

VLSI on-chip interconnection performance simulations and measurements

by D. C. Edelstein
G. A. Sai-Halasz
Y.-J. Mii

We examine electrical performance issues associated with advanced VLSI semiconductor on-chip interconnections or "interconnects." Performance can be affected by wiring geometry, materials, and processing details, as well as by processor-level needs. Simulations and measurements are used to study details of interconnect and insulator electrical properties, pulse propagation, and CPU cycle-time estimation, with particular attention to potential advantages of advanced materials and processes for wiring of high-performance CMOS microprocessors. Detailed performance improvements are presented for migration to copper wiring, low- ϵ dielectrics, and scaled-up interconnects on the final levels for long-line signal propagation.

Introduction

As circuit density steadily increases in advanced logic and memory chips, and as transistor performance improves with each new chip generation, electrical performance of the on-chip wiring becomes increasingly significant. Raising the level of integration brings more circuit interconnections onto a chip, and although the device and wire dimensions are decreasing, the maximum chip size has actually been increasing. The parasitic loading of the

longest lines will thus increase, and the time is rapidly approaching when on-chip wiring delays will reach or exceed 50% of the cycle time for the fastest logic chips unless special measures are taken. The purpose of this paper is to study the impacts and limitations of wiring design and materials as they may relate to the performance of future semiconductor integrated circuits.

The growing importance of VLSI on-chip interconnects has been recognized by the semiconductor industry and the scientific research community, as indicated by the increasing volume of work presented in related conferences and journals [1-4]. The semiconductor industry has proposed significant improvements to ULSI interconnect technology in the most recently issued *National Technology Roadmap for Semiconductors* [5], and both industry and academia have worked strongly to advance the art of VLSI interconnect technology in the past few years. The majority of this recent work can be divided into several categories. The first of these addresses chip fabrication and unit processes: for example, how to increase the number of wiring levels, decrease dimensions, improve reliability and yield, and/or reduce cost through novel tools or processes. A representation of the current state of the art is described in [6], and includes reactive-ion-etched (RIE) aluminum-alloy lines, tungsten interlevel studs, and chemical-mechanical polishing (CMP) for global and local planarization of insulator and metal levels.

©Copyright 1995 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/95/\$3.00 © 1995 IBM

There is also a body of recent work devoted to bringing in new interconnect materials, such as insulators with lower dielectric constants than the industry-standard SiO_2 ($\epsilon \approx 4.0$), and metals with lower resistivities than the standard aluminum alloys ($\rho \approx 3\text{--}4 \mu\Omega\text{-cm}$). This work ranges from basic studies of materials properties in thin films to reliability of fully integrated multilevel structures. An example of such advanced work is reported in [7] which achieves multilevel integration of copper interconnects with a low-dielectric-constant polyimide, again using CMP for planarization and interconnect patterning. For most work on advanced materials, little, if any, data have been obtained to estimate the consequences of incorporating these materials at a system level.

A third recently emphasized category for VLSI interconnects comprises electrical performance modeling and measurements, and examination of the effect of interconnect parasitics on circuits and systems. Reference [8], for example, contains an extensive and thorough development of such issues. In order to handle the complexity of estimating CPU system performance, the interconnects are typically treated in a statistical and generalized fashion in terms of their electrical parameters, geometrical structure, and layout. The potential impact that modifications of the chip fabrication processes might have on the final interconnect parameters is not considered. Conversely, very detailed modeling and measurements of electrical behavior on actual interconnects have been performed, but here it is very difficult to make general system-level conclusions. The main purpose of this latter work is to verify the materials properties, or reconcile the modeling with the measurement. In this paper, we attempt to present a fairly comprehensive view of ultimate system performance as affected by the on-chip interconnects, as charted through details of the materials, measurements, and fabrication techniques that may be required.

Interconnect resistances and capacitances

The on-chip interconnect resistance and capacitance contribute to the CMOS gate delay, which can be approximated [8, 9] by

$$T_g = R_d(C_d + f_g C_r) + f_g R_d C_w l_w + 0.4 R_w C_w l_w^2 + 0.7 R_w C_r l_w, \quad (1)$$

where T_g is the overall gate delay, R_d and C_d are the driver FET (field-effect transistor) output resistance and capacitance, C_r is the receiver FET switching capacitance, R_w , C_w , and l_w are the wire resistance per unit length, capacitance per unit length, and length, and f_g is the gate fan-out. This equation approximates gate delay in which driver resistance is large compared to wire resistance, where capacitive shielding due to wire resistance [10] does not play a role. In general, the device parameters are not

constant with time, and delays are nonlinear functions of driver resistance relative to wire resistance. A consideration of these factors would require a more sophisticated treatment, as in [10], but is not necessary for a general discussion of the wire-related delays.

The first term on the right-hand side (RHS) of Equation (1) is the intrinsic delay associated with charging the driver and receiver loads. The second RHS term is the time required to charge the wire(s) with finite driver current; it is proportional to the interconnect capacitance and the driver resistance (which is inversely proportional to driver width). In principle, the driver could be widened to minimize this term (as in custom design), assuming that it is permitted by the budgeted power and area. The third RHS term is the distributed delay of the interconnect load itself, commonly referred to as the interconnect "RC delay." The fourth RHS term is the time it takes to charge the receiver input through the wire resistance; it becomes most significant for large-fan-out circuits fed by long wires, as in clock distribution networks.

It is useful to define a critical interconnection length l_c where the RC delay is half of the total gate delay [8]. Wire nets can then be considered "long" or "short" with respect to how they compare to l_c , and possibly treated accordingly. For future high-performance CMOS circuits, $l_c \approx 2\text{--}3 \text{ mm}$, so the number of long wire nets on a chip will be very significant. Because long wires strongly affect performance by their resistance, as in the fourth RHS term, this should be minimized (for example, by increasing the wire cross section). Short wires add delay primarily through capacitive loading of minimum-sized devices, as in the second RHS term; the wires should be reduced in cross section to minimize their capacitance. A wiring hierarchy has therefore been proposed [11, 12] in which all wires shorter than l_c are routed on lower-lying, minimum-pitch levels, and the relatively smaller number of long wires with scaled-up wire thickness and pitch are placed on upper levels. This is discussed further in the section on system performance.

It is important to look at the detailed behavior of the interconnect resistance and capacitance. This has been discussed frequently, but seldom in relation to the fabrication processes and non-primary materials (such as metal cladding or dielectric etch stops). While behavior can be estimated generally and even fairly accurately with simple assumptions about the interconnect electrical properties, for higher accuracy or comparisons of chip fabrication processes, these details are important and have a measurable impact. In this section, we present modeling of the interconnect resistance and capacitance as functions of materials and geometries, particularly applicable to those fabrication processes that have been demonstrated in our laboratories.

Resistances

Calculated resistances for aluminum- and copper-based interconnects are shown in **Figure 1**. The curves show "effective resistivity" (ρ_{eff}) as a function of linewidth down to $0.15 \mu\text{m}$, for lines with an aspect ratio of 1.0 and differing fabrication processes. The CMOS design rule which roughly corresponds to these linewidths is indicated on the top scale, and extends down to the $0.1\text{-}\mu\text{m}$ generation. The effective resistivity $\rho_{\text{eff}} = RA/l$ accounts for the parallel contributions of the principal metal, as well as any cladding made with other metals. This, rather than the principal-metal resistivity, is the more appropriate quantity to consider, since both patterning and capacitances are defined by the overall conductor dimensions, which include the cladding.

For a given linewidth and cladding thickness, ρ_{eff} depends on the principal metal, the cladding geometry, and any alloying reactions among the metals involved. The cladding geometry is determined by electromigration and corrosion concerns as well as the fabrication process. The two primary fabrication processes considered here are "metal-RIE" (or Al-RIE) and "damascene"; both are described in detail in companion papers in this issue. In metal-RIE, a blanket thin-film stack of Al alloy and cladding is deposited and then etched in a reactive plasma through a photoresist stencil. A dielectric is then deposited so that it fills the gaps between the lines as well as above them. In the damascene technique [13, 14], a blanket dielectric film is first deposited and patterned, and then cladding and base metal are deposited to fill the pattern. The excess metal (overburden) is then removed by chemical-mechanical polishing (CMP) to reveal the embedded interconnects. Both processes are used in the current state-of-the-art VLSI interconnects [6] which have Al-RIE wires and W-damascene studs. For copper, which is difficult to pattern by RIE, the damascene process is used exclusively to pattern lines and studs [7, 15, 16]. Because of the order of patterning and film deposition, metal-RIE leaves cladding at the tops and/or bottoms of the lines, while damascene lines have cladding on the bottoms and two sides. This is illustrated by the insets in Figure 1.

Figure 1 shows curves for Al(Cu) interconnects made by RIE or the damascene process, with "current" (solid line) or "future" (dashed line) cladding layers. The current Al-RIE process as practiced at IBM has bottom and top layers of Ti which react with the Al(Cu) during sintering steps to produce TiAl_3 , with a resistivity of $\sim 32 \mu\Omega\text{-cm}$ [17] that is ~ 3.5 times the original Ti thickness [18]. The future curve assumes that this reaction is somehow inhibited, while Ti is still available to reduce interfacial contact resistance (by reduction of metal oxides) and increase electromigration resistance. The future scheme may be overly optimistic, since the thick TiAl_3 is lost as an

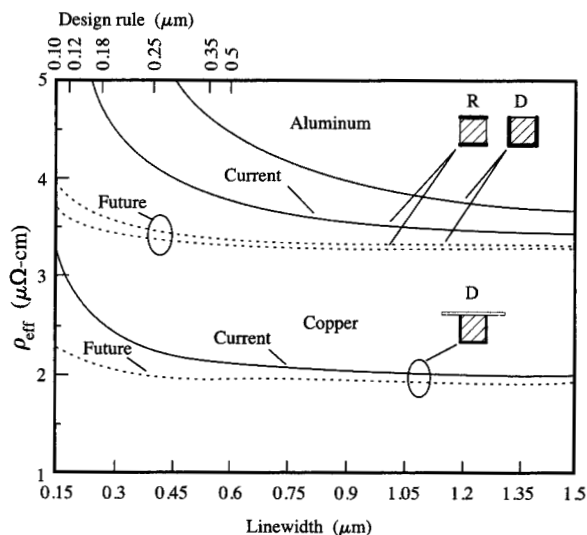


Figure 1

Effective resistivity (ρ_{eff}) vs. linewidth for square-cross-section aluminum and copper interconnects. Upper scale shows approximate CMOS design-rule generation. Insets indicate metal-RIE (R) vs. damascene (D) methods, and current (solid) or future (dashed) curves assume standard or thin cladding.

important redundant conductor to carry current in the event of an electromigration-induced failure.

The cladding for the copper interconnects is also assumed to have a high resistivity, but does not react with copper and serves only as an adhesion layer and diffusion barrier, and so can be made very thin. Many possible choices for this cladding have been studied [19–22]. The cladding is not required as a redundant conductor of current, since copper has such a large electromigration resistance [16, 23, 24], and it is not required to reduce contact resistance, which is typically very small for copper. Figure 1 shows curves for current and future copper wires, where the future cladding thickness has been reduced to $100 \text{ \AA}$. Although films this thin of some liner materials have been proven to perform well as copper diffusion barriers [22], it may be overly optimistic to assume that an acceptably low defect density will still be obtained. The curves in Figure 1 approach the base metal resistivities for large dimensions; these are the values typically obtained in our interconnect processes: $3.2 \mu\Omega\text{-cm}$ for sintered Al(0.5 at.% Cu), and $1.85 \mu\Omega\text{-cm}$ for Cu.

It is clear from Figure 1 that Cu offers a very significant reduction in resistance compared to Al, i.e., more than 40%. This is maintained from currently demonstrated to advanced future dimensions (cf. design rule generation on

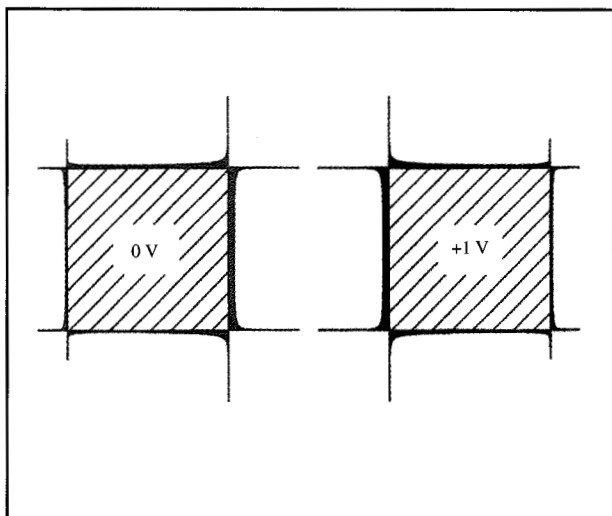


Figure 2

Calculated charge density (colored regions) on the surfaces of two infinite lines of square cross section and unity spacing, when charged to a potential difference of 1 V.

top axis). Even Cu with the current cladding compares favorably to future-cladded aluminum down to 0.25- μm linewidths, roughly corresponding to the 0.18- μm -design-rule generation projected to arrive in the early 2000s [5]. For lines of this size, thermal stress-voiding and electromigration may be paramount [25], and Cu would have a distinct advantage because of its superior mechanical properties. Furthermore, for logic chips with many wiring levels, such small dimensions might not be maintained at the upper levels (where resistance would be more important on the longer lines), so more of the resistance advantage of Cu could be recovered.

The resistance improvement obtained with Cu should improve CPU cycle time primarily by reducing long-line delays (since short interconnects have much lower resistance than the driver output impedance), but this advantage can also be shifted toward reducing interconnect capacitance by scaling down Cu wire cross sections relative to the original Al ones. This will also reduce crosstalk and improve manufacturability by reducing the wiring aspect ratios. This scaling should be justified by the exceptionally high electromigration resistance of good-quality Cu interconnects [24], which would support much higher current densities for the same lifetime at operating temperatures. If the technology allows, even further reduction of Cu cross sections could be made in cases where the wire resistance is not important. Thus, it appears that cladded Cu VLSI and ULSI interconnects

will maintain significant performance advantages over aluminum-based ones through the foreseeable future.

• Capacitances

The nature of the capacitance of high-aspect-ratio interconnects can be elicited with the help of **Figure 2**. This shows a graphical representation of the charge density induced on the surface of two square-cross-section lines (which extend into and out of the page) when charged to a potential difference of one volt, as calculated numerically using a two-dimensional capacitance program [26]. This shows that for typical on-chip interconnects, the induced charge density is substantially different from that of an ideal, parallel-plate (p-p) capacitor, which has uniform charge only on the two opposing conductor surfaces. Here mutually repulsive forces push the charge to the corners, until reverse forces build up to compensate. The result is large fringing fields emanating from the corners of the lines. In contrast to the ideal p-p case, electric fields also extend outside the two opposing surfaces, giving rise to additional surface charge on the outer faces. Because of the additional charge supplied to achieve a one-volt potential, by the relation $q = CV$ the capacitance is higher (almost 3 $\times$) than the p-p formula would predict. In fact, even at an aspect ratio of 4.0 (still at spacing of 1.0), the capacitance is 46% higher than that for the p-p calculation. This type of comparison has been studied more extensively in [27, 28].

Since on-chip interconnect capacitance is strongly determined by fringing fields, and since the geometry of the interconnects is in general three-dimensional, numerical rather than analytic simulation becomes necessary. The behavior of interconnect capacitances in various situations was extensively studied [28] using 2D [26] and 3D [29] numerical programs. Besides showing general trends, this work may provide useful capacitance values for present and future design generations, since these values remain constant for uniform scaling of the structure cross sections. Some of the salient results are presented here. For example, the presence of excess charge density in the wire corners prompts one to consider modifying their structure, e.g., by rounding the wires. Capacitances were compared for pairs of wires at constant pitch, as cross sections were scaled to reduce resistances. In the vicinity of unit dimensions for wire width (or diameter), thickness, and spacing, rounding the wires does not help substantially. As the metal area is increased at constant pitch, scaling up the height of rectangular wires clearly showed less increase in capacitance than increasing the radius of round wires or the width of rectangular ones. This underscores the trend in VLSI/ULSI technology to increase interconnect aspect ratios as pitches are decreased, in order to stave off rising resistances and current densities.

When more interconnects are brought in from upper and lower levels, charge on the central wires redistributes itself from the sides to the tops and bottoms of the wires. However, the additional charge that must be added to maintain the potential difference increases the total capacitance. The crosstalk component C_h/C_t (h = horizontal, t = total) drops proportionately much faster than C_t rises [28], until the vertical spacing between interconnects is comparable to the horizontal spacing. For interconnects of unity dimension and spacing, as in Figure 2, C_h/C_t drops from 100% to 50% by the addition of ground planes at four-unit spacing above and below, while C_t rises by only 38%. For ground planes at unity spacing, C_h/C_t drops to 20%, but C_t is now a factor of 2 higher than without the ground planes. On the basis of this behavior, interlevel distances somewhat larger (e.g., $2\times$) than the metal line thicknesses are recommended for VLSI/ULSI multilevel interconnects, to reduce the crosstalk without too large an increase in total capacitance. The difficulties presented by this approach are the high aspect ratios for interlevel via patterning and filling, and the trend toward increasing these ratios even further for future ULSI generations.

To study trends in multilevel interconnect capacitances, geometrical models such as that shown in Figure 3 were used. Here alternating, orthogonally running planes of wires (M1, M2, etc.) are separated by dielectric passivation. Levels M1 (bottom level) and M2 have all available wiring channels occupied, or full track occupancy, by metal wires; and M3 is either 67% or 100% occupied for various cases studied. The substrate and higher-lying levels are represented by infinite ground planes. Demarcation lines indicate breakup in overlap and crossing regions, planes of reflection symmetry, and periodic extension. Segments are further broken up into subintervals (not indicated). The structure is embedded in the principal passivating dielectric (not shown for clarity), and additional thin layers of other dielectrics may occur both below and above each wiring plane. For example, Al-RIE in SiO_2 has no additional dielectrics, whereas copper in SiO_2 [16] or polyimide [7] has either one or two thin layers of Si_3N_4 , respectively. The model assumes fully planarized wiring levels such as are present in VLSI chips which use chemical-mechanical polishing (CMP). Nonplanar topographies are much more difficult to model with the present program, but preliminary work indicates a substantial capacitance increase (e.g., $>20\%$) for nonplanarized geometries. Reduced capacitance is thus a performance benefit of CMP technology. As is widely recognized, CMP allows vertical stacking of constant-pitch levels using vertical-sidewall vias, as opposed to the scaled-up pitches required with tapered vias. This increases wirability and allows chip-size reduction.

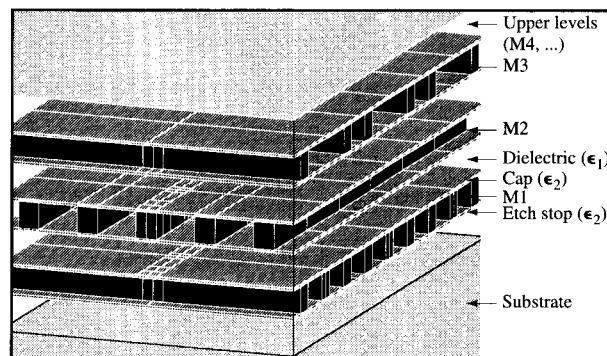


Figure 3

Three-dimensional model used for capacitance calculations, which includes three levels of planarized wiring (M1, M2, M3), ground planes for substrate and upper levels (M4, etc.), uniform dielectric (ϵ_1 , not shown for clarity), and optional thin cap and etch-stop planes of a different dielectric (ϵ_2). From [28], reproduced with permission.

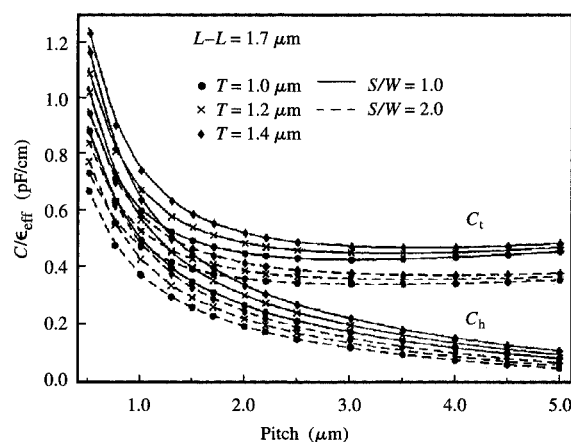


Figure 4

Normalized horizontal (C_h) and total (C_t) capacitances using the model of Figure 3. Results are shown for different wire thicknesses (T), interline-spacing/linewidth ratios (S/W), and one level-level ($L-L$) spacing. From [28], reproduced with permission.

Capacitances are calculated for the central wire on M2 relative to other wires on M2 (C_h), and to wires on M1, M3, and the ground planes. No conductors are assumed to be floating, so that worst-case capacitances are presented. An example of 3D wiring capacitance trends is shown in Figure 4, with curves for C_h and C_t for pitches (P) ranging

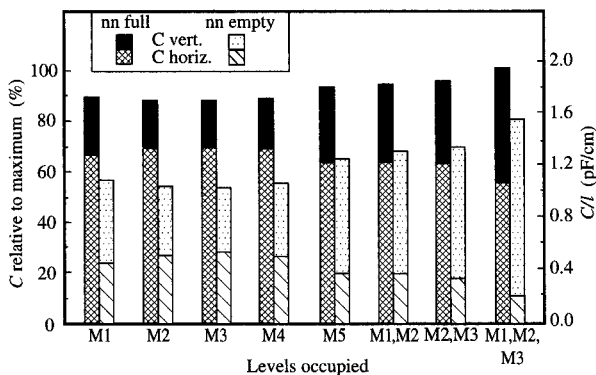


Figure 5

Three-dimensional capacitances vs. wiring track and level occupancy. Substrate ground plane and M6 wiring are assumed for all cases. Bars in each pair show horizontal and vertical capacitances for full (nn full) or empty (nn empty) nearest-neighbor wiring tracks adjacent to the wire in question.

from 0.5 to 5 μm , metal thickness (T) from 1.0 to 1.4 μm , interlevel spacing ($L-L$) of 1.7 μm , and two ratios of interline spacing/linewidth (S/W) of 2.0 (dashed lines) and 1.0 (solid lines). The curves show capacitance per unit length normalized to the effective relative dielectric constant of the insulator system (ϵ_{eff}), and so are more generally applicable to other insulator systems. The values S/W of 2.0 and 1.0 reflect typical practice for bipolar ECL and CMOS logic chips, respectively. Although the wire thicknesses shown here are more applicable to the former chip type, future generations of high-speed CMOS will be increasingly affected by wiring resistance, and will also require fairly thick wires. Since capacitances remain constant for uniform scaling of the model cross section, projections to future ULSI dimensions at similarly high aspect ratios may be made.

A number of trends are indicated in Figure 4. First, a broad, nearly flat minimum exists for wire aspect ratios of 1.0 or less. For example, this value is ~ 1.8 – 2.0 pF/cm for CMOS wires ($S/W = 1.0$, $\epsilon_{\text{eff}} = 4.0$) in a SiO_2 dielectric. For Al(Cu) wiring with Ti cladding (cf. Figure 1, $\rho_{\text{eff}} \approx 3.6 \mu\Omega\text{-cm}$), the RC product (RC delay) would be 650 ps/cm (260 ps/cm^2) at room temperature. At 85°C (a typical operating temperature), these numbers would increase by 27%, and at 77 K they might drop by a factor of 3–5. The capacitances in Figure 4 increase rapidly for higher aspect ratios and submicron design rules ($P < 2 \mu\text{m}$), since the wire thickness is not being scaled. C_i also starts increasing at large pitches because of the large parallel-plate interaction with the vertical levels, which are not moving away. This is quite relevant to current chip designs, where

certain long wires must be increased in width to reduce resistance. Because of this effect, the RC product cannot be reduced beyond 2 – $3\times$ by widening the wires [9]. In the figure, C_h increases rapidly through the flat C_i region, and reaches 50% of C_i near aspect ratios of 1.0. Future ground-rule scaling trends will lead to crosstalk components of 65% or more, unless new materials are selected which allow a reduction in the scaling trend. Among other general observations made from this modeling are the following: The central M2 wire is effectively shielded by its nearest neighbors, and thus has virtually no coupling to the ground planes or the second-nearest neighbors; and the vertical capacitances are only $\sim 10\%$ less than what they would be if M1 and M3 were solid ground planes. As is shown in the next figure, capacitances of wiring on dense levels will not be greatly reduced even by substantial depopulation of over- and underlying wiring.

Figure 5 shows results of 3D modeling to study the effects of wire loading on the central wire, as this wire is placed on different levels. Here all levels have 100% track occupancy unless otherwise noted, and the wiring dimensions used are appropriate for the $0.35\text{-}\mu\text{m}$ CMOS design-rule generation. As in Figure 3, the model includes a ground plane for the substrate; in contrast, however, the upper ground plane is replaced by actual wiring on the M6 level only. The stacked-bar heights represent total capacitance as constituted by horizontal (hatched fill) and vertical (shaded fill) components, as a function of wiring track occupancy above, below, and adjacent to the central wire. The two bars in each pair show capacitances for either fully loaded (left bar, “nn full”) or empty nearest-neighbor tracks (right bar, “nn empty”) on the same level as the central wire, which can be on any level from M1 to M5. The left vertical scale shows percent values relative to the fully loaded case, which for SiO_2 dielectric would be 1.91 pF/cm, as seen on the right vertical scale. It is seen that for the “nn full” case, regardless of the presence or absence of interconnects on adjacent levels (except at substrate and M6), or the level on which the central wire resides, C_i varies by only 12%. The crosstalk varies more, but remains in the 50–70% range. Omission of wires from nearest-neighbor tracks lowers C_i somewhat more (21–35%), mainly through the large drop in C_h which is partially compensated by a rise in the vertical component. The primary point of the results in Figure 5 is that as long as there is *some* metal above, below, or beside a high-aspect-ratio interconnect (i.e., the usual case on a chip), the total capacitance can only be reduced to $1/2$, or, more typically, $2/3$ of the fully loaded value, and then only at the expense of track occupancy on that level. Deleting neighboring wires is an effective way to reduce crosstalk, again at the expense of wiring density. These results have implications for interconnect routing and optimization as

well; fairly coarse assumptions can be used for extracting routed interconnect parameters, and changing vertical loading is not a very viable option for reducing capacitances.

Besides layout and dimensions, detailed dielectric structure affects 3D wiring capacitances. For example, in damascene processes, the inclusion of even very thin high-dielectric planes at the top and/or bottom of each wiring level causes a disproportionately high increase in capacitance, due to their location in the high fringing field areas. The model of Figure 3 was used to calculate 3D capacitances versus variation in the principal dielectric (ϵ_1) and thin cap and etch-stop dielectric planes (ϵ_2) above and below the wires. Figure 6 shows resulting "effective" relative dielectric constants $\epsilon_{\text{eff}} = C(\epsilon_1, \epsilon_2)/C(\epsilon_1 = \epsilon_2 = 1.0)$ for different ratios of dielectric constants ϵ_2/ϵ_1 vs. volumetric proportion of the ϵ_2 material, $V_{\epsilon_2}/(V_{\epsilon_1} + V_{\epsilon_2})$. The chosen ratios represent some realistic materials possibilities, such as $\text{Si}_3\text{N}_4/\text{polyimide}$ or $\text{SiO}_2/\text{fluoropolymer}$ (2.414); $\text{Si}_3\text{N}_4/\text{SiO}_2$ (1.8); $\text{SiO}_2/\text{polyimide}$ (1.414); and $\text{SiBN}/\text{polyimide}$ (1.2) [30]. It is clear that the curves are much higher than volume-proportional weighting of the two dielectrics would predict. [This would yield straight lines emanating from the origin and intersecting the curves at $V_{\epsilon_2}/(V_{\epsilon_1} + V_{\epsilon_2}) = 1.0$.] This figure shows the consequence of a large capacitance of even a very thin high-dielectric material, if it is present in the fringing-field regions of the wires. Unfortunately, there may not be obvious choices for low- ϵ materials which perform the needed functions as well. For example, Si_3N_4 is a very good Cu diffusion barrier, whereas SiO_2 is not. Both have excellent RIE selectivity for O_2 plasmas which may be needed to etch very low- ϵ materials such as polymers. However, even if low- ϵ SiO_2 were used to cap a fluoropolymer (with $\epsilon = 2.3$), the SiO_2 cap would cause the dielectric capacitance to increase considerably. If a Si_3N_4 film were required to cap Cu interconnects, and an SiO_2 film were used for the lower RIE stop, the advantage of using that polymer would be nearly eliminated, e.g., versus a glass with $\epsilon \approx 3.5$ and no cap.

Finally, a general comparison is made across potential ULSI interconnect technologies for 3D capacitances and interconnect RC products. This is shown in Figure 7, and is similar to previous comparisons presented in [28, 31]. Here the 0.35- μm CMOS-generation 3D model was used, as above. Wiring with advanced materials such as polyimide (PI, $\epsilon = 3.0$), fluoropolymer (FP, $\epsilon = 2.3$), and copper is compared to the standard Al(Cu)/Ti wiring in PECVD SiO_2 (Ox, $\epsilon = 4.1$), which has a capacitance of 1.9 pF/cm and an RC constant of 3.2 ns/cm². The examples using Cu include cap and etch-stop films of $\epsilon = 4$ and thicknesses less than 0.1 μm , for all cases except Cu/Ox, where one Si_3N_4 cap is assumed. The Al/PI case assumes a single thin SiO_2 cap on top of the polyimide. Examples are

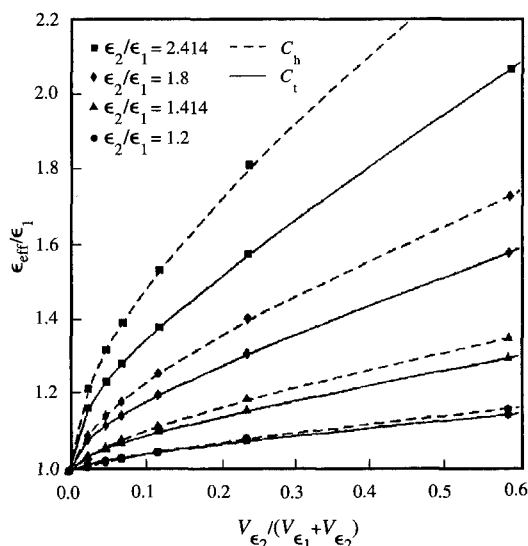


Figure 6

Effective relative dielectric constant (ϵ_{eff}) normalized to principal dielectric (ϵ_1) vs. volume ratio of thin cap and etch-stop planes (ϵ_2 , cf. Figure 3). Horizontal (dashed lines) and total (solid lines) components are shown for various material combinations, ϵ_2/ϵ_1 . From [28], reproduced with permission.

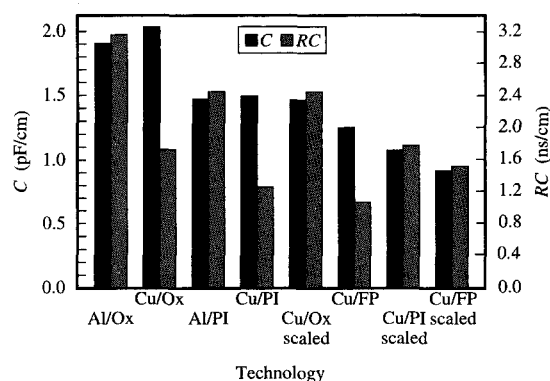


Figure 7

Three-dimensional capacitance and RC for multilevel 0.5- μm interconnects made in various technologies. Ox = SiO_2 ($\epsilon = 4.1$), PI = polyimide ($\epsilon = 3.0$), and FP = fluoropolymer ($\epsilon = 2.3$). Appropriate dielectric caps and etch stops are included in the calculations. In three cases the copper line cross sections are scaled by the Cu/Al resistance ratio.

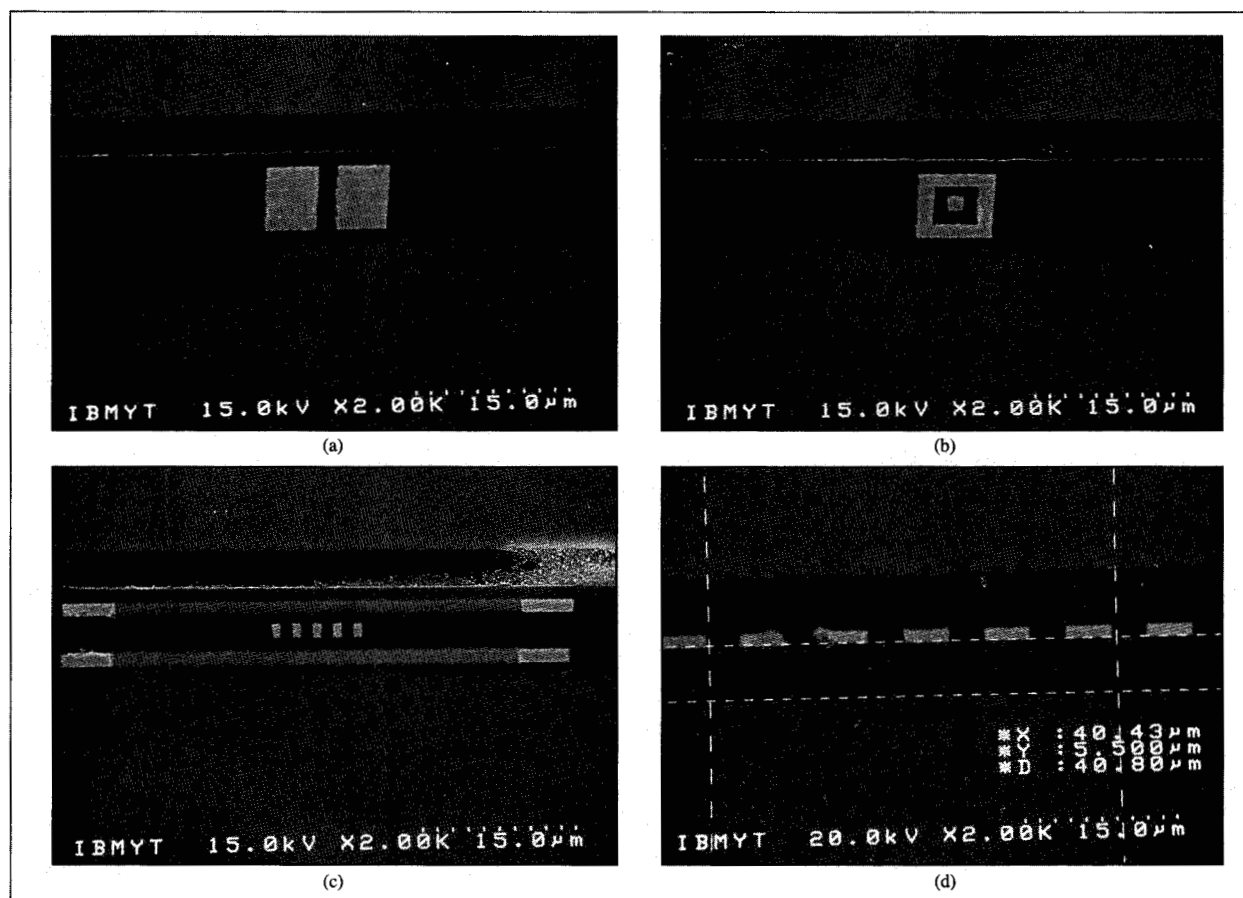


Figure 8

Cross-sectional SEM micrographs of multilevel damascene copper/polyimide interconnects: (a) balanced coplanar waveguide; (b) coaxial line; (c) five-wire transmission line; (d) interdigitated comb capacitor.

also shown for Cu wiring with cross sections scaled to meet the original Al wire resistance. It is shown that scaled Cu in SiO_2 provides as much drop in capacitance as Al in a dielectric of 3.0 (25% drop). The use of scaled Cu, instead of Al, wiring is equivalent to advancing the interconnect dielectric by one technology generation. Ultimately, unscaled or scaled Cu/FP offers a $3\times$ drop in RC or a $2\times$ drop in capacitance, respectively. With hierarchically scaled wiring levels, as described in the section on system performance, both advantages could be utilized to optimize performance.

Measurements

In order to verify modeling and projections of interconnect parasitics, electrical measurements must be performed on interconnect structures at the relevant dimensions, and with the actual materials and processing history to be

implemented. This requires additional techniques beyond the typical resistance, capacitance, and loaded ring-oscillator measurements. Examples of measurement methods developed to address this requirement [32] are presented in this section, including dielectric anisotropy and high-frequency loss, and propagation of picosecond electrical pulses on micron-scale thin films and interconnects.

Figure 8 shows cross-sectional SEM micrographs of some structures built for these measurements. The Cu/polyimide damascene process [7] was used to fabricate these single- and multilevel interconnects, which include (a) balanced coplanar waveguides, (b) coaxial transmission lines, (c) multi-wire transmission lines with crossed-line loading, and (d) interdigitated comb capacitors. Wafers were processed through one, two, and three wiring levels (with no, one, or two via levels) to allow a variety of measurements at different stages of processing. Wafer

substrates include silicon of various doping levels, and fused silica ("quartz"), the latter to eliminate conductive substrate effects on propagation and capacitances.

• *Polyimide dielectric anisotropy and loss*

Interdigitated comb capacitors were used to measure 2D and 3D capacitances, to verify the modeling, to check for influence of processing on material properties, and to extract in-plane dielectric constants of $\sim 1\text{-}\mu\text{m}$ insulator films. These films may be inhomogeneous when processed in structures [33], or may possess anisotropic susceptibilities. For example, in [33] comb capacitor measurements were used in conjunction with 2D capacitance modeling to extract the dielectric constant of PECVD oxides deposited in the narrow gaps between the capacitor metal lines. It was found that certain conventional deposition processes yielded material with a substantially higher dielectric constant (e.g., 5.0) between the lines, compared to the blanket-film values (e.g., 4.0). Evidently the chemistry of the CVD process is altered in the high-aspect-ratio gaps during filling, and this is relevant to standard industry manufacturing processes. In the present work, the in-plane dielectric constants ($\epsilon_{\parallel}$) of $\sim 1\text{-}\mu\text{m}$ BPDA-PDA and PMDA-BTFF [34] polyimides, and parylene films were obtained. (There is interest in these low- ϵ insulators for their potential use in VLSI/ULSI chips.) This work was motivated by predictions of an in-plane value for BPDA-PDA much higher than the more easily measured out-of-plane ($\epsilon_{\perp}$) dielectric constant value of 2.9–3.0, based on optical index of refraction measurements. It was surmised that this was due to a high degree of ordering of the polymer chains in thin films, for polymer chains which possess anisotropic susceptibilities [35].

Obtaining the in-plane dielectric constant is facilitated by use of the Cu damascene process. By means of this process, Cu interconnects are embedded in the pre-cured and patterned polyimide film with tops and bottoms flush with the film/substrate and film/air interfaces. During capacitance measurement, electric field lines in the polymer are identically horizontal, while the curved fringing-field lines are present only in the substrate and air, which have known, isotropic dielectric values. Thus only the in-plane (x - y) dielectric constant $\epsilon_{\parallel}$ of the unknown material is sampled. If the damascene process were not used, polyimide would have to be applied to fill gaps in pre-patterned wires, and some molecules would orient to the vertical interfaces. The fringing-field lines (with x , y , and z -components) would be present in the polymer. This would produce a completely different measurement result, which would not necessarily agree with extrapolations from optical index of refraction measurements in blanket films, where the x - y and z -components of birefringence are separately determined.

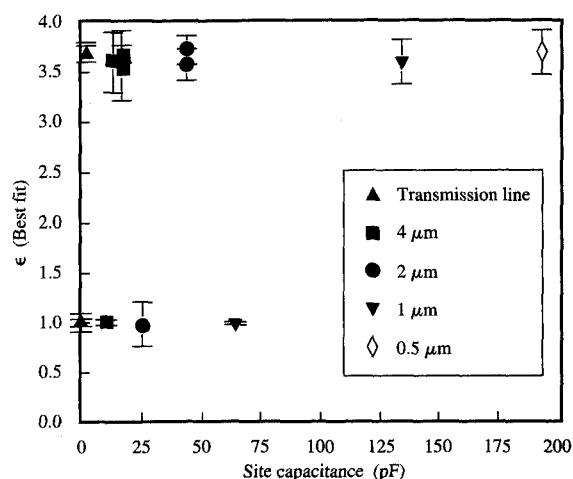


Figure 9

In-plane dielectric measurement of $1.2\text{-}\mu\text{m}$ BPDA-PDA polyimide films using damascene comb capacitors with the indicated gap spacings, and a one-level balanced coplanar transmission line. Upper (lower) data show dielectric constant ϵ extracted from capacitances before (after) removal of the polyimide. From [32], reproduced with permission; © 1993 IEEE.

Numerical 2D capacitance modeling [26] was used to extract $\epsilon_{\parallel}$ and errors from measured capacitances. Accuracy is further increased by obtaining an expression relating $\epsilon_{\parallel}$ to the difference between two capacitance measurements: $\epsilon_{\parallel} = f(C - C_0)$, where C_0 has a known dielectric such as air between the lines. In the metal-RIE structures, capacitance is simply measured before and after deposition of the unknown dielectric. In the damascene case, capacitance is measured first for the unknown case, and then again after selectively removing the dielectric in question. Modeling is then used to obtain the sensitivities of $\epsilon_{\parallel}$ to variations in physical dimensions. SEM cross-sectioning with accurate measurements of linewidths and film thicknesses is used to provide model input and error estimation.

Figure 9 shows data obtained with the above procedure for $1.2\text{-}\mu\text{m}$ -thick films of BPDA-PDA polyimide. The in-plane dielectric constant and error bars are shown for comb capacitors with indicated interline spacings. All were fabricated on insulating quartz wafers to eliminate coupling to a conductive substrate such as Si. The capacitance was remeasured after the polymer was removed, and the model was verified at $\epsilon = 1.0$. The results show $\epsilon_{\parallel} = 3.65 \pm 0.06$ in the kHz-MHz regime, independent of gap spacing (as expected, since the polymer was cured before patterning).

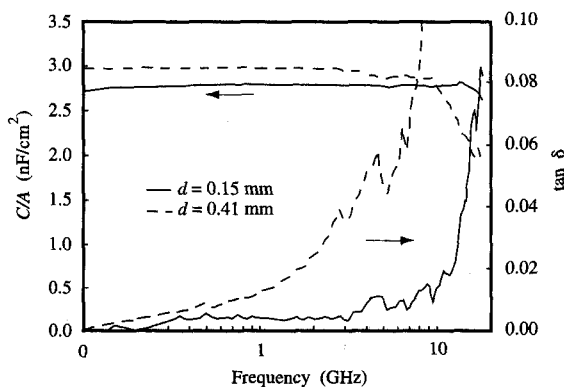


Figure 10

Dielectric loss ($\tan \delta$, solid lines) and capacitance/area (dashed lines) for BPDA-PDA metal-insulator-metal capacitors of two diameters (d).

The combined error is better than 2%, and includes contributions from the capacitance measurement, site and wafer variation, metrology, and modeling. The result agrees with predictions based on in-plane index of refraction measurements in the optical [35], THz [36], and GHz [37] regimes, and is 25% higher than the out-of-plane value $\epsilon_{\perp} = 2.9$, which can be measured in blanket films on Si with metal dot capacitors. Similar work has obtained $\epsilon_{\perp}/\epsilon_{\parallel}$ of 2.6/2.7 for PMDA-BTFFB and 2.4/2.4 for parylene, this time when the dielectric was deposited and cured to fill gaps in metal-RIE comb capacitors.

Using the above BPDA-PDA dielectric constants, 3D modeling finds an effective dielectric constant of 3.6 for dense submicron multilevel wiring with thin Si_3N_4 etch-stop layers. This polyimide was originally chosen for interconnect integration on the basis of its high stiffness, high thermal stability, and (assumed) low dielectric constant; the stiffness is related to high molecular ordering, and in turn (in this case) to high dielectric anisotropy. Thus, the advantage of replacing SiO_2 with BPDA-PDA in a damascene structure is minimal. Other polyimides, such as PMDA-BTFFB, have since been introduced. These may have mechanical properties and thermal stability similar to those of BPDA-PDA, but with dielectric constants below 3.0 [34] that are much more isotropic.

Dielectric loss was also obtained in 1- μm films to ~ 10 GHz, as shown in **Figure 10**. Here metal-insulator-metal vertical capacitors were formed by depositing and curing the polyimide on metallized, high-conductivity Si substrates, followed by evaporation of gold dots with various diameters. Commercial coplanar microwave probes

and a vector network analyzer were used to measure the complex, frequency-dependent reflection coefficient (s_{11}), from which the impedance is obtained. The experimental resolution in the loss constant ($\tan \delta$) was 0.002. **Figure 10** shows extracted capacitance (normalized to capacitor area) and $\tan \delta$ of 1.2- μm BPDA-PDA polyimide under ambient conditions, for 0.15- and 0.41-mm-diameter capacitors. The loss is flat and low (≤ 0.004) up to frequencies where parasitic inductances of the system start to dominate. The onset frequency for this rapidly rising loss was found to scale inversely with capacitor size, as evidenced by the two sets of data in **Figure 10**. This implicates parasitic LRC resonance of the sample and probing system, as opposed to dielectric loss, which should not depend on capacitor size. Without further modeling and measurement beyond the parasitic resonance, meaningful data are obtained only below this point. The measured loss in the flat region is low enough not to affect on-chip signal propagation, especially on relatively lossy on-chip interconnects [38], and this measurement covers the range of frequencies relevant to signals on advanced CMOS digital computer chips, at least for clock frequencies up to a few hundred MHz. Thus, we have shown with the preceding measurements that it is possible to obtain all necessary dielectric properties of advanced materials in thin films and structures pertinent to interconnects on high-performance VLSI chips. The dielectric constant results may be combined with modeling data from the previous section to predict present and future interconnect capacitances.

• Picosecond pulse propagation

In principle, complete knowledge of the interconnect electrical parameters and accurate modeling should allow prediction of pulsed (ac) behavior on computer chips. However, simplifying assumptions are often necessary to make the modeling tenable, and it is prudent to obtain measurement verification of any critical predictions. Beyond this, it is of scientific interest to explore the limits of high-frequency measurements on very small structures, such as on-chip interconnects. To this end, very high-frequency pulse-propagation data were obtained from some of the transmission lines represented in **Figure 8**. The measurements were performed with 35- and 3-ps-rise-time electrical pulses. The former were achieved with a commercial time-domain system and high-frequency probes, and the latter with a novel optoelectronic sampling system and probes developed here [39, 40] and described below. The 3-dB bandwidths for the two techniques are 10 GHz and 100 GHz, respectively. **Figure 11** shows time-domain reflectometry (TDR) data and simulations for one-level Cu/PI balanced coplanar waveguides (cf. **Figure 9**) on both quartz and 2- $\Omega\text{-cm}$ Si substrates. Two lengths, 5.5 and 11 mm, were measured for each. Modeling was

performed for the simpler case of quartz substrates using ASTAP [41], which allows for distributed and skin effects. Excellent agreement between modeling and measurement is obtained for the waveguide on quartz when distributed effects are included. Two other simulations show results neglecting skin effects and resistance in the 11-mm-long metal. As shown, the skin effect increases the propagation delay, while the resistance gives significant attenuation and some damping of ringing oscillations. These lines have dimensions characteristic of long-run signal lines on VLSI chips (5 μ m wide, 2- μ m gap), and the measurement bandwidth is also relevant for high-performance processors. Clearly, distributed effects and skin effects show up significantly on these interconnects, and must be included in any relevant circuit simulations.

When measured on Si substrates, there is large coupling of the probe pads and transmission line to the conductive substrate; this gives rise to the negative initial dip and degraded rise time of the data. Also there is a slow-wave effect [42, 43] from inductive coupling to a thick, lossy Si conductor, causing $\sim 2\times$ increased delay, and more attenuation. The relationship between the slow-wave effect and substrate doping at larger line dimensions and similar frequencies has been studied for microstrip lines [43], where it was pointed out that maximal slow-wave effects happen to occur at typical CMOS Si substrate doping levels, in the frequency range of interest for high-performance VLSI applications. However, typical CMOS chips do not generate electric fields in the substrate as in the microstrip configuration, and the long interconnects likely to be affected by additional propagation delay might reside on upper levels of the chip, and could thus be shielded from the substrate by lower-lying wiring. In order to test the validity of these assumptions, high-frequency probing on multilevel VLSI interconnects will be necessary. This would be especially appropriate for multiple, parallel, and simultaneously switching wires such as in a data path, where inductive coupling effects would be combined among the wires [44].

The same waveguides were measured with the 3-ps, 100-GHz optoelectronic system, and data are shown in Figure 12. This system uses single picosecond pulses from a mode-locked laser ("pump" pulses) to momentarily short out a photoconductive gap between dc-biased transmission-line conductors fabricated on a movable wafer probe. Each pump pulse creates an electrical transient which propagates down the probe and onto the device under test. Reflections are picked up by the same probe, and transmitted signals by an identical probe contacting the other end of the sample. A second laser pulse ("probe" pulse), optically delayed from the pump pulse in a controllable manner, momentarily shorts one of the transmission lines to a ground conductor. By measuring the photocurrent through this detection circuit

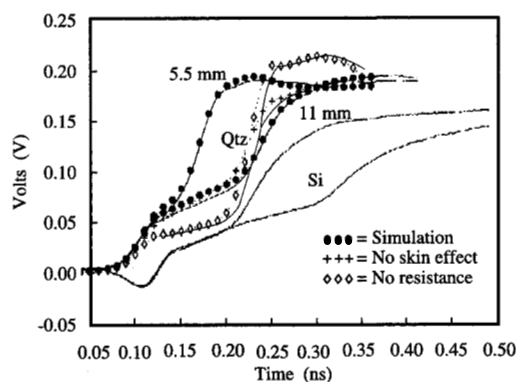


Figure 11

Time-domain reflectometry data (solid lines) with 35-ps pulses on one-level copper/polyimide coplanar waveguides of lengths 5.5 and 11 mm, built on fused-silica (Qtz) and Si substrates. Simulations for both Qtz films are shown with distributed effects ($\bullet$). For the 11-mm film, simulations are also shown neglecting skin effect (+), and neglecting resistance ($\diamond$). From [32], reproduced with permission; © 1993 IEEE.

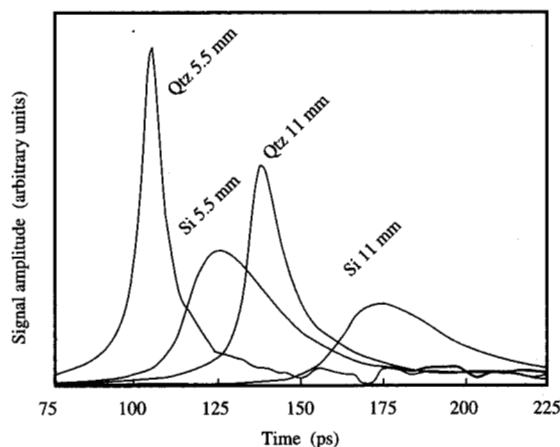


Figure 12

Time-domain transmission on the lines measured in Figure 11, using 3-ps pulses from an optoelectronic probing system. (Time zero is arbitrary.) From [32], reproduced with permission; © 1993 IEEE.

as a function of pump-probe time delay, a sampling measurement is obtained which is the convolution of the sample response with the laser pulse autocorrelation. The system achieves a 3-ps time resolution and a 10- μ V sensitivity after two minutes of data acquisition, for a

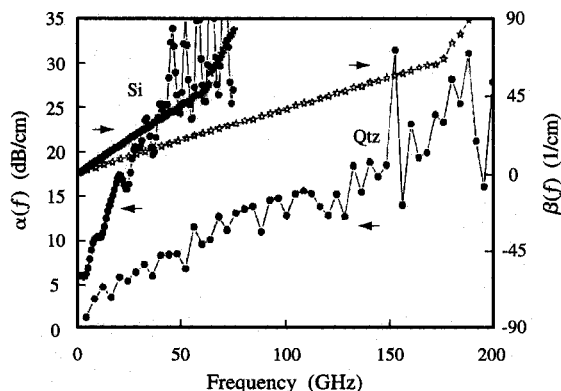


Figure 13

Propagation constant $\Gamma(f) = \alpha(f) + j\beta(f)$ obtained using Equation (2) and data from Figure 12. From [32], reproduced with permission; ©1993 IEEE.

300-ps measurement scan. Figure 12 shows time-domain transmission (TDT) data for the waveguides measured in Figure 11 (the time zero is arbitrary here). The frequency content of the data extends to nearly 200 GHz, which is more than adequate for digital as well as microwave circuit applications. A new technique [45] was used to analyze these data. This technique obtains the complex propagation constant $\Gamma = \alpha + j\beta$ from TDT data on two lengths of otherwise similar lines, effectively canceling out probe and pad parasitics. Multiple reflections can be windowed out of the time-domain data, thereby avoiding the complicating effect that reflections have on analysis of conventional frequency-domain measurements. After windowing, the data are Fourier-transformed, and the attenuation (α) and phase (β) constants are obtained from the ratio and difference, respectively, of components from the two measurements as

$$\Gamma(f) \equiv \alpha(f) + j\beta(f) = -\frac{1}{(l_1 - l_2)} \ln \frac{A_1(f)}{A_2(f)} + j \frac{\phi_1(f) - \phi_2(f)}{(l_1 - l_2)}, \quad (2)$$

where f is frequency, and A_i and ϕ_i are respectively the amplitude and phase of Fourier-transformed TDT data from lines of length l_i with $l_1 > l_2$. The relationships between components of $\Gamma(f)$ and the transmission-line impedances are derived in [45] as well.

Figure 13 shows results of this analysis on the data of Figure 12. These curves completely describe the transmission-line ac behavior, and can be used directly to propagate arbitrary input waveforms through arbitrary

lengths of such lines. The attenuation constant for the quartz substrate shows a small narrow feature at 154 GHz which is believed to arise from deliberately designed small notches spaced 1 mm apart on the transmission line. These serve as tick marks for the propagation delay and confirm a value of $\tau_d = 65$ ps/cm. This delay is predicted by modeling of the Cu/PI transmission line on the $\epsilon = 3.85$ quartz substrate with the in-plane polyimide dielectric measurement of Figure 9 (resulting in $\epsilon_{\text{eff}} = 3.80$), and the expression $\tau_d = (c/\sqrt{\epsilon_{\text{eff}}})^{-1}$. The value of 65 ps/cm is also obtained from the $\beta(f)$ curve as $\tau_d(f) = \beta(f)/2\pi f$, and is constant in the frequency range for valid data (limited by signal/noise to below 175 GHz). The $\alpha(f)$ data should be dominated by metal resistance (including skin effect), and do not seem to show any unusual behavior due to dielectric loss ($\tan \delta$), although no attempt was made to extract $\tan \delta$ from the data. The value of $\tan \delta$ for this polyimide was measured in this frequency range in [37], and was below the experimental limit of 0.01. Thus the agreement among various measurements in this section, as well as those made by other authors, verifies the novel measurement methods used here, and shows uniform dielectric behavior of this polyimide to extremely high frequencies. Figure 13 also shows results from the Si substrate sample. Here, frequencies beyond 50 GHz are capacitively shunted to the substrate (i.e., the sample is a low-pass filter), and the propagation velocity obtained from the $\beta(f)$ curve is 1/2 that on the quartz substrate because of the slow-wave effect. The loss is significantly higher than for the same Cu interconnects on quartz because of the conductive effects of the Si substrate.

Data for the multilevel transmission lines, including the coaxial line, have also been obtained with both 10-GHz and 100-GHz systems, although not presented here. Propagation delays, cross-coupling, and even- and odd-mode propagation have been observed, but are much more difficult to model for the multiconductor systems. For two-conductor systems such as the coaxial interconnect and multilevel balanced coplanar waveguide [Figures 8(a) and 8(b)], only single-mode propagation is supported, and the propagation is readily shown to match the modeling predictions, as in the case above. These measurements indicate that full characterization of interconnect structures in the VLSI size range is possible in time and frequency domains, to well beyond the highest-performance digital logic chips envisioned in silicon.

System performance

Once the details of interconnect parameters have been characterized, the final stage of analysis is to determine their impact on the performance of the system, or central processing unit (CPU). The motivation for this becomes especially apparent for future high-performance ULSI microprocessors. For current CPUs with cycle times

around 10 ns, wire RC delays (typically less than 1 ns) are barely noticeable. However, these wire delays will remain no matter how fast the active devices become, and will constitute an unacceptably high proportion of the few-ns cycle time of a large-chip CPU unless a different approach is taken, especially with regard to interconnect resistance [11, 12]. To study this in detail, we use recently devised CPU performance estimator routines [9, 12, 46, 47] which are based on wiring statistics and details of the processor size, architecture, and wiring configuration.

Two sets of CPU performance estimator routines, [12, 46, 47] and [9], are largely similar in approach and lead to similar trends and conclusions despite a few differences in assumptions. Similarities exist with earlier cycle-time estimators (e.g., [8]), which make predictions by considering the chip technology, design, and architecture parameters. In [8], calculations are based on the total number of logic gates and average-length interconnects, and do not account for long interconnects, fan-out, or custom-sized gate widths. The more recent models are more detailed, though they still rely on some assumptions in selecting parameters for a generic CPU model. The description here pertains to the model first described in [12].

• Model description and assumptions

The cycle-time model incorporates, to first order, all of the major factors which would influence high-performance processors. Although results presented here are derived for single-chip (uni-) processors, they are also applicable to multichip CPUs with chip crossings and package interconnects. (It is assumed that even for massively parallel computing, there will always be a premium on the highest-performing uniprocessor.) Results here are also limited to CMOS processors, although a comparison has been made of the ultimate limits of bipolar (ECL) and CMOS CPUs [46, 47]. Emphasis is placed on performance limits at the high end, as opposed to low-power or minimal-chip-size CPUs of lower performance and complexity.

The model accounts for the number of circuits, the size of the memory arrays, the chip size (and number), the number of wiring levels, wiring parameters, propagation effects, input-output needs, lithography design rules, gate length, gate oxide thickness, power supply voltage, and room-temperature or liquid-nitrogen-temperature (LN_2) operation. Circuit timings are calculated with simple, linearized equations, whose coefficients capture the essential properties of device scaling. The model has some features which distinguish it from previous cycle-time estimators, including several effects based on empirical experience. These include accounting for multiple nets and fan-out, individual pitches, capacitances, and resistances at each wiring level, finite signal time-of-flight, empirical correction factors between theoretical and experimental

wiring results, the effect of wiring channel blockage between levels, increased loading of a critical path, space consumed by power and ground routing, and alternative wiring schemes such as repeaters, wide wires, and "fat" wires (defined below).

In order to estimate CPU cycle times at a given complexity, the chip size must be determined. This in turn requires a detailed assessment of the needed chip area for the devices and wiring. The total needed wire length is obtained from wiring statistics [48–50]. Wiring statistics involve empirical factors, which are not yet well established for CMOS CPUs. The parameter with the most influence is the Rent exponent [51, 52]. In the modeling it is assumed that a CPU chip consists of a small number of building blocks, each with a high internal Rent exponent, but the interconnections between the blocks are governed by a smaller exponent. Many cases were investigated to determine the influence of differing choices of the Rent exponent. It was found that this variation has practically no effect on performance, but does strongly influence power consumption.

The first step used in estimating wiring needs is to determine the average net length, $\bar{L}_{net}$, from the expression [50]

$$\bar{L}_{net} = f_t P (1 + 0.1 \log N_c) [1 + 0.3(f_g - 1)] N_c^{(R_i - 0.5)}, \quad (3)$$

where f_t is a numerical "fudge" factor, P is the circuit pitch (or average distance between circuits), N_c is the number of circuits on the chip, f_g is the average gate fan-out, and R_i is the internal Rent exponent. The value of f_t was found empirically to be ~ 0.6 [50]. Wires longer than about half a chip-edge length do not follow the statistics of the average wires. The number of long wires is obtained more empirically, on the basis of experience with the signal I/O count behavior when the CPU is divided into a small number of chips. In the model it is assumed that if the number of chips in the processor is decreased by a factor of 2, half of the off-chip connections which become absorbed on the new chip end up as long wires. The total wire length needed by the chip is then obtained as the sum of the average nets, $N_c \bar{L}_{net}$, plus the long wires. The available length for wiring on the chip must then be obtained. First the total track length in the logic area is calculated from the pitches and wire levels. Power and ground are assumed to take up 20% of each level for CMOS. Next, the effect of blockages is computed (see below). The fractions lost to power/ground and track blockages are subtracted from the total length to obtain the track length available for interconnections. Finally, the constraint is imposed that the needed wiring length must be less than 40% of the available track length for the chip to be wirable. (This figure is based on the wire length that

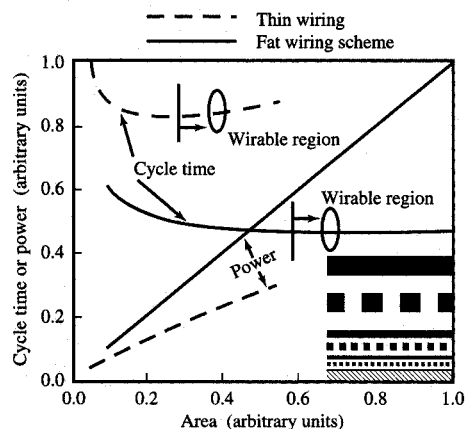


Figure 14

Cycle time and power consumption for a high-end CMOS CPU vs. chip size (area), when wired with standard thin wiring (dashed lines) or the fat-wire scheme (solid lines), as depicted in the inset. Demarcations show lower limits of chip area based on assumed wiring statistics, below which the chips may not be wirable.

is likely to be achieved with good wiring tools.) This obtains the minimum-sized wirable CPU.

The blockage of channels on one level from upper levels must be considered in assessing CPU wirability. Via connections from upper levels block tracks in all levels below them, down to the substrate. If all the pitches are identical, it is estimated from empirical data that a level blocks 12–15% of the channels on all levels below it. This implies that CPU chips cannot be made indefinitely smaller by adding wiring levels (six or seven would be the most), and there is a minimum chip size defined by the interconnections [53].

Once the processor size has been obtained, the area occupied by the memory cells is subtracted; 50% of the remaining area is assumed usefully available (because of layout constraints) to connect the logic circuits. The area to be populated with devices must accommodate long-line off-chip and on-chip drivers, and possibly repeaters as well. Both the on- and off-chip drivers are inverters with unity fan-in and fan-out, with p-MOS gates twice as wide as n-MOS gates. The widths of these drivers are calculated in relation to the characteristic impedance of the interconnects. After allowance is made for the drivers and receivers, the remaining area is divided evenly among two-input NAND gates. Finally, on the basis of allotted area and the assumed design rules, device widths are made to be as large as possible in order to maximize performance.

The next step is to calculate gate delay for an assumed critical path in the CPU. For this the device characteristics

are estimated as a function of design-rule technology. Detailed circuit simulations were carried out at 0.25- μm gate length using the ASTAP program [41], and a performance projection curve was obtained by numerical fitting of experimental large-signal transconductance data [54, 55] for gate lengths from 0.28 μm to 0.07 μm . For short lines, gate delays can then be calculated from Equation (1) above. However, for long, low-RC lines in a critical path, transmission-line effects begin to surface. (If RC delays dominate, the transmission-line characteristics are damped out [38].) In a SiO_2 insulator, the minimum propagation delay is $\tau = \sqrt{\epsilon/c} = 67$ ps/cm, but a more realistic result based on the calculated impedances might be [38] $\tau = \sqrt{LC} = 85$ ps/cm. For chip-length lines, this delay may be significantly longer than the switching time of drivers in the considered technologies; thus, the finite signal propagation speed must be taken into account.

In the modeling this is handled by sizing drivers for an internal impedance of about half of the wire characteristic impedance. This ensures adequate voltage on the wire while minimizing power and noise. To account for the heavier-than-average loading in the critical paths, the 0.3 factor multiplying the f_g term in Equation (3) is doubled to 0.6. Finally, resistive and time-of-flight delays, calculated separately, are combined with the capacitive and intrinsic delay components. From extensive ASTAP modeling, it was found that the best empirical formula for adding time-of-flight delay (τ_{tof}) and RC delay (τ_{RC}) to obtain total wire delay (τ_w) is [47] $\tau_w = (\tau_{\text{tof}}^{1.6} + \tau_{\text{RC}}^{1.6})^{1/1.6}$.

• Results

The model CPU under investigation consists of one million two-way NAND gates with fan-out $f_g = 2$. (Choosing two-way NANDs as basic building blocks serves the modeling purposes well, but does not imply that real processors would be implemented in this manner.) This system comprises four million FET transistors for logic circuits and over twelve million in the memory cells. The critical path assumed here contains 30 NAND-gate stages. All but one of the stages are connected with the average wire length of the critical paths, and the remaining stage drives a chip-edge-length wire through an inverter buffer. If there is no package delay, the inverter buffer and long wire delays are counted twice. Circuit driving capability is maximized through the choice of device widths, as in custom design.

Figure 14 shows cycle time and power as a function of chip area used to accommodate the number of circuits stated above, assuming 0.35- μm CMOS-generation ground rules. This illustrates the general situation in high-performance CMOS processors. It is presented in relative, arbitrary units (a.u.), since this behavior is qualitatively the same for all large systems when the CPU can fit onto one or two chips. To the left of the arrows, the processor may or may not be wirable. This boundary is characterized

by an internal Rent exponent of 0.55. The dashed lines show data for conventional, minimum-pitch, and thickness wiring on six levels, where long wires are widened as much as possible to reduce RC delay. Optimized repeaters are also added to reduce the long-line delays. It turns out that the cycle times represented for this CPU do not approach the potential offered by the high-performance devices, because of the dominance of the resistive delay of the chip-length interconnect in the critical path. The fact that the cycle time immediately begins to rise as area is increased in the wirable region indicates that the CPU is limited by the long-line delay. Repeaters do help reduce long-line delays, since there they decrease the dependence of delay on wire length from quadratic to linear [8, 11]. Still, at some point, even this linear dependence will dominate. Cycle times with optimized repeaters were also studied in [9] and [31]. For the 0.35- μm generation, performance with repeaters was improved by less than 10% compared to the wide-wire approach, and this required six repeaters per long line. Adding a large number of repeaters would entail additional power consumption, loss of Si chip area, loss of wiring efficiency, and added design complexity.

To avoid the long-line RC problem, a better approach has been suggested [11, 12], and that is to provide two kinds of wires for high-performance processors. First, thin wires are provided that serve the vast majority of circuits. These wires are typically internal to functional blocks, at most 1–3 mm in length but with much shorter average length, scaled to the minimum pitch available, and predominantly on the first two or three wiring levels. These wires are mainly responsible for making the chip wirable by providing a sufficient number of circuit interconnections. For these “short” wires, RC delay may not be appreciable, although the $R_o C_w$ term in Equation (1) may still contribute measurably to circuit delay and power. Thus, the main way to tailor these wires to increase performance is to lower their capacitance. Second, there is a need for long wires, where density is secondary to delay considerations. These interconnections were previously part of the package but are now integrated onto the chip. They run between distant parts of the chip, and are from several mm to chip-edge length. They must be designed so that their time for signal propagation is as small a fraction as possible of the CPU cycle time. It follows that the cross section of these wires and insulators cannot follow minimum ground rules. For example, in [9] it was shown that widening thin wires can only reduce RC by at most 3 $\times$, which is not nearly enough. This implies a split in the needs for low-power or cost-driven chips vs. complex high-performance chips with respect to wiring requirements.

The inset of Figure 14 depicts such a proposed wiring scheme (“fat wires”) [12, 46, 47], which was implemented

in the model to generate the solid-line curves. This scheme uses a scaled hierarchy of x - y plane-pairs, with dimensions scaled uniformly such that capacitance remains constant while resistance and RC decrease inversely as the cross-sectional area increases. In the model, these pitches are freely adjusted to optimize performance. Although it might seem that using fat wires wastes too many wiring channels, the difference is not as large as it might appear. The blockage between levels of varying pitch scales in proportion to pitch, so a fat-wire level on top of the chip provides less wiring capability than one at a fine pitch, but a fine-pitch level at the top would affect the wirability of all of the lower levels more severely.

The fat-wire approach (solid lines) reduces the cycle time in this case by nearly a factor of 2 over the thin wide-wire approach (dashed lines), without many of the penalties associated with repeaters, and the cycle time is now close to its unloaded delay. As opposed to the thin wide-wire cycle-time curve, the fat-wire curve is seen to decrease with increasing logic area in the wirable region. This is because the driver widths are scaled with chip area, while the interconnection load scales as the square root of the area, so more current is available to drive the loads [47]. The advantage gained by increasing area is eventually reversed by finite signal time-of-flight, and the cycle time would actually begin to rise because of this if the figure were extended. The main impact of larger chip areas is increased power consumption. Across the wirable region, the cycle time remains practically unchanged, but the power nearly doubles. Power is proportional to switching capacitance (C_{sw}) and clock frequency (f) as $P = C_{sw} V^2 f$. The thin wide-wire chip uses significantly less power, in part because minimum pitch levels allow a smaller wirable chip (therefore a smaller wire load), but mainly because the clock frequency is so much lower. With the fat-wire approach, multi-hundred-MHz CPUs should be possible on multi- cm^2 , fully populated chips at these ground rules, and ultimately clock frequencies approaching 1 GHz might be achievable at room temperature in future CMOS generations [47].

The fat-wire scheme is a design option which would tend to remove the dependence of performance on wire resistance, and thus increase the choice of wiring material. At this point it becomes necessary to reevaluate the benefits of changing to Cu and low- ϵ dielectrics. Figure 15 shows results from [31], where the model was used to compare performance for different wiring materials (metal/ ϵ , symbols) with the use of repeaters and wide, thin wires where needed (solid lines) or the fat-wire scheme (dashed lines). Instead of chip size, the number of allowed wiring levels is indicated. In this work a smaller chip was modeled, so the impact on the long lines is somewhat less, and scaling of Cu cross sections was not assumed. In [31], calculations were also shown for cases where neither fat

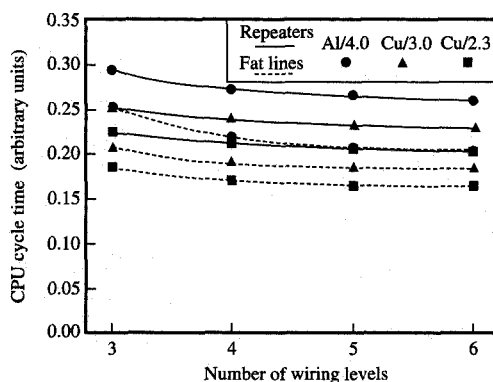


Figure 15

Normalized cycle time vs. number of needed wiring levels for a CMOS CPU wired with different materials (symbols, metal/ ϵ), and use of optimized repeaters (solid lines) or fat wires (dashed lines). From [31], reproduced with permission; © 1993 IEEE.

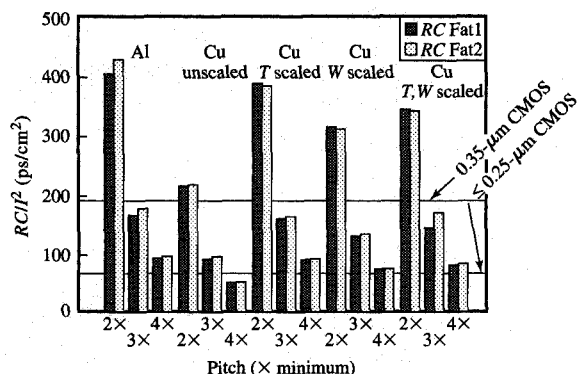


Figure 16

RC constants from 3D modeling for two levels of fat wires (Fat1, Fat2) on top of minimum-pitch levels vs. fat-wire pitch (2 $\times$, 3 $\times$, 4 $\times$ minimum pitch of 1.4 μ m). Wires are Al with square cross section, Cu at same dimensions, or Cu scaled in thickness (T), width (W), or both (T,W) to match Al-wire resistance. Interlevel spacing is scaled equal with line thickness. Target RC values (horizontal lines) might be necessary to wire large, ultrahigh-performance CPUs at the indicated CMOS design rules.

wires nor repeaters were assumed. Very substantial improvements (e.g., 2 $\times$) are made by introducing copper and low- ϵ dielectric without fat wires, but the ultimate cycle times are not competitive with either the repeater or the fat-wire schemes. Once fat wires are used to remove

the dependence on long-wire resistance, performance is improved primarily by reducing the capacitances of average-length lines. At this point, reducing capacitance by 40% and RC by 60% was still able to reduce the cycle time by 17%. If Cu wires on the lower levels were scaled to reduce capacitance by another 25%, the overall performance might improve by nearly 30%.

• Fat-wire technology

The dimensions of fat wires necessary to meet the above performance goals may tax the interconnect fabrication technology. Although the pitches and widths are readily achieved, patterning high-aspect-ratio wires and vias with such large film thicknesses may be beyond the practical limits of current processes and tooling for metal-RIE, metal filling of vias, SiO₂-RIE, and SiO₂ CMP planarization. Furthermore, contact must be made from the lower fat-wire level to a minimum-pitched level through a thick interlevel dielectric, requiring either landing pads (which block channels) or a very aggressive via aspect ratio. Although the performance premium on low metal resistivity is lifted by fat wires, differences in fabrication methods and needed wire sizes for a given RC value will ultimately determine which metal can be used for these interconnects.

Figure 16 shows 3D calculations of RC constants for several sizes of two-level Al and unscaled or scaled Cu fat wires on top of dense, minimum-pitched wiring. The unscaled lines are square to minimize RC, with equal width, space, and interlevel separation of half the indicated pitch. Three pitches are chosen at two, three, and four times the minimum pitch of 1.4 μ m. Final dielectric passivation is also included in the model, and is respectively either nonplanar or planar for metal-RIE or damascene processes. The cases for Cu wires include unscaled, and scaled in thickness (T), width (W), or both (T,W) to match the Al resistance. (Since the interlevel spacing is also scaled with line thickness, full reduction of capacitance is not realized in this scaling.) Pairs of bars show RC for lower (Fat1) and upper (Fat2) wiring levels. Two horizontal dashed lines indicate approximate RC targets for cross-chip fat wires needed to achieve ultimate cycle times for a 2-cm, six-level-metal CMOS CPU at the indicated design-rule generations.

We observe that Al fat wires come close to the target for the 0.35- μ m-generation chip at 4.2- μ m pitch, and for future generations at 5.6- μ m pitch. However, the current processes and tooling used for Al may not readily support fabrication of unity-aspect-ratio 2.8- μ m lines, spaces, and vias and associated nonplanarities. Cu wires achieve these RC targets at a one-step reduction of pitch, or without pitch reduction, for example, by scaling T to substantially reduce all film thicknesses by $\sim 2\times$. Alternatively, linewidth scaling for Cu offers the largest decrease in

capacitance (and thus driver power), and a very large decrease in crosstalk, which is a critical issue for long parallel-running lines. From a fabrication standpoint, there are no known limits encountered in the damascene process at these particular dimensions [56].

Concluding remarks

We have studied trade-offs in on-chip interconnect design from a process, materials, and performance standpoint for future high-end CMOS uniprocessors, and have linked these in detail to possible future interconnect technologies that are currently being investigated. We have also shown that electrical measurements with high resolution, accuracy, and very high frequency can be made on these interconnects to characterize dielectric and propagation behavior in the on-chip environment. In some cases, the measurements were shown to agree quite precisely with electrical modeling, which allows one to perform further projections based on the modeling alone.

It was shown that Cu interconnects may offer almost a $2\times$ drop in resistance compared to Al through the foreseeable future, where realistic cladding and processes were taken into account. Judicious scaling of Cu, permitted by its high electromigration resistance to support higher current densities, allows as much capacitance reduction for Cu/SiO₂ as with Al/low- ϵ alternatives, which are currently receiving much attention from the industry. Bringing in low- ϵ dielectrics in conjunction with Cu might allow a $3\times$ reduction of RC on unscaled wiring levels, a $2\times$ reduction of capacitance on scaled levels, and possibly more than 20% CPU performance improvement after other wiring design options have been exhausted. As other factors for improvement begin to saturate, this becomes a very significant quantity. Furthermore, any material change that can decrease the capacitance of all the interconnects will drop the power (at the same performance) very significantly, first by reducing the wire load on the drivers, and then by allowing both drivers and repeaters to shrink while still maintaining adequate device-to-wire load ratio. Power reduction for high-performance CPUs will increase in importance as clock frequencies increase, and power is always an issue for lower-performance CPUs such as those in battery-powered portable systems.

Large, complex future processors will suffer a severe loss in performance if long-line resistances are not reduced substantially, for example by means of fat wires. The fat-wire scheme reduces the RC problem to coping with time-of-flight delays, which for CMOS is a much less severe restriction on performance. If fat wires are adopted, metal resistivity will not have as major an impact on CPU speed as the circuits approach their intrinsic, unloaded delays. This near-ultimate performance would be achievable with Al fat wires, if permitted by the fabrication technology.

However, fat wires are much more accessible with Cu because of the smaller dimensions required and the greater extendability of the damascene process. In either case, on the basis of the estimations we have presented, it seems fair to say that silicon CMOS technology has very exciting prospects for growth in performance, so long as careful attention is paid to the interconnects.

References

1. *Proceedings of the 12th International VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1995.
2. *Proceedings of the Symposium on VLSI Technology*, IEEE Electron Device Society and Japan Society of Applied Physics, 1994.
3. *Proceedings of the International Electron Devices Meeting*, IEEE Electron Device Society, 1994.
4. *MRS Bull.* **19** (1994).
5. *The National Roadmap for Semiconductors*, Semiconductor Industry Association, San Jose, CA, 1994.
6. C. W. Koburger III, W. F. Clark, J. W. Adkisson, E. Adler, P. E. Bakeman, A. S. Bergendahl, A. B. Botula, W. Chang, B. Davari, J. H. Givens, H. H. Hansen, S. J. Holmes, D. V. Horak, C. H. Lam, J. B. Lasky, S. E. Luce, R. W. Mann, G. L. Miles, J. S. Nakos, E. J. Nowak, G. Shahidi, Y. Taur, F. R. White, and M. R. Wordeman, "A Half-Micron CMOS Logic Generation," *IBM J. Res. Develop.* **39**, 215 (1995).
7. B. Luther, J. F. White, C. Uzoh, T. Cacouris, J. Hummel, W. Guthrie, N. Lustig, S. Greco, N. Greco, S. Zuhoski, E. Colgan, S. Mathad, L. Saraf, E. J. Weitzman, C.-K. Hu, F. Kaufman, M. Jaso, L. P. Buchwalter, S. Reynolds, C. Smart, D. Edelstein, E. Baran, S. Cohen, C. Knoedler, J. Malinowski, J. Horkans, L. Deligianni, J. Harper, P. Andricacos, J. Paraszczak, D. J. Pearson, and M. Small, "Planar Copper Polyimide Back End of the Line Interconnections for ULSI Devices," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1993, p. 15.
8. H. B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addison-Wesley Publishing Co., Reading, MA, 1990.
9. Y. Mii, "Performance Consideration for the Scaling of Sub-Micron On-Chip Interconnections," *Proc. SPIE* **1805**, 332 (1992).
10. F. Dartu, N. Menezes, J. Qian, and L. T. Pillage, "A Gate-Delay Model for High-Speed CMOS Circuits," *Proceedings of the 31st ACM/IEEE Design Automation Conference*, 1994, p. 576.
11. P. M. Solomon, "The Need for Low Resistance Interconnects in Future High-Speed Systems," *Proc. SPIE* **947**, 104 (1988).
12. G. A. Sai-Halasz, "Directions in Future High-End Processors," *ICCD Digest*, p. 230 (1992).
13. M. Chow, W. L. Guthrie, and F. B. Kaufman, "Method of Forming Fine Conductive Lines, Patterns, and Connectors," U.S. Patent 4,702,792, 1987.
14. C. W. Kaanta, S. G. Bombardier, W. J. Cote, W. R. Hill, G. Kerszykowski, H. S. Landis, D. J. Poindexter, C. W. Pollard, G. H. Ross, J. G. Ryan, S. Wolff, J. E. Cronin, "Dual Damascene: A ULSI Wiring Technology," *Proceedings of the 8th International IEEE VLSI Multilevel Interconnect Conference*, Santa Clara, CA, 1991, p. 123.
15. C.-K. Hu, M. B. Small, F. Kaufman, and D. J. Pearson, "Copper-Polyimide Wiring Technology for VLSI Circuits," *Mater. Res. Soc. Symp. Proc. VLSI V*, 369 (1990).
16. D. C. Edelstein, "Performance and Integration of Copper VLSI/ULSI Interconnects," *Proceedings of the American Vacuum Society 41st National Symposium*, 1994, p. 123.

17. C.-K. Hu, N. J. Mazzeo, and C. Stanis, "Electromigration in Two-Level Bamboo Grain Structure Al(Cu)/W Interconnections," *Mater. Chem. Phys.* **35**, 95 (1993).
18. J. Tardy and K. N. Tu, "Solute Effect of Cu on Interdiffusion in Al₃Ti Compound Films," *Phys. Rev. B* **32**, 2070 (1985).
19. C.-K. Hu, S. Chang, M. B. Small, and J. E. Lewis, "Diffusion Barrier Studies for Cu," *Proceedings of the 3rd International IEEE VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1986, p. 181.
20. E. Kolawa, P. J. Pokela, J. S. Reid, J. S. Chen, R. P. Ruiz, and M. A. Nicolet, "Sputtered Ta-Si-N Diffusion Barriers in Cu Metallizations for Si," *IEEE Electron Device Lett.* **12**, 321 (1991).
21. S.-Q. Wang, "Barriers Against Copper Diffusion into Silicon and Drift Through Silicon Dioxide," *MRS Bull.* **19**, 30 (1994).
22. T. Nakano, H. Ono, T. Ohta, T. Oku, and M. Murakami, "Diffusion Barrier Properties of Transition Metals and Their Nitrides for Cu Interconnections," *Proceedings of the 11th International VLSI Multilevel Interconnection Conference*, 1994, p. 407.
23. C.-K. Hu and B. Luther, "Electromigration in Two-Level Interconnects of Cu and Al Alloys," *Mater. Chem. Phys.* **41**, 1 (1995).
24. C.-K. Hu, B. Luther, F. B. Kaufman, J. Hummel, C. Uzoh, and D. J. Pearson, "Cu Interconnection Integration and Reliability," *Thin Solid Films* **262**, 84 (1995).
25. C.-K. Hu, K. P. Rodbell, T. D. Sullivan, K. Y. Lee, and D. P. Bouldin, "Electromigration and Stress-Induced Voiding in Fine Al and Al-Alloy Thin-Film Lines," *IBM J. Res. Develop.* **39**, 465 (1995, this issue).
26. W. T. Weeks, "Calculation of Coefficients of Capacitance of Multiconductor Transmission Lines in the Presence of a Dielectric Interface," *IEEE Trans. Microwave Theory Tech.* **MTT-18**, 35 (1970).
27. A. E. Ruehli and P. A. Brennan, "Accurate Metallization Capacitances for Integrated Circuits and Packages," *IEEE J. Solid-State Circuits* **SC-8**, 289 (1973).
28. D. C. Edelstein, "3-D Capacitance Modeling of Advanced Multilayer Interconnection Technologies," *Proc. SPIE* **1389**, 352 (1990).
29. J. F. Janak, D. D. Ling, and H.-M. Huang, "C3DSTAR: A 3D Wiring Capacitance Calculator," *Proceedings of the IEEE International Conference on Computer-Aided Design*, 1989, p. 530.
30. M. Maeda, T. Makino, E.-I. Yamamoto, and S. Konaka, "A Low-Permittivity Interconnection Using an SiBN Interlayer," *IEEE Trans. Electron. Devices* **36**, 1610 (1989).
31. J. Paraszczak, D. Edelstein, S. Cohen, E. Babich, and J. Hummel, "High Performance Dielectrics and Processes for ULSI Interconnection Technologies," *Proceedings of the IEEE International Electron Devices Meeting*, 1993, p. 261.
32. D. C. Edelstein, M. Scheuermann, and L. M. Geppert, "Picosecond Propagation on Multilevel Copper/Polyimide Back End of the Line Interconnections," *Proceedings of the 10th International VLSI Multilevel Interconnection Conference*, 1993, p. 511.
33. G. C. Schwartz, Y.-S. Huang, and W. J. Patrick, "The Effective Dielectric Constant of Silicon Dioxides Deposited in the Spaces Between Adjacent Conductors," *J. Electrochem. Soc.* **139**, L118 (1992).
34. J. W. Labadie, H. Lee, D. Boese, D. Yoon, W. Volksen, P. Brock, Y. Cheng, M. Ree, and K. R. Chen, "Polyimides with Reduced Dielectric Anisotropy," *Proceedings of the 43rd IEEE Electronic Components Technology Conference*, 1993, p. 327.
35. D. Boese, D. Y. Yoon, J. D. Swalen, and J. F. Rabolt, "Chain Orientation and Anisotropies in Optical and Dielectric Properties in Thin Films of Stiff Polyimides," *J. Polym. Sci. B, Polym. Phys.* **30**, 1321 (1992).
36. M. Ree, K.-J. Chen, D. P. Kirby, N. Katzenellenbogen, and D. Grischkowsky, "Anisotropic Properties of High-Temperature Polyimide Thin Films: Dielectric and Thermal-Expansion Behaviors," *J. Appl. Phys.* **72**, 2014 (1992).
37. W. Robertson, G. Arjavalingam, G. Hougham, G. Kopcsay, and D. Edelstein, "Microwave Properties of Free-Standing Dielectric Films," *Electron. Lett.* **28**, 62 (1992).
38. A. Deutsch, G. V. Kopcsay, V. A. Ranieri, J. K. Cataldo, E. A. Galligan, W. S. Graham, R. P. McGouey, S. L. Nunes, J. R. Paraszczak, J. J. Ritsko, R. J. Serino, D. Y. Shih, and J. S. Wilczynski, "High Speed Signal Propagation in Lossy Transmission Lines," *IBM J. Res. Develop.* **34**, 601 (1990).
39. M. Scheuermann, R. Sprik, J.-M. Halbout, P. A. Moskowitz, and M. Ketchen, "Ultra-High Bandwidth Detachable Optoelectronic Probes," *Proc. Picosec. Electron. Optoelectron.* **4**, 22 (1989).
40. D. C. Edelstein, R. B. Romney, and M. Scheuermann, "Rapid Programmable 300 ps Optical Delay Scanner and Signal-Averaging System for Ultrafast Measurements," *Rev. Sci. Instrum.* **62**, 579 (1991).
41. *ASTAP—Advanced Statistical Analysis Program*, Order No. SH20-1118-0, available through IBM branch offices.
42. H. Guckel, P. A. Brennan, and I. Palöcz, "A Parallel-Plate Waveguide Approach to Microminiaturized, Planar Transmission Lines for Integrated Circuits," *IEEE Trans. Microwave Theory Tech.* **MTT-15**, 468 (1967).
43. H. Hasegawa, M. Furukawa, and H. Yanai, "Properties of Microstrip Lines in Si-SiO₂ System," *IEEE Trans. Microwave Theory Tech.* **MTT-19**, 869 (1971).
44. D. A. Priore, "Inductance on Silicon for Sub-Micron CMOS VLSI," *Proceedings of the IEEE Symposium on VLSI Circuits*, 1993, p. 17.
45. A. Deutsch, G. Arjavalingam, and G. V. Kopcsay, "Characterization of Resistive Transmission Lines by Short-Pulse Propagation," *IEEE Micr. Guided Wave Lett.* **2**, 25 (1992).
46. G. A. Sai-Halasz, "High End Processor Trends and Limits," *Proceedings of the International Conference on Advanced Microelectronic Devices and Processing*, Sendai, Japan, 1994, p. 753.
47. G. A. Sai-Halasz, "Performance Trends in High-End Processors," *Proc. IEEE* **83**, 20 (1995).
48. W. E. Donath, "Placement and Average Interconnection Lengths of Computer Logic," *IEEE Trans. Circ. & Syst.* **CAS-26**, 272 (1979).
49. W. E. Donath, "Wire Length Distribution for Placements of Computer Logic," *IBM J. Res. Develop.* **25**, 152 (1981).
50. W. R. Heller, C. G. Hsi, and W. F. Mikhail, "Wireability—Designing Wiring Space for Chips and Chip Packages," *IEEE Design & Test of Computers* **1**, 43 (1984).
51. B. S. Landman and R. L. Russo, "On a Pin or Block Relationship for Partitions of Logic Graphs," *IEEE Trans. Computers* **C-20**, 1469 (1971).
52. W. E. Donath, "Equivalence of Memory to 'Random Logic,'" *IBM J. Res. Develop.* **18**, 401 (1974).
53. W. E. Pence and J. P. Krusius, "The Fundamental Limits for Electronic Packaging and Systems," *IEEE Trans. Components, Hybrids & Manuf. Technol.* **CHMT-10**, 176 (1987).
54. G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. A. Rishton, and E. Ganin, "High Transconductance and Velocity Overshoot in NMOS Devices at the 0.1 μ m-Gate-Length Level," *IEEE Electron Device Lett.* **EDL-9**, 464 (1988).

55. G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. A. Rishton, E. Ganin, T. H. P. Chang, and R. H. Dennard, "Experimental Technology and Performance of 0.1- μ m-Gate-Length FETs Operated at Liquid-Nitrogen Temperature," *IBM J. Res. Develop.* **34**, 452 (1990).
56. D. C. Edelstein, "Advantages of Copper Interconnects," *Proceedings of the 12th International VLSI Multilevel Interconnection Conference*, Santa Clara, CA, 1995, p. 301.

Received February 13, 1995; accepted for publication August 11, 1995

Daniel C. Edelstein *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (EDELSTN at YKTVMV, edelstn@watson.ibm.com).* Dr. Edelstein received the B.S., M.S., and Ph.D. degrees in applied physics from Cornell University in 1982, 1986, and 1990, respectively. His thesis work concentrated on ultrafast laser spectroscopy of III-V semiconductors, and novel ultrafast source and measurement techniques. There he invented the femtosecond optical parametric oscillator and the optical delay scanner, both of which later became commercial products. From 1982 to 1989, he worked at or was on educational leave from the IBM Thomas J. Watson Research Center, where he has been a research staff member since 1989. At IBM, he has worked on various aspects of advanced interconnection technology, including process and materials development, electrical modeling, experimental measurements, and the development of advanced and novel structures. He currently leads a project on the integration, performance, and reliability of copper VLSI/ULSI interconnects.

George A. Sai-Halasz *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (SAI at YKTVMV, sai@watson.ibm.com).* Dr. Sai-Halasz graduated in physics in 1966 from Roland Eötvös University, Budapest, Hungary. He received the Ph.D. in physics from Case Western Reserve University in 1972. From 1972 to 1974 he held a postdoctoral fellowship at the Physics Department of the University of Pennsylvania. He joined the IBM Thomas J. Watson Research Center in 1974. Dr. Sai-Halasz has made seminal experimental and theoretical contributions to fields ranging from quantum solids and nonequilibrium superconductivity to the physics and device aspects of semiconductor superlattices. He was a co-originator and the primary theoretical exponent of the Type II superlattice, and he introduced a statistical model for predicting radiation-induced soft-error rates in VLSI circuits. He was the manager and technical leader of the first demonstration of FET technology at 0.1 μ m and below. His most recent work is in the area of systems and high-end CMOS processor custom design. Dr. Sai-Halasz spent two years in various IBM business and product divisions, and served on government and industry advisory panels. He is a Fellow of the IEEE.

Yuh-Jier Mii *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (MII at YKTVMV, mii@watson.ibm.com).* Dr. Mii received the B.S. degree in electrical engineering in 1982 from the National Taiwan University, Taipei, Taiwan, and the Ph.D. in electrical engineering from the University of California at Los Angeles in 1990. His thesis work included the introduction of infrared modulators using the strongly enhanced Stark effect of intersubband transitions in asymmetric quantum wells, and the investigation of hole intersubband absorption in SiGe/Si quantum wells. From 1990 to 1991, he worked at the AT&T Bell Laboratories in Murray Hill, New Jersey, on high-electron mobility SiGe/Si heterostructures and low-defect-density epitaxy of SiGe buffer layers on Si substrates. In 1991 Dr. Mii became a research staff member at the IBM Thomas J. Watson Research Center, where he has worked on advanced interconnections, bi-CMOS, and deep-submicron CMOS technologies. Dr. Mii left IBM in 1995.