Interconnect Limits on Gigascale Integration (GSI) in the 21st Century

JEFFREY A. DAVIS, RAGURAMAN VENKATESAN, ALAIN KALOYEROS,
MICHAEL BEYLANSKY, SHUKRI J. SOURI, KAUSTAV BANERJEE, MEMBER, IEEE,
KRISHNA C. SARASWAT, FELLOW, IEEE, ARIFUR RAHMAN, MEMBER, IEEE,
RAFAEL REIF, FELLOW, IEEE, AND JAMES D. MEINDL, FELLOW, IEEE

Invited Paper

Twenty-first century opportunities for GSI will be governed in part by a hierarchy of physical limits on interconnects whose levels are codified as fundamental, material, device, circuit, and system. Fundamental limits are derived from the basic axioms of electromagnetic, communication, and thermodynamic theories, which immutably restrict interconnect performance, energy dissipation, and noise reduction. At the material level, the conductor resistivity increases substantially in sub-50-nm technology due to scattering mechanisms that are controlled by quantum mechanical phenomena and structural/morphological effects. At the device and circuit level, interconnect scaling significantly increases interconnect crosstalk and latency. Reverse scaling of global interconnects causes inductance to influence on-chip interconnect transients such that even with ideal return paths, mutual inductance increases crosstalk by up to 60% over that predicted by conventional RC models. At the system level, the number of metal levels explodes for highly connected 2-D logic megacells that double in size every two years such that by 2014 the number is significantly larger than ITRS projections. This result emphasizes that changes in design, technology, and architecture are needed to cope with the onslaught of wiring demands. One potential solution is 3-D integration of transistors, which is expected to significantly improve interconnect performance. Increasing the number of active layers, including the use of separate layers for repeaters, and optimizing the wiring network, yields an improvement in interconnect performance of up to 145% at the 50-nm node.

Manuscript received February 15, 2000; revised October 1, 2000.

J. A. Davis, R. Venkatesan, and J. D. Meindl are with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

A. Kaloyeros and M. Beylansky are with the Center for Advanced Thin Film Technology and Department of Physics, The State University of New York (SUNY) at Albany, Albany, NY USA.

S. J. Souri, K. Banerjee, and K. C. Saraswat are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9505 USA.

A. Rahman and R. Reif are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Publisher Item Identifier S 0018-9219(01)02068-0.

***Keywords*—*Crosstalk, epitaxial growth, interconnections, modeling, scattering, technology forecasting, thin films, thin-film transistors, transmission lines, wafer bonding, wiring.*

I. Introduction

The International Technology Roadmap for Semiconductors (ITRS) projects that by 2011 over one billion transistors will be integrated into a single monolithic die [1]. The wiring system of this billion-transistor die will deliver power to each transistor, provide a low-skew synchronizing clock to latches and dynamic circuits, and distribute data and control signals throughout the chip. The resulting design and modeling complexity of this GSI multilevel interconnect network is enormous such that over 10^{17} coupling inductances and capacitances throughout a nine-to-ten-level metal stack must be managed. A seminal paper [2] focuses on the transistor limits for a GSI system; therefore, this paper will address the limits that on-chip interconnects place on a GSI system design in the 21st century.

Interconnect limits potentially threaten to decelerate or halt the historical progression of the semiconductor industry because the miniaturization of interconnects, unlike transistors, does not enhance their performance. Scaling transistors to the nanometer regime is plagued with many challenges, such as drain-induced-barrier lowering (DIBL), quantum mechanical gate tunneling, mobility degradation, and reliability problems due to random placement of dopant atoms in a host silicon lattice [1], but once overcome MOSFET channel scaling will enhance intrinsic gate delay [1]. For instance, scaling MOSFET channel length from 1000 to 100 nm to 35 nm dramatically reduces the intrinsic MOSFET switching time as seen in Table 1. Scaling interconnects into the

0018–9219/01$10.00 © 2001 IEEE

Table 1
Interconnect and Transistor Scaling Properties

Technology	MOSFET switching delay (t_d=CV/I)	Intrinsic delay of ***minimum scaled*** 1mm interconnect	Intrinsic delay of ***reverse-scaled*** 1mm interconnect	W/F
1.0μm(Al, SiO_2)	~20ps	~5ps	~5ps	1
0.1μm(Al, SiO_2)	~5ps	~30ps	~5ps	~ 1.5
35nm(Cu, low k)	~2.5ps	~250ps	~5ps	~ 4.5

nanometer regime is also plagued with many challenges, such as resistivity degradation, material integration issues, high-aspect ratio via and wire coverage, planarity control, and reliability problems due to electrical, thermal, and mechanical stresses in a multilevel wire stack [1], and once these challenges are overcome, minimum interconnect scaling will still degrade interconnect delay. For example, Table 1 also illustrates that the intrinsic interconnect delay of a 1-mm length interconnect at the 35-nm technology node overwhelms the transistor delay by two orders of magnitude.

A potential solution to this interconnect dilemma is to reverse scale longer semiglobal and global interconnects such that they have "fat" cross-sectional dimensions [3], [4]. This strategy enhances interconnect performance, but at the expense of wire density. For example, to balance the interconnect delay of a 1-mm interconnect length with the transistor switching delay, the wire size at the 35-nm generation must be almost five times larger than the minimum lithographic size as seen in Table 1. Because die area is directly related to cost, the area penalties of the reverse scaled strategies could hinder the exponential reduction in cost per function that has propelled semiconductor technology over the past several decades.

The central thesis of this paper is that in the 21st century opportunities for GSI will be governed in part by a hierarchy of physical limits on interconnects whose levels are codified as fundamental, material, device, circuit, and system [2], [6]. In Section II, fundamental limits are derived from the basic axioms of electromagnetic, communication, and thermodynamic theories. In Section III, material limits are determined by the transformation of bulk properties of metallic interconnects as they are scaled into the nanometer regime. In Section IV, device limits deal directly with the problems of interconnect miniaturization and provide a rationale for reverse-scaling strategies. New metrics for crosstalk with and without on-chip inductive effects are presented. At the circuit level in Section V, the impact of transistor driver output resistance on interconnect performance and crosstalk is investigated. Finally, in Section VI, system limits imposed by reverse-scaled multilevel interconnect networks are investigated using a compact wire-length distribution model to predict the wiring requirements of future GSI products. Wire area limits of reverse-scaled multilevel networks in a two-dimensional (2-D) planar transistor process are projected, and the opportunity for three-dimensional (3-D) integration of transistors is rigorously explored to help alleviate interconnect delay and density problems.

II. Fundamental Limits

This discourse on interconnect limits begins through examination of several of the most basic principles that govern the physical world. The limits discussed in this section are immutable and are unchanged through the use of advanced materials, sophisticated device structures, inventive circuit techniques, or novel instruction set architectures. These limits, therefore, are defined as fundamental and will irrevocably limit interconnect performance, energy dissipation, and signal integrity in the 21st century.

A. *Performance Limits*

The role of GSI global interconnects is to transmit binary switching events that are generated from constituent computational elements. The fundamental limit, therefore, on interconnect performance is set by the shortest delay between a binary switching event in a transmitter and a binary transition detected at a receiver. To determine the shortest possible delay, the communication channel connecting the transmitter to the receiver is assumed to be a perfect noise-free lossless interconnect.

The maximum transmission speed is limited by the speed of an electromagnetic wave propagating in free space and is a well-known quantity derived from Maxwell's equations [7]. Assuming that free space surrounds a lossless interconnect, then the Helmholz equations, which are derived from Maxwell's equations, describe the propagation of electric and magnetic fields. A key result obtained from the Helmholz equation is that the free-space wave propagation speed c_o is given by

$$c_o = \frac{1}{\sqrt{\mu_o \varepsilon_o}} \tag{1}$$

where μ_o and ε_o are, respectively, the permeability and the permittivity of free space. The latency τ in communicating a binary transition event from the transmitter to the receiver must be greater than

$$\tau \geq \frac{L}{c_o} \tag{2}$$

where L is the transmission distance.

This fundamental limit is clearly represented in the reciprocal length squared versus time delay plane as seen in Fig. 1 after [2]. The region to the left of the line with a slope of negative two in logarithmic scaling in this plane is a forbidden region of interconnect operation.

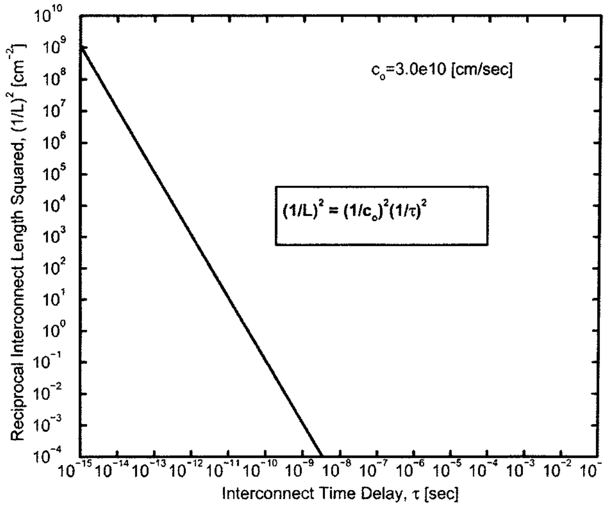


Fig. 1. Fundamental performance limit set by the electromagnetic propagation in free space.

B. Energy Limits

The second fundamental limit is based upon Shannon's communication theorem for the maximum capacity of a communication channel. The expression for the maximum capacity of a communication channel with a white Gaussian thermal noise source is given by [8]

$$C \leq B \log_2\left(1 + \frac{P_s}{kTB}\right) \tag{3}$$

where

- C maximum channel capacity measured in bits/s;
- P_s average signal power of the input;
- kTB Johnson thermal noise power delivered to a matched load [8];
- B bandwidth of the receiver;
- k Boltzmann's constant ($=1.38 \times 10^{-23}$ J/K);
- T temperature ($\sim$300 K) [8].

Assuming that the average energy per bit is $P_s/C = E_{\rm bit}$, then solving for $E_{\rm bit}$ in (3) gives

$$E_{\rm bit} = kT\,\frac{B}{C}\,(2^{C/B} - 1). \tag{4}$$

Setting the derivative of (6) equal to zero or $dE_{\rm bit}/d(C/B) = 0$ and employing L'Hospital's rule gives

$$E_{\rm bit}(C/B \rightarrow 0) = E_{\rm bit}(\min) = (\ln 2)kT. \tag{7}$$

Note that $(C/B) \rightarrow 0$ is tantamount to calculating the energy transfer of an infinitely long bit or a single binary transition. If the energy transferred during a binary transmission on an interconnect is less than $kT \ln 2$, then the binary transition cannot be differentiated from thermal noise regardless of advanced error-correcting encoding techniques.

This energy also sets a lower limit on low-swing interconnect buses. In the limit, the smallest swing of an interconnect bus is set by the quantization of charge. The minimum switching potential of a single electron interconnect is set at

$$V_{dd,\,\min} = \frac{2E}{q} = \frac{kT2\ln 2}{q} \approx 0.036\ [\mathrm{V}]. \tag{5}$$

C. Noise Limits

In digital circuits an important metric of a binary transition is its potential swing, and in the presence of thermal noise this potential is perturbed from its nominal value. The best metric for this perturbation is the standard deviation of thermal noise voltage across a resistor, which is derived by Nyquist [2] to be

$$\sigma_n = \sqrt{4kTBR} \tag{6}$$

where

- k Boltzmann's constant (1.38×10^{-23} J/K);
- T temperature (=300 K);
- B bandwidth of the receiver;
- R resistance of the interconnect load.

The most statistically significant deviation of the potential at the end of the line is defined by (6). The interconnect noise floor, therefore, is set by the thermal noise fluctuation across a load with a resistance equal to the characteristic impedance of free space. Assuming that $T_{\rm rise}$ is the reciprocal receiver bandwidth, this fundamental limit is

$$V_n = \left(\frac{\mu_o}{\varepsilon_o}\right)^{1/4}\sqrt{\frac{4kT}{T_{\rm rise}}}. \tag{7}$$

III. Material Limits

Device feature sizes are crossing a critical physical threshold below which the performance of extremely narrow interconnect lines is controlled primarily by: 1) the properties of their surfaces and interfaces, as driven by one- and two-dimensional scattering effects; and 2) the characteristics of their impurity and defect densities, as governed by the type and distribution of grain boundaries, dislocations, and junctions. This transition represents a major show stopper in the successful development of the material and process (M&P) technologies necessary to ensure maximum signal transmission in sub-50-nm device nodes through reduced resistance $\times$ capacitance (RC) time delay. In particular, the physics of resistivity behavior in extremely fine conductor lines represents a daunting and potentially insurmountable challenge that needs to be understood and resolved in order to ensure the extendibility of today's chip architecture below the 50-nm device node.

In this respect, the resistivity of thin-film conductors is given by [9], [10]

$$\rho(\text{thin film}) = \rho(\text{thermal}) + \rho(\text{extrinsic}) \tag{8}$$

where ρ(thermal) is the contribution due to electron–phonon "coupling" (i.e., electronic interactions with thermally induced lattice vibrations), and ρ(extrinsic) is the contribution

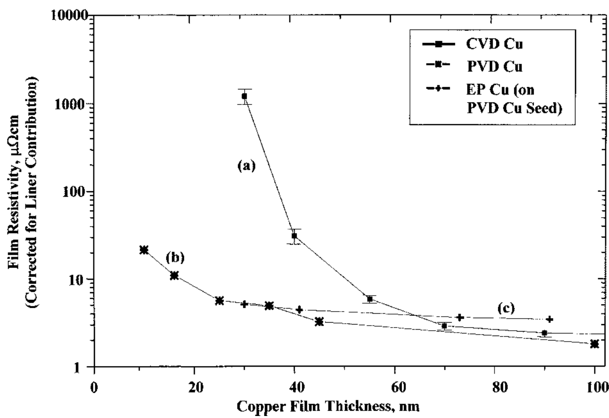


Fig. 2. Resistivity r_{Cu} (thin film) as function of thickness for blanket (unpatterned) polycrystalline Cu films by: (a) TCVD from the source precursor $\mathrm{Cu^{I}}$(hfac)(tmvs), (b) collimated sputtering, and (c) electrochemical deposition (ECD). Due to the significantly reduced Cu thickness investigated, experimental resistivity values were corrected for liner contributions by using the suitable approximation (from [4]): Cu(thin film) $= 4.53t[R_{\mathrm{eff}}R_{\mathrm{liner}}/(R_{\mathrm{liner}} - R_{\mathrm{eff}})]$ Where t, R_{eff}, and R_{liner} represent, respectively, the thickness of the Cu layer, the effective resistance of the Cu/liner stack, and the resistance of the liner.

from electron scattering by impurities, defects, grain boundaries, and film surface and interface, as given by

$$\rho(\text{extrinsic}) = \rho(\text{defect}) + \rho(\text{impurity}) + \rho(\text{grain boundary}) + \rho(\text{surface/interface}). \quad (9)$$

For illustration purposes, Fig. 2 plots the resistivity ρ_{Cu} (thin film) as function of thickness for blanket (unpatterned) polycrystalline copper thin films deposited on 9-nm-thick tantalum nitride (TaN_x) by [11]: 1) thermal chemical vapor deposition (TCVD) from the source precursor $\mathrm{Cu^{I}}$(hfac)(tmvs), where hfac = hexafluoroacetylacetonate and tmvs = trimethylvinylsilane; 2) collimated sputtering; and 3) electrochemical deposition (ECD). As expected, the total resistivity ρ_{Cu} (thin film) in all three cases was observed to increase with decreasing film thickness, with the rate of increase exhibiting significant dependence on the deposition technique due to morphological and textural differences between the corresponding three types of Cu films.

The increased resistivity with thickness reduction is attributed *in part* to surface roughness induced scattering effects [12], which are caused predominantly by the island-like morphology of polycrystalline Cu films, i.e., films where surface roughness is on the order of or larger than film thickness [13]. These effects tend to play an increasingly more pronounced role as the polycrystalline film becomes thinner. This trend is documented in Fig. 3, which displays the relative surface roughness (surface grain size), plotted as percent of film thickness, for TCVD-grown polycrystalline Cu films deposited on tantalum nitride and tungsten nitride [11]. In this study, surface grain size and associated root-mean-square surface roughness were determined by atomic force microscopy (AFM) and focused-ion-beam scanning electron microscopy (FIB-SEM).

More specifically, Fig. 4(a) displays the island-like surface morphology of a thinner, 35-nm-thick TCVD Cu film on TaN_x. In contrast, Fig. 4(b) shows an appreciably smoother surface morphology for a thicker, 60-nm-thick, TCVD Cu on the same liner material. The islands become increasingly discontinuous with further reduction in film thickness. Their boundaries act as progressively higher potential barriers, thus leading to a gradual rise in resistivity. Finally, below a critical thickness, a matrix of completely disconnected nuclei is formed, with the associated resistivity becoming infinite. The value of this critical thickness is strongly dependent on the mechanisms of Cu film nucleation and growth, as driven by the nature and characteristics of thin-film formation in CVD, sputtering, and ECD processing, and the surface chemistry, morphology, and texture of the underlying liner material.

Over the years, various theoretical treatments were only partially successful at modeling the dependence of resistivity on surface roughness for ultrathin metallic films [14]. In particular, Elsom *et al.* [15] developed a numerical model for the rise in resistivity as function of decreased thickness for Cu films with island-like morphology. Unfortunately, the model was limited to cases where the island size was larger than the bulk mean free path for electron scattering in Cu, a limitation that severely restricts the applicability of the model to sub-50-nm interconnect lines, as discussed below.

Elimination of surface roughness induced scattering effects requires the development of M&P solutions that combine the ability to "nanoengineer" film morphology and texture, with the implementation of predictive models using comprehensive theoretical treatments, to grow epitaxial Cu/liner interconnect stacks with atomically smooth surfaces and interfaces. These solutions include the identification of epitaxial "zero thickness" liner materials that are closely lattice-matched to Cu, and the development of atomically tailored, interfacially controlled processing methodology, such as atomic layer CVD technologies. They also involve the use of atomically engineered zero-thickness interfacial layers, such as surfactants, which act as a "wetting" layer that ensures the availability of a high density of surface nucleation sites and reduces the nucleation barrier to Cu formation. The desired outcome is to eliminate island-type morphology through the achievement of a Frank van der Merve, layer by layer, Cu growth [16].

For illustration purposes, Fig. 5 plots the resistivity as function of thickness for blanket polycrystalline Cu thin films on TaN_x and indium-seeded TaN_x. The two sets of Cu films were grown using identical processing conditions. The use of indium (In) as surfactant led to a significant reduction in total resistivity as compared to the case where no In was employed. This behavior is attributed to the role of the surfactant layer in reducing the activation barrier to Cu nucleation and growth, leading to films with appreciably smoother surface morphology, as documented in the FIB-SEM micrographs of Fig. 6. The selection of a surfactant must, however, satisfy a stringent set of requirements, including that its thickness must be restricted to a few monolayers in order to maximize space availability for the actual copper conductor. In addition, it must be

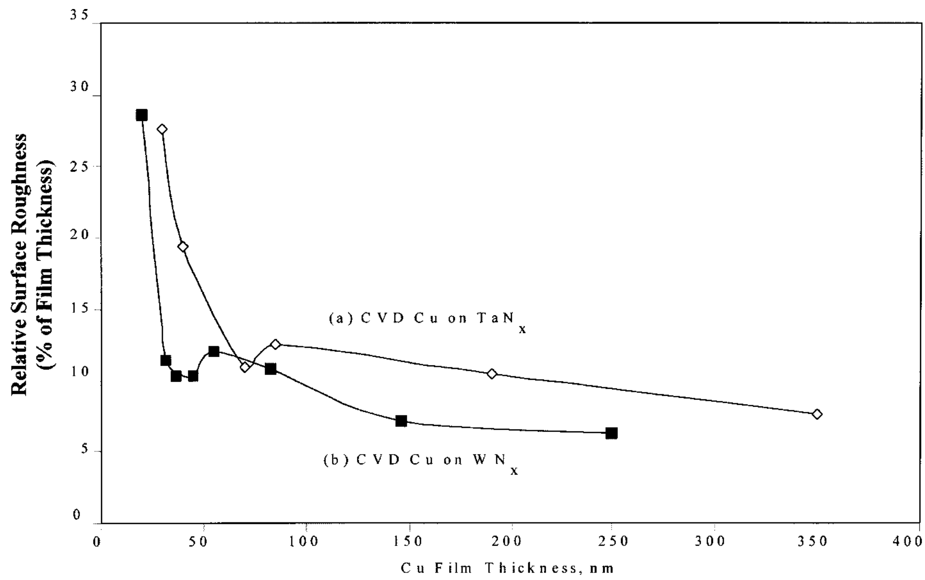


Fig. 3. Relative surface roughness (surface grain size), plotted as percent of film thickness, for TCVD-grown polycrystalline Cu films deposited on tantalum nitride and tungsten nitride.

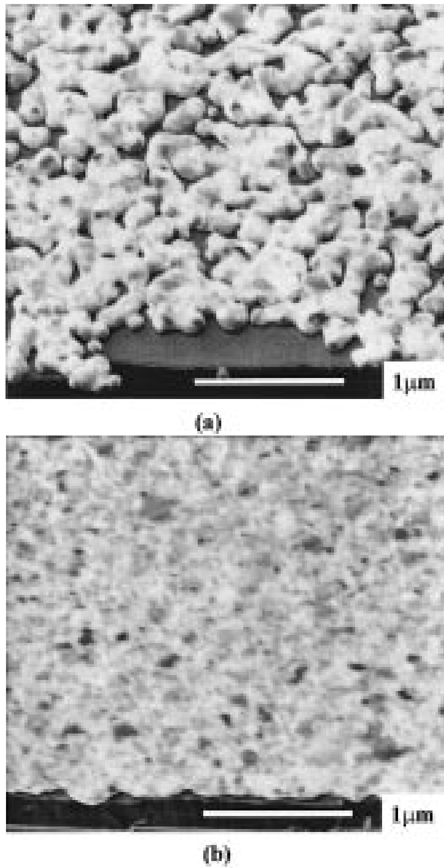


Fig. 4. FIB-SEM micrographs of the surface morphology of: (a) 35-nm-thick TCVD Cu films on TaN_x and (b) 60-nm-thick TCVD Cu on the same liner material.

mechanically, thermally, and structurally stable under typical semiconductor fabrication flows, and preferably retain its as-deposited chemical and compositional integrity. In the case a Cu alloy is formed, however, the inclusion of the surfactant material must be limited to extremely small concentrations within the Cu matrix and must not induce any unacceptable increase in the overall effective resistance of the resulting Cu alloy [17].

Apart from surface roughness induced scattering, the increased resistivity with thickness reduction is also caused by surface and interface induced scattering phenomena. The latter become predominant in films where the thickness t is on the order of or smaller than the bulk mean free path λ for electron scattering in the corresponding metal [18]. As can be seen in Table 2 [8], which displays the bulk mean free path for selected metals of interest, surface scattering effects are expected to become predominant in sub-50-nm Cu lines. Interestingly, the resulting rise in the overall resistivity of progressively narrower conducting lines could potentially produce equivalent conductivity characteristics in aluminum, tungsten, and copper-based interconnects. This possibility could have significant implications in terms of the selection of most appropriate material systems for gigascale metallization schemes.

A number of theoretical treatments have already been developed for the effects of grain boundary and surface scattering on thin-film resistivity [14], [15], [18]. Sambles combined key elements of these treatments, which are almost universally based on the semiclassical scattering model, into a comprehensive expression for the general case of a film with different roughness profiles at its surface and interface. In this expression, the ratio of bulk resistivity to thin-film resistivity is given by

$$[\rho(\text{bulk})/\rho(\text{thin film})] = G(\alpha) - F(p, p_s, p_i, \gamma). \quad (10)$$

The first term accounts for grain boundary scattering with

$$\alpha = (R_G.\lambda)/(1 - R_G)\langle D\rangle \quad (11)$$

where R_G is the grain boundary reflection coefficient and $\langle D\rangle$ is the average grain size. The second term is known as

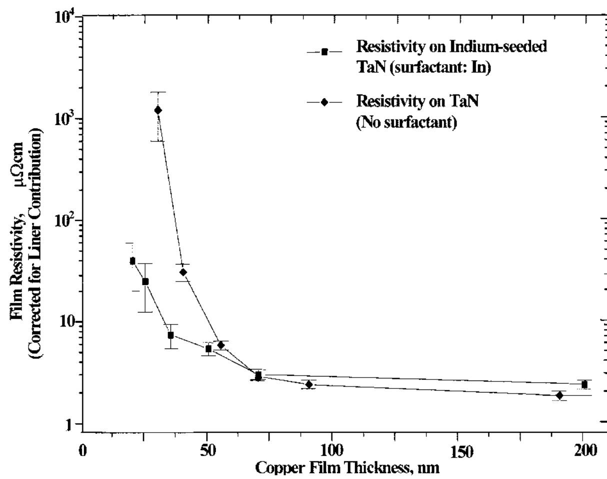
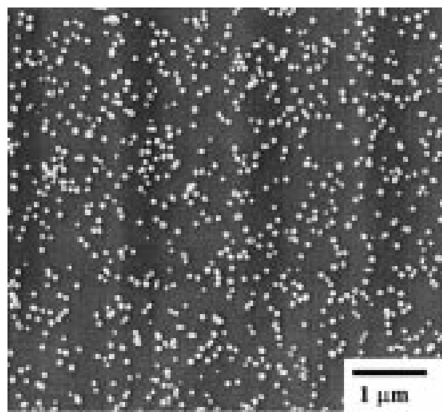


Fig. 5. Resistivity as function of thickness for blanket polycrystalline Cu thin films on: (a) TaN_x and (b) In-seeded TaN_x.



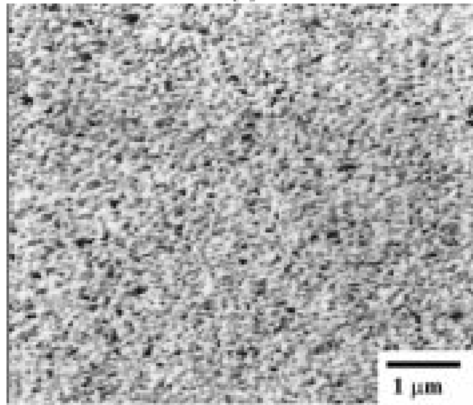


Fig. 6. FIB-SEM micrographs of the surface morphology of TCVD Cu films on: (a) TaN_x and (b) In-seeded TaN_x.

the Fuchs modified term and accounts for surface scattering effects. In this term, p is the probability for specular electron scattering from the film surface and interface, while p_s and p_i are the roughness profiles of, respectively, the surface and interface. The coefficient γ is the ratio of film thickness t to the mean free path λ. The model thus predicts that the reduction in surface scattering effects requires the development of M&P solutions that maximize specular electron scattering.

Table 2
Electron Mean Free Paths for Selected Metals of Interest (From [18])

Metal	Mean Free Path λ (nm)
Copper (Cu)	39.3
Aluminum (Al)	14.9
Tungsten (W)	14.2
Silver (Ag)	52.7
Gold (Au)	35.5

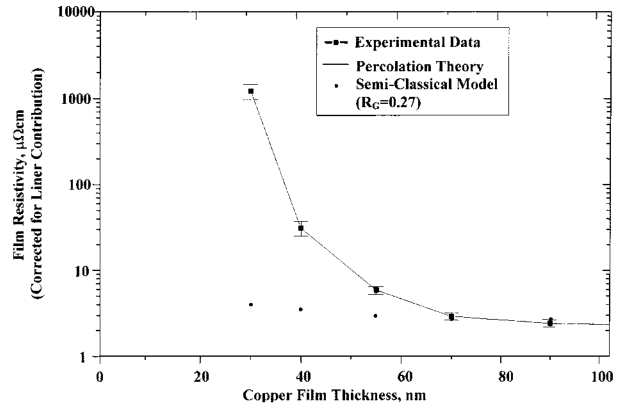


Fig. 7. Comparison of experimental Cu resistivity profiles versus thickness for TCVD Cu on TaN_x with predictions of the Sambles semiclassical scattering model and percolation theory.

The model was found to be in excellent agreement with experimental resistivity measurements for film thickness above 50 nm, as shown in Fig. 7. This agreement was achieved by using a grain boundary reflection coefficient R_G of 0.27. This value is low and implies that the TCVD Cu films are pure and dense, with the contribution to film resistivity from grain boundary induced scattering effects being minimal. The model was in serious disagreement with the experiment for film thickness below 50 nm. This discrepancy is expected and is attributed to the fact that a basic assumption in the derivation of the semiclassical model is that surface roughness is smaller than film thickness. Clearly, this assumption is not applicable to ultrathin Cu films, which are characterized by a more "island-like" morphology. As a result, percolation theory was successfully applied to model resistivity behavior in sub-50-nm Cu lines.

Percolation theory is a statistical theory that describes the properties of any given randomly assembled system near the point where it changes from a macroscopically disconnected to a connected one [19]. This point is called a percolation threshold and the overall system properties are expected to change drastically near this threshold. This approach is highly applicable to the case of ultrathin conductors,

especially in view of the random character of island formation and grain agglomeration that is typically observed in thin-film growth. Percolation theory describes the resistivity of such an ultrathin conductor system as a random resistor network. The obvious choice for a percolation threshold in this case is the critical thickness (t_c), above which the film becomes continuous. Film resistivity drops sharply near the critical thickness where Cu islands merge together and form a backbone for electron transport. Several system properties are found to obey the so-called scaling laws near the percolation threshold. In particular, the scaling law for system resistivity states that resistivity is proportional to the difference between the fraction of a substrate area covered by the film and the critical substrate area coverage at the percolation threshold [20].

It has been shown that in case of random nucleation area coverage is proportional to film thickness, so the final expression for film resistivity is given by

$$\rho(\text{thin film}) - \rho(\text{bulk}) \propto (t - t_c)^{-m} \tag{12}$$

where m is a critical exponent, which is equal to 1.3 for 2-D systems. As shown in Fig. 7, this value of the critical exponent yielded excellent fit with the experimental data. The fit yielded a value of 29.7 nm for the critical thickness t_c, which is highly consistent with experimental observations, thus providing additional proof to the accuracy of the percolation theory fit.

Subsequent theoretical modeling efforts will center on analytical and numerical calculations of surface scattering in finite-size topographies with emphasis on one-dimensional (1-D) to 2-D crossover effects. Resulting findings will be coupled to experimental resistivity measurements in ultranarrow conducting lines to establish baseline metrics for the dependence of 2-D grain boundary and surface scattering behavior on device feature size. The net projected outcome is the development and optimization of M&P solutions that can grow epitaxial Cu/liner interconnect stacks with atomically smooth surfaces and interfaces, while maximizing specular electron surface scattering in ultranarrow interconnect lines.

IV. Device Limits

Interconnect device limits in this section will probe the inherent attributes of wires free from the effects of transistors. To investigate interconnect device limits in the 21st century, basic interconnect structures are presented in this section to elucidate performance and noise limits on interconnects.

A. Performance Limits

1) Resistance and Capacitance (RC) Effects: Unlike the transistor, interconnect performance is not enhanced through miniaturization. This result is presented most succinctly using a distributed RC network to model a single global on-chip interconnect. The latency of this interconnect is given by the distributed RC time delay (assuming $rcL^2 > L/v$) as

$$\tau = rcL^2 \tag{13}$$

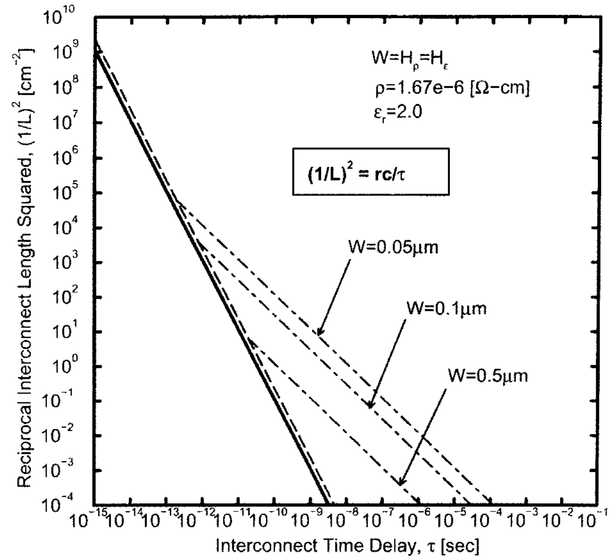


Fig. 8. Scaling effects on interconnect time delay limits.

where

- r distributed resistance per unit length;
- c distributed ground capacitance per unit length;
- L interconnect length;
- ν speed of electromagnetic wave propagation.

Using a simple parallel plate model for the parasitic capacitance per unit length of the interconnect, the interconnect delay in (13) becomes

$$\tau = \frac{\rho\varepsilon}{H_\varepsilon H_\rho} L^2 \tag{14}$$

where

- ρ resistivity of the conductor;
- ε permittivity of the insulator;
- H_ρ thickness of the metal conductor;
- H_ε dielectric thickness.

The interconnect latency metric in (14) clearly reveals the scaling properties of global interconnects. Ideal scaling of all wire dimensions, including length, results in no reduction in delay. Furthermore, because transistor numbers and die sizes are increasing with each new technology generation, global interconnect lengths are increasing, which results in significant interconnect performance degradation [3]. Scaling effects on interconnect latency have been rigorously investigated [2], [3] and are illustrated most effectively in the reciprocal length squared versus time delay plane seen in Fig. 8 after [2]. The diagonal in this plane is a locus of constant distributed RC product, and interconnect operation is forbidden to the left of each locus for interconnects with a smaller cross-sectional dimension $H_\rho = H_\varepsilon$. This plot reveals that reverse scaling of interconnects of global interconnect dimensions (H_ρ, H_ε) reduces interconnect latency [3], [4].

2) Inductance Effects: Reverse-scaling methodologies reduce RC delay, but at gigahertz clock frequencies reverse-scaling necessitates the inclusion of self-inductance in global signal interconnects, clock lines, and power distribution networks. Inductance introduces unique challenges for

each type of interconnect. For example, variations in return path currents for each leg of a balanced clock tree (BCT) network produces variation in interconnect delay and reflection characteristics [21]. Inductance in power distribution networks produces voltage transients that are dependent on the number of simultaneously switched devices. As representative of on-chip inductance issues, this analysis will concentrate on the influence of inductance on global signal interconnects. For global clock and signal interconnects, gigahertz chip designers must provide controlled current return paths to reduce on-chip inductive effects. To investigate aggressive interconnect limits, therefore, perfect return path currents are assumed in this paper using ideal ground planes.

Assuming negligible skin effect, the telegrapher's equation describes the transient voltage along a single interconnect. On-chip interconnect modeling is complicated by the fact that high-density global wires must include both inductance and resistance such that neither quantity is a perturbation to a well-known LC or RC solution. The complete solution to the telegrapher's equation, therefore, is succinctly and efficiently given by a series of modified Bessel functions in

$$V_{\text{fin}}(L,t) = 2V_{\text{gen}}(L,t,0) + 2\sum_{n=1}^{q}\sum_{i=0}^{n}\sum_{j=0}^{j_{\max}} \cdot \frac{n(n-1+j)!}{i!j!(n-i)!}(-1)^i\Gamma^{n-i+j} \cdot V_{\text{gen}}((2n+1)L,t,i+j) \quad (15)$$

where

$$V_{\text{gen}}(x,t,m) = V_{dd}e^{-(rt/2l)} \cdot \left(\frac{Z_o}{R_{tr}+Z_o}\left(\frac{t-x\sqrt{lc}}{t+x\sqrt{lc}}\right)^{m/2} I_m\left[\frac{r}{2l}\sqrt{t^2-\left(x\sqrt{lc}\right)^2}\right] + \frac{1}{2}\sum_{k=1}^{k_{\max}}\left(\frac{t-x\sqrt{lc}}{t+x\sqrt{lc}}\right)^{(1/2)(k+m)} \cdot I_{(k+m)}\left[\frac{r}{2l}\sqrt{t^2-\left(x\sqrt{lc}\right)^2}\right](4-\Gamma^{k-1}(\Gamma+1)^2) \right) \cdot u_o\left(t-x\sqrt{lc}\right) \quad (16)$$

where $I_i(-)$ is a ith-order modified Bessel function, L is the interconnect length, t is time, l, r, and c are the distributed inductance, resistance, and capacitance per unit length, respectively, Γ is the reflection coefficient at the source, q is the current reflection number given by $q = \langle 0.5(t/(x\sqrt{lc}) + 1.0)\rangle - 1.0$, the notation $\langle x \rangle$ is defined as the decimal truncation of x (i.e., $\langle 2.7 \rangle = 2$), and $j_{\max}$ and $k_{\max}$ are determined to obtain the desired accuracy of solution (in the limit they both go to infinity) [23].

Using a near wave-front approximation to (15) and a distributed RC model in [24], the 50% time delay of a single interconnect device with the inclusion of inductance can be approximated by

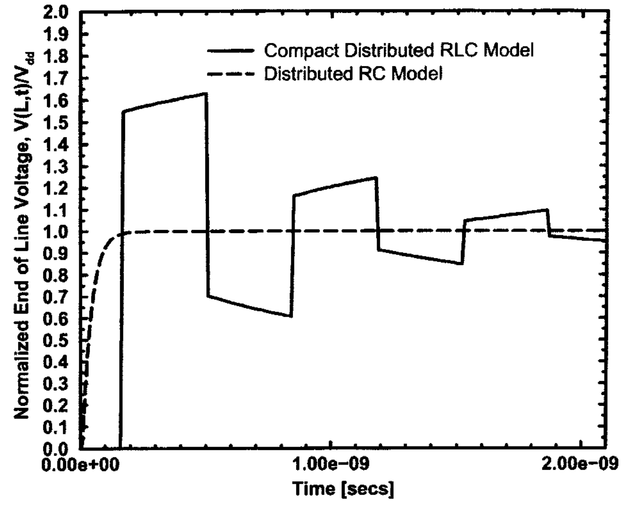


Fig. 9. Comparison of distributed rc and distributed rlc model for a global interconnect with $Z_o = 266.5$ W, $r = 37.87$ W/cm, $L = 3.6$ cm, $R_{tr} = 0$.

$$\tau = \frac{L}{\sqrt{lc}}u_o\left(2\sqrt{\frac{l}{c}}\ln(4) - rL\right) + (0.377rcL^2)u_o\left(rL - 2\sqrt{\frac{l}{c}}\ln(4)\right) \quad (17)$$

where $u_o(-)$ is a step function. Inductance effects for this interconnect become significant when

$$rL \leq 2.77Z_o. \quad (18)$$

The effect of inductance on a high-speed global interconnect is illustrated by comparing the transient response of an on-chip copper interconnect ($H_\rho = H_\varepsilon = W = 2$ μm) using distributed RC models with the compact distributed rlc model in (15). As seen in Fig. 9, the distributed RC model does not capture transient reflections and underestimates the time delay of this aggressive on-chip interconnect design. Moreover, significant overshoot at the end of this interconnect is not predicted with distributed RC models. Overshoot in this aggressively scaled interconnect in Fig. 9 is almost 70% higher than the supply voltage.

B. Crosstalk Limits

1) Resistance and Capacitance (RC) Effects: Even in high-speed GSI multilevel interconnect networks, distributed RC models are still needed to determine the transient behavior of local and semiglobal interconnects and, therefore, are used to investigate the limits on crosstalk for shorter high-speed interconnects. Local interconnects, which make up the majority of on-chip interconnects [25], will continue to scale to minimum feature size dimensions to maximize wire density. An existing distributed RC interconnect model with a step-response excitation voltage predicts that the peak crosstalk (at the load of the quiescent line), V_n, between the two parallel wires is length, scaling, and material independent for homogeneous dielectrics [24].

The finite switching time of a interconnect driver, however, must be considered to fully understand crosstalk limits for local and semiglobal interconnects. Using a ramp-response ideal voltage source driving an active line parallel to a quiescent line, the complete solution to this peak noise voltage at the load end of the quiescent line is given by [26]

$$\frac{V_{\text{peak}}}{V_{dd}} = 1.05\,\frac{16}{\pi^3}\,\frac{rc_m L^2}{T_{\text{rise}}}\left(\exp\left[\frac{1.04}{\left(\frac{4}{\pi^2}\right)}\,\frac{T_{\text{rise}}}{rc_{grd}L^2}\right]-1\right) \cdot \left(\frac{1-\exp\left[\frac{1.04}{\left(\frac{4}{\pi^2}\right)}\,\frac{T_{\text{rise}}}{rc_{grd}L^2}\right]}{1-\exp\left[\frac{1.04}{\left(\frac{4}{\pi^2}\right)}\,\frac{T_{\text{rise}}}{r(c_{grd}+2c_m)L^2}\right]}\right)^{-((c_{grd}+2c_m)/2c_m)} \tag{19}$$

where c_m is the mutual capacitance between wires and c_{grd} is the ground capacitance of each wire. Assuming that the driver switching time, $T_{\text{rise}} \gg \tau$, is slower than the interconnect step response, τ in (13), then the peak crosstalk voltage increases with the *square* of interconnect length and is given by

$$\frac{V_n}{V_{dd}} = 0.51\,\frac{rc_m L^2}{T_{\text{rise}}}. \tag{20}$$

Using simple parallel plate models for the mutual capacitance transforms (20) into

$$\frac{V_n}{V_{dd}} = \frac{0.51}{T_{\text{rise}}}\,\frac{\rho\varepsilon L^2}{WS}. \tag{21}$$

The salient observation derived from (21) is the scaling dependence of peak crosstalk voltage. Fig. 10 illustrates that minimum scaling of wire dimensions of a 1-mm length interconnect from 1 μm to 50 nm *drastically increases peak crosstalk* at the load end of the quiescent line. For example, for a 1-ns risetime the peak noise voltage to switching potential ratio increases almost three orders of magnitude from approximately 0.0002 to 0.1 when scaling this interconnect from 1 μm to 50 nm. The diagonals in this plot of peak crosstalk voltage to binary switching potential ratio versus source voltage rise time in Fig. 10 are loci of constant resistance and mutual capacitance product. The peak noise voltage with device level models increases with the *inverse square* of the device dimension $W = S$. Physically this occurs because minimum wire scaling increases wire resistance, which hinders discharge of crosstalk currents on the quiescent line.

As seen in Fig. 10, there are two distinct crosstalk regions. In the first region, crosstalk is at a maximum when the total interconnect delay is limited by the intrinsic step-response interconnect delay (i.e., $\tau > T_{\text{rise}}$) and is described in [24] and given by

$$\frac{V_n}{V_{dd}} = \frac{1}{2}\,\frac{c_m}{c_{grd}+c_m}. \tag{22}$$

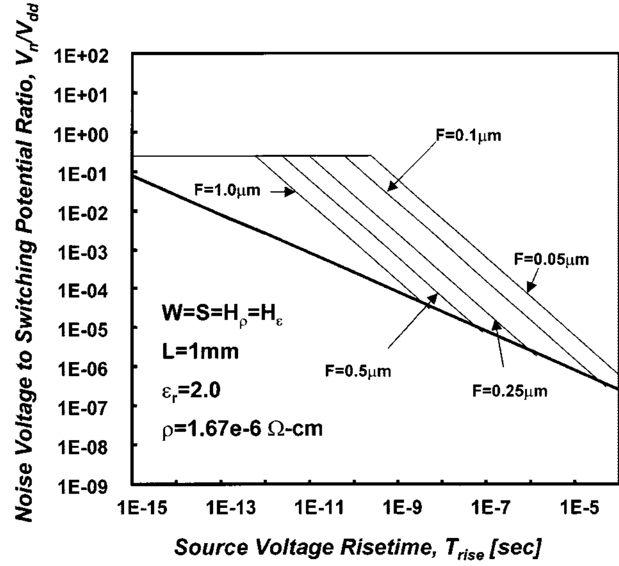


Fig. 10. Effects of interconnect scaling on crosstalk for a semiglobal interconnect.

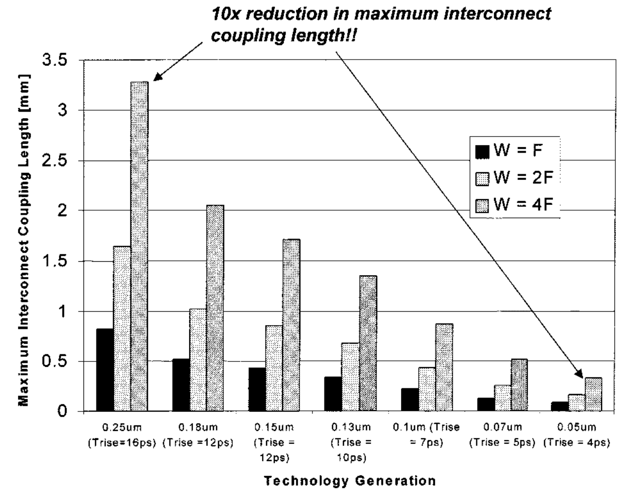


Fig. 11. Maximum interconnect coupling length of local and semiglobal interconnect at which crosstalk begins to exceed 10% of switching potential.

Crosstalk is reduced in the second region when the intrinsic driver switching time dominates the step-response interconnect delay ($T_{\text{rise}} > \tau$) and is described by (20). As the MOSFET switching time decreases and intrinsic interconnect delay increases [1] as illustrated in Table 1, crosstalk problems will infest the multilevel wiring network and dramatically increase the number of local and semiglobal interconnects with high crosstalk. Fig. 11 illustrates the interconnect length at which the peak noise voltage is 10% of the supply voltage for each ITRS generation over the next 15 years with wire dimensions equal to F, 2F, and 4F. The maximum coupling length decreases almost an order of magnitude by 2014, which will drastically increase the number of interconnects with significant crosstalk.

2) Inductance Effects: With the advent of multigigahertz clock frequencies, another serious challenge for the GSI designer is on-chip interconnect inductance. Just as with interconnect performance, this parasitic has its greatest effect

on reverse-scaled high-speed global interconnects. To determine aggressive crosstalk limits on inductance for the device level, just as in the previous section, an ideal ground plane is used to provide a low-impedance path for return currents. In addition, it is also assumed that the finite switching time of a MOSFET only slightly affects long global interconnect crosstalk and, therefore, is ignored.

Assuming negligible skin effect, the telegraphers equation for two symmetric lines is used to describe the transient response along two coupled interconnects and is given by

$$\frac{\partial^2}{\partial x^2}\begin{bmatrix} V_A \\ V_Q \end{bmatrix} = \begin{bmatrix} l_s & l_m \\ l_m & l_s \end{bmatrix}\begin{bmatrix} c_{grd}+c_m & -c_m \\ -c_m & c_{grd}+c_m \end{bmatrix}\frac{\partial^2}{\partial t^2}\begin{bmatrix} V_A \\ V_Q \end{bmatrix} + r\begin{bmatrix} c_{grd}+c_m & -c_m \\ -c_m & c_{grd}+c_m \end{bmatrix}\frac{\partial}{\partial t}\begin{bmatrix} V_A \\ V_Q \end{bmatrix} \quad (23)$$

where

- V_A voltage along the active line;
- V_Q voltage along the quiescent line;
- l_s self-inductance of each line;
- l_m mutual inductance between each line.

Empirical expressions for the capacitance [27] and inductance matrices [28] are used for parasitic estimation. The transient response along the quiescent line is calculated using the compact distributed rlc expression

$$V_Q(L,t) = \tfrac{1}{2}(V_{\text{fin}}(L,t,l=l_s+l_m,c=c_{grd}) - V_{\text{fin}}(L,t,l=l_s-l_m,c=c_{grd}+2c_m)) \quad (24)$$

where V_{fin} is defined in (15).

Effects of mutual inductance pose significant limitations on peak crosstalk reduction. Using (22) and (24), Fig. 12 shows the length dependence of crosstalk with and without the inclusion of inductance on two coupled lines with negligible source impedance ($R_{tr} \ll rL$). Using the distributed RC models with a step-response voltage in [24] the crosstalk is length independent; however, with the inclusion of inductance a strong nonlinear length dependence of crosstalk emerges as seen in Fig. 12. For $R_{tr} \ll rL$, the distributed rlc crosstalk is roughly 60% higher than that predicted by RC models. The expression for this maximum crosstalk voltage with the inclusion of inductance, which is derived from (24), is given by [23]

$$\frac{V_{\text{peak, max}}}{V_{dd}} = \left(\frac{c_{grd}}{c_{grd}+2c_m}\right)^{c_{grd}/2c_m}\left(\frac{c_{grd}}{c_{grd}+2c_m}\right)\frac{2c_m}{c_{grd}} \approx \frac{\pi}{4}\frac{c_m}{c_{grd}+c_m}. \quad (25)$$

The peak crosstalk is approximately $2\pi/4 = 1.57$ times larger than predicted by a distributed RC model in [24].

To help control crosstalk gigahertz interconnect network ground planes or dedicated ground wires maybe necessary for the suppression of unpredictable crosstalk caused by inductance. For distributed RC and RLC high-speed global interconnects, (25) reveals that providing ground planes sufficiently close to interconnect structures can be an effective strategy for controlling crosstalk. For local, semiglobal, and global interconnects, further reduction in crosstalk can be achieved by increasing wire spacing.

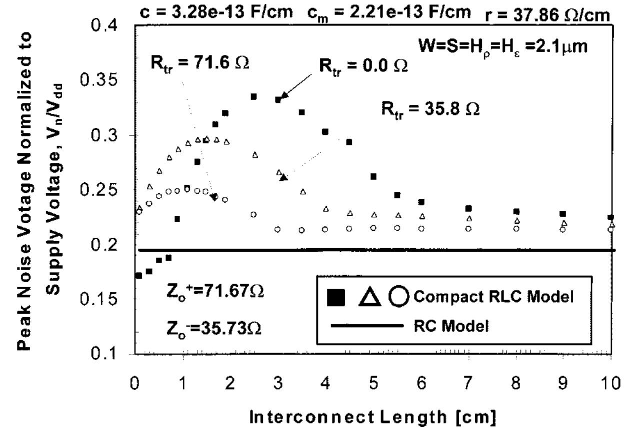


Fig. 12. Nonlinear length dependence of crosstalk for various driver resistance of 0.0 Ω, 35.8 Ω, and 71.6 Ω.

V. Circuit Limits

To gain insight into interconnect circuit limits, simple models that retain only the essence of the problem under attack are engaged. To this end, a transistor is modeled as an equivalent resistance in series with an ideal voltage source that drives an active interconnect in isolation or in proximity to an identical quiescent wire. In addition, the *limits* to reducing circuit delay and crosstalk are determined through the use of ideal current return paths for each interconnect structure. Such assumptions clearly elucidate the effects of source resistance on interconnect performance and crosstalk. The key conclusion of this section is that transistor output resistance exacerbates interconnect circuit delay and crosstalk.

A. Circuit Delay Limits

The effects of delay can be approximated using a near wave-front approximation to a Bessel function expansion similar to (15) and a distributed RC model after [24]. Uniting these two models and assuming that the wire capacitance dominates the transistor input capacitance ($cL \gg C_{tr}$), the approximate time for the transient voltage of an interconnect load to reach χV_{dd} is given by

$$\tau = \frac{L}{\sqrt{lc}}u_o\left(Z_o\left(\frac{2}{\chi}e^{-(rL/2Z_o)}-1\right)-R_{tr}\right) + \left(\left(0.1+0.4\ln\left(\frac{1}{1-\chi}\right)\right)rcL^2+\ln\left(\frac{1}{1-\chi}\right)R_{tr}cL\right) \cdot u_o\left(R_{tr}-Z_o\left(\frac{2}{\chi}e^{-(rL/2Z_o)}-1\right)\right) \quad (26)$$

where $Z_o = \sqrt{l/c}$ and R_{tr} is the equivalent transistor output impedance. The 90% (i.e., $\chi = 0.9$) interconnect latency limit for a very "fat" global wire ($rL \ll R_{tr}$) is given by

$$\tau = 2.3R_{tr}cL \quad (27)$$

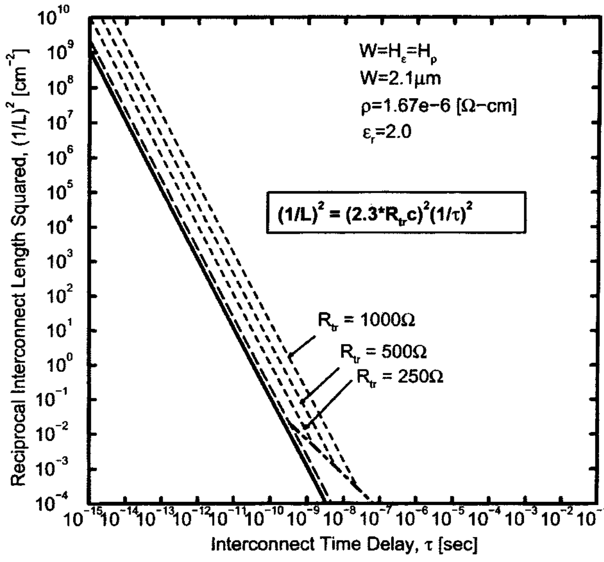


Fig. 13. Circuit limits for distributed rc models.

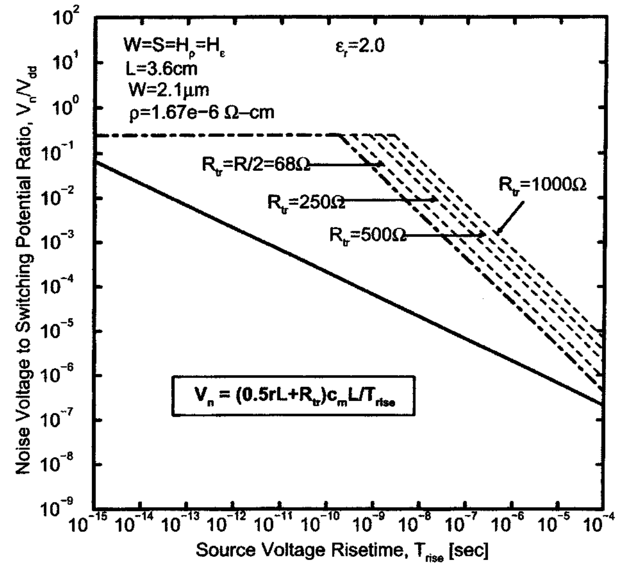


Fig. 14. RC circuit limits on peak crosstalk.

which is approximately valid *only* when the $R_{tr} > Z_o(2.22e^{-(rL/2Zo)} - 1)$. The detrimental effects of driver resistance on the interconnect latency are elucidated in the reciprocal length square versus time delay plane after [2] in Fig. 13. The circuit limit in Fig. 13 approaches the speed of a propagating electromagnetic wave when

$$R_{tr} = Z_o\left(2.22e^{-(rL/2Z_o)} - 1\right) \approx 1.22Z_o \tag{28}$$

where the approximation holds for very small values of the wire resistance. In general, the driver resistance that minimizes both wire delay ($\tau \approx L/v$) and overshoot is given by

$$R_{tr} = Z_o\left(\frac{2}{\chi}\,e^{-(rL/2Z_o)} - 1\right) \tag{29}$$

which is valid as long as $rL < 2Z_o \ln(2/\chi)$. Once this condition is violated, time-of-flight operation is unachievable because the line resistance significantly attenuates fasting rising "lc" transients, and the ideal driver resistance for minimum delay approaches zero.

B. Crosstalk Limits

For interconnect circuits in a GSI multilevel network that have a delay that is dominated by the driver switching time, the extra driver resistance increases the peak noise voltage at the end of a quiescent line according to the following approximation:

$$\frac{V_{\text{peak}}}{V_{dd}} \approx \frac{1}{T_{\text{rise}}}(0.51rL + 1.02R_{tr})c_m L. \tag{30}$$

Using (30) for the condition when $T_{\text{rise}} > \tau$ and the model in [24] for $\tau > T_{\text{rise}}$, this crosstalk limit using distributed RC models is plotted in Fig. 14 for $R_{tr} = 68\ \Omega$, $250\ \Omega$, $500\ \Omega$, and $1000\ \Omega$. The $\tau > T_{\text{rise}}$ region crosstalk remains approximately unchanged as predicted by [24, (22)]. Increasing the source resistance in the $T_{\text{rise}} > \tau$ region, however, substantially increases peak crosstalk at the load end of a quiescent line. In the latter region, larger driver resistance increases peak crosstalk voltage because extra resistance diminishes the ability of the quiescent line to quickly discharge crosstalk currents.

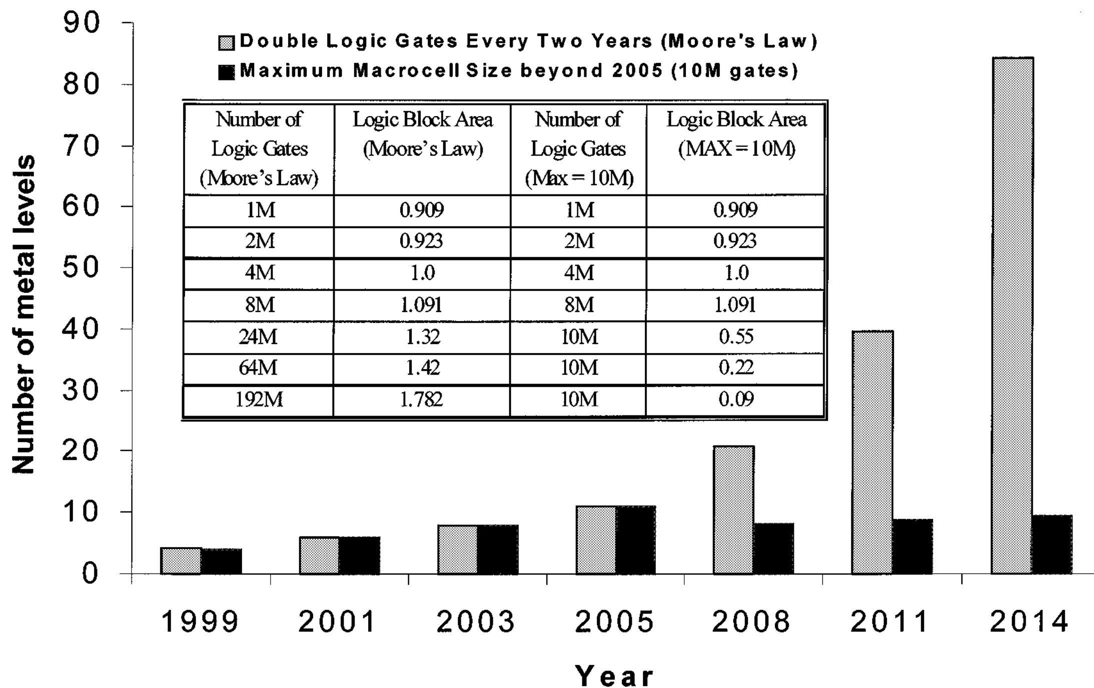
For high-speed global interconnects where the finite driver rise time is negligible and the cumulative interconnect resistance is on the order of the lossless characteristic impedance of the interconnect, inductive effects must be included to fully understand the effects of driver resistance on interconnect circuit limits. The central thesis of this section is partially violated with high-speed global lines because increasing the driver resistance suppresses inductive effects. For example, using a complete solution to telegrapher's equation without skin effect, a complete series solution similar to (24) is used to plot the peak crosstalk voltage at the end of a quiescent line in Fig. 12 ($R_{tr} = 35.8\ \Omega$, $R_{tr} = 71.6\ \Omega$). The extra driver resistance suppresses crosstalk in the nonlinear inductance region, but has negligible effects in the resistance limited region ($rL \gg Z_o$). The penalty for adding extra source resistance, however, is a possible increase in interconnect circuit delay of the active line.

VI. System Limits

System limits are the most nebulous and difficult to project because of the difficulty in generic modeling of future GSI processors. However, a stochastic interconnect distribution model, which has been verified with real microprocessors [25], is used in this section to explore the limitations that reverse-scaled multilevel interconnect networks impose on a GSI system.

A. 2-D Integration Limits

Using a complete wire length distribution in [25] and the ITRS [1] provides a unique opportunity to project the number of metals levels for highly connected logic megacells. A highly connected logic block is defined as a statistically homogeneous array of logic gates in which a well-established empirical relationship know as Rent's Rule



Number of Logic Gates (Moore's Law)	Logic Block Area (Moore's Law)	Number of Logic Gates (Max = 10M)	Logic Block Area (MAX = 10M)
1M	0.909	1M	0.909
2M	0.923	2M	0.923
4M	1.0	4M	1.0
8M	1.091	8M	1.091
24M	1.32	10M	0.55
64M	1.42	10M	0.22
192M	1.782	10M	0.09

Fig. 15. Projection of the number of metal levels over the next 15 years with the assumption that historical trends remain.

describes the input–output (I/O) requirements of arbitrarily sized megacells. The wiring distribution of a 2-D megacell is based upon Rent's Rule [29] and is given in [25].

The complete wiring distribution along with interconnect performance and noise models are used to construct the architecture on a GSI multilevel wiring network. In this network it is assumed that interconnects on adjacent metal levels in a multilevel network are routed orthogonally. The wire dimensions on each orthogonal wiring pair are calculated to insure that the latency of the longest interconnect does not exceed 90% of the clock period, and each pair of levels is occupied with interconnects by equating the required interconnect area to the available interconnect area. To determine the absolute limits on system signal integrity, it is assumed that ultrahigh-speed designs have low-impedance ground planes that are inserted between each orthogonal pair of wire levels to control the vast number of coupling inductances in an unshielded GSI multilevel interconnect network.

This stochastic wiring distribution is used to illustrate the limitations of historical approaches to microprocessor and ASIC design. Starting with the assumption that one million highly connected logic gates are contained in a logic megacell for 1999, the number of metal levels is projected over the next 15 years by doubling the number of highly connected logic gates in a megacell every two years. Logic megacell areas for projected designs are calculated by using the projected transistor densities, minimum feature size, and clock frequencies outlined in the ITRS [1]. As seen in Fig. 15, the number of required metal levels approaches unrealistic values beyond 2005. In fact, the number of projected levels at 2014 is almost an order of magnitude larger than the number of levels prescribed by the ITRS at 2014. As an alternative to Moore's Law scaling, for example, Fig. 15 also shows that saturating the maximum number of highly connected gates at a value around 10 M keeps the number of metal levels per megacell to a controllable number through 2014. Without significant changes to traditional microprocessor or ASIC 2-D transistor technologies, design methodologies, or architectures, Fig. 15 suggests that interconnect limits could undermine Moore's law.

B. 3-D Integration Opportunities

Interconnect delays are increasingly dominating IC performance due to increases in chip size and reduction in the minimum feature size [30]. In spite of new materials like Cu with low-k dielectric interconnect delay is expected to be substantial below 130-nm technology node, thereby severely limiting chip performance [31]. Therefore, the need exists for alternative technologies to overcome this problem. One such promising technique is 3-D ICs with multiple active Si layers. 3-D integration (schematically illustrated in Fig. 16) to create multilayer Si ICs is a concept that can significantly alleviate interconnect delay problems, increase transistor packing density and reduce chip area. Each Si layer in the 3-D structure can have multiple layers of interconnect. Each of these layers are connected together with vertical interlayer interconnects (VILICs) and common global interconnects as shown schematically in Fig. 16. In a 3-D structure a large number of long horizontal interconnects commonly used in 2-D structures can be replaced by short vertical interconnects. Additionally, the 3-D architecture offers extra flexibility in system design, placement, and routing. For instance, logic gates on a critical path can be placed very close to each other using multiple active layers. This would result in reduced chip footprint leading to a significant reduc-

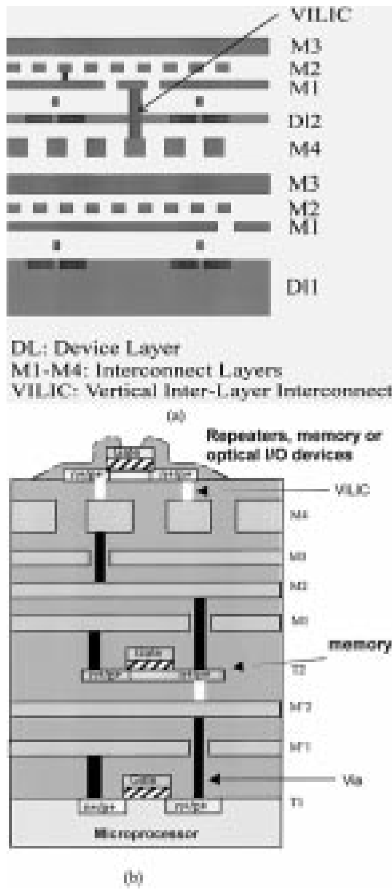


Fig. 16. (a) Schematic representation of 3-D integration based on wafer-bonding approach. Device layer 1, Dl1, is generally a bulk Si layer; device layer 2, Dl2, can be a thinned Si or SOI layer. (b) Alternative approach to 3-D integration based on recrystallization or epitaxial growth.

tion in RC delay and can greatly enhance the performance of logic circuits [32], [33]. This technology can also be exploited to build systems on a chip, by placing circuits with different voltage and performance requirements in different layers. One such example is to have logic circuits in the first Si layer and then have memory circuits in the second layer to realize distributed memory systems in a microprocessor.

1) Performance Estimation of 3-D ICs: A 3-D solution seems an obvious answer to the interconnect delay problem. Since chip size directly affects the interconnect delay, therefore by creating a second active layer, the total chip footprint can be reduced, thus shortening critical interconnects and reducing their delay. In modern logic circuits the chip size is not just limited by the cell size, but also limited by how much metal is required to connect the cells. The transistors on the silicon surface are not actually packed to maximum density but are spaced apart to allow metal lines above to connect one transistor or one cell to another. The metal required on a chip for interconnections is determined not only by the number of gates, but also by other factors such as architecture, average fan-out, number of I/O connections, routing complexity, etc. Therefore, it is not obvious that by using a 3-D structure, the chip size will be reduced. In this work we study the possible effects of 3-D integration on chip area and performance by

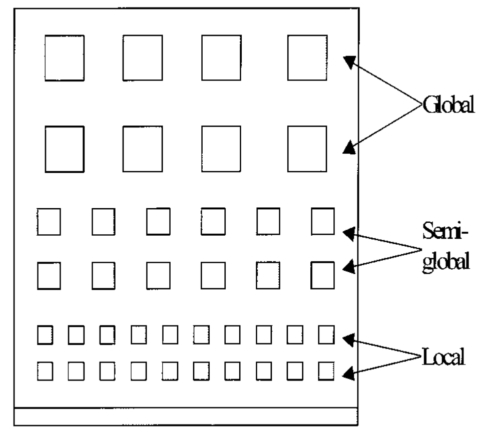


Fig. 17. A three-tier interconnection structure.

modeling the optimal distribution of the metal interconnect lines.

To better understand how a 3-D design will affect the amount of metal wires required for interconnections we applied a stochastic approach for estimating wiring requirements derived for a 2-D structure [25], [34] and modified it for 3-D ICs to quantify effects on interconnect RC delay. Using a three-tier interconnection structure (local, semiglobal, and global), illustrated in Fig. 17, the semiglobal tier pitch that minimizes the wire limited chip area is determined. The maximum interconnect length on any given tier is determined by the interconnect delay criteria. The methodology presented in [25] can be extended easily to derive the wire-length distribution of a 3-D IC. The wire-length distribution and the interconnect delay criteria can be used for tradeoff analysis between 2-D and 3-D ICs. The 3-D interconnect scheme being considered for our analysis is shown in Fig. 16(a).

a) Wire-length distribution: In deriving the 3-D wire-length distribution, instead of a hierarchical partitioning approach [35], we use a nonhierarchical partitioning [25]. Since it is not apparent how Rent's parameters should change as 2-D integrated circuits are mapped into three dimensions, we assume that the same Rent's parameters are applicable to both 2-D and 3-D implementation of an integrated circuit. A more elaborate description of this methodology is described elsewhere [33], [36]. To derive the point-to-point wire-length distribution of an integrated circuit of random logic networks with N_t transistors, the integrated circuit is partitioned into N logic gates, where $N = N_t/\phi$; ϕ is a function of the average fan-in (f.i.) and fan-out (f.o.) in the system [4]. The average separation between the adjacent logic gates is called gate pitch, and it is equal to $(A_c/N)^{1/2}$, where A_c is the die area.

Following the methodology presented in [25], the point-to-point wire-length distribution of 3-D IC is given by

$$f_{3\mathrm{D}}(l) = \Gamma M_{3\mathrm{D}}(l) I_{3\mathrm{D}}(l) \tag{31}$$

where

Γ	normalization constant;
$M_{3\mathrm{D}}(l)$	number of gate pairs separated by length ℓ;
$I_{3\mathrm{D}}(l)$	number of point-to-point interconnects between these gate pairs.

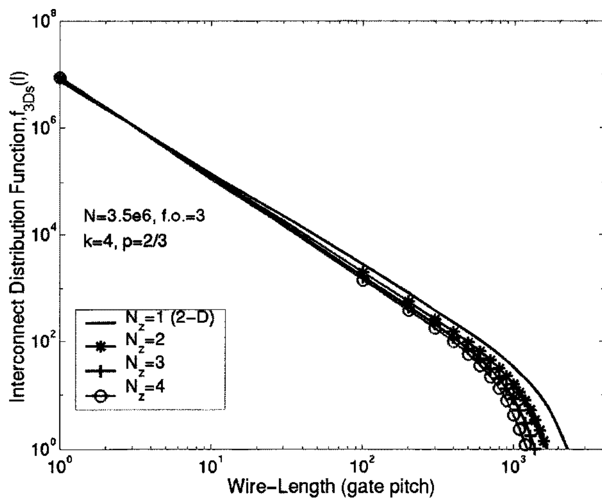


Fig. 18. Wire-length distribution of 2-D and 3-D integrated circuits for symmetrical interconnect schemes, $f_{3\mathrm{D}s}(l)$. N_z is the number of device layers.

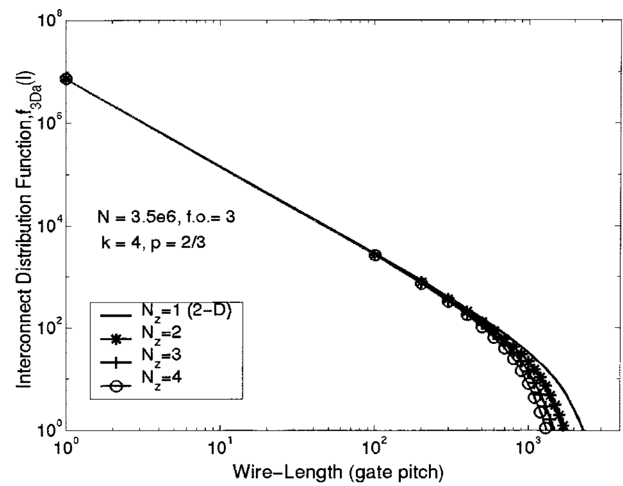


Fig. 19. Wire-length distribution of 2-D and 3-D integrated circuits for asymmetrical interconnect schemes, $f_{3\mathrm{D}a}(l)$. N_z is the number of device layers.

The value of Γ is estimated such that the total number of point-to-point interconnects in a 2-D or 3-D IC is conserved. $M_{3\mathrm{D}}(l)$ is estimated by taking into account the equidistant gate pairs located within a device layer and between device layers [33]. $I_{3\mathrm{D}}(l)$ is estimated by applying Rent's rule where the source and sink gate pairs, connected by a wire, can be located on the same or different device layers [33]. In our analysis, two limiting cases of the 3-D wire-length distribution are considered. In the ***symmetric interconnection*** scheme, for any source logic gate, the sink logic gate can be located on the same or other device layers, and there is a comparable number of interconnections between gate pairs on the same and different device layers. In the ***asymmetric interconnection*** scheme, we assume the number of interconnections between the logic gates on different device layers is negligible compared to the number of interconnections within the device layers.

The wire-length distributions for homogeneous random logic networks in 2-D and 3-D ICs are shown in Figs. 18 and 19. In a 3-D IC, as more device layers are added, the wire-length distribution becomes narrower resulting in fewer and shorter semiglobal and global wires. In both 3-D interconnect schemes, the average and total wire lengths are shorter. However, a symmetric interconnection scheme results in shorter average and total wire lengths compared to an asymmetric interconnection scheme.

b) Simulation results: Using the wire-length distribution and the interconnect delay criteria, some interesting tradeoff analysis can be performed between 2-D and 3-D ICs. For example: 1) chip area can be estimated for fixed clock frequency; 2) clock frequency can be estimated for fixed chip area; or 3) number of interconnect levels can be estimated for fixed chip area and clock frequency. Simulation results of some of these tradeoff analyses are presented here.

To estimate the clock frequency, we use a critical path model that has a logic depth of 15. The logic gates are approximated by NAND gates with fan-in and fan-out of three. We assume all the logic gates drive average length wires, while one logic gate drives a chip-edge length wire [4]. We assume the chip area is interconnect limited, and it is estimated by equating the available chip area with the required chip area [34]. The available chip area is a function of the number of device layers, the chip/die size, total number of interconnect layers, and the wiring efficiency in each interconnect layer. The required chip area is the product of the wiring pitches and the total wire length of local, semiglobal and global wires. The wiring efficiency model presented in [4] can be extended to estimate the wiring efficiency of 3-D ICs. To make a fair comparison between different 2-D and 3-D technologies, we introduce a cost/complexity function. We define a cost function, c.f. $\sim(m+n_b)$, where m is the number of interconnect levels per device layer, and $n_b = N_z - 1$ is the number of interdevice layer bonding steps, and N_z is the number of device layers. For example, in a 2-D IC c.f. $\sim$6 implies that there are six interconnect levels. For the same cost function in a 3-D IC with two device layers, there are five interconnect levels/device layer and one bonding step.

The input parameters of our analysis are presented in Table 3. These parameters are consistent with the technology requirement for microprocessors in 0.18-μm technology node [37]. The clock frequency is estimated by keeping the total chip area, $N_z A_c$, fixed and applying the cost constraint. The simulation results are shown in Fig. 20. The improvement in clock frequency in a 3-D IC results from the reduction in interconnect delay of the average length and chip-edge length wires due to their shorter wire-lengths and larger wiring pitch. The total wire length in a 3-D IC is shorter than that of a 2-D IC. Since the wiring area is proportional to $pitch \times total\ wire\ length$, for comparable available wiring area, the wiring pitch in a 3-D IC can be increased to reduce the interconnect delay. In a 3-D IC, due to the constant cost function, c.f. $\sim(n_b + m)$, fewer interconnect levels per device layer are available as more device layers are integrated. Wiring area is also reduced due to the via blockage of VILICs. Based on our modeling

Table 3
Input Parameters for Random Logic Network/Microprocessor Applications [37]

PHYSICAL PARAMETERS	VALUE
Number of Transistors, N_t	21 million
Equivalent Number of Logic Gates, N	3.5 million
Assumed Rent's Exponent, p	2/3
Assumed Rent's Coefficient, k	4
Number of Wiring Levels	6
Interconnect's Aspect Ratio	1.5
Inter -Level Dielectric Constant, Polymer	$\varepsilon_r = 2.5$
Metal resistivity, Cu	1.67 ($\mu\Omega$–cm)

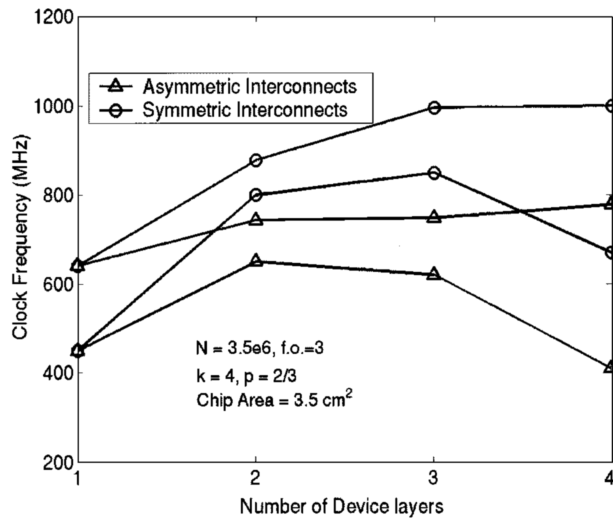


Fig. 20. Clock frequency of 2-D and 3-D IC as a function of number of device layers for fixed chip area and cost function constraints. Simulation results of clock frequency are presented for both with and without repeaters inserted in the long wire. The width/length ratio of the transistors in the critical path is five.

approach, there is an optimum number of device layers that can be integrated profitably to improve the clock frequency. For the example being considered, it appears to be three to four.

To estimate the impact of 3-D integration on chip area, another set of tradeoff analyses can be performed. In this case the clock frequency and the cost function are kept constant, and the total chip area is estimated. The required chip area of 2-D and 3-D ICs for 450-MHz clock frequency, and c.f. $\sim$6 is shown in Fig. 21. Assuming the interconnect delay is proportional to $wire\text{-}length^2/pitch^2$, for similar interconnect delay constraint, since the wire length in a 3-D IC is shorter, the wiring pitch can be reduced. Both the shorter wire length and the flexibility to reduce the wiring pitch for fixed clock frequency constraint lead to the lower chip area in a 3-D IC.

The analysis presented so far was for a 180-nm 3-D technology for a fixed cost function. Next we extend this analysis to study the effect of scaling the technology to smaller feature size, increasing the number available metal layers and active Si layers. In the next set of analyses, the 3-D interconnect scheme being considered is shown in Fig. 16(b). However,

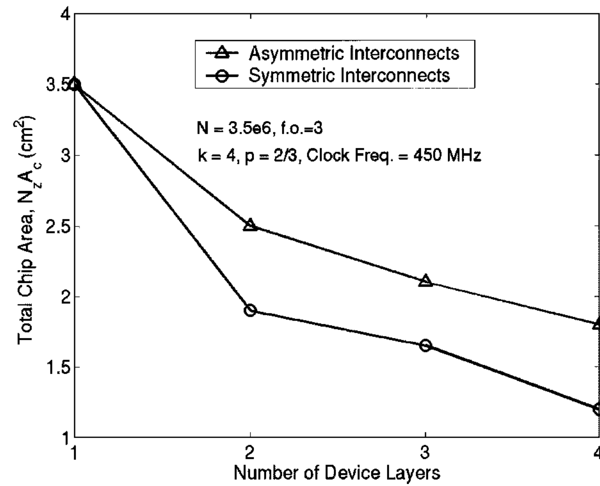


Fig. 21. Simulation results of total chip area for fixed clock frequency and cost constraint. The total chip area is given by $NzAc$, where Nz is the number of device layers and Ac is the die area.

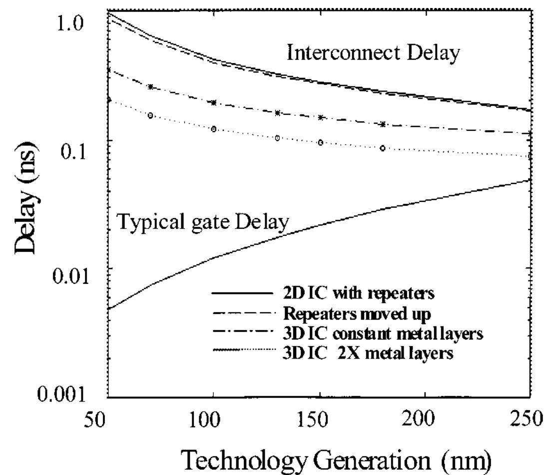


Fig. 22. Interconnect delay limits IC performance with scaling. Moving repeaters to upper active tiers reduces interconnect delay by 9%. 3-D (two active layers) shows significant delay reduction (64%). Increasing the number of metal levels in 3-D reduces interconnect delay by a further 35%.

similar analysis can be carried out for other approaches to 3-D integration as well.

Interconnect delay as a function of technology is calculated (Fig. 22) using data projected by the NTRS for 2-D ICs. Also shown are delays for 3-D ICs with two active layers, where wire pitches are increased to match the 2-D IC areas, calculated using the 3-D chip area estimation model described above. Interconnect delay is reduced by 64% as a result. In all these calculations the number of metal levels is conserved between 2-D and 3-D ICs. This assumption can be relaxed such that each active layer in 3-D ICs may have its own associated lower metal tiers with a universal global tier used for connecting the active-layer networks. The total number of metal layers is thus increased in this 3-D case.

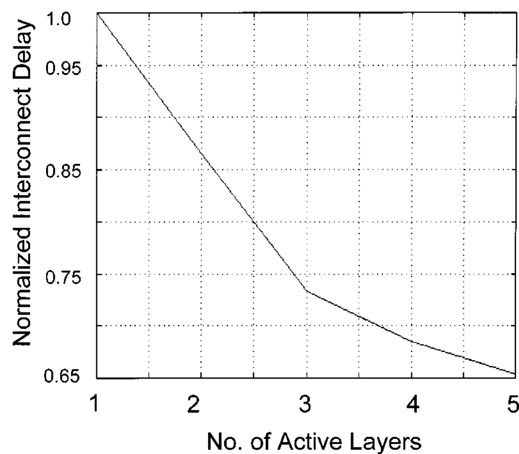


Fig. 23. Signal delay for multiple active Si layers normalized to single layer delay for worst case scenario, shown for 50-nm node.

In estimating chip area, the metal requirement is calculated from the obtained wire-length distribution. The total metallization requirement is appropriately divided among the available metal layers in the corresponding technology. Thus in the example shown in Fig. 17, the local tier has three metal layers, the semiglobal one and the global two. However, the chip area is determined by the resulting area of the local tier as it is the most densely packed. Consequently, higher tiers are routed within a larger area. The resulting delays are also shown in Fig. 22. At the 50-nm node the delay improvement is an additional 35%.

Fig. 23 compares the interconnect delay for up to five active layers for the 50-nm node. In this calculation only 10% of the interblock wires are assumed vertical and the number of metal layers is conserved. Delay is shown to improve with an increase in the number of active layers, however, with diminishing returns. This is due to the increase in the remaining lateral interblock wires as a fraction of the total wiring requirement with increasing number of active layers.

2) 3-D Technology Options: Although the concept of 3-D integration was demonstrated as early as in 1979 [38], it largely remained a research curiosity, since IC performance was device limited. However, with the growing menace of RC delay in recent times, this technology is being viewed as a potential alternative that can not only maintain chip performance well beyond the 130-nm node, but also inspire a new generation of circuit design concepts. Presently, there are several possible fabrication technologies that can be used to realize multiple layers of active area (single crystal Si or recrystallized poly-Si) separated by interlayer dielectrics (ILDs) for 3-D circuit processing. A brief description of these alternatives is given below. The choice of a particular technology for fabricating 3-D circuits will depend on the requirements of the system, since the circuit performance is strongly influenced by the electrical characteristics of the fabricated devices as well as on the manufacturability and process compatibility with the relevant 2-D technology.

Beam Recrystallization: A very popular method for fabricating a second silicon layer on top of an existing substrate is to deposit polysilicon and fabricate thin-film transistors (TFT). To enhance the performance of TFTs, an intense laser or electron beam is used to induced recrystallization of the polysilicon. This technique however may not be very practical for 3-D devices because of the high temperature involved during melting of the polysilicon and also due to difficulty in controlling the grain size variations. Beam recrystallized polysilicon films also suffer from lower carrier mobilities and unintentional impurity doping. However, high- performance TFTs fabricated using low temperature processing, and even low-temperature single-crystal Si TFTs have been recently demonstrated [39], [40] that can be employed to fabricate advanced 3-D circuits.

Processed Wafer Bonding: Another alternative is to bond two fully processed wafers, on which chips are fabricated on the surface including some interconnects, such that the chips completely overlap [41]. Vias are etched to electrically connect both chips after metallization. A backside of the bonded pair can be back-etched to allow for further processing or the bonding of more pairs in this vertical fashion. Other advantages of this technology lie in the similar electrical properties of devices on all active levels and the independence of processing temperature since all chips can be fabricated separately and later bonded. The major limitation of this technique is its lack of precision (best case alignment ± 3 μm), which restricts the interchip communication to global metal lines. However, for applications where each chip is required to perform independent processing before communicating with its neighbor this technology can prove attractive.

Silicon Epitaxial Growth: Another technique for forming additional Si layers is to etch a hole in a passivated wafer and epitaxially grow a single crystal Si seeded from open window in the ILD. The silicon crystal grows vertically and then laterally, to cover the ILD [42]. In principle, the quality of these fabricated devices can be as good as those fabricated underneath on the wafer surface since the grown layer is single crystal with few defects. However, the high temperatures (1000 °C) involved in this process cause significant degradation in the quality of devices on lower layers. Also this technique cannot be used over metallization layers. Low-temperature silicon epitaxy using ultrahigh-vacuum chemical vapor deposition (UHV-CVD) has been recently developed [43]. However, this process is not very attractive for batch processing.

Solid Phase Crystallization (SPC): As an alternative to high-temperature epitaxial growth, low-temperature deposition and crystallization of amorphous silicon, which passivates the lower active layer devices, can be employed. The amorphous film can be randomly crystallized to form a polysilicon film. TFT performance can be enhanced by eliminating grain boundaries. For this purpose, local crystallization can be induced using low-temperature processes such as using patterned seeding of Germanium [44], or by using metal-induced lateral crystallization (MILC) [45], [46]. This technique offers the flexibility of creating multiple active layers that are compatible with current processing

environments, and recent results prove the feasibility of building high-performance TFTs at low processing temperatures that can be compatible with lower level metallization [47]. MILC, for example, can be used to build repeaters above metal lines. It is found that the electrical characteristics of these TFTs are approaching the single crystal SOI devices [48].

3) Concerns in 3-D Circuits:

a) Thermal issues: An extremely important issue in 3-D ICs is heat dissipation [49]. Thermal effects are already known to significantly impact interconnect and device reliability in present 2-D circuits. The problem is expected to be exacerbated by the reduction in chip size, assuming that same power generated in a 2-D chip will now be generated in a smaller 3-D chip, resulting in a sharp increase in the power density. Analysis of thermal problems in 3-D circuits is therefore necessary to comprehend the limitations of this technology, and also to evaluate the thermal robustness of different 3-D technology options.

It is well known that most of the heat energy generated in integrated circuits arises due to transistor switching. This heat is typically conducted through the silicon substrate to the package and then to the ambient by a heat sink. With multilayer device designs, devices in the upper layers will also generate a significant fraction of the heat. Furthermore, all the active layers will be insulated from each other by layers of dielectrics (LTO, HSQ, polyimide, etc.), which typically have much lower thermal conductivity than Si [50], [51]. Hence, the heat dissipation issue can become even more acute for 3-D ICs and can cause degradation in device performance and reduction in chip reliability due to increased junction leakage, electromigration failures, and acceleration of other failure mechanisms. However, initial analysis indicates that thermal problems in 3-D circuits can be alleviated by optimizing the interconnect capacitance, chip frequency and the area.

b) Interconnect capacitance and crosstalk: In 3-D devices an additional electrical coupling between the top layer metal of the first active layer and the devices on the second active layer would be present [52]. This needs to be addressed at the circuit design stage. However, for deep submicrometer technologies, the aspect ratio of interconnects is approximately 1.5–2. Thus, line-to-line capacitance is the dominant portion of the overall capacitance. Therefore, the presence of an additional silicon layer on top of a metal level will not affect the capacitance per unit length of these lines. For technologies with very small aspect ratio, the change in interconnect capacitance due to the presence of an additional silicon layer would be significant, as reported in [52].

VII. Conclusion

Twenty-first century interconnect limits are codified into fundamental, material, device, circuit and system limits. At the fundamental level, electromagnetic wave velocity will limit the performance of overly aggressive designs of high-speed synchronous die-edge-length interconnects. In addition, the absolute minimum energy per binary transition for reduced swing low-power interconnects is limited to $kT\ln(2)$ according to Shannon's communication theorem. At the material level, the resistivity of wire conductors increases *substantially* in sub-50-nm technology. This increase is primarily controlled by the scattering mechanisms due to the properties of the surfaces and interfaces of copper films, as driven by 1- and 2-D scattering effects. This limit requires the development and optimization of M&P solutions that can grow epitaxial Cu/liner interconnect stacks with atomically smooth surfaces and interfaces, while maximizing specular electron surface scattering in ultranarrow sub-50-nm interconnect lines. At the device level, both minimum and reverse scaling strategies have a pronounced effect on interconnect crosstalk limits. Minimum interconnect scaling significantly increases crosstalk on many GSI local and semiglobal interconnects, and it is shown that the coupling length at which significant crosstalk ($\leq 0.1V_{dd}$) occurs could decrease by an order of magnitude over the next 15 years. Reverse scaling of global interconnects causes inductance to significantly influence on-chip interconnect transients. Even with ideal return path conditions, mutual inductance increases crosstalk by up to 60% over that predicted by conventional RC models. At the circuit level, transistor driver output impedance in distributed RC interconnects circuits only exacerbates interconnect performance and crosstalk limits for semiglobal and local interconnects. When inductance is important ($R_{tr} + rL \approx Z_o$), careful driver design helps reduce overshoot and inductive crosstalk, but potentially at the cost of excess circuit delay. Finally, at the system level the continued historical approaches to chip design are scrutinized. Using 2-D integration of transistors and technology projections from the ITRS, the number of metal levels explodes for highly connected logic megacells that double in size every two years. Beyond 2005, the number of metal levels predicted with a stochastic wiring distribution model reaches unattainable values such that by 2014 the number of metal levels is almost an order of magnitude larger than what is projected by the ITRS. This result emphasizes that substantial changes in design methodologies, technologies, and architectures are needed to cope with the onslaught of wiring demands. One possible solution to this problem that is highlighted in this paper is the feasibility of 3-D integration of transistors. It has been demonstrated that interconnect performance is significantly improved by using 3-D ICs. By increasing the number of active layers, including the use of separate layers for repeaters, and optimizing the wiring network, these results predict an improvement in interconnect performance of up to 145% at the 50-nm node. This modeling is also conservative, leaving room for further improvement, as optimization of logic block placement and connectivity is considered. Some of the major concerns for 3-D circuits are power dissipation and the associated thermal effects and additional complexity introduced in fabrication technology.

References

[1] Semiconductor Industry Association, *International Technology Roadmap for Semiconductors: 1999*. Austin, TX: SEMATECH, 1999.

[2] J. Meindl, "Low power microelectronics: Retrospective and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, Apr. 1995.

[3] H. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

[4] G. Sai-Halz, "Performance trends in high-end processors," *Proc. IEEE*, vol. 83, pp. 20–36, Jan. 1995.

[5] R. Keyes, "Physical limits in semiconductor electronics," *Science*, vol. 195, pp. 1231–1235, Mar. 1977.

[6] J. Meindl, "Interconnect limits on ultra-large-scale integration," *VLSI 1985*, pp. 13–19, 1985.

[7] J. Jackson, *Classical Electrodynamics*. New York: Wiley, 1975.

[8] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–356, July 1948.

[9] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Philadelphia, PA: Saunders College, 1976, pp. 133–141.

[10] C. R. Pichard, C. R. Tellier, and A. J. Tosser, "Decomposition of polycrystalline metallic film resistivity from Mathiessen's rule," *J. Physics*, vol. D12, no. L101, 1979.

[11] A. E. Kaloyeros, M. Bylansky, and E. T. Eisenbraun, *Proc. 1999 Annu. Rev. Interconnect Focus Center*, Dec. 14–15, 1999.

[12] H.-D. Liu, Y.-P. Zhao, G. Ramanath, S. P. Murarka, and G.-C. Wang, submitted for publication.

[13] G. Ramaswamy, A. K. Raychaudhuri, J. Goswami, and S. A. Shivashankar, "Large deviation from Mathiessen's rule in chemical vapor deposited copper films and its correlation with nanostructure," *J. Phys.*, vol. D30, no. L5, 1997.

[14] J. R. Sambles, "The resistivity of thin metal films—some critical remarks," *Thin Solid Films*, vol. 106, no. 321, 1983.

[15] K. C. Elsom and J. R. Sambles, "Macroscopic surface roughness and the resistivity of thin metal films," *J. Phys.*, vol. F11, no. 647, 1981.

[16] M. H. Grabow and G. H. Gilmer, *Semiconductor-Based Heterostructures*, M. L. Green *et al.*, Ed. Warrendale, PA: Metallurgical Soc., 1986.

[17] S. P. Murarka, *Metallization: Theory and Practice for VLSI and ULSI*. Boston, MA: Butterworth-Heinemann, 1993, pp. 47–49.

[18] K. Fuchs, "The conductivity of thin metallic films according to the electron theory of metals," in *Proc. Cambridge Philosophical Soc.*, vol. 34, 1938.

[19] M. Sahimi, *Applications of Percolation Theory*. New York: Taylor and Francis, 1994, p. 258.

[20] V. Korobov, M. Leibovich, and Y. Shapiro, "Structure and conductance evolution of very thin indium oxide films," *Appl. Phys. Lett.*, vol. 65, no. 2290, 1994.

[21] P. J. Restle and A. Deutsch, "Designing the best clock distribution network," in *1998 Symp. VLSI Circuits Dig. Tech. Papers*, 1998, pp. 2–5.

[22] R. E. Collin, *Foundations for Microwave Engineering*. New York: McGraw-Hill, 1992.

[23] J. A. Davis and J. D. Meindl, "Compact distributed RLC models for multilevel interconnect networks," in *1999 VLSI Symp. Technology Dig. Tech. Papers*, pp. 165–166.

[24] T. Sakurai, "Closed-form expression for interconnection delay, coupling, and crosstalk in VLSI's," *IEEE Trans. Electron Devices*, vol. 40, pp. 118–124, Jan. 1993.

[25] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire length distribution for gigascale integration (GSI) Part I: Derivation and validation," *IEEE Trans. Electron Devices*, vol. 45, pp. 580–589, March 1998.

[26] J. A. Davis and J. D. Meindl, "Length, scaling, and material dependence of crosstalk between distributed RC interconnects," in *Proc. 1999 Int. Interconnect Technology Conf.*, May 1999, pp. 227–229.

[27] J. Chern, J. Jurang, L. Arledge, P. Li, and P. Yang, "Multilevel metal capacitance models for CAD design synthesis systems," *IEEE Electron Device Lett.*, vol. 13, pp. 32–34, Jan. 1992.

[28] C. Paul, *Analysis of Multi-Conductor Transmission Line*. New York: Wiley, 1994.

[29] Landmann and Russo, "On a pin versus block relationship for partitions of logic paths," *IEEE Trans. Comput.*, vol. C-20, pp. 1469–1479, Dec. 1971.

[30] K. C. Saraswat and F. Mohammadi, "Effect of interconnection scaling on time delay of VLSI circuits," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 645–650, Apr. 1982.

[31] M. T. Bohr, "Interconnect scaling—The real limiter to high performance ULSI," *IEDM Tech. Dig.*, pp. 241–244, 1995.

[32] S. J. Souri and K. C. Saraswat, "Interconnect performance modeling for 3D integrated circuits with multiple Si layers," in *Int. Interconnect Technology Conf. Proc.*, 1999, pp. 24–26.

[33] A. Rahman, A. Fan, J. Chung, and R. Reif, "Wire-length distribution of three-dimensional integrated circuits," in *Int. Interconnect Technology Conf. Proc.*, 1999, pp. 233–235.

[34] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI)—Part II: Applications to clock frequency, power dissipation, and chip size estimation," *IEEE Trans. Electron Devices*, vol. 45, Mar. 1998.

[35] D. Stroobandt and J. V. Campenhout, "Estimating interconnect lengths in three-dimensional computer systems," *IEICE Trans.*, vol. E80-D, pp. 1024–1031, 1997.

[36] A. Rahman, A. Fan, and R. Reif, "System-level performance evaluation of three-dimensional integrated circuits,", submitted for publication.

[37] Semiconductor Industry Association, *National Technology Roadmap for Semiconductors: 1997*. Austin, TX: SEMATECH, 1997.

[38] K. F. Lee, J. F. Gibbons, K. C. Saraswat, and T. I. Kamins, "Thin film MOSFET fabricated in laser-annealed polycrystalline silicon," *J. Appl. Phys. Lett.*, pp. 173–175, 1979.

[39] M. A. Crowder, P. G. Carey, P. M. Smith, R. S. Sposili, H. S. Cho, and J. S. Im, "Low-temperature single crystal Si TFT's fabricated on Si-films processed via sequential lateral solidification," *IEEE Electron Device Lett.*, vol. 19, no. 8, pp. 306–308, 1986.

[40] A. Kohno, T. Sameshima, N. Sano, M. Sekiya, and M. Hara, "High performance poly-Si TFT's fabricated using pulsed laser annealing and remote plasma CVD with low temperature processing," *IEEE Trans. Electron Devices*, vol. 42, no. 2, pp. 251–257, 1995.

[41] A. Fan, A. Rahman, and R. Reif, "Copper wafer bonding," *Electrochem. Solid State Lett.*, vol. 2, no. 10, pp. 534–536, 1999.

[42] G. W. Neudeck, S. Pae, J. P. Denton, and T. Su, "Multiple layers of silicon-on-insulator for nanostructure devices," *J. Vac. Sci. Technol. B*, vol. 17, no. 3, pp. 994–998, 1999.

[43] H.-Y. Lin, C.-Y. Chang, T. F. Lei, J.-Y. Cheng, H.-C. Tseng, and L.-P. Chen, "Characterization of polycrystalline silicon thin film transistors fabricated by ultrahigh-vacuum chemical vapor deposition and chemical mechanical polishing," *Jpn. J. Appl. Phys. 1, Regul. Rap. Short Notes*, vol. 36, no. 7A, pp. 4278–4282, July 1997.

[44] V. Subramanian and K. C. Saraswat, "High-performance germanium-seeded laterally crystallized TFT's for vertical device integration," *IEEE Trans. Electron Devices*, vol. 45, no. 9, pp. 1934–1939, 1998.

[45] S.-W. Lee and S.-K. Joo, "Low temperature poly-Si thin film transistor fabrication by metal-induced lateral crystallization," *IEEE Electron Device Lett.*, vol. 17, no. 4, pp. 160–162.

[46] S. Y. Yoon, S. K. Kim, J. Y. Oh, Y. J. Choi, W. S. Shon, C. O. Kim, and J. Jang, "A high-performance polycrystalline silicon thin-film transistor using metal-induced crystallization with Ni solution," *Jpn. J. Appl. Phys. 1, Regul. Rap. Short Notes*, pp. 7193–7197, Dec. 1998.

[47] A. R. Joshi and K. C. Saraswat, "Sub-micron thin film transistors with metal induced lateral crystallization," in *Proc. 196th Meeting Electrochemical Soc.*, Honolulu, HI, 1999.

[48] S. Jagar, M. Chan, M. C. Poon, H. Wang, M. Qin, P. K. Ko, and Y. Wang, "Single grain thin-film-transistor (TFT) with SOI CMOS performance formed by metal-induced-lateral-crystallization," *IEDM Tech. Dig.*, pp. 293–296, 1999.

[49] S. A. Kuhn, M. B. Kleiner, P. Ramm, and W. Weber, "Thermal analysis of vertically integrated circuits," *IEDM Tech. Dig.*, pp. 487–490, 1995.

[50] K. E. Goodson and Y. S. Ju, "Heat conduction in novel electronic films," in *Annu. Rev. Mater. Sci.*, vol. 29, 1999, pp. 261–293.

[51] K. Banerjee, A. Amerasekera, G. Dixit, and C. Hu, "The effect of interconnect scaling and low-k dielectric on the thermal characteristics of the IC metal," *IEDM Tech. Dig.*, pp. 65–68, 1996.

[52] S. A. Kuhn, M. B. Kleiner, P. Ramm, and W. Weber, "Interconnect capacitances, crosstalk, and signal delay in vertically integrated circuits," *IEDM Tech. Dig.*, pp. 487–490, 1995.

Jeffrey A. Davis received the B.E.E., M.S.E.E, and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology in 1993, 1997, and 1999, respectively.

He joined the faculty at Georgia Tech as an Assistant Professor in 1999. He is currently serving as the Co-Program Chair for the 2001 System Level Interconnect Prediction Workshop and is the Faculty Chair for the ECE student–faculty committee at Georgia Tech. He has coauthored more than 30 conference and journal publications during the past five years. His current research interests are in the areas of high-speed interconnect modeling and optimization, high-speed interconnect networks, optimal multilevel interconnect architectures, and interconnect-centric design methodologies for future GSI processors.

Dr. Davis received the best student paper award at the 1999 International Interconnect Technology Conference.

Raguraman Venkatesan was born in Sindri, Biharstate, India, in 1976. He received the B.Tech. and M.S. degrees in electrical engineering from the Indian Institute of Technology, Bombay, and the Georgia Institute of Technology, Atlanta, in 1998 and 2000, respectively.

He is currently a Ph.D. student in the Gigascale Integration Group at the Georgia Institute of Technology. His research interests include designing optimal multilevel wiring networks and inductance modeling.

Alain Kaloyeros received the Ph.D. degree in experimental condensed matter physics from the University of Illinois, Urbana-Champaign, in 1987.

He is Professor of Physics and Executive Director of the University of Albany Institute for Materials and its affiliated centers, including the New York State Center for Advanced Tin Film Technology (CAT), the New York State Center of Excellence in Nanoelectronics, NanoFab 200 and the Energy and Environmental Technologies Applications Center (EETAC). He was also recently appointed Founding Dean of the School of NanoSciences and Materials at UAlbany. He has authored and coauthored more than 100 articles and contributed to seven books on topics pertaining to the science and technology of advanced semiconductor and optoelectronics thin film materials, vapor phase processes, and high-resolution X-ray, electron, and photon-based characterization and metrology. He holds nine U.S. patents.

Dr. Kaloyeras is a past recipient of the NSF Presidential Young Investigator (PYI) Award, the NSF Research Initiation Award (RIA), the Albany Foundation Academic Laureate Award, and the Citizen of the University Award.

Michael Beylansky received the B.S degree in chemistry from Moscow University, Russia, in 1988, and the Ph.D degree in physical chemistry from the University of Illinois, Chicago, in 1998.

From 1988 to 1993, he worked at Moscow State University on synthesis and surface analysis of type IV-VI semiconductor heterostructures. He did his post-doctoral work on copper interconnects at the Center for Advanced Technology at the State University of New York (SUNY) at Albany. He joined IBM Corporation Microelectronics Division, East Fishkill, NY, in 2000, where he is Staff Engineer/Scientist in DRAM thin-film process development.

Shukri J. Souri received the B.A. degree in engineering science from Oxford University, U.K., and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1992 and 1994, respectively. He is currently pursuing the Ph.D. degree from the Electrical Engineering Department, Stanford University.

From 1994 to 1997, he was a Member of Technical Staff with the Corporate Research Division of Raychem Corporation, Menlo Park, CA, where he worked on electroceramic semiconducting and ferroelectric materials and devices for circuit protection applications. His research interests include 3-D IC performance modeling and interconnect network architecture. He has also worked on 3-D integration using advanced seeding and crystallization techniques. He co-invented a number of U.S. patents and has several publications in his areas of interest.

Kaustav Banerjee (Member, IEEE) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1999.

Since March 1999, he has been with Stanford University, Stanford, CA, as a Research Associate at the Center for Integrated Systems. He also works as a technical consultant in the EDA industry. His research interests include signal integrity, reliability and performance optimization issues in high-performance VLSI and high-frequency (RF) mixed-signal applications. He is also interested in all aspects of integrated heterogeneous circuits and systems including System-on-Chip designs. At Stanford, Dr. Banerjee leads an interdisciplinary research team of ten doctoral students. As part of the MARCO Interconnect Focus Center at Stanford, he is actively involved in the research of 3-D ICs. He is also involved in several collaborative research initiatives with other leading Universities. He co-advises doctoral students in the Electrical Engineering Departments of the University of Southern California, Los Angeles, and the Swiss Federal Institute of Technology, Lausanne, Switzerland. He has held several summer research positions at the Semiconductor Process and Device Center of Texas Instruments Inc., Dallas, during 1993-1997. He has authored or coauthored more than 40 research publications in archival journals and refereed international conferences and has presented numerous invited talks and tutorials. At present, he serves on the organizing committee of the International Symposium on Quality Electronic Design (ISQED), and on the technical program committees of the EOS/ESD Symposium, ISQED, and the International Reliability Physics Symposium (IRPS).

Krishna C. Saraswat (Fellow, IEEE) received the B.E. degree in Electronics and Telecommunications in 1968 from Birla Institute of Technology and Science, Pilani, India, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1969 and 1974, respectively.

During 1969-1970, he worked on microwave transistors at Texas Instruments, Dallas, and since 1971, he has been with Stanford University, Stanford, CA, where he is presently a Professor of Electrical Engineering and Associate Director of the NSF/SRC Engineering Research Center for Environmentally Benign Semiconductor Manufacturing. During 1996-97 he was the Director of the Integrated Circuits Laboratory at Stanford. He is working on a variety of problems related to new and innovative materials, device structures, and process technology of silicon devices and integrated circuits. His special areas of interest are thin-film MOS transistors (TFTs) on insulator for 3-D multilayer ICs, thin-film technology for VLSI interconnections and contacts, process and equipment modeling, ultrathin MOS gate dielectrics, rapid thermal processing, and development of tools and methodology for simulation and control of a manufacturing technology. His group has developed several simulators for process, equipment and factory performance simulations, such as SPEEDIE for etch and deposition simulation, SCOPE for IC factory performance simulations, and a thermal simulator for RTP equipment design. Currently, he is also involved in the development of an interconnect process simulator. He has authored or coauthored more than 350 technical papers.

Prof. Saraswat is a Member of The Electrochemical Society and The Materials Research Society. He received the Thomas D. Callinan Award by The Electrochemical Society in May 2000 for his contributions to the dielectric science and technology. He was co-editor of the IEEE TRANSACTIONS ON ELECTRON DEVICES during 1988–1990.

Arifur Rahman (Member, IEEE) was born in Dhaka, Bangladesh. He received the B.S. degree from Polytechnic University, NY, in 1994 and the M.S. degree from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 1996, both in electrical engineering. Currently, he is pursuing the Ph.D. degree at MIT.

His Ph.D. dissertation work is in modeling system performance and technology requirements of three-dimensional integrated circuits. His research interests are device physics and interconnect modeling.

Mr. Rahman is a Member of Tau Beta Pi.

Rafael Reif (Fellow, IEEE) received the ingeniero electrico degree from Universidad de Carabobo, Valencia, Venezuela, in 1973, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1975 and 1979, respectively.

From 1973 to 1974, he was an Assistant Professor at Universidad Simon Bolivar, Caracas, Venezuela. In 1978, he became a Visiting Assistant Professor at the Department of Electrical Engineering, Stanford University. In 1980, he joined the Massachusetts Institute of Technology, Cambridge, MA, where he is currently a Professor in the Department of Electrical Engineering, and Computer Science, and the Associate Department Head for Electrical Engineering. Professor Reif was the Director of MIT's Microsystems Technology Laboratories (MTL) for the period 1990–1999. He is presently working on future interconnect technologies, and on environmentally benign replacement chemistries for microelectronics fabrication.

Dr. Reif held the Analog Devices Career Development Professorship of MIT's Department of Electrical Engineering and Computer Science, and was awarded the IBM Faculty Fellowship of MIT's Center for Materials Science and Engineering from 1980 to 1982. He also received a United States Presidential Young Investigator Award in 1984. He is a Member of Tau Beta Pi, the Electrochemical Society, and the American Physical Society.

James D. Meindl (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA.

He is the Director of the Joseph M. Pettit Microelectronics Research Center and has been the Joseph M. Pettit Chair Professor of Microelectronics at Georgia Institute of Technology since 1993. He was Senior Vice President for Academic Affairs and Provost of Rensselaer Polytechnic Institute from 1986 to 1993. He was with Stanford University from 1967 to 1986 as the John M. Fluke Professor of Electrical Engineering, Associate Dean for Research in the School of Engineering, Director of the Center for Integrated Systems, Director of the Electronics Laboratories and Founding Director of the Integrated Circuits Laboratory.

Dr. Meindl is a Fellow of the American Association for the Advancement of Science and a Member of the American Academy of Arts and Sciences and the National Academy of Engineering and its academic advisory board. He received a Benjamin Garver Lamme Medal from ASEE, an IEEE Education Medal, an IEEE Solid-State Circuits Medal, and an IEEE Beatrice K. Winner Award. He has also been awarded the IEEE Electron Devices Society's J. J. Ebers Award, the Hamerschlag Distinguished Alumnus Award, Carnegie Mellon University, the 1999 SIA University Research Award, and the IEEE Third Millennium Medal.