*Seventh International Dielectrics and Conductors for ULSI Multilevel Interconnection Conference (DCMIC), Santa Clara, CA, March 5-9, 2001, pp. 38-50.*

# Trends for ULSI Interconnections and Their Implications for Thermal, Reliability and Performance Issues

**(Invited Paper)**

Kaustav Banerjee
Center for Integrated Systems
Stanford University, Stanford, CA 94305
kaustav@stanford.edu

**Abstract**

This paper provides an overview of ULSI interconnect scaling trends and their implications for thermal, reliability and performance issues simultaneously. It shows how interconnect scaling requirements for deep sub-micron (DSM) technologies cause increasing thermal effects. The paper then examines the impact of thermal effects on both interconnect design and electromigration (EM) reliability. Specifically, it discusses the impact of thermal effects on the allowable current density limits. Furthermore, it also discusses how thermal and reliability constrained current density limits may conflict with those obtained through purely performance based criterion. Additionally, high-current interconnect design rules for ESD and I/O circuits are also examined.

## 1 Introduction

As VLSI technology scales, interconnects are becoming the dominant factor determining system performance and power dissipation [1], [2], [3]. The ever-increasing demand for speed and functionality of Si-based advanced high performance chips has caused aggressive scaling of ICs beyond 0.25-μm minimum feature size. These technology nodes, commonly referred to as *deep sub-micron*, (DSM) allow VLSI circuits to meet the required device density and various circuit performance specifications. This trend has resulted in a dramatic reduction of the interconnect metal pitch and increased the number of metallization levels to accommodate the increasing number of wired circuits per chip. A schematic cross-section of a multilevel interconnect scheme is shown in Figure 1. This aggressive interconnect scaling has resulted in increasing current densities and associated thermal effects.

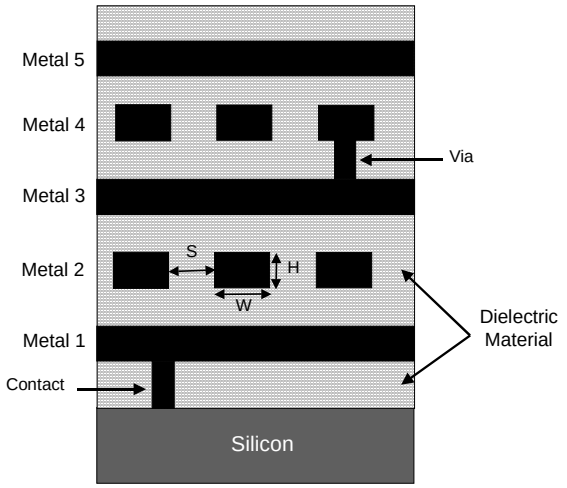



Figure 1. A schematic cross section of a multi-level interconnect scheme employed in present VLSI circuits. The metal pitch is defined as ($W + S$) and the aspect ratio is defined as ($H/W$).

Thermal effects are an inseparable aspect of electrical power distribution and signal transmission through the interconnects in VLSI circuits due to self-heating (or Joule heating) caused by the flow of current [4]. Thermal effects impact interconnect design and reliability in the following ways. Firstly, they limit the maximum allowable RMS current density, $j_{RMS\text{-}max}$ (since the RMS value of the current density is responsible for heat generation) in the interconnects, in order to limit the temperature increase. Secondly, interconnect lifetime (reliability) which is limited by electromigration (EM), has an exponential dependence on the inverse metal temperature [5]. Hence, temperature rise of metal interconnects due to self-heating phenomenon can also limit the maximum allowed average current density, $j_{avg\text{-}max}$, since EM capability is dependent on the average current density [6]. Thirdly, thermally induced open circuit metal failure under short-duration high peak currents including electrostatic discharge

(ESD) is also a reliability concern, [7] and can introduce latent EM damage [8] that has important reliability implications.

Additionally, low dielectric constant (low-k) materials are being introduced as an alternative intra- and inter- level insulator to reduce interconnect capacitance (therefore delay) and cross-talk noise to enhance circuit performance [9], [10], [11]. These materials can further exacerbate thermal effects owing to their lower thermal conductivity than silicon dioxide [12].

Furthermore, it is important to understand how thermal and reliability constraints may conflict with the performance optimization steps employed at the circuit level. Hence, thorough analysis of thermal effects in DSM interconnects is necessary to comprehend their full impact on circuit design, accurately model their reliability, and provide thermally safe design guidelines for various technologies.

In the next sub section interconnect scaling requirements as per [13] are examined and their implications for thermal effects are discussed briefly.

### 1.1 Trends in Interconnect Scaling and Implications for Thermal Effects

As VLSI circuits continue to be scaled aggressively, a rapid increase in functional density and chip size is observed as shown in Figure 2. This has resulted in increasing number of interconnect levels and reduction in interconnect pitch in order to realize all the inter-device and inter-block communications. Figure 3 shows this trend with interconnect levels increasing further in the near future, from 6 levels at the 250 nm node to 9 levels at the 50 nm node. This increase in the number of interconnect levels causes the upper most interconnect layers to move further away from the Si substrate making heat dissipation more difficult. Additionally, decreasing interconnect pitch will cause increased thermal coupling. Furthermore, the critical dimensions of contacts and vias are also decreasing with scaling as shown in Figure 4 resulting in higher current densities in these structures. Compounded with the introduction of low-k dielectrics as alternative insulators, whose thermal conductivities are also much lower than that of silicon dioxide (Figure 5), it is envisioned that thermal effects in interconnects can potentially become another serious design constraint.

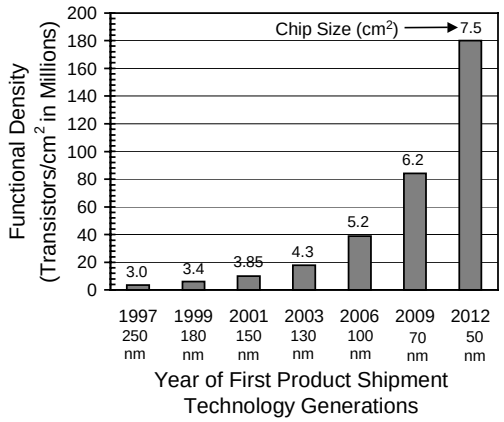



Figure 2 Present and projected functional density and chip size of microprocessors for different technology nodes and year of first product shipment. The functional density is expected to increase from 3.7 million in 1997 to 180 million in 2012.

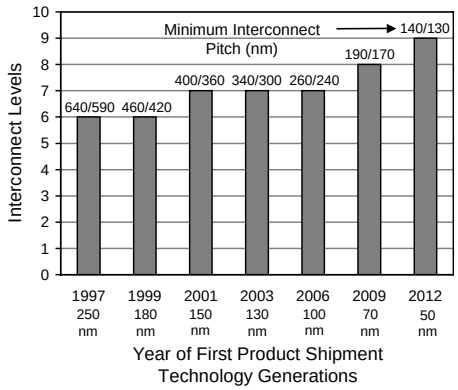



Figure 3. Present and projected interconnect levels and minimum contacted/non-contacted interconnect metal pitch of logic circuits for different technology nodes and year of first product shipment.

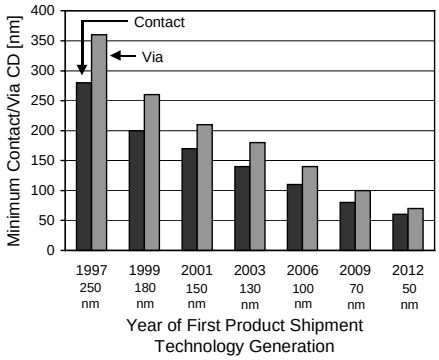



Figure 4. Present and projected minimum contact/via critical dimensions (CD) of logic circuits for different technology nodes and year of first product shipment.

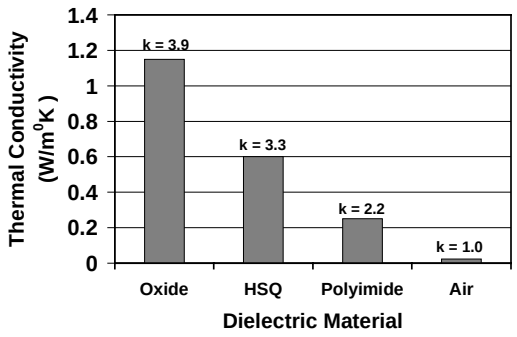



Figure 5. Thermal conductivity and approximate dielectric constants (k) of some insulating materials used (or being introduced) in high performance circuits.

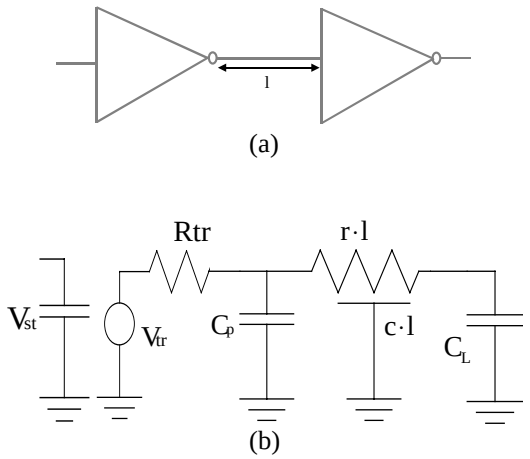



Figure 6 (a) An interconnect of length *l* between two buffers (b) The equivalent RC circuit. $V_{st}$ is the voltage at the input capacitance that controls the voltage source $V_{tr}$. $R_{tr}$ is the driver transistor resistance, $C_p$ is the output parasitic capacitance and $C_L$ is the load capacitance of the next stage, *r* and *c* are the interconnect resistance and capacitance per unit length respectively.

A simple analysis to demonstrate the implications of technology scaling on technology performance is now presented [4]. This analysis is instructive since these in turn have important implications on thermal effects and on reliability requirements for interconnects. Consider an interconnect segment of length *l* between two inverters as shown in Figure 6 (a). Figure 6 (b) shows the equivalent distributed RC representation. Here the inverter on the left that is driving the interconnect is represented as a voltage source controlled by the voltage $V_{st}$ at the input capacitance. $R_{tr}$ is the equivalent transistor resistance, $C_p$ is the parasitic capacitance composed mainly of the drain capacitance of the transistors. $C_L$ is the load capacitance or the input capacitance of the second inverter. Also, *c* and *r* are the capacitance and resistance per unit length of the interconnect line.

Using this simple model the performance of a technology can be most simply summarized by the overall delay time $\tau$ for a signal, which can be expressed as,

$$\tau = R_{tr}\left(C_p + C_L\right) + \left(R_{tr}c + rC_L\right) l + \frac{1}{2} rcl^2 \quad (1)$$

$R_{tr}$, $C_p$, and $C_L$ are functions of the transistor design and the circuit design *W/L* ratios used. The interconnect resistance per unit length, *r*, depends on the resistivity of the interconnect metal while *c* depends on the dielectric constant of the surrounding insulating material, interconnect metal pitch, interconnect geometry, and underlying insulator thickness [14]. The three terms on the right hand side of (1) represent the intrinsic gate delay, the load delay, and the interconnect delay contributions, respectively. The factor of half in the interconnect delay arises due to the distributed nature of the interconnect. A more general expression for $\tau$ will be provided later.

Now if *s* is the scaling parameter, which can be defined as the ratio of the feature size at a newer technology node to the feature size at an older reference technology node, then *s* is always less than one. There are two simplified scaling scenarios for interconnects:

- Scale metal pitch (*W* + *S*) at constant metal thickness (*H*) (see Figure 1)
- Scale metal pitch and the metal thickness

Under the first scenario, the load delay, line delay and current density will scale as *1/s*, $1/s^2$, and *1/s* respectively. Under the second scenario, the load delay, line delay and current density will scale as $1/s^2$, $1/s^2$, and $1/s^2$ respectively. In either case, it can be seen that interconnect delays begin to dominate technology performance, and that current density increases with scaling.

The interconnect scaling requirements drive several technology enhancements. Use of low-k materials lowers the interconnect capacitance per unit length *c* (particularly intra-level) in (1) and therefore lowers the total interconnect delay

($0.5rcl^2$). Lower interconnect capacitance also helps in minimizing cross-talk noise. Furthermore, lower interconnect capacitance also helps in reducing the dynamic power dissipation ($P_{dyn}$) during the switching of gates in digital circuits, which can be estimated by

$$P_{dyn} = \frac{1}{2}\alpha C V^2 f \qquad (2)$$

where $\alpha$ is the activity factor or switching probability, $C$ is the total capacitance, $V$ is the power supply voltage and $f$ is the clock frequency of the circuit.

Hence, it is the minimization of interconnect capacitance that is driving the introduction of low-k dielectric materials. Similarly, lower interconnect resistance and higher current density requirements drive the use of new metallization (namely Cu). Since these low-k materials also have lower thermal conductivity than silicon dioxide (Figure 5), heat dissipation becomes even more difficult. Thus VLSI technology scaling has important implications on thermal effects as discussed in this section. The various trends in technology scaling that causes increased thermal effects in interconnects can be summarized as:

- Increasing current density
- Increasing number of interconnect levels
- Introduction of low-k dielectric materials
- Increased thermal coupling

### 1.2 Implications of Thermal Effects on Reliability and Performance

In current VLSI interconnect designs, current density design rules are typically based on EM lifetimes [6]. However, the actual current densities are determined by the interconnect parameters (resistance and capacitance per unit length and the length of the interconnect) and the strength of the buffer which is driving the interconnect length. In a typical design, long interconnects, which can potentially have large current densities, are usually split into buffered segments to improve performance (signal slew and delay) [15]. The signal line length and buffer sizes are optimized to give maximum performance for a given technology. Signal lines longer than the optimum length would increase interconnect delay while buffer sizes larger than the optimum would result in an increase in buffer delay. Moreover, increasing buffer size would increase the current density of the signal line connected to the output and may also cause excessive power dissipation. Hence, it is important to quantify whether the associated current densities in signal lines optimized for maximum performance also meet the EM design limits.

As mentioned earlier, in DSM technologies, low-k materials invariably have poor thermal properties and therefore the use of such materials can significantly impact the EM design limits. On the other hand, the use of Copper, which has lower resistivity, alleviates the problem to some extent. It is therefore important to quantify whether EM reliability or performance is the dominant factor determining the optimal signal line length in various low-k/Cu based interconnect systems.

Recently a methodology for quantitative comparison of the impact of EM, thermal effects and performance on the optimization of the signal line length has been formulated [16]. The methodology has been applied to various low-k/Cu interconnect systems [17]. The next two sections deal with this methodology. In Section 2, effect of self-heating on EM is briefly discussed followed by an analysis of the self-consistent design approach that takes quasi-2D heat conduction into consideration. The modified self-consistent approach is then used to quantify the impact of new interconnect materials (Cu and low-k dielectrics) on allowed current density limits. Section 3 introduces the methodology for computing the current densities from performance considerations only, and then provides a direct comparison between the reliability and performance based current density design limits [17].

## 2 Interconnect Reliability

### 2.1 Influence of Self-Heating on EM

EM lifetime reliability of metal interconnects is modeled by the well known Black's equation [5] given by,

$$TTF = A\, j^{-n}\, exp\left(\frac{Q}{k_B T_m}\right) \qquad (3)$$

where TTF is the time-to-fail (typically for 0.1% cumulative failure). $A$ is a constant that is

dependent on the geometry and microstructure of the interconnect, $j$ is the DC or average current density. The exponent $n$ is typically 2 under normal use conditions, $Q$ is the activation energy for grain-boundary diffusion and equals ~ 0.7 eV for Al-Cu, $k_B$ is the Boltzmann's constant, and $T_m$ is the metal temperature. The typical goal is to achieve 10 year lifetime at 100 $^0$C, for which (3) and accelerated testing data produce a design rule value for the acceptable current density, $j_0$, at the reference temperature $T_{ref}$. However, this design rule value does not comprehend self-heating.

The effect of self-heating can be analyzed from the following: The metal temperature, $T_m$ in (3) is given by,

$$T_m = T_{ref} + \Delta T_{self-heating} \qquad (4)$$

and,

$$\Delta T_{self-heating} = (T_m - T_{ref}) = I_{rms}^2 R R_\theta \quad (5)$$

where $T_{ref}$ is typically taken as ~100 $^0$C, $\Delta T_{self\text{-}heating}$ is the temperature rise of the metal interconnect due to the flow of current, $R$ is the interconnect resistance, and $R_\theta$ is the effective thermal impedance of the interconnect line to the chip. $I_{rms}$ is the RMS current for a time varying current waveform, or the DC current for a constant current stress. It can be observed from (3) and (4), that as self-heating increases, the metal temperature increases, and hence the EM lifetime decreases exponentially. Therefore, it is important to accurately account for self-heating in (3).

## 2.2 Self-Consistent Interconnect Design Analysis

In this sub-section the formulation of the self-consistent solutions [18], [16] for allowed interconnect current density is summarized, and then applied to analyze low-k/Cu interconnects in the next sub-section. The $\Delta T_{self\text{-}heating}$ in interconnects given by (5) can be written in terms of the RMS current density as,

$$j_{rms}^2 = \frac{(T_m - T_{ref}) K_{ins} W_{eff}}{t_{ins} t_m W_m \rho_m(T_m)} \qquad (6)$$

Here $t_m$ and $W_m$ are the thickness and width of interconnect metal line, and $\rho_m(T_m)$ is the metal resistivity at temperature $T_m$. Note that the thermal impedance $R_\theta$ in equation (5) has been expressed as,

$$R_\theta = \frac{t_{ins}}{K_{ins} L W_{eff}} \qquad (7)$$

Here $t_{ins}$ is the total thickness of the underlying dielectric, $K_{ins}$ is the thermal conductivity normal to the plane of the dielectric, and $L$ is the length of the interconnect. In this expression for the thermal impedance, $W_{eff}$ has been modeled as the effective width of the metal line taking quasi-2D heat conduction [16] into consideration from experimental data for high aspect ratio lines.

Now, in order to achieve an EM reliability lifetime goal mentioned earlier, we must have the lifetime at any ($j_{avg}$) current density and metal temperature $T_m$, equal to or larger than the lifetime value (eg. 10 year) under the design rule current density stress $j_0$, at the temperature $T_{ref}$. This value of $j_0$ is dependent on the specific interconnect metal technology. Therefore we have,

$$\frac{exp\left(\frac{Q}{k_B T_m}\right)}{j_{avg}^2} \geq \frac{exp\left(\frac{Q}{k_B T_{ref}}\right)}{j_0^2} \qquad (8)$$

Using the relationship between $j_{avg}$, $j_{peak}$, $j_{rms}$, and $r$ for a rectangular unipolar pulse, ($j_{avg} = r\, j_{peak}$, & $j_{rms} = r^{0.5} j_{peak}$) we have after eliminating $j_{peak}$,

$$\frac{j_{avg}^2}{j_{rms}^2} = r \qquad (9)$$

Substituting for $j_{rms}^2$ from (6) and $j_{avg}^2$ from (8) in (9) we get the self-consistent equation given by,

$$r = j_0^2 \left( \frac{exp\left(\frac{Q}{k_B T_m}\right)}{exp\left(\frac{Q}{k_B T_{ref}}\right)} \right) \frac{t_{ins} t_m W_m \rho_m(T_m)}{(T_m - T_{ref}) K_{ins} W_{eff}} \qquad (10)$$

Note that this is a single equation in the single unknown temperature $T_m$. Once this self-consistent temperature is obtained from (10), the corresponding maximum allowed $j_{rms}$ and $j_{peak}$ can be calculated from (6) and the current density

relationships given above. The self-consistent equation given by (10) for unipolar pulses is also valid for more general time varying waveforms with an effective duty cycle $r_{eff}$ [19].

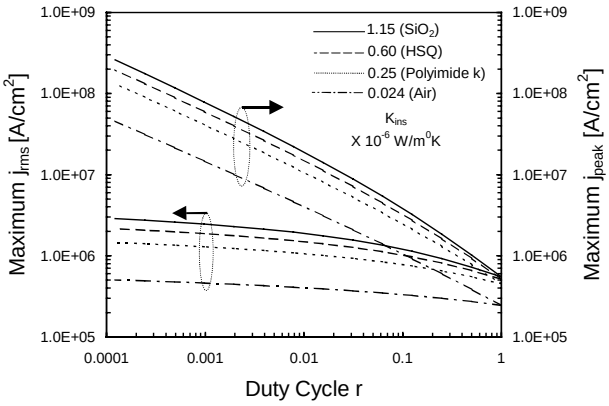



Figure 7. Self-consistent solutions for maximum allowed values of $j_{rms}$ and $j_{peak}$ for Metal 6 in 0.25-µm technology for different values of $K_{ins}$. Interconnect metal is Cu with $\rho_m(T_m) = 1.67 \times 10^{-6}$ [1 +6.8 x $10^{-3}$ $^0C^{-1}$ ($T_m$ - $T_{ref}$)] Ω-cm. The activation energy is assumed to be same as that for AlCu.

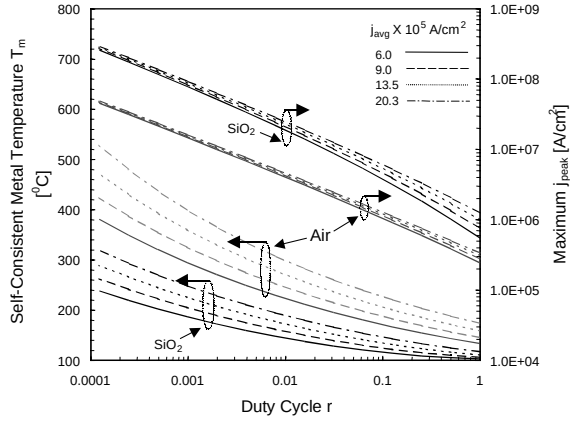



Figure 8. Self-consistent metal temperature and the maximum allowed $j_{peak}$ for Metal 6 in a 0.25-µm Cu technology, as a function of duty cycle for different values of $j_{avg}$ (or $j_0$) for $SiO_2$ and air as the dielectric.

## 2.3 Impact of New Materials on Current Density Limits

Allowed interconnect current densities are expected to be strongly influenced by low-k materials, which cause increased Joule-heating due to their lower thermal conductivity [12], [20]. Therefore we begin by analyzing the effect of introducing new interconnect and dielectric

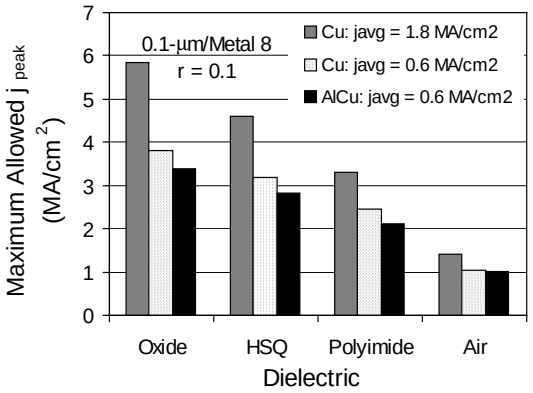



Figure 9. Comparison of the maximum allowed $j_{peak}$ values for AlCu and Cu lines with two different $j_{avg}$ values for Metal 8 of a 0.10-µm technology shown for different dielectric materials.

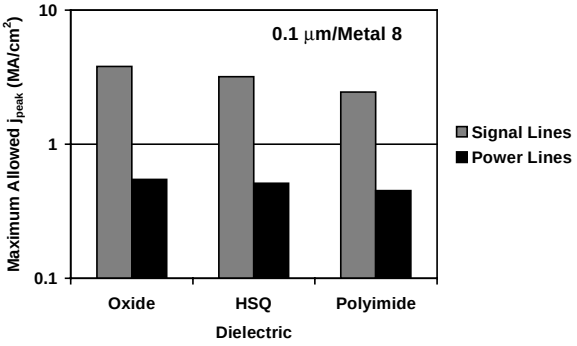



Figure 10. Comparison of the maximum allowed $j_{peak}$ values for signal ($r$ = 0.1) and power lines ($r$ = 1.0) for Metal 8 of a 0.10 µm technology shown for different dielectric materials. The lower values of $j_{peak}$ in power lines is because of their higher temperature rise resulting from carrying DC current.

materials on allowable current density limits. In Figure 7 the self-consistent values of $j_{rms}$ and $j_{peak}$ are plotted as a function of the duty cycle $r$, for different dielectrics. It can be observed that $j_{rms}$ and $j_{peak}$ decrease significantly as dielectrics with lower thermal conductivity are introduced. For small values of $r$, $j_{rms}$ varies very slowly with $r$. This is due to increased Joule-heating. In Figure 8 we plot $j_{peak}$ and $T_m$ as a function of $r$ for $SiO_2$ and air as the dielectric for different values of the design current density, $j_{avg}$. It can be observed that $j_{avg}$ does not change $j_{peak}$ appreciably, and as $r$ reduces, the increase in $j_{peak}$ with $j_{avg}$ becomes negligible. Furthermore, for a dielectric material with low thermal conductivity such as air, $j_{peak}$ is almost independent of $j_{avg}$. This indicates that introduction of better interconnect materials

becomes increasingly ineffective in increasing $j_{peak}$ for dielectrics with poor thermal properties. Figure 9 summarizes the impact of low-k dielectrics on allowed $j_{peak}$ for both Cu and AlCu interconnect systems. It can be observed that the difference between the maximum allowed $j_{peak}$ for AlCu and Cu interconnects reduces as dielectric materials with poor thermal properties are introduced. For the specific case of air as the dielectric, the $j_{peak}$ values are very similar for AlCu and Cu. This is demonstrated for two different values of $j_{avg}$ for Cu, one value identical to that of AlCu and the other three times higher than this value, which accounts for the improved EM performance in Cu. We also compared current carrying capabilities of signal and power lines and the results are shown in Figure 10. Consistent with Figure 7, it can be observed that the $j_{peak}$ values are about an order of magnitude lower for the power lines due to increased thermal effects.

## 3 Interconnect Performance

### 3.1 Implications of Performance Optimization on Signal Line Current Densities

As a next step we demonstrate a methodology for computing current density from performance considerations only [16], [17]. Consider an interconnect of length $l$ between two buffers. The schematic representation is shown in Fig. 6(a). Fig. 6(b) shows an equivalent RC circuit for the system. The voltage source ($V_{tr}$) is assumed to switch instantaneously when voltage at the input capacitor ($V_{st}$) reaches a fraction $x$, $0 \le x \le 1$ of the total swing. Hence the overall delay of one segment is given by:

$$\tau = b(x) R_{tr} (C_L + C_P) + b(x) (c R_{tr} + r C_L) l + a(x) r c l^2 \quad (11)$$

where $a(x)$ and $b(x)$ only depend on the switching model, i.e., $x$. For instance, for $x$=0.5, $a$=0.4 and $b$=0.7 [21]. If $r_0$, $c_0$ and $c_p$ are the resistance, input and parasitic output capacitances of a minimum sized inverter respectively then $R_{tr}$ can be written as $r_0 / s$ where $s$ is size of the inverter in multiples of minimum sized inverters. Similarly $C_P = s\ c_p$ and $C_L = s\ c_0$. If the total interconnect of length $L$ is divided into n segments of length $l = L/n$, then the overall delay is given by,

$$T_{delay} = n\tau = \frac{L}{l} b(x) r_0 (c_0 + c_p) + b(x) \left( c \frac{r_0}{s} + s r c_0 \right) L + a(x) r c l L \quad (12)$$

It should be noted in the above equation that $s$ and $l$ appear separately and therefore $T_{delay}$ can be optimized separately for $s$ and $l$. The optimum values of $l$ and $s$ are given as:

$$l_{opt} = \sqrt{\frac{b(x) r_0 (c_0 + c_p)}{a(x) r c}} \quad (13)$$

$$s_{opt} = \sqrt{\frac{r_0\, c}{r\, c_0}} \quad (14)$$

Note that $s_{opt}$ is independent of the switching model, i.e., $x$.

Since, for deep sub-micron technologies, a significant fraction of interconnect capacitance, $c$, is contributed by coupling and fringing capacitances to neighboring lines as shown in Figure 11, a full 3D-capacitance extraction using SPACE3D [22] for signal lines at various metal levels was used to obtain the values of $c$ for SPICE simulations.

This inverter-interconnect structure is used as a delay stage in a multi-stage ring oscillator and the current waveforms and current densities along the interconnect are be obtained. In practice, the input capacitance $C_L$ of the inverter is almost constant but the output resistance $R_{tr}$ and output parasitic capacitance $C_P$ are bias dependent and therefore change during the output transition. Therefore accurate values of optimal interconnect length and buffer size need to be determined by SPICE simulation. For this, we take advantage of the fact that the optimal interconnect length does not depend on the buffer size. Therefore, we first set the buffer size to an appropriate value and sweep the interconnect length and find the optimum length which minimizes the ratio of the ring oscillator stage delay and interconnect length. Using this optimum length, we subsequently sweep buffer sizes and find the optimum buffer size, which minimizes the stage delay. This allows us to obtain the values of $l_{opt}$ and $s_{opt}$ taking into account the bias dependence of transistor resistances and capacitances and the switching model.

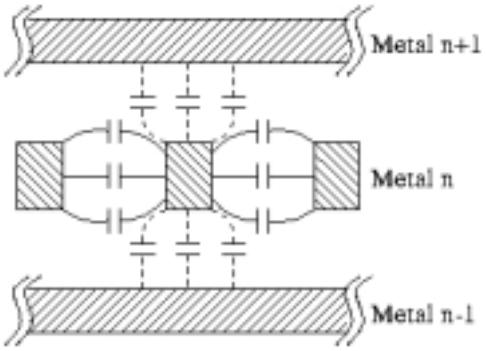



Figure 11. Typical Interconnect Structure.

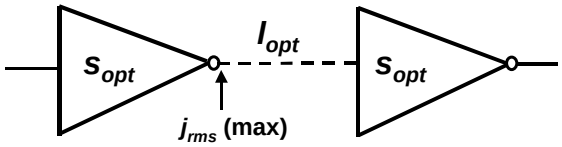



Figure 12. Optimized buffer and interconnect segment showing location of maximum current density.

This analysis is carried out for every metal layer in the two technologies under study.

Note that due to the distributed nature of the interconnect, the maximum current density occurs close to the buffer output (Figure 12). Hence, we need to verify whether this maximum current density, which is obtained from performance considerations ($j_{\text{-performance}}$) only, also meets the EM current density limits ($j_{\text{-reliability}}$) obtained earlier using the self-consistent approach.

The interconnect current waveform in the metal lines for 0.25-µm and 0.1-µm technologies, as obtained from SPICE simulations, were found to be bipolar as expected (Figure 13). It should be noted that (3) represents the EM lifetime reliability equation for a unipolar pulse or dc current. The EM lifetime under bipolar stress conditions is known to be higher than that under the unipolar case. In this work, the unipolar EM lifetime reliability equation is used as a worst case limit.

Also, the relative rise and fall skew was found to be same across both technologies. From our simulations it was observed that drivers and interconnects optimized using (13) and (14), maintain good slew rates for rising and falling transitions for all the metal layers and across both technologies with an effective duty cycle ($r_{eff} = j_{avg}^2 / j_{rms}^2$) of 0.12 ± 0.01 as shown in Table 1.

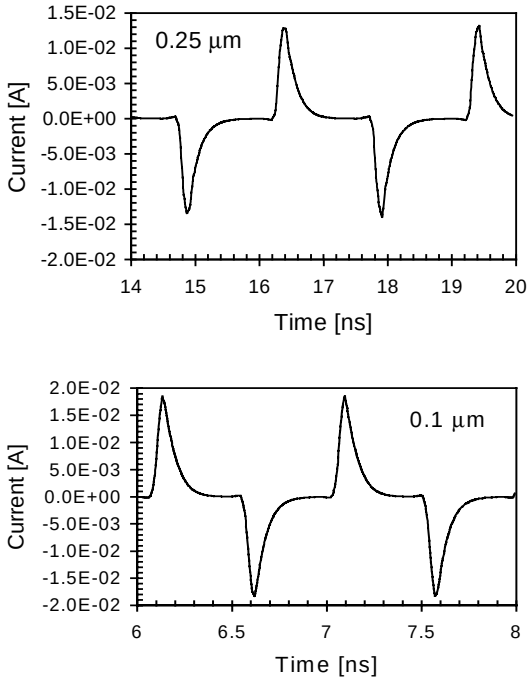



Figure 13. Current waveforms in the top layer metal lines obtained from SPICE simulations for the 0.25 µm and the 0.1 µm technologies.

Since for the signal lines the current waveform is symmetric and bipolar, $j_{avg}$ and $j_{rms}$ are computed over half the time-period to obtain $r_{eff}$. Table 1 also lists the maximum values of the average and peak current densities.

In the above analysis it was assumed that the line capacitance per unit length is constant. In an actual design this is not necessarily the case. Consider the interconnect structure shown in Figure 11. The total interconnect capacitance $C_{total}$ can be viewed as a sum of capacitance to lines on the same metal layer $C_{neighbor}$ and lines on other metal layers $C_{other}$ (shown by solid and dashed lines respectively in Figure 11) $C_{neighbor}$ and $C_{other}$ consist of both parallel plate and fringing capacitances. Lines on adjacent metal layers are typically routed orthogonal to each other. Hence the total capacitance between metal lines at two adjacent layers is very small. On the other hand the capacitance to neighboring lines is large. This is specially the case for deep sub-micron technologies where the aspect ratio of the lines is greater than 1 and as a result, ($C_{neighbor}$ / $C_{total}$) is very high (0.7 to 0.9). The effective capacitance of the line ranges from ($2C_{neighbor}+C_{other}$) to $C_{other}$ depending upon whether the neighboring line

| Layer | $l_{opt}$ (mm) | $s_{opt}$ | $j_{avg}$ ($x10^5 A/cm^2$) | $j_{peak}$ ($x10^5 A/cm^2$) | $r_{eff}$ |
|---|---|---|---|---|---|
| 1 | 2.51 | 75 | 3.06 | 27.02 | 0.11 |
| 2 | 2.41 | 80 | 2.91 | 26.38 | 0.11 |
| 3 | 4.79 | 157 | 1.67 | 14.69 | 0.11 |
| 4 | 5.48 | 148 | 1.32 | 13.48 | 0.10 |
| 5 | 12.5 | 494 | 0.61 | 5.18 | 0.12 |
| 6 | 13.9 | 454 | 0.52 | 4.69 | 0.11 |

(a)

| Layer | $l_{opt}$ (mm) | $s_{opt}$ | $j_{avg}$ ($x10^5 A/cm^2$) | $j_{peak}$ ($x10^5 A/cm^2$) | $r_{eff}$ |
|---|---|---|---|---|---|
| 1 | 0.85 | 36 | 5.13 | 38.99 | 0.13 |
| 2 | 0.9 | 40 | 5.10 | 39.48 | 0.13 |
| 3 | 1.57 | 97 | 2.73 | 20.93 | 0.13 |
| 4 | 1.79 | 85 | 2.39 | 18.27 | 0.13 |
| 5 | 4.47 | 276 | 1.28 | 9.78 | 0.13 |
| 6 | 4.68 | 249 | 1.21 | 9.00 | 0.13 |
| 7 | 8.8 | 500 | 0.49 | 3.67 | 0.13 |
| 8 | 9.58 | 522 | 0.44 | 3.5 | 0.12 |

(b)

Table 1. Optimized interconnect length ($l_{opt}$), buffer size ($s_{opt}$), corresponding average and peak current densities and duty cycle ($r$) for (a) 0.25-µm Cu technology with insulator dielectric constant = 3.3 and (b) 0.1-µm Cu technology with insulator dielectric constant = 2.0. The optimum repeater size given by $s_{opt}$ means that the width of the NMOS and PMOS for this inverter are obtained by scaling the width of the corresponding transistors in a minimum sized inverter by a factor of $s_{opt}$.

signals are switching in the opposite direction or the same direction.

Furthermore, it is observed from simulation and can be shown that the interconnect current remains almost the same if the load capacitance of a buffer changes. The rise and fall time will get affected significantly but the interconnect current does not change appreciably if the buffer drive strength is unchanged. However, for the slower transition, $r$ increases and for the faster transition, $r$ decreases (typical values observed in simulation are 0.2 for slower transition and 0.05 for the faster transition).

### 3.2 Comparisons between Reliability and Performance Based Current Density Limits

Figure 14 shows the comparison of $j_{avg\text{-}performance}$ with the values of $j_{avg\text{-}reliability}$ for a 0.25-µm technology based on [13]. It can be observed that $j_{avg\text{-}performance}$ is always lower than $j_{avg\text{-}reliability}$ for all the dielectrics. However, for low-k dielectric materials we find that the difference between $j_{avg\text{-}reliability}$ and $j_{avg\text{-}performance}$ reduces. This is due to the fact that these dielectrics have lower

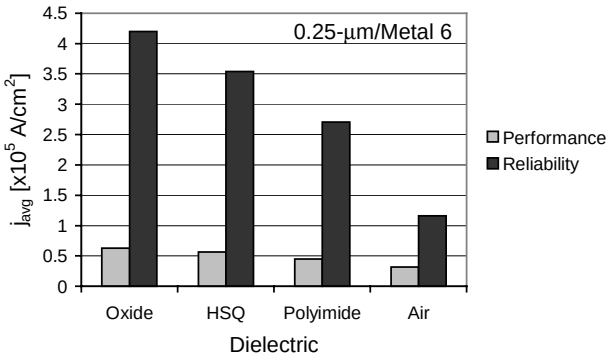



Figure 14. Comparison of $j_{avg}$ values obtained from reliability and electrical performance considerations for Metal 6 of a 0.25-µm Cu technology. $r_{eff}$ = 0.11.

thermal conductivity than oxide which leads to greater interconnect Joule-heating and therefore lower allowable current densities.

Finally, it should be noted that the $j_{avg\text{-}reliability}$ values generated above were for an isolated line. In real ICs with multiple layers of interconnect, the Joule-heating of interconnect lines is known to be significantly more severe due to thermal coupling between neighboring lines [23]. Figure 15 shows comparison of $j_{avg\text{-}performance}$ with the values of $j_{avg\text{-}reliability}$ for a 0.1-µm technology including the $j_{avg\text{-}reliability}$ values for realistic interconnect structures using finite element simulations to demonstrate an optimistic scenario of thermal coupling using various dielectrics.

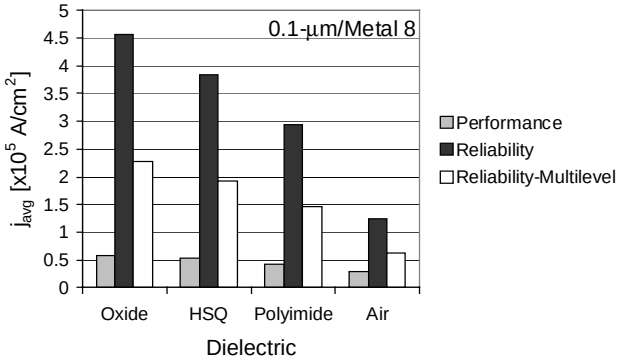



Figure 15. Comparison of $j_{avg}$ values obtained from reliability and performance considerations for Metal 8 of a 0.1-µm Cu technology. $r_{eff}$ = 0.12. The effect of thermal coupling in realistic multilevel interconnect arrays on the $j_{avg}$ values is also shown here for various dielectric materials.

It can be observed that the $j_{avg\text{-}performance}$ values remain lower than the $j_{avg\text{-}reliability}$ values even after thermal coupling in realistic structures is taken into account. However, in this simulation it is assumed that $r_{eff}$ = 0.12 at all times, which can actually increase (to ~ 0.2), if neighboring lines switch in the opposite direction as discussed earlier. This increase in $r_{eff}$ will further lower $j_{avg\text{-}reliability}$. Hence, impact of switching states of signal lines on allowed current density design rules is important.

## 4 Thermal Effects under High-Current Stress Conditions

In this section we will briefly address interconnect design issues for high-current robustness. Apart from normal circuit conditions, ICs also experience high current stress conditions, the most important of them being the electrostatic discharge (ESD), that causes accelerated thermal failures [24]. Semiconductor industry surveys indicate that ESD is the largest single cause of failures in ICs. ESD is a high current short-time scale phenomena that can lead to catastrophic open circuit failures, and to latent damage [25].

Interconnect failure due to ESD is becoming an important issue as VLSI scaling continues. As the number of wired gates ($G$) per chip increases (see Figure 2) the number of I/O pins ($N_P$) also increases as per Rent's Rule [26], which is given by

$$N_P = K \cdot G^{\beta} \qquad (15)$$

where $K$ is the average number of I/Os per gate, and $\beta$ is the Rent exponent that can vary from 0.1 to 0.7. As a consequence of this increasing I/O pins, the package floor planning is changing from peripheral package connections to array grids in order to accommodate the increased I/O pin count [27]. In the array architecture, the interconnect widths between external pads and the ESD structures must decrease to preserve chip wirability and to prevent timing delays in critical paths and in the receiver and driver networks. This trend can increase the susceptibility of interconnects to ESD failure. Furthermore, technology scaling and the transition to new interconnect and dielectric materials necessitates a growing need to comprehend the high current behavior of these structures and analyze their failure mechanisms in order to provide robust design guidelines.

In [7] it has been shown that the critical current density for causing open circuit metal failure in AlCu interconnects is ~ 60 MA/cm$^2$. A short time scale high-current failure model for designing robust interconnects to avoid thermal failure under high peak current conditions is also formulated in [7]. This model can be used to determine the critical current for open circuit metal failure in terms of the pulse width and the line width as shown in Figure 16. These curves provide design guidelines for ESD protection circuit and I/O interconnects. For example, a typical Human Body Model (HBM) ESD pulse, which can be described as a ~100 ns/1.3 A event [24], would require the width of the metal line to be greater than ~2.5 μm. Similarly for an electrical overstress (EOS) event of 1 μs the minimum line width should be more than ~7 μm. Hence a safe design guideline for these stress conditions would be ~10 μm. A study from IBM has shown that the model can also be applied to design damascence Cu interconnects [27].

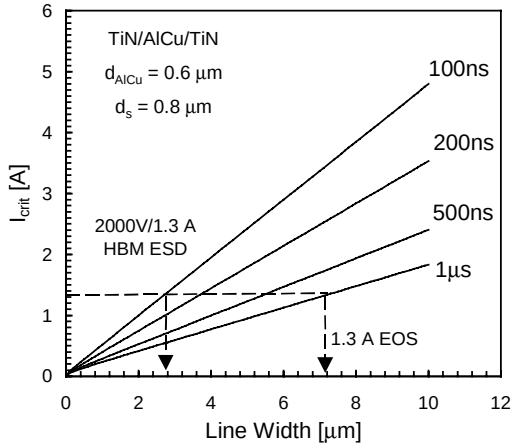
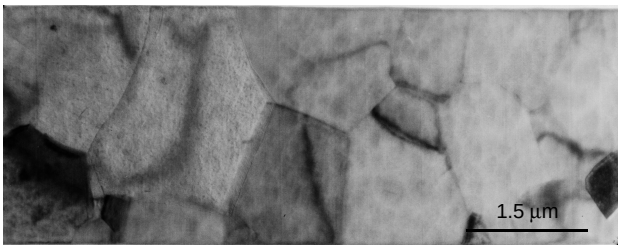


Figure 16. Design rules for high-current robustness [7].

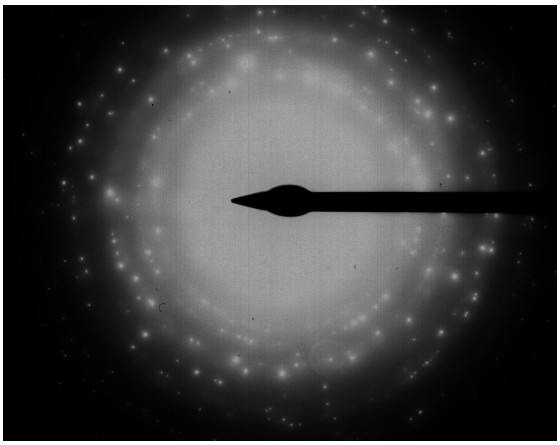
As mentioned earlier, interconnects can also suffer latent damage if the lines resolidify after melting and this has been shown to degrade the EM lifetime [8]. Figure 17(a) shows a TEM micrograph of an unstressed 3.0 μm wide AlCu line. It can be observed that the grain size is ~ 1.5 μm. A TEM diffraction analysis result is also shown in Figure 17(b). The diffraction pattern, which consists of *diffraction spots*, confirms the information from the TEM micrograph, that the unstressed lines have small number of large grains. The sparse spatial formation of the diffraction

spots is typical for a polycrystalline material with small number of large grains. Figure 18(a) shows a TEM image of a line (segment) that has been stressed by a high current pulse but has not shown any physical damage or change in line resistance. The change in the microstructure is clearly visible. The diffraction patterns shown in Figure 18(b) are ring shaped, which is due to the presence of a large number of grains with random orientations, indicating that the defect areas have smaller grain size. Such grain size reduction will result from rapid resolidification from a molten state.

The model presented in [7] can also be used to avoid this latent reliability hazard. These interconnect design rules must be obeyed for high current robustness.

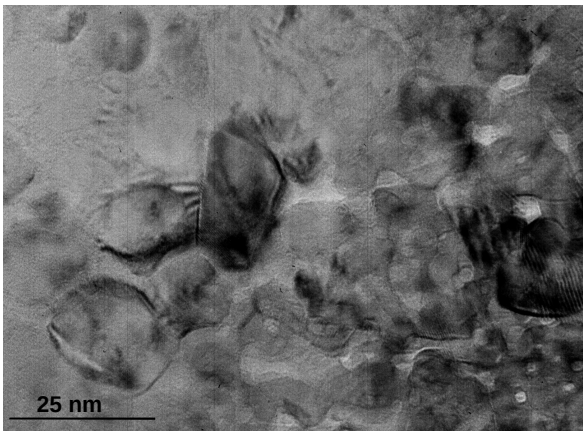



(a)

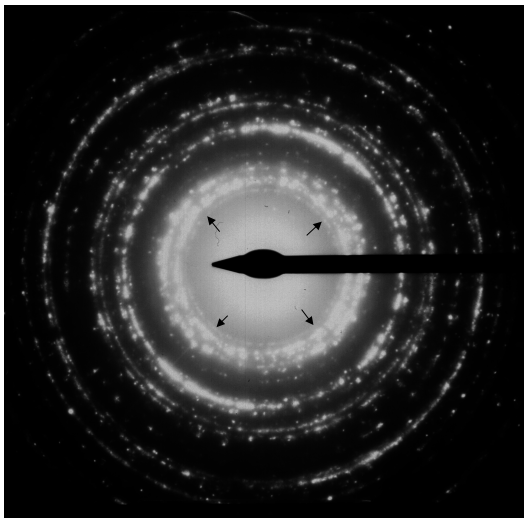


(b)

Figure 17 a) TEM micrograph of an unstressed AlCu line with the b) corresponding TEM diffraction pattern. All materials above and below the AlCu line (including the TiN layers) were removed by ion-milling process.




(a)



(b)

Figure 18 a) Microstructure of the metal within resolidified sections showing small grain sizes b) TEM *diffraction rings* validating the presence of a large number of small grains with random orientations. The innermost faint ring (shown by the arrows) indicates the appearance of a new phase.

## Summary

ULSI scaling requirements are causing increased thermal effects in interconnects due to increasing metallization levels, current density and thermal coupling, along with the introduction of low-k dielectric materials with poor thermal conductivity. Coupled analysis of EM, thermal effects and interconnect performance optimization is necessary to quantify their impact on current density limits and to understand various tradeoffs between technology, reliability and performance issues. A methodology, which allows

determination of both reliability and performance based current density limits, was discussed. This technique can be effectively applied to study the impact of technology scaling and performance optimization on interconnect reliability. For the low-k/Cu interconnect systems (based on NTRS), it was shown that as long as point-to-point interconnect performance can be optimized, EM design limits for those signal lines will be satisfied.

Additionally, high-current design rules for interconnects in ESD and I/O circuits were discussed, and a latent damage mechanism that can cause overall product reliability degradation was also examined.

In conclusion, thermal effects are a growing problem for deep sub micron interconnects and require careful considerations. In fact, a recent study that employed full-chip thermal analysis using data from the *International Technology Roadmap for Semiconductors*, showed that the temperature of the global metal lines are increasing inspite of the total chip power density remaining more or less constant across various technology nodes (180 nm to 50 nm) [28]. If reliability based current density limits begin to conflict with those based on performance, employing dummy thermal vias might provide a solution for alleviating thermal problems [29].

## Acknowledgments

The author would like to acknowledge support from the MARCO Interconnect Focus Center at Stanford and the SRC.

## References

[1] M. T. Bohr, “Interconnect scaling-the real limiter to high performance ULSI,” *Tech. Dig. IEDM*, 1995, pp. 241-244.

[2] W. J. Dally, “Interconnect limited VLSI architecture,” *IITC*, 1999, pp. 15-17.

[3] S. J. Souri, K. Banerjee, A. Mehrotra, and K. C. Saraswat, "Multiple Si layer ICs: Motivation, performance analysis, and design implications," *ACM Design Automation Conference*, 2000, pp. 213-220.

[4] K. Banerjee, "Thermal effects in deep sub-micron VLSI interconnects," *Tutorial Notes*, *IEEE International Symposium on Quality Electronic Design*, 2000.

[5] J. R. Black, “Electromigration – A brief survey and some recent results,” *IEEE Trans. Electron Devices*, vol. ED-16, pp. 338-347, 1969.

[6] B. K. Liew, N. W. Cheung, and C. Hu, “Projecting interconnect electromigration lifetime for arbitrary current waveforms,” *IEEE Trans. Electron Devices*, vol. 37, pp. 1343-50, 1990.

[7] K. Banerjee, A. Amerasekera, N. Cheung and C. Hu, “High-current failure model for VLSI interconnects under short-pulse stress conditions,” *IEEE Electron Device Lett.*, vol. 18, No. 9, pp. 405-407, 1997.

[8] K. Banerjee, D. Y. Kim, A. Amerasekera, C. Hu, S. S. Wong, and K. E. Goodson, "Microanalysis of VLSI interconnect failure modes under short-pulse stress conditions," *IRPS*, 2000, pp. 283-288.

[9] J. Ida et al, “Reduction of wiring capacitance with new low dielectric SiOF interlayer film for high speed/low power sub-half micron CMOS,” *Tech. Dig. VLSI Symp.*, 1994, pp. 59-60.

[10] *MRS Bulletin*, October 1997.

[11] B. Shieh, K. C. Saraswat, J.P. McVittie, S. List, S. Nag, M. Islamraja, R.H. Havemann, "Air-Gap formation during ILD deposition to lower interconnect capacitance," *IEEE Electron Device Lett.*, vol. 19, no. 1, pp. 16-18, 1998.

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

[13] The National Technology Roadmap for Semiconductors-NTRS, 1997 http://www.itrs.net/ntrs/publntrs.nsf

[14] T. Sakurai and K. Tamaru, “Simple formulas for two- and three-dimensional capacitances,” *IEEE Trans. Electron Devices*, vol. ED-30, no. 2 pp. 183-185, 1983.

[15] J. Culetu, C. Amir, and J. McDonald, “A practical repeater insertion method in high speed VLSI circuits,” *ACM Design Automation Conference*, 1998, pp. 392-395.

[16] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu, "On thermal effects in deep sub-micron VLSI interconnects," *ACM Design Automation Conference*, 1999, pp. 885-891.

[17] K. Banerjee, A. Mehrotra, W. Hunter, K. C. Saraswat, K. E. Goodson, and S. S. Wong, "Quantitative projections of reliability and performance for low-k/Cu interconnect systems," *IRPS*, 2000, pp. 354-358.

[18] W. R. Hunter, “Self-consistent solutions for allowed interconnect current density – Part I: Implications for technology evolution,” *IEEE Trans. Electron Devices*, vol. ED-44, pp. 304-309, 1997.

[19] W. R. Hunter, “Self-consistent solutions for allowed interconnect current density – Part II: Application to design guidelines,” *IEEE Trans. Electron Devices*, vol. ED-44, pp. 310-316, 1997.

[20] K. E. Goodson and Y. S. Ju, “Heat conduction in novel electronic films,” *Annu. Rev. Mater. Sci*., 29: pp. 261-293, 1999.

[21] R. H. J. M. Otten and R. K. Brayton, “Planning for performance,” *Proc. 35th Annual Design Automation Conference*, 1998, pp. 122-127.

[22] “Physical design modeling and verification project (SPACE),” *http://cas.et.tudelft.nl/research/space.html*

[23] S. Rzepka, K. Banerjee, E. Meusel, and C. Hu, “Characterization of self-heating in advanced VLSI interconnect lines based on thermal finite element

simulation," *IEEE Trans. on Components, Packaging and Manufacturing Technology-Part A*, vol. 21, No. 3, pp. 1-6, 1998.

[24] C. Duvvury and A. Amerasekera, "ESD: A pervasive reliability concern for IC technologies," *Proceedings of the IEEE*, Vol. 81, No. 5, pp. 690-702, 1993.

[25] C. Duvvury and A. Amerasekera, "State-of-the-art issues for technology and circuit design of ESD protection in CMOS ICs," *Semiconductor Science. and Tech*., pp. 833-850, 1996.

[26] B.S. Landman and R.L. Russo, "On a pin versus block relationship for partitions of logic graphs," *IEEE Trans. Computers*, vol. C-20, no. 12, 1971.

[27] S. H. Voldman, "ESD robustness and scaling implications of aluminum and copper interconnects in advanced semiconductor technology," *Proc. EOS/ESD Symposium*, 1997, pp. 316-329.

[28] S. Im and K. Banerjee, "Full chip thermal analysis of planar (2-D) and vertically integrated (3-D) high performance ICs," *Tech. Dig. IEDM*, 2000, pp. 727-730.

[29] T-Y Chiang, K. Banerjee, K. C. Saraswat, "Effect of via separation and low-k dielectric materials on the thermal characteristics of Cu interconnects," *Tech. Dig. IEDM*, 2000, pp. 261-264.