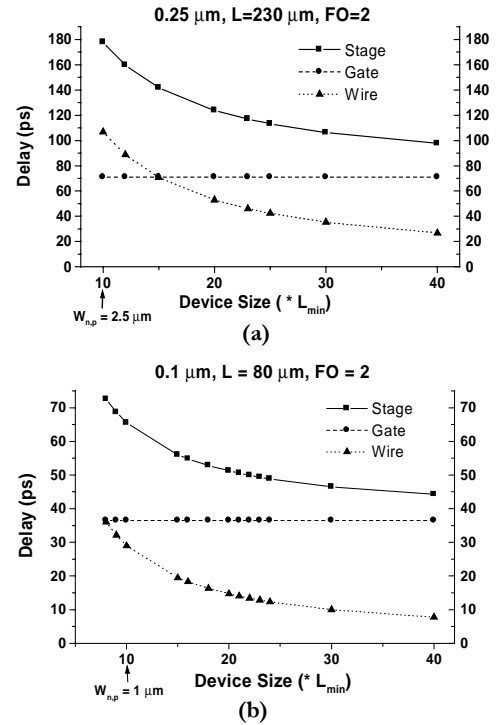


Figure 5. (a) Device size determination for 0.25 μm , FO = 2 (b) Same plot for 0.1 μm illustrates drop in interconnect delay

Previous studies reporting rises in interconnect delay have tended to focus on a fixed line length. Due to shrinking gate pitches, local wirelengths are expected to shrink with process scaling. As demonstrated in table 4 we forecast a decrease in average wiring capacitance by a factor of 12 from 0.25 to 0.05 μm due to shorter lines and low-k dielectrics. This point underlies our conclusion that interconnect delay will not dominate within 50K gate modules of future designs.

However, the scaling down of wirelength does not hold at the global level. In contrast, these signals must necessarily get longer in order to ensure connectivity in larger chips. Due to the use of functional clustering (keeping modules that need to communicate often near one another), extremely long global wires ($L \geq$ chip edge length) can be minimized and excluded from critical paths. Recent study has suggested that typical global wires might be closer to half the chip edge length [26]. We use this as an approximate starting point for a global wire in 0.25 μm ($L = 1$ cm) and scale it up by 15% for each process shrink. Simulation results show that at a constant buffer width (fixed I_{dsat}), delay decreases from process to process mainly due to the drop in voltage swing. Low-k dielectrics cancel out the 15% expected increase in wirelength so that delay varies roughly with V_{DD} according to $CV/2I$.

A comparison between the approaches in 7.1.2 and 7.1.3 is deferred until the impact of noise on delay can be considered. We will also extend our model from the delay of a single stage to the delay of an entire critical path.

7.2 Noise

The impact of noise on system performance was described in section 4.2.2. In this part of the analysis, we modify the results of section 7.1.3 to include delay deterioration effects and also discuss the impact of crosstalk noise on DSM designs. The general result of delay deterioration is increased stage delays due to higher effective capacitance and larger power dissipation due to bigger drivers. Also, the portion of total delay attributable to interconnect is increased. Signal integrity is viewed as a reliability problem. It places limitations on the routability of signals on lower-level, minimum-pitch wiring. Increasing wiring pitch to accommodate crosstalk reduces layout densities, which can be a problem in wire-limited designs.

7.2.1 Critical path

We begin by creating a generic critical path model that will allow us to track ASIC clock frequencies through scaling. From empirical design data, we have determined that in 0.35 μm ASIC designs a critical path typically consists of about 14 stages. While we use a 2-input NAND as the gate type for each stage, we allow for different fan-out conditions, which are determined from fan-out distributions of the design data. Our model is broken down into blocks of 8, 3, 2, and 1 stages with fan-outs of 1, 2, 3, and 4 respectively. In addition, we include a global wire routed on the top metal layer, which is buffered to reduce delay. Delay deterioration effects are not considered on the global wire and a fixed buffer size is assumed throughout scaling. Finally, 10% timing overhead is allotted for the critical path due to clock skew, process variation (both device and interconnect), and other phenomenon. This modeled critical path closely reflects the characteristics of a typical path in a design.

7.2.2 Results

The noise analysis uses worst-case neighboring wire switching activity to determine new delay values. The worst-case occurs when two adjacent wires simultaneously switch in the opposite direction as the victim line. Since there is no wiring component to t_{gates} , the in-

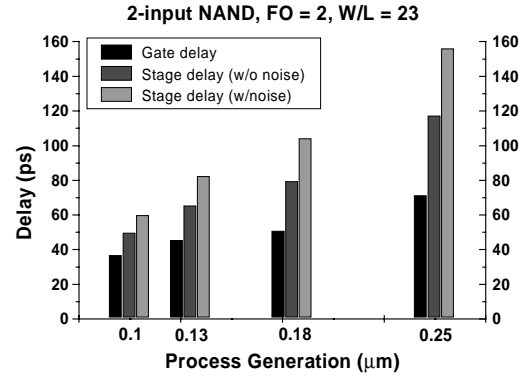


Figure 6. Evolution of stage delay relative to a fixed gate delay with and without noise considerations

trinsic gate delay will remain the same. However, as anticipated the larger effective interconnect capacitance due to Miller effect results in larger t_{wire} values. Results are shown in figure 6, which illustrates the relationship between gate delay and stage delay for various processes. Clearly the presence of delay deterioration increases the portion of total delay attributable to interconnect. However, we still foresee a drop in the ratio of interconnect delay to total delay even considering noise effects. For example, considering noise effects t_{wire} comprises 55% of the total delay at 0.25 μm but only 39% at 0.1 μm . These values are far from the 80% forecasts commonly reported and reflect the impact of shrinking average wirelengths on a chip.

In general, delay deterioration yields approximately an 80% increase in t_{wire} . This number corresponds very closely with the contribution of coupling capacitance to total line capacitance, which is about 75% throughout scaling. To compensate for the larger effective capacitance the optimal driver size is increased. These larger devices contribute to enhanced power dissipation and may also reduce layout density if drivers are made large enough. For a fixed optimal device size of $W/L = 23$, we incorporate the new delay numbers into the critical path model to determine expected ASIC clock frequencies in DSM. Global wire delay is taken from simulations with a 2-stage buffering system and a fixed stage width of 100 μm . Results are presented in table 5 and demonstrate similar trends as [1] with a 100-150 MHz performance increase. Isolating the logic delay component of the critical path, we compare its evolution normalized to 0.25 μm with results from table 4. We find that the inverse of logic delay scales as 1.5, 1.94, and 2.63 here compared to the analytical results of 1.61, 2.18, and 3.07. The discrepancies are due mainly to the inclusion of junction and overlap capacitances as well as delay deterioration effects. Nonetheless, the first-order analysis of section 7.1.2 yields qualitatively correct results that are within 10 to 15% of the more rigorous approach of this section.

Process (μm)	Logic Delay (ps)	Buffer Delay (ps)	Clock Skew/ Process Variation (ps)	Frequency (MHz)	NTRS Frequency (MHz)
0.25	1975	205	242	413	300
0.18	1320	150	163	612	500
0.13	1035	120	128	780	700
0.10	750	110	96	1045	900

Table 5. ASIC performance predictions including delay deterioration effects

7.2.3 Crosstalk

Crosstalk noise, or signal integrity, can be considered a reliability issue. Problems that can be caused by crosstalk include functional errors due to false switching and enhanced device stress when bootstrapping occurs ($V_{ds} > V_{DD}$). Crosstalk is of the utmost concern when using dynamic logic families such as domino [27]. As a reliability problem, the most straightforward way to deal with signal integrity is the generation of accurate design rules. For instance, bounds may be set on the amount of tolerable crosstalk noise (e.g. 20% of V_{DD}). From this constraint, analytical models [28] can be used to define a critical line length for different metal layers and driver scenarios. This parameter, L_{noise} , is then compared to L_{delay} , which is defined as the maximum line length that can be used on a given metal layer before buffering becomes beneficial [29]. We have found that for typical driver conditions, crosstalk is a more severe restriction on routing in lower level metals than delay (i.e. $L_{noise} < L_{delay}$). In addition, L_{noise} is also typically smaller than the dimensions of a 50K gate module within which it is desirable to connect the majority of gates with metals 1 and 2. The conclusions of this discussion are that noise can be a limiting factor in routing at the lower metal layers, which may lead to a loss in routing density due to possible increases in pitch, shielding wires, or the need to route in higher layers.

7.3 Power

7.3.1 Importance of power

Low-power designs, especially microprocessors, have received a large amount of attention recently as portable and wireless applications gain marketshare. Also, even in the highest performance designs power has become an issue since the extremely high frequencies being attained (close to 1 GHz) can easily lead to power dissipation in the many tens of watts. Dissipation of this amount of power requires heat sinks, resulting in higher costs and potential reliability problems. In this section, we discuss the reasons why power has become a significant issue and describe the 3 types of power consumption and how they can be expected to scale with CMOS processes.

In high-performance ASIC's there are three main reasons why power dissipation is rising. First, the presence of larger numbers of devices and wires integrated on a larger chip results in an overall increase in the total capacitance found on a design. Second, the drive for higher performance leads to increasing clock frequencies and dynamic power is directly proportional to the rate of charging capacitances (in other words, the clock frequency). Finally, the use of scaled voltages to improve reliability, decrease delay, and ironically, drop power consumption, leads to an increase in leakage current. While decreasing supply voltages does result in significant power savings overall, the static, or standby, current increases which can be detrimental to low-activity designs that require very little standby power consumption. An excellent overview of power issues in CMOS ULSI circuits is given in [30] along with a discussion of power modeling techniques in [31].

7.3.2 Dynamic power

Dynamic power consumption occurs as a result of charging capacitive loads at the output of gates. These capacitive loads are in the form of wiring capacitance, junction capacitance, and the input (gate) capacitance of fan-out gates. The expression for dynamic power was given previously and trends for several of the variables have already been discussed. Switching activity is a difficult parameter to estimate although it can frequently be approximated in the 0.1 to 0.2 range with reasonable accuracy.

In order to determine the impact of CMOS scaling on dynamic power consumption, we develop a simplified model of a 50K gate module which may exist in future ASIC's. Given such information

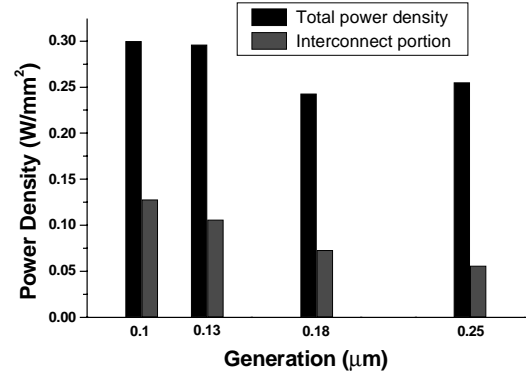


Figure 7. Evolution of dynamic power density with scaling

as packing density (devices/cm²), wiring pitches, average device size (taken from section 7.2.2), and routing density (metal occupancy), we calculate the dynamic power density through process scaling. We do this by first estimating the size of a module, then calculating the interconnect and device components of the total capacitance. Figure 7 summarizes the results and shows that power density is not increasing appreciably despite the rise in clock frequency. This analysis implies that dynamic power dissipation will increase approximately proportionally to the chip area. It should be noted that power dissipation in the clock network, off-chip drivers, and memory blocks are excluded from this analysis of a simple standard cell module.

7.3.3 Static power

The dominance of CMOS in modern circuit design is due in large part to its lack of static power consumption. This perceived benefit is becoming less true as voltages are scaled in order to limit dynamic power. Ideally, when the gate voltage of a MOSFET is below V_t there is negligible conduction. However, a small amount of leakage current flows at these conditions due to the inability of the gate to completely turn off the conducting channel. A good approximation of the amount of static current is given by ($T = 50^\circ \text{C}$) [4]:

$$I_{static} = 10 \frac{\mu A}{\mu m} \cdot W \cdot 10^{\frac{-V_t}{95mV}} \quad (1)$$

It is seen that leakage current is an exponential function of threshold voltage. The need for scaling V_t to maintain current drive has been discussed and results in a marked rise in leakage current as we move into DSM. Leakage currents in the range of nA/μm become serious when considering the large integration levels in ULSI. Static power consumption is given by $P_{static} = I_{static} V_{DD}$. Table 6 calculates the leakage power density for a 50K gate module. There is a 2500X increase from 0.25 to 0.1 μm, demonstrating that leakage power is becoming a larger component of total power consumption which should not be ignored. The rapid rise in P_{static} calls for the use of multiple- V_t processes or novel circuit techniques to limit standby power consumption. Finally, the exponential relationship between static power and V_t is important since variation in V_t can be significant. Devices exhibiting V_t 's at the lower tolerance

Process (μm)	V _{DD} (V)	V _t (V)	W _{device} (μm)	I _{static} (μA / block)	Block area (mm²)	P _{static} (mW / mm²)
0.25	2.5	0.625	5.75	3.03	2.5	0.003
0.18	1.8	0.450	4.1	151.8	1.43	0.191
0.13	1.5	0.375	3	677.3	0.83	1.22
0.1	1.2	0.3	2.3	3198	0.5	7.67

Table 6. Scaling of static power consumption within a 50K gate module

limit of a process will exhibit considerably more leakage current than a nominal device. In the same manner, static current is strongly dependent on temperature, with high operating temperatures resulting in significantly worsened MOSFET subthreshold characteristics. Poor control of either V_t or operating temperature may lead to wild fluctuations in static power consumption.

7.3.4 Short-circuit power

The final component of power dissipation is short-circuit power. Finite rise and fall times at the input of gates means that both the pull-up and pull-down networks of a CMOS gate are conducting simultaneously for a short period of time. During this time, current is flowing between V_{DD} and ground, resulting in short-circuit power dissipation. Research in this area has demonstrated that well-designed circuits exhibit short-circuit power that is less than 20% of the dynamic component, with 5 to 10% a more typical value [32]. Well-designed circuits strive to maintain reasonable input and output rise times so that short-circuit current cannot flow for an appreciable amount of time. In summary, short-circuit power dissipation is a manageable portion of the total power budget and can be approximated as 10% of the dynamic power consumption.

8. Summary and Methodological Implication

Ultimately we are concerned with the impact of DSM effects on future design methodologies. Our results indicate that interconnect delay will be small (<25%) in blocks of 50K gates and will remain reasonable (<40%) even when pessimistic noise considerations are introduced. These results presume that *lines are adequately driven to compensate for capacitive loads and noise effects*. Thus, design flows shown in figures 1 and 2 may be used in blocks of 50K gates if they can ensure accurate timing and noise analysis and sufficient cell sizing. It appears that blocks of 100K gates will also be manageable although power dissipation penalties will begin to be significant at that point. Beyond 100K gates, size of these blocks is limited by several factors dealing with interconnect. For instance, by making modules too large wirelengths increase due to connectivity requirements within the block. Longer wires translate to larger devices to drive the wires, sacrificing area and power. As a result we envision that future integrated circuits will be implemented hierarchically with large macro-blocks of approximately 50K to 100K gates. Significant functionality, such as a 32-bit microprocessor, can be implemented in such a block. In designs with 10^7 logic gates, this translates into a well-defined layout with 100 to 200 modules in the core area.

9. Future Work

Parts of this work form the basis for the new Berkeley Advanced Chip Performance Calculator (BACPAC). This system-level performance model improves upon previous work in [22,33] by incorporating enhanced analytical models for delay, noise, power, and area while also addressing the rising system-on-a-chip hierarchy of future ASIC's and microprocessors. The model is available at: <http://www-device.eecs.berkeley.edu/~dennis/BACPAC>

10. Acknowledgements

The authors gratefully acknowledge the design data supplied by Andres Teene of Symbios Corporation. Discussions of technology forecasting with Phil Fisher, Andrew Kahng, Steve Trimberger, and the Berkeley Nexsis team were very helpful. In particular, Amit Mehrotra and Sunil Khatri helped develop the strawman technology.

References

- [1] Semiconductor Industry Association, *National Technology Roadmap for Semiconductors*, 1997.
- [2] S. Devadas, A. Ghosh, and K. Keutzer, *Logic Synthesis*, McGraw-Hill, 1994.
- [3] N. Sherwani, *Algorithms for VLSI Physical Design Automation*, Kluwer, 1995.
- [4] C. Hu, "Device and Technology Impact on Low Power Electronics," in *Low Power Design Methodologies*, ed. Jan Rabaey, Kluwer, pp. 21-35, 1996.
- [5] D. Edelstein *et al.*, "Full copper wiring in a sub-0.25 μm CMOS ULSI technology," *Proc. of IEDM*, pp. 773-6, 1997.
- [6] S. Venkatesan *et al.*, "A high-performance 1.8V, 0.2- μm CMOS technology with copper metallization," *Proc. of IEDM*, pp. 769-72, 1997.
- [7] L. Su, *et al.*, "A high-performance 0.08 μm CMOS," *Proc. of VLSI Symposium on Technology*, pp. 12-13, 1996.
- [8] M. Rodder, *et al.*, "A 0.1 μm gate length CMOS technology with 30A gate dielectric for 1-1.5V applications," *Proc. of IEDM*, pp. 223-226, 1997.
- [9] K. Rahmat, O.S. Nakagawa, S-Y. Oh, and J. Moll, "A scaling scheme for interconnect in deep submicron processes," *Proc. of IEDM*, pp. 245-8, 1995.
- [10] M. Miyamoto, T. Takeda, and T. Furusawa, "High-speed and low-power interconnect technology for sub-quarter-micron ASIC's," *IEEE Transactions on Electron Devices*, pp. 250-256, Feb. 1997.
- [11] E.M. Zielinski, *et al.*, "Damascene integration of copper and ultra-low-k xerogel for high performance interconnects," *Proc. of IEDM*, pp. 936-938, 1997.
- [12] D. Sylvester, C. Hu, O.S. Nakagawa, and S-Y. Oh, "Interconnect scaling: signal integrity and performance in future high-speed CMOS designs," *Proc. of VLSI Symposium on Technology*, pp. 42-3, 1998.
- [13] F. Dartu, and L. Pileggi, "Calculating worst-case gate delays due to dominant capacitance coupling," *Proc. of DAC*, pp. 46-51, 1997.
- [14] G. Yee, R. Chandra, V. Ganesan, and C. Sechen, "Wire delay in the presence of crosstalk," *Proc. of TAU*, pp. 170-175, 1997.
- [15] C. Hu, "Gate oxide scaling limits and projection," *Proc. of IEDM*, pp. 319-322, 1996.
- [16] N. Rohrer, *et al.*, "A 480MHz RISC microprocessor in a 0.12 micron Leff CMOS technology with copper interconnects," *Proc. of ISSCC*, pp. 240-1, 1998.
- [17] J. Montanaro, *et al.*, "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor," *IEEE Journal of Solid-State Circuits*, pp. 1703-1714, Nov. 1996.
- [18] BSIM3 version 3.1, user's manual, UC-Berkeley, 1997.
- [19] R. Payne, "Metal pitch effects in deep submicron IC design," *Electronic Engineering*, pp. 45-7, Jul. 1996.
- [20] T.R. Bednar, R.A. Piro, D.W. Stout, L. Wissel, and P.S. Zuchowski, "Technology-migratable ASIC library design," *IBM Journal of Research and Development*, pp. 377-385, Jul. 1996.
- [21] S-P. Jeng *et al.*, "Implementation of low-dielectric constant materials for ULS circuit performance improvement," *Proc. of Symposium on VLSI Technology, Systems, and Applications*, pp. 164-168, 1995.
- [22] G.A. Sai-Halasz, "Performance trends in high-performance processors," *Proc. of the IEEE*, pp. 20-36, Jan. 1995.
- [23] A. Deutsch, *et al.*, "Modeling and characterization of long on-chip interconnections for high-performance microprocessors," *IBM Journal of Research and Development*, pp. 547-567, Sept. 1995.
- [24] RAPHAEL user's manual, version 4.0, TMA, 1997.
- [25] P. Fisher and R. Nesbitt, "The test of time: Clock cycle estimation and test challenges for future microprocessors," *IEEE Circuits and Devices Magazine*, pp. 37-44, Mar. 1998.
- [26] P. Zarkesh-Ha, J.D. Meindl, "Stochastic net length distributions for global interconnects in a heterogeneous system-on-a-chip," *Proc. of VLSI Symposium on Technology*, pp. 44-5, 1998.
- [27] D.A. Carlson, R.W. Castolino, and R.O. Mueller, "Multimedia extensions for a 550-MHz RISC microprocessor," *IEEE Journal of Solid-State Circuits*, pp. 1618-1624, Nov. 1997.
- [28] O.S. Nakagawa, D. Sylvester, J.G. McBride, and S-Y. Oh, "Closed-form modeling of on-chip crosstalk noise in deep-submicron ULSI interconnect," *Hewlett-Packard Journal*, pp. 39-45, Aug. 1998.
- [29] R. Otten, "Global wires: harmful?," *Proc. of ISPD*, pp. 104-109, 1998.
- [30] A.P. Chandrakasan, S. Sheng, and R.W. Broderson, "Low-power CMOS digital design," *Proc. of the IEEE*, pp. 473-484, Apr. 1992.
- [31] D. Liu and C. Svensson, "Power consumption estimation in CMOS VLSI chips," *IEEE Journal of Solid-State Circuits*, pp. 663-670, Jun. 1994.
- [32] H.J.M. Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," *IEEE Journal of Solid-State Circuits*, pp. 468-473, Aug. 1984.
- [33] H.B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addison-Wesley, 1990.