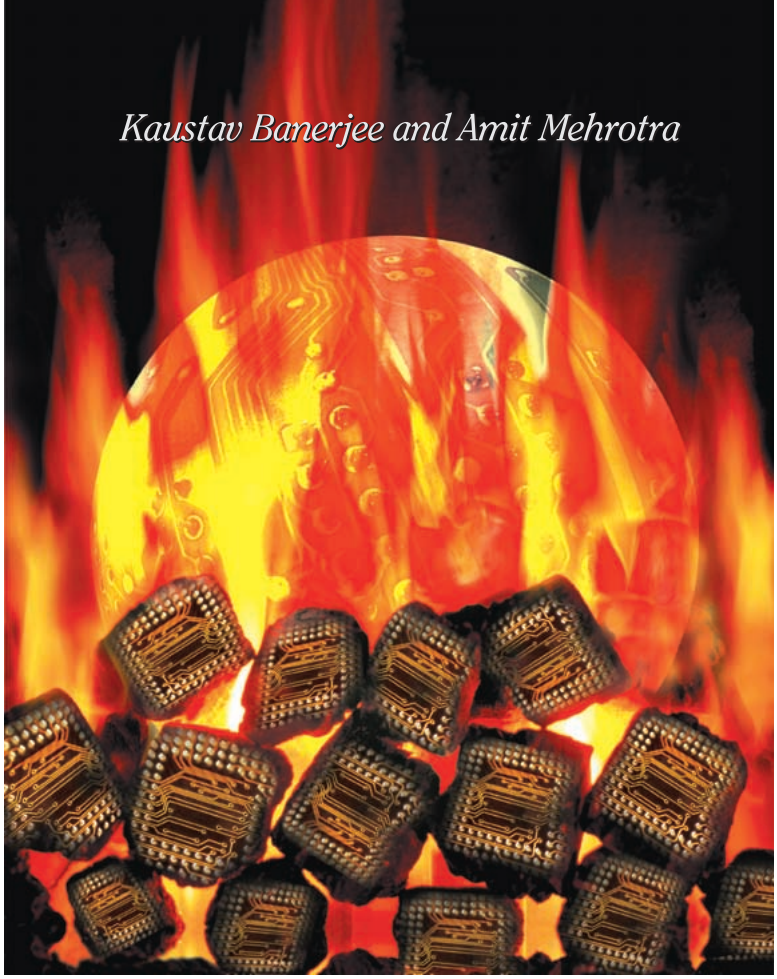
Kaustav Banerjee and Amit Mehrotra



Global (Interconnect) Warming

Understanding Thermal Effects in VLSI Interconnects Is Important for Handling Performance and Reliability Issues as the Technology Continues to Scale

This article presents a comprehensive analysis of the thermal effects in advanced high-performance interconnect systems arising due to self-heating under various circuit conditions, including electrostatic discharge (ESD). Technology (Cu, low-k, etc.) and scaling effects on the thermal characteristics of the interconnects, and on their electromigration (EM) reliability, have been analyzed simultaneously, which have important implications for providing robust and aggressive deep sub-micron (DSM) interconnect design guidelines. The analysis takes into account the effects of increasing interconnect (Cu) resistivity with decreasing line dimensions and the effect of a finite barrier metal thickness. Furthermore, the impact of these thermal effects on the design (driver sizing) and optimization of the in-

 8755-3996/01/$10.00 ©2001 IEEE

terconnect length between repeaters at the global-tier signal lines are investigated. Finally, the reliability implications for minimum-sized vias in optimally buffered signal nets will also be quantified.

Aggressive interconnect scaling has resulted in increasing current densities and associated thermal effects.

Overview

As VLSI technology scales, interconnects are becoming the dominant factor determining system performance and power dissipation [1, 2]. Additionally, interconnect reliability due to *electromigration* and *thermal effects* is fast becoming a serious design issue particularly for long signal lines [3-5]. Thermal effects are an inseparable aspect of electrical power distribution and signal transmission through the interconnects due to *self-heating* (or *Joule heating*) caused by the flow of current. Current flow in a VLSI interconnect causes a power dissipation of I^2R, where I is the current through the interconnect and R is the line resistance. Since the interconnects, especially the global-tier interconnects, are far away from the substrate, which is attached to the heat sink, the heat generated due to this I^2R power dissipation cannot be efficiently removed and therefore causes an increase in interconnect temperature. This phenomenon is referred to as Joule heating or self-heating. Even though this I^2R power dissipation is not a major portion of the total chip power dissipation, since this power is dissipated by the interconnects, which are separated from the substrate by a dielectric that has very low thermal conductivity, it can cause significant temperature rise in the interconnect. In fact, it has been recently shown that interconnect Joule heating in advanced technology nodes can strongly impact the magnitude of the maximum temperature of the global lines despite negligible changes in chip power density [6], which will, in turn, strongly affect the EM lifetime of the interconnect.

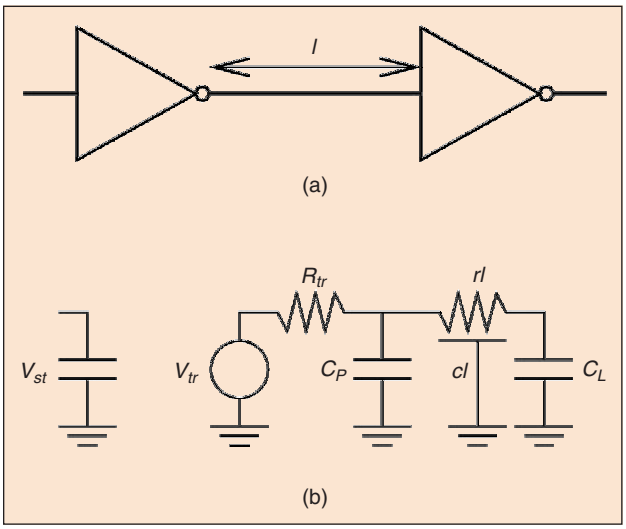
The ever-increasing demand for speed and functionality in silicon-based VLSI systems has caused aggressive scaling of devices, reduction in the interconnect pitch, and increase in metallization levels. This aggressive interconnect scaling has resulted in increasing current densities [7] and associated thermal effects. Furthermore, low dielectric constant (low-k) materials are being introduced as an alternative insulator to silicon dioxide to reduce interconnect capacitance (therefore delay) and cross-talk noise to enhance circuit performance [8]. These materials can further exacerbate thermal effects owing to their poor thermal properties [9].

Apart from normal circuit conditions, ICs also experience high-current stress conditions, the most important of them being ESD, which causes accelerated thermal failures [10]. Semiconductor industry surveys indicate that ESD is the largest single cause of failures in ICs [11]. Interfacing between multiple power supply chips, as well as between multiple power supply blocks within a chip, also causes high-current conditions at the I/O circuitry [12]. Interconnects are also known to experience similar stress conditions during testing for latch-up robustness. The need to understand high-current behavior of VLSI circuits has increased in importance due to their continuous scaling.

Trends in VLSI Scaling and Implications for Thermal Effects

In this section various interconnect technology scaling trends will be discussed to illustrate their impact on thermal effects. Even though VLSI circuits continue to be scaled aggressively, rapid increase in functional density has resulted in a steady increase in chip size. This has resulted in an increasing number of interconnect levels and reduction in interconnect pitch in order to realize all the inter-device and inter-block communications. The number of interconnect metal levels is projected to increase from six levels at the 180-nm node to nine levels at the 50-nm node [13]. Furthermore, the critical dimensions of contacts and vias are also decreasing with scaling, 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, it is envisioned that thermal effects in interconnects can potentially become another serious design constraint.

A simple analysis to demonstrate the implications of technology scaling on technology performance is now presented. This analysis is instructive since these implications, 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 Fig. 1(a).



1. Interconnect of length l between two identical inverters.

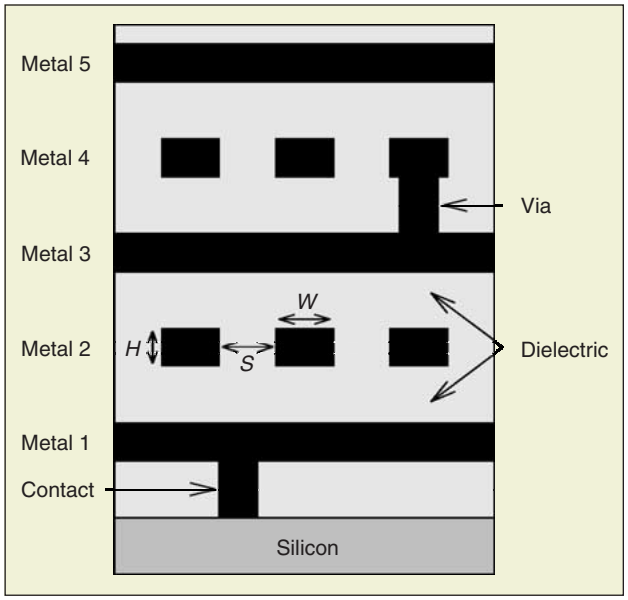
Figure 1(b) shows the equivalent distributed RC representation. Here the inverter on the left that is driving the interconnect is represented as a voltage source V_{tr} controlled by the voltage V_{st} at the input capacitance. R_{tr} is the equivalent transistor resistance and C_p is the parasitic capacitance 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 delay time t_d for a logic signal, which is given by the following equation [14]:

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

where 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 [15]. The first term on the right-hand side of Eq. (1) represents the intrinsic gate delay while the other two terms represent 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 [16].

Let $s < 1$ be the scaling parameter, which is defined as the ratio of the feature size at a newer technology node to the feature size at an older reference technology node. There are two simplified scaling scenarios for interconnects:

- scale metal pitch $W + S$ at constant metal thickness H (Fig. 2) and
- scale metal pitch and the metal thickness.



2. A schematic cross-section of a multilevel interconnect scheme employed in present VLSI circuits. The metal pitch is defined as W + S and the aspect ratio is defined as H / W.

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 all scale as $1/s^2$. 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 Eq. (1) and therefore lowers the delay component $R_{tr}cl + \frac{1}{2}rcl^2$. 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 [16]

$$P_{dyn} = \frac{1}{2}\alpha CV^2 f \quad (2)$$

where α 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, 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 cause increased thermal effects in interconnects can be summarized as:

- increasing current density;
- increasing number of interconnect levels;
- introduction of low-k dielectric materials; and
- increased thermal coupling.

Impact on Interconnect Reliability and Design

Thermal effects impact interconnect design and reliability in the following ways. First, they limit the maximum allowable RMS current density, $j_{\text{rms}-\text{max}}$ [see Eq. (9)] in the interconnects, in order to limit the temperature increase. Second, interconnect lifetime (reliability), which is limited by EM, has an exponential dependence on the inverse metal temperature [17]. Hence, temperature rise of metal interconnects due to the self-heating phenomenon can also limit the maximum allowed average current density $j_{\text{avg}-\text{max}}$, since EM capability is dependent on the average current density [18]. Third, thermally induced open-circuit metal failure under short-duration high peak currents including ESD is also a reliability concern [19]. As the number of wired gates G per chip increases, the number of I/O pins, N_P also increases, as per Rent's Rule [16], which is given by

$$N_P = KG^{\beta} \quad (3)$$

where K is the average number of I/Os per gate and β is the Rent exponent that can vary from 0.1 to 0.7. As a consequence of this increase of 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 [20]. 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. Additionally, ESD events can introduce latent EM damage, which has important reliability implications [21, 22].

A detrimental effect of using low-k dielectric materials is that the metal temperature increases, which makes these lines more susceptible to high-current failures.

Presently, interconnect design rules are not generated in a self-consistent manner [13, 23]; i.e., EM and self-heating are not simultaneously considered for generating EM lifetime guidelines for interconnects. Recently, Hunter [24] solved the EM lifetime equation for Al-Cu, and the one-dimensional (1-D) heat equation, in a self-consistent manner that comprehended both EM and self-heating simultaneously. Rzepka et al. [25] have carried out detailed simulations of self-heating in multilevel interconnects using finite element analysis. Their study concluded that, in the near future, multilevel interconnect arrays will be affected by self-heating more severely. As deep sub-micron interconnect technologies are rapidly evolving, with the introduction of Cu and low-k dielectrics, there is an increasing need to understand their effects simultaneously on the thermal characteristics of these interconnects and on their EM reliability, in order to provide robust design guidelines. Furthermore, it is not clear whether thermal constraints conflict with the performance optimization steps employed at the circuit level. Hence, a 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.

Scope of this Study

The analysis presented in this article examines the self-consistent solutions for allowed interconnect current density for technologies up to 50 nm involving Cu and various low-k materials as per the International Technology Roadmap for Semiconductors (ITRS) [13] and compares it with the current densities obtained by performance optimization methodologies. Both unipolar current waveforms, which are present in power supply lines, and bipolar current waveforms, which are present in signal lines, have been considered in this analysis that simultaneously comprehends electromigration and Joule heating. Since in the ITRS roadmap, the line widths of even global wires scale with technology, one needs to consider the metal resistivity increase with technology scaling due to increased electron scattering from the interfaces [26] and because of a greater fraction of interconnect area being consumed by metal barrier. Also for global lines, the interconnect cross-section area is large and therefore the rms current is fairly large. However, this current also flows through the contacts and vias that connect the global metal lines to silicon substrate. In this work we also quantify thermal effects in vias and show that if minimum-sized vias are used for optimal buffering of global interconnects, the lines may not be a reliability concern but the current densities in the minimum sized vias will cause reliability concerns. Finally, design guidelines for high-current robustness of interconnects used in the ESD protection and I/O circuitry are also examined in detail.

Preliminaries

Average, RMS, and Peak Current Densities

Circuit designers are typically provided with the maximum allowable values for three interconnect current densities. These are the average current density j_{avg}, the RMS current density j_{rms}, and the peak current density j_{peak}. These quantities are defined as follows. The peak current density is simply the current density corresponding to the peak current of the waveform

$$j_{peak} = \frac{I_{peak}}{A} \tag{4}$$

where A is the cross-sectional area of the interconnect. The average current density is defined as

$$j_{avg} = \frac{1}{T}\int_0^T j(t)\,dt \tag{5}$$

where T is the time period of the current waveform. The RMS current density is defined as

$$j_{rms} = \sqrt{\frac{1}{T}\int_0^T j^2(t)\,dt}. \tag{6}$$

For a fixed temperature, EM lifetime of interconnects is known to be determined by j_{avg} [18]. Self-heating is determined by j_{rms}. Presently, high-performance interconnect design is based on the specified limits for the maximum values of the average, RMS, and peak current densities [23]. However, as pointed out earlier, they are not self-consistent; i.e., these values do not simultaneously comprehend the two temperature-dependent mechanisms: EM and self-heating.

Electromigration

EM is the transport of mass in metals under an applied current density and is widely regarded as a major wear-out or failure mechanism of VLSI interconnects [17]. When current flows through the interconnect metal, an electronic wind is set up opposite to the direction of current flow. These electrons, upon colliding with the metal ions, impart sufficient momentum and displace the metal ions from their lattice sites, creating vacancies. These vacancies condense to form voids that result in increase of interconnect resistance or even open-circuit conditions [27]. EM lifetime reliability of metal interconnects is modeled by the well-known Black's equation [17], given by

$$TTF = A^* \, j^{-n} \exp\left(\frac{Q}{k_B T_m}\right) \tag{7}$$

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, and the exponent n is typically 2 under normal use conditions. The activation energy Q in narrow (< 1 μ) copper lines is dominated by surface transport [28] and is ~ 0.5 eV [29], k_B is the Boltzmann's constant, and T_m is the metal temperature. Therefore, the EM lifetime of copper interconnects decreases as line width decreases, as opposed to AlCu lines where EM lifetimes are known to improve with line scaling [30]. The typical goal is to achieve a 10-year lifetime at 100 °C, for which Eq. (7) and accelerated testing data produce a design rule value for the acceptable current density at T_{ref}, j_0.

Self-Heating

The effect of self-heating can be analyzed from the following. The metal temperature T_m in Eq. (7) is given by

$$T_m = T_{\text{ref}} + \Delta T_{\text{self-heating}} \tag{8}$$

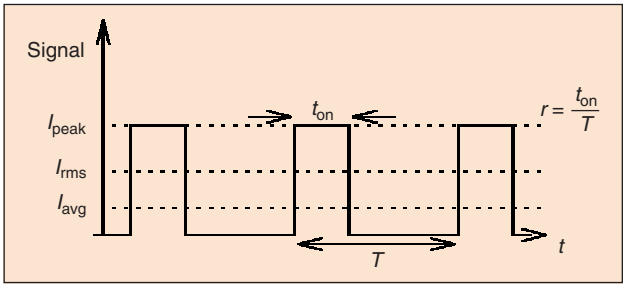
and, under steady-state thermal conditions,

$$\Delta T_{\text{self-heating}} = (T_m - T_{\text{ref}}) = \frac{1}{T}\int_0^T I^2 R R_\theta \, \mathrm{d}t = I_{\text{rms}}^2 R R_\theta \tag{9}$$

where T_{ref} is the reference chip (silicon junction) temperature and is typically taken as 100 °C to 120 °C, $\Delta T_{\text{self-heating}}$ is the temperature rise of the metal interconnect due to the flow of current, R is the interconnect resistance, and R_θ is the thermal impedance of the interconnect line to the substrate. This equation assumes that the frequency of interconnect current is much greater than the inverse of thermal time constant (a few MHz), which implies that the metal temperature has very small variations about T_m. The reference temperature of the chip T_{ref} results primarily due to the total power dissipation in the chip and consists of switching, leakage, short-circuit, and static power dissipation [31]. T_{ref} can be estimated from the following equation [6]:

$$T_{\text{ref}} = T_0 + R_n\left(\frac{P}{A}\right) \tag{10}$$

where T_0 is the chip ambient temperature, P is total power dissipation, and A is the chip area. Here R_n represents the substrate (Si) layer plus the package thermal resistance. Using Eq. (10), R_n for the present technology node (180 nm) can be calculated. Assuming the same value for R_n, the die temperatures at other technology nodes can be estimated using Eq. (10). R_n is dominated by the package thermal resistance [6]. Thus, both EM and self-heating are temperature-dependent effects, and as self-heating increases, EM lifetime decreases exponentially according to Eq. (7).



3. A unipolar pulsed waveform illustrating various current definitions.

Coupled Electromigration and Self-Heating Analysis

Interconnects can be broadly classified into two categories: signal lines and power lines (clock lines and data busses are special cases of signal lines). They differ in that currents in signal lines are bidirectional (or bipolar) [18], while those in power lines are usually unidirectional (or unipolar). We now present the coupled EM and self-heating analysis for unipolar and bipolar pulses separately.

Unipolar Current Stress Condition

We first consider unipolar current waveforms for some illustrative analysis. Using the definitions in the "Preliminaries" section, for the three current densities, one can easily show that

$$j_{\text{avg}} = r j_{\text{peak}} \tag{11}$$

and

$$j_{\text{rms}} = \sqrt{r}\, j_{\text{peak}} \tag{12}$$

where r is the duty cycle defined as t_{on} / T in Fig. 3.

We now introduce the formulation of the self-consistent solutions [24] for allowed interconnect current density and then apply them to analyze Cu interconnects. $\Delta T_{\text{self-heating}}$ in interconnects given by Eq. (9) can be written in terms of the RMS current density as

$$j_{\text{rms}}^2 = \frac{(T_m - T_{\text{ref}}) K_{\text{ins}} W_{\text{eff}}}{t_{\text{ins}} t_m W_m \rho_m (T_m)}. \tag{13}$$

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_θ in Eq. (9) has been expressed as

$$R_\theta = \frac{t_{ins}}{K_{ins} L W_{eff}}. \tag{14}$$

This expression for the thermal impedance is based on a quasi-2-D heat conduction model with

$$W_{eff} = W_m + 0.88 t_{ins}$$

valid for $W_m / t_{ins} > 0.4$ and is accurate to within 3% [32]. 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. It can be deduced from Eq. (14) that, under steady-state Joule heating, as t_{ins} increases, the temperature rise increases, which results in higher temperatures for global-tier metal lines.

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

$$\frac{\exp\left(\frac{Q}{k_B T_m}\right)}{j_{avg}^2} \geq \frac{\exp\left(\frac{Q}{k_B T_{ref}}\right)}{j_0^2}. \tag{15}$$

From Eqs. (11) and (12) we have, after eliminating j_{peak},

$$\frac{j_{avg}^2}{j_{rms}^2} = r. \tag{16}$$

Substituting for j_{rms}^2 from Eq. (13) and j_{avg}^2 from Eq. (15) in (16), we get the self-consistent equation given by

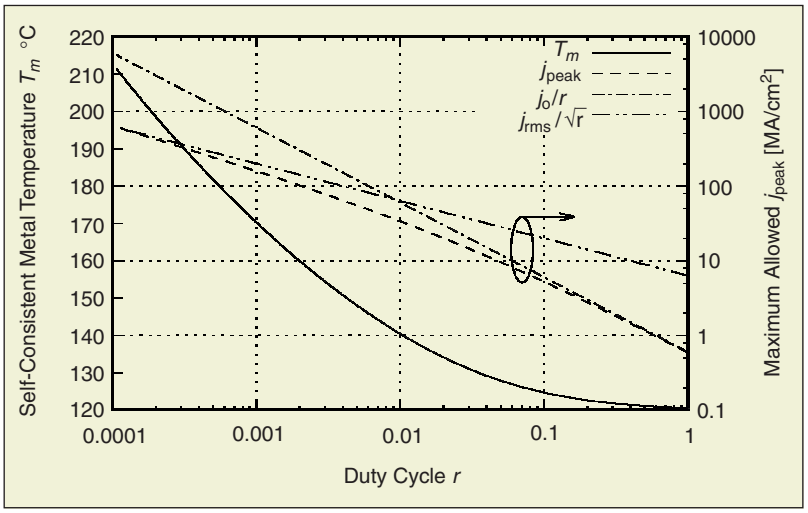
$$r = j_0^2 \frac{\exp\left(\frac{Q}{k_B T_m}\right)}{\exp\left(\frac{Q}{k_B T_{ref}}\right)} \frac{t_{ins} t_m W_m \rho_m(T_m)}{(T_m - T_{ref}) K_{ins} W_{eff}}. \tag{17}$$

Note that this is a single equation in the single unknown temperature T_m. Once this self-consistent temperature is obtained, the corresponding maximum allowed j_{peak} and j_{rms} can be calculated from Eqs. (12) and (13). The self-consistent equation given by Eq. (17) for unipolar pulses is also valid for more general unipolar time varying waveforms with an effective duty cycle r_{eff} [33].

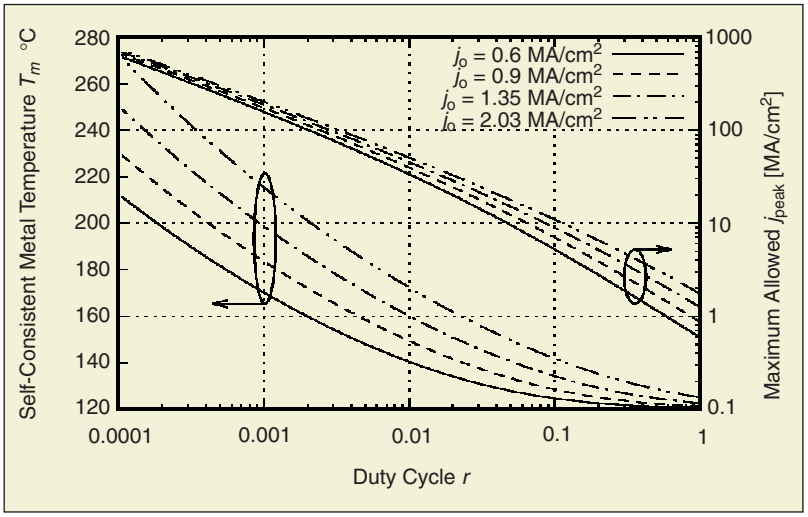
One of the consequences of the self-consistent equation is that, for a certain j_0, as r decreases the self-consistent temperature and the maximum allowed j_{peak} increases. This effect is shown in Fig. 4 for Cu interconnects. Second, it can be observed that as r decreases, the ratio

$$\frac{j_{peak(self-consistent)}}{j_{peak(without\ self-heating,\ i.e.,\ the\ line\ labeled\ j_0/r)}}$$

decreases monotonically. At $r = 10^{-2}$, the self- consistent j_{peak} is nearly two times smaller than the j_{peak} obtained from EM constraint only; i.e., without self-heating. From the EM lifetime relation given by Eq. (7), this implies that if a design used only the average (EM) current as the design guideline without comprehending self-heating, it could have a lifetime nearly four times



4. Self-consistent solutions of T_m and j_{peak} for metal 6 of a 180-nm technology node with $j_0 = 0.6\ MA/cm^2$, $\rho_m(T_m) = 2.7 \times 10^{-6}[1 + 6.8 \times 10^{-3}/K \times (T_m - T_{ref})]\ \Omega$-cm, $T_{ref} = 120\ °C$ and $Q = 0.5\ eV$. The two dotted lines indicate j_{peak} based on (a) $j_{peak} = j_0 / r$ and (b) $j_{peak} = j_{rms} / \sqrt{r}$.



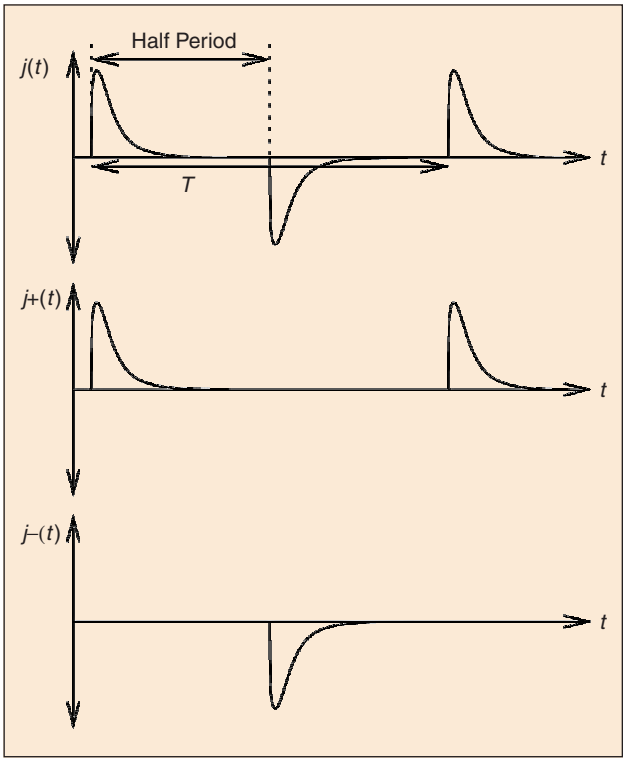
5. Self-consistent analysis showing the dependence of T_m and j_{peak} on j_0 for metal 6 of a 180-nm technology node. Other parameters are the same as in Fig. 4.

smaller than the reliability requirement. Third, it can be observed from Fig. 5 that, as j_0 is increased, the self-consistent interconnect metal temperature T_m increases as expected from Eq. (17). However, the maximum allowed j_{peak} does not increase much for values of $r < 10^{-2}$. That is, j_0 becomes increasingly ineffective in increasing j_{peak} as the duty cycle r decreases. From the above analysis, it apparently seems that EM capability might cease to be the dominant driver of interconnect technology evolution and thermal effects will limit the maximum allowed j_{peak}.

The self-heating of interconnect lines within a 3-D array could be significantly more severe due to thermal coupling between neighboring lines.

Bipolar Current Stress Condition

As pointed out earlier, contrary to the current in power supply lines, current waveform in signal lines is both positive and negative. The average current j_{avg} as defined in Eq. (5) is zero for these signal lines. Therefore, for signal lines, the electromigration problem is not as severe as lines with unidirectional current flow. However, Joule heating is dependent on the RMS current density, which is not altered. Therefore, an expression similar to Eq. (17) needs to be derived for the bipolar current case. Hunter [33] derived an expression for the self-consistent metal temperature for square bipolar currents. However, the current in a VLSI interconnect is not a square waveform and therefore we need to derive the expression for the self-consistent metal temperature for a general bipolar waveform [34].



6. A bipolar pulsed waveform illustrating various current definitions.

Figure 6 shows the bipolar current density waveform $j(t)$ found in typical signal lines. Let $j_+(t)$ denote the positive excursions of the current density and $j_-(t)$ denote the negative excursions of the current density. Therefore,

$$j(t) = j_+(t) + j_-(t).$$

Obviously $j_+(t)$ and $j_-(t)$ are mutually exclusive in t.

For the case of an arbitrary bipolar ac signal, detailed studies have shown that the effective ac value of current density responsible for EM, $j_{EM\ bipolar}$, is given by the average current recovery (ACR) model [33] as follows:

$$j_{EM\ bipolar} = j_{ACR} = \frac{1}{T}\left\{\int_0^T |j_+(t)|\,dt - R\int_0^T |j_-(t)|\,dt\right\} \tag{18}$$

where j_{ACR} is the current density of the ACR model and R is a recovery parameter (< 1) of the ACR model. It heuristically accounts for the degree of healing of EM void damage that occurs when the current direction changes [33].

Now, similar to the unipolar case, in order to achieve the EM reliability lifetime goal mentioned above, we must have the lifetime in any ($j_{EM\ bipolar}$) current density and metal temperature T_m equal to or larger than the lifetime value (e.g., 10 years) under the design rule current density stress j_0 at the temperature T_{ref}. Therefore, we must have

$$\frac{\exp\left(\frac{Q}{k_B T_m}\right)}{j^2_{EM\ bipolar}} \geq \frac{\exp\left(\frac{Q}{k_B T_{ref}}\right)}{j_0^2}. \tag{19}$$

From Eq. (19) it follows that

$$j_{EM\ bipolar} \leq j_0 \exp\left(\frac{Q}{2k_B T_m} - \frac{Q}{2k_B T_{ref}}\right). \tag{20}$$

For an arbitrary bipolar pulse the average current model for EM should be replaced by the ACR model; hence, it follows that

$$j_{ACR} \leq j_0 \exp\left(\frac{Q}{2k_B T_m} - \frac{Q}{2k_B T_{ref}}\right). \tag{21}$$

The current waveforms in signals lines are invariably *symmetric bipolar* waveforms. A symmetric bipolar waveform is defined as

$$\int_0^T |j_+(t)|\,dt = \int_0^T |j_-(t)|\,dt.$$

From now on we will concentrate on developing the self-consistent equation for maximum allowed j_{peak} for symmetric bipolar current waveforms. For these waveforms, Eq. (18) can be rewritten as

$$j_{ACR} = (1-R)\frac{1}{T}\int_0^T |j_+(t)|\mathrm{d}t. \tag{22}$$

Substituting Eq. (22) in Eq. (21) with the equality sign we have

$$\frac{1}{T}\int_0^T |j_+(t)|\mathrm{d}t = \frac{j_0}{1-R}\exp\left(\frac{Q}{2k_BT_m} - \frac{Q}{2k_BT_{ref}}\right). \tag{23}$$

Define the average current as

$$j_{avg\ bipolar} = \frac{1}{T}\int_0^T |j(t)|\mathrm{d}t. \tag{24}$$

It follows that for a symmetric bipolar case

$$j_{avg\ bipolar} = \frac{2}{T}\int_0^T |j_+(t)|\mathrm{d}t = \frac{2}{T}\int_0^T |j_-(t)|\mathrm{d}t. \tag{25}$$

Using Eq. (23),

$$j_{avg\ bipolar} = \frac{2j_0}{1-R}\exp\left(\frac{Q}{2k_BT_m} - \frac{Q}{2k_BT_{ref}}\right). \tag{26}$$

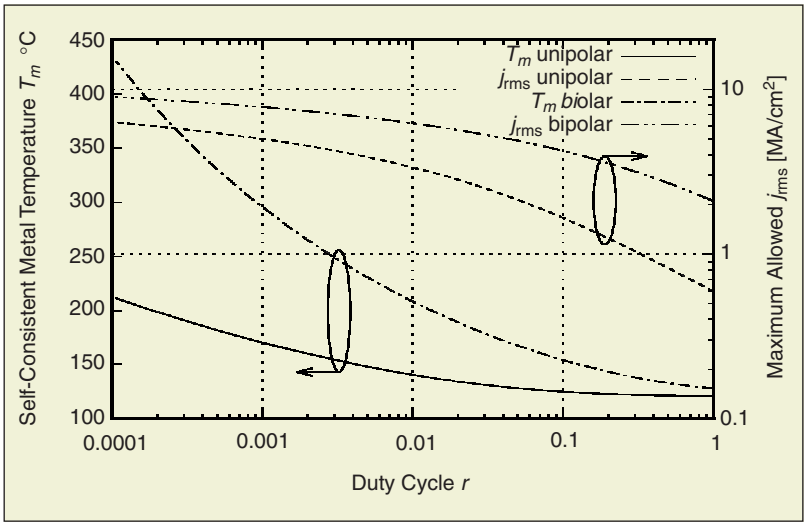
Define the equivalent duty cycle as

$$r = \frac{j^2_{avg\ bipolar}}{j^2_{rms}}. \tag{27}$$

Therefore, from Eqs. (26) and (13),

$$r = \frac{4j_0^2\exp\left(\frac{Q}{k_BT_m} - \frac{Q}{k_BT_{ref}}\right)t_{ins}t_mW_m\rho_m(T_m)}{(1-R)^2(T_m - T_{ref})K_{ins}W_{eff}}. \tag{28}$$

Figure 7 compares the maximum allowed j_{rms} and metal temperature T_m for unipolar and bipolar pulses for $R = 0.5$. Note that j_{peak} will depend on the actual shape of the bipolar waveform and therefore j_{rms} is used for comparison. Equivalent duty cycle is defined in Eq. (27). Note that if only half-period was considered (see Fig. 6), the effective r for the bipolar case as defined in Eq. (27) would be the same as the unipolar case. From Figure 7 we observe that the maximum allowed j_{rms} increases from the unipolar case. However, as the effective duty factor decreases, the maximum allowed j_{rms} values for the two case become very similar. For instance, for $r = 1$, the allowed j_{rms} for the bipolar case is three times higher than the allowed j_{rms} for the unipolar case, while at $r = 10^{-4}$, this factor is only 1.2.



7. Self-consistent solutions of T_m and j_{peak} for metal 6 of a 180-nm technology node for unipolar and bipolar pulses with $R = 0.5$. Other parameters are the same as in Fig. 4.

Quasi-2-D Heat Conduction

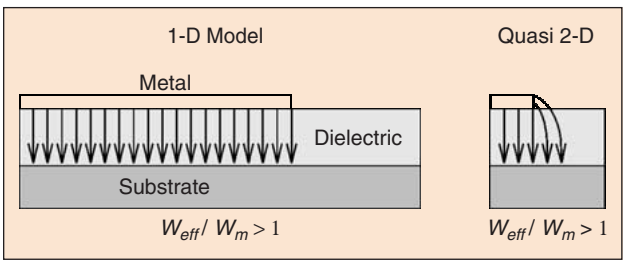
We now analyze the effects of line-width scaling on the thermal impedance defined by Eq. (14), which is accurate to within 3% for W_m / t_{ins} 0.4 [32]. In current DSM technologies, the ratio W_m / t_{ins} gets even smaller (especially for the top-level metal lines), and W_{eff}, which is meant to account for the extra heat conduction from the sides (illustrated in Fig. 8), must be obtained for real DSM interconnect structures experimentally. We will first express W_{eff} in a more general form:

$$W_{eff} = W_m + \phi t_{ins}. \tag{29}$$

Here ϕ is introduced as the heat-spreading parameter that must be extracted from data. In this work we will use $\phi = 0.88$ as the heat-spreading parameter. For DSM technologies our estimates for maximum allowed j_{peak} may therefore be somewhat conservative.

Reliability-Based Limits of Interconnect Current Densities

We now consider the technology specifications for the VLSI technology process ranging from 180 nm to 50 nm as per the ITRS [13] shown in Table 1. The interconnect metal is assumed to be Cu. We then solve for the self-consistent metal temperature T_m and the corresponding maximum allowed j_{rms} and j_{peak} values.



8. Quasi-2-D heat conduction.

Table 1. ITRS interconnect parameters for 180-nm to 50-nm technologies. All dimensions are in nm. Pitch is twice the width for all cases. *n* is the number of layers in the current tier. t_{ild} is the inter-layer dielectric thickness. ε_r is the dielectric constant.

Tech.	ε_r	k_{ins} ($W/(m–K)$)	Local Tier				Semiglobal Tier				Global Tier			
			n	w_m	t_m	t_{ild}	n	w_m	t_m	t_{ild}	n	w_m	t_m	t_{ild}
180	3.75	1.05	2	250	350	350	2	320	640	672	2	525	1155	1260
130	3.1	0.54	3	182.5	273.75	273.75	2	232.5	511.5	488.25	2	382.5	956.25	1032.75
100	1.9	0.19	3	132.5	225.25	225.25	3	170	408	374	2	280	756	784
70	1.5	0.12	3	92.5	175.75	125.75	3	120	300	276	3	195	546	565.5
50	1.25	0.07	3	65	136.5	136.5	3	82.5	222.75	198	3	137.5	398.75	412.5

Before presenting the self-consistent solutions, we must point out an important caveat, which is the distinction between *thermally long* and *thermally short* metal lines. Since interconnect leads are typically connected to the diffusion (or another metal lead) through a contact (or via), the temperature at the end regions of these lines is generally lower than the temperature in regions far away from the ends. Under steady-state conditions, the governing heat equation for an isolated metal line can be expressed as

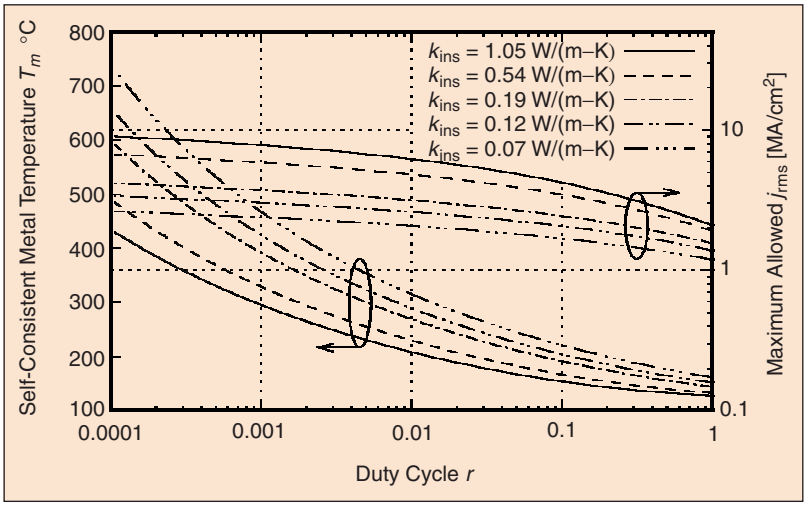
$$\frac{d^2T}{dx^2} - \frac{T - T_0}{\lambda^2} + \frac{\rho_m j_{rms}^2}{K_m} = 0$$

where T_0 is the temperature at the ends of the line, K_m is the thermal conductivity of the interconnect metal, and λ is the characteristic thermal diffusion length, which is given by [35]:

$$\lambda = \sqrt{\frac{t_m K_m t_{ins}}{K_{ins}\left(1 + 0.88\frac{t_{ins}}{W_m}\right)}}$$

where λ is of the order of 1 μm-20 μm. Metal lines that are much longer than λ are termed thermally long while those that are close to the value of λ are termed thermally short.

We will primarily focus on the analysis of the thermally long lines, which give rise to worst case scenarios. We will not concentrate on thermal effects for small intra-block interconnects. Their lengths are usually of the same order of magnitude as the thermal characteristic length. Also, these nets are typically routed on the first few levels of metal, which are closer to the silicon substrate, and hence the thermal problem is not as severe. Long interconnects that could be much larger than the thermal characteristic length will therefore form the main focus of the following analysis. These long lines usually realize the inter-block communication and hence, for delay purposes, they are routed on the higher, less resistive top-few layers of metal. However, these metal layers are usually far away from the silicon substrate and hence can potentially experience thermal problems. Therefore, self-consistent solutions were obtained only for the top metal layers.



9. Self-consistent solutions for maximum allowed j_{rms} and T_m for metal 6 in a 180-nm technology for a symmetric bipolar waveform for $R = 0.5$ for different dielectrics.

We begin by analyzing the effect of introducing new dielectric materials on reliability. In Fig. 9 the self-consistent values of T_m and j_{rms} are plotted as a function of duty cycle (r) for different dielectrics but using the top-level metal dimensions of the 180-nm technology node for symmetric bipolar current waveforms. It can be observed that j_{rms} decreases significantly as dielectrics with lower thermal conductivity are introduced. For small values of r, j_{rms} varies very slowly with r, thereby signifying the increasing importance of self-heating.

This indicates that introduction of better interconnect materials to improve performance becomes increasingly ineffective in increasing j_{rms} for dielectrics with poor thermal properties. Another detrimental effect of using low-k dielectric materials is that the metal tempera-

ture increases, as is evident in Fig. 9, which makes these lines more susceptible to high-current failures [3, 9].

Figure 10 plots the maximum allowed j_{rms} for the top-level metal lines for the ITRS technology nodes for power supply and ground lines. The duty factor therefore is assumed to be 1, and Eq. (17) is used to determine T_m and j_{rms}. Recall that for $r = 1$, $j_{peak} = j_{rms} = j_{avg}$ for unipolar current waveforms. Figure 11 plots the maximum allowed j_{rms} for the top-level metal lines for the ITRS technology nodes for symmetric bipolar waveforms. The duty factor r is taken to be 0.3 for all technologies except for the 50-nm technology, for which $r = 0.44$. The reason for this choice is explained in the next section. It is evident from Fig. 11 that j_{rms} shows a slight decreasing trend as the technology scales. It is instructive to contrast this result with Fig. 9, which shows j_{rms} variation by *only* considering different dielectric materials while the interconnect geometry is not scaled. From Fig. 9 it can be observed that if interconnect geometry is not scaled, maximum allowed j_{peak} decreases by a factor of 2.25 from 180-nm technology to 50-nm technology node due to decreasing thermal conductivity of dielectrics. If scaled interconnects as per ITRS specifications are considered, j_{rms} reduces by only 16.8%. Therefore, for isolated lines, scaling of interconnect geometries helps to offset the detrimental effects of using dielectric materials with poor thermal properties.

Performance-Based Interconnect Current Densities

As a next step we outline a methodology for computing current density from performance considerations only. Consider an interconnect of length l between two buffers. The schematic representation is shown in Fig. 1. Figure 1(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:

$$\tau = b(x)R_{tr}(C_L + C_P) + b(x)(cR_{tr} + rC_L)l + a(x)rcl^2 \quad (30)$$

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 $b = 0.7$ [36]. 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 $C_P = sc_p$ and $C_L = sc_0$. If the total interconnect of length L is divided into n segments of length $l = L / n$, then the overall delay is given by

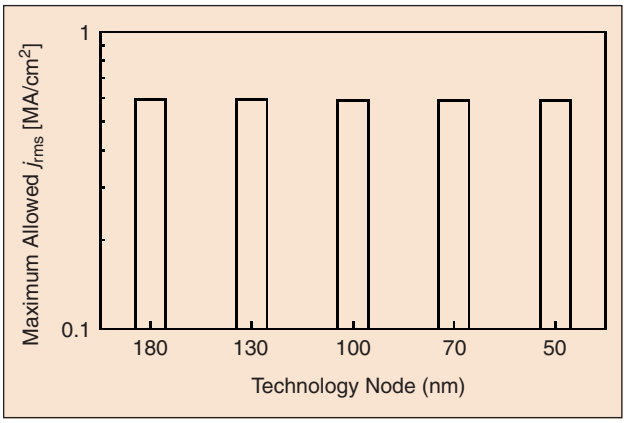
$$\begin{aligned} T_{delay} &= n\tau \\ &= \frac{L}{l}b(x)r_0(c_0 + c_p) + b(x)(c\frac{r_0}{s} + src_0)L + a(x)rclL. \end{aligned} \quad (31)$$

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

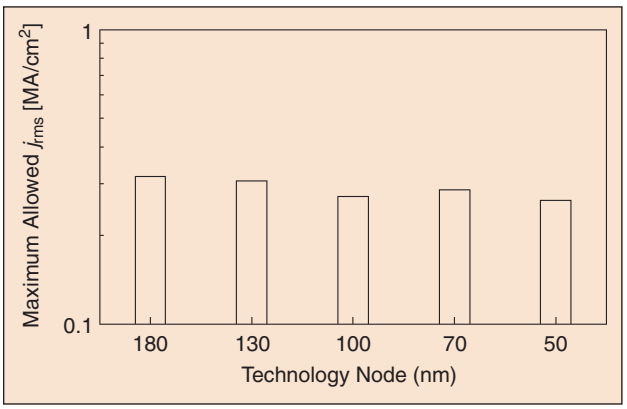
$$l_{opt} = \sqrt{\frac{b(x)r_0(c_0 + c_p)}{a(x)rc}} \quad (32)$$

$$s_{opt} = \sqrt{\frac{r_0 c}{rc_0}}. \quad (33)$$

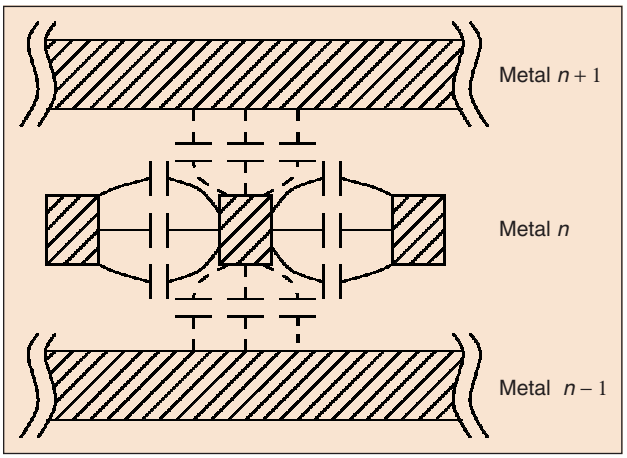
Note that s_{opt} is independent of the switching model; i.e., x.



10. Maximum allowed j_{rms} for top-layer metal of ITRS technologies for dc currents.



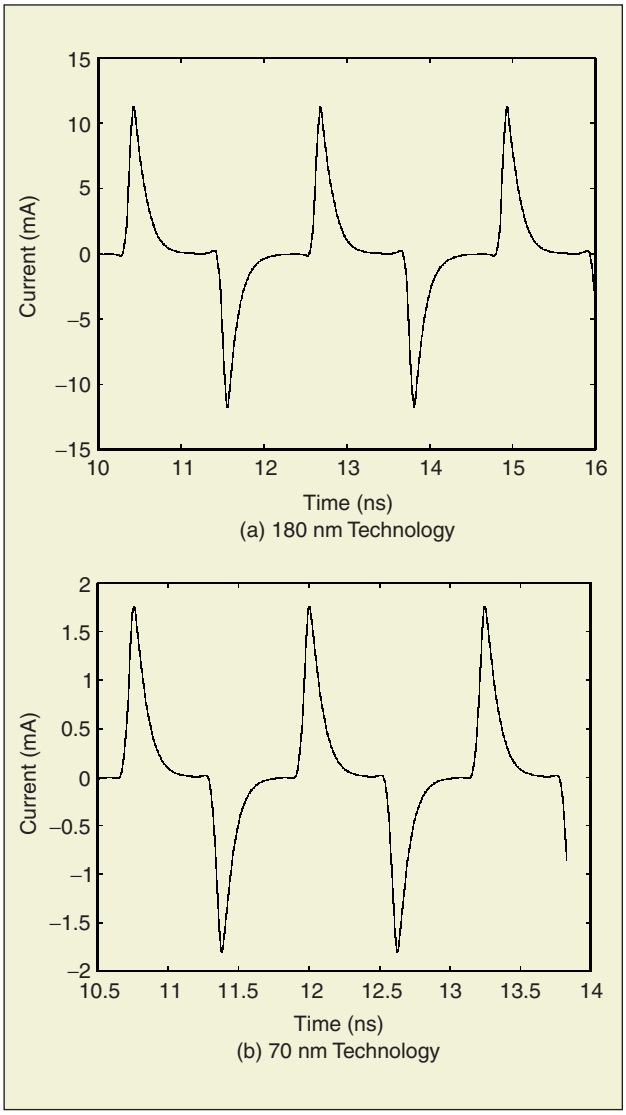
11. Maximum allowed j_{rms} for top-layer metal of ITRS technologies for symmetric bipolar currents.



12. Typical metal interconnect structure.

Table 2. Optimized interconnect and buffer parameters for global tier for various ITRS technology nodes at 120 °C.

Tech.	l_{opt} (mm)	s_{opt}	j_{rms} (MA/cm^2)	$j_{avg\ bipolar}$ (MA/cm^2)	Duty Cycle r
180	3.33	174	0.622	0.339	0.297
130	2.5	151	0.66	0.368	0.311
100	2.22	110	0.61	0.34	0.311
70	1.32	82	0.66	0.373	0.319
50	1.06	53	0.46	0.302	0.431



13. Current waveforms in the top-layer metal lines obtained from SPICE simulation.

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 Fig. 12, we performed a full 3-D-capacitance extraction using FASTCAP [14] for signal lines at various metal levels to obtain the values of c.

This inverter-interconnect structure can be simulated using SPICE to obtain the interconnect current waveforms along the interconnect. In order to compare the current densities using SPICE simulations and those computed in Fig. 11, an appropriate value of r needs to be chosen for simulations. Since these lines are mostly global-tier interconnects, they are expected to carry signals at almost every clock cycle unless the block they are communicating to is powered down. Therefore, this inverter-interconnect structure is used as a delay stage in a multistage ring oscillator and the current waveforms and current densities along the interconnect are obtained. These current waveforms and Eq. (27) are used to determine the equivalent duty factor. Table 2 summarizes these values for various ITRS technologies.

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 voltage dependent and therefore change during the output transition. Therefore, accurate values of optimal interconnect length and buffer size need to be determined by SPICE simulations. 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 that 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 that minimizes the ratio of the stage delay and interconnect length. 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.

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

Also, the relative rise and fall slope was found to be the same across all technologies. From our simulations it was observed that drivers and interconnects optimized using Eqs. (33) and (32) maintain good slew rates for rising and falling transitions across all technologies, except the 50-nm node, with an effective duty cycle $\left(r = j^2_{\text{avg bipolar}} \ / \ j^2_{\text{nms}}\right)$ of 0.31 ± 0.01, as shown in Table 2. r is higher for the 50-nm technology node because the transistors don't switch off completely (according to BSIM3 models) and therefore the interconnect waveform is affected. Since for the signal lines the current waveform is symmetric and bipolar, $j_{\text{avg bipolar}}$ is computed over half the time period to obtain r. The interconnect waveforms for the top-layer metal lines for 180-nm and 70-nm technology nodes are shown in Fig. 13. It can be observed that the relative rise and fall skew is the same across both technologies.

Figure 14 shows the comparison of $j_{performance}$ with the values of $j_{reliability}$ for various technology nodes. It can be observed that $j_{performance}$ is always lower than $j_{reliability}$ for all technologies. This implies that the optimum interconnect length of an isolated signal line is determined solely by performance considerations.

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 Fig. 12. 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 Fig. 12.) $C_{neighbor}$ 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 especially 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 signals are switching in the opposite direction or the same direction.

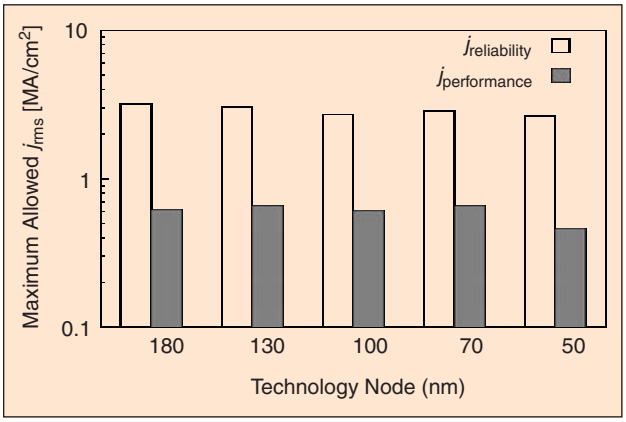
Furthermore, it is observed from simulation and can be shown that the interconnect current remains almost the same if the load capacitance of a buffer changes. The rise and fall time will be affected significantly, but the interconnect current does not change appreciably if the buffer drive strength is unchanged. However, for the slower transition r increases, and for the faster transition r decreases (typical values observed in simulation are 0.4 for slower transition and 0.2 for the faster transition).

From Fig. 7, it can be observed that $j_{rms-reliability}$ also does not change appreciably for this range of duty factors. Hence, the observations made in Fig. 14 will also not change much if the interconnect capacitance changes with signal patterns.

Effect of Interconnect Dimension on Copper Resistivity

The experiments in the previous section were carried out assuming that the metal resistivity does not change with line width. In an actual VLSI interconnect, metal resistivity starts increasing as the minimum dimension of the metal line becomes comparable to the *mean free path* of the electrons (i.e., interconnect metal is considered to be a *thin-film*) [34]. This is because surface scattering starts having a non-negligible contribution to the resistivity compared to the contribution due to bulk scattering. Furthermore, surface scattering also reduces the thermal coefficient of resistivity of material.

Another effect that is responsible for increased resistivity is the presence of barrier material for copper interconnects (see Fig. 18). Since the resistivity of the barrier material is extremely high compared to copper, it can be assumed that all the current is carried by copper. Therefore, the effective area through which the current conduction takes place is reduced, or, equivalently, the effective resistivity of the metal line of the same drawn dimension is in-



14. Comparison of j_{rms} values obtained from reliability and electrical performance considerations for top-layer metal for various ITRS technology nodes.

Table 3. Resistivity and temperature coefficient ratios for the global tier metals for various technologies. All dimensions in nm. Barrier thickness of 10 nm assumed for all technology nodes.

Tech.	w	$\frac{\rho}{\rho_{0\ thin-film}}$	$\frac{\rho}{\rho_{0\ barrier}}$	$\frac{\rho}{\rho_{0\ eff}}$	$\frac{\alpha}{\alpha_0}$
180	525	1.0162	1.0487	1.0657	0.9527
130	382.5	1.0224	1.0663	1.0902	0.9353
100	280	1.0308	1.0914	1.1250	0.9122
70	195	1.0448	1.1351	1.1859	0.8752
50	137.5	1.0646	1.2003	1.2779	0.8263

Table 4. Optimized interconnect and buffer parameters for global tier for various ITRS technology nodes taking into account increased resistivity due to surface scattering and barrier materials.

Tech.	l_{opt} (mm)	s_{opt}	j_{rms} (MA/cm^2)	$j_{avg\ bipolar}$ (MA/cm^2)	r
180	3.0	179	0.633	0.344	0.296
130	2.4	146	0.643	0.353	0.302
100	2.12	96	0.559	0.311	0.310
70	1.2	82	0.626	0.361	0.332
50	0.99	48	0.400	0.263	0.432

creased. This becomes more of a problem as metal lines scale since it is very difficult to scale the thickness of the barrier material.

It is envisioned that thermal effects in interconnects can potentially become another serious design constraint.

According to [26] the resistivity ρ of a thin-film metal can be expressed in terms of bulk resistivity ρ_0 as

$$\frac{\rho_0}{\rho} = 1 - \frac{3}{2k}(1-p)\int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5}\right)\frac{1-e^{-kx}}{1-pe^{-kx}}\,dx$$

where $k = d/\lambda_{mfp}$, d is the smallest dimension of the film (in our case, the width), λ_{mfp} is the bulk mean free path of electrons, and p is fraction of electrons that are elastically reflected at the surface. For copper, $p = 0.47$ and $\lambda_{mfp} = 421$ Å at 0 °C [37]. Moreover, since the temperature alters the mean free path of the electrons, the temperature coefficient α of the thin film of metal is also different from its bulk value. α can be related to the bulk temperature coefficient α_0 as [26]:

$$\frac{\alpha}{\alpha_0} = \frac{1 - \frac{3}{k}(1-p)\int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5}\right)\frac{1-e^{-kx}}{1-pe^{-kx}}\,dx}{1 - \frac{3}{2k}(1-p)\int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5}\right)\frac{1-e^{-kx}}{1-pe^{-kx}}\,dx} + \frac{\frac{3}{2}(1-p)^2\int_1^\infty \left(\frac{1}{x^2} - \frac{1}{x^4}\right)\frac{e^{-kx}}{(1-pe^{-kx})^2}\,dx}{1 - \frac{3}{2k}(1-p)\int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5}\right)\frac{1-e^{-kx}}{1-pe^{-kx}}\,dx}$$

The resistivity and temperature coefficient ratios for the global-tier metals for various technologies is given in Table 3. Table 4 shows $l_{opt}, s_{opt}, j_{rms}, j_{avg\ bipolar}$ and r values recalculated taking into account increased resistivity due to surface scattering and the presence of barrier materials. Figure 15 shows the $j_{rms\ reliability}$ and $j_{rms\ performance}$ considering the increased metal resistivity and decreased thermal coefficient of resistivity due to decreasing metal line width. It can be seen that both $j_{rms\ reliability}$ and $j_{rms\ performance}$ reduce slightly as compared to the case when bulk values are used.

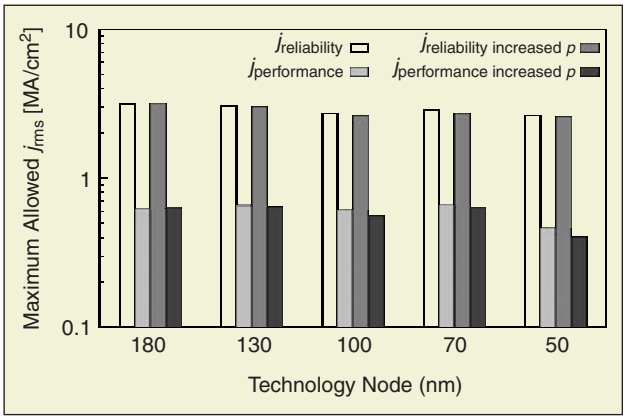
Another effect that can cause metal resistivity to increase is skin effect. This is normally observed at high frequencies at which the current gets confined almost entirely to a very thin sheet at the surface of the conductor. The thickness of this sheet, known as the skin depth, determines the effective cross-sectional area of the conductor and its resistance. The skin depth (δ) is given by [38]:

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \qquad (34)$$

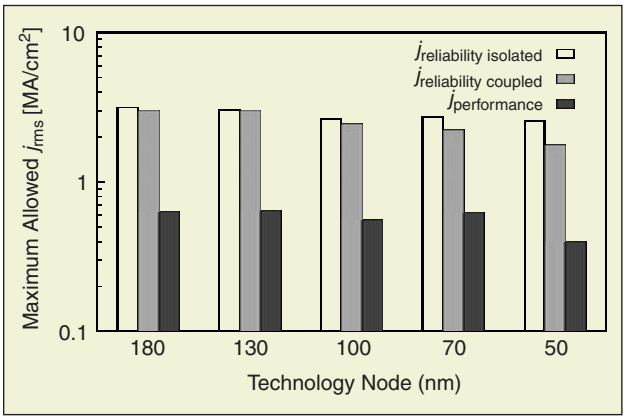
Here, the frequency $f = \omega/2\pi$ and μ and σ are the permeability and conductivity of the interconnect material, respectively. It was found that for skin effect to start impacting the line resistance, the thickness of Cu lines needs to be larger than 2.089 μm for 1 GHz and 0.661 μm for 10 GHz signals. Since the dimensions of the global Cu lines used in this study were << 1μm, skin effect is not expected to impact the resistivity of these lines, at least for frequencies up to 10 GHz.

Thermal Coupling in 3-D Interconnect Arrays

In this section we will briefly address the issue of thermal coupling in VLSI interconnects. Our self-consistent analysis of self-heating and EM effects presented above was based on single isolated interconnect lines. In a real IC there are densely packed layers of interconnect lines that form a 3-D array. The self-heating of interconnect lines within such an array could be significantly more severe due to thermal coupling between neighboring lines [25]. The heat-flow analysis for such struc-



15. Comparison of j_{rms} values obtained from reliability and electrical performance considerations for top-layer metal for various ITRS technology nodes with and without taking into account increased resistivity due to surface scattering and barrier materials.



16. Maximum allowed j_{rms} for top-layer metal of ITRS technologies for symmetric bipolar currents with and without thermal coupling from neighboring lines. Also shown is the j_{rms} from performance considerations.

tures is complicated and must involve numerical simulation techniques such as finite element method.

The RMS current density can be empirically shown to obey the following relationship

$$j_{\text{rms}}^2 \propto \frac{T_m - T_{\text{ref}}}{\rho_m(T_m)} \tag{35}$$

and the proportionality constant κ (which is independent of the interconnect material) can be obtained empirically from the finite element analysis. In our work we use the results presented in [6], which include coupling between interconnects at all metal layers.

Substituting Eqs. (35) and (26) in Eq. (27), we can obtain another self-consistent equation similar to Eq. (28)

$$r = \frac{4 j_0^2 \exp\left(\frac{Q}{k_B T_m} - \frac{Q}{k_B T_{\text{ref}}}\right)\rho_m(T_m)}{(1-R)^2 (T_m - T_{\text{ref}})\kappa}. \tag{36}$$

From this we can calculate the maximum allowed j_{rms} for densely packed metal lines. Figure 16 compares this with the j_{rms} obtained from Fig. 15 for isolated metal lines. We find that the maximum allowed j_{rms} reduces for the 3-D case, where all the metal lines are heated with equal current load in all the leads. Moreover, with technology scaling, the percentage reduction in j_{rms} due to coupling from neighboring lines increases, therefore $j_{\text{rms reliability}}$ comes closer to $j_{\text{rms performance}}$ as the technologies scale.

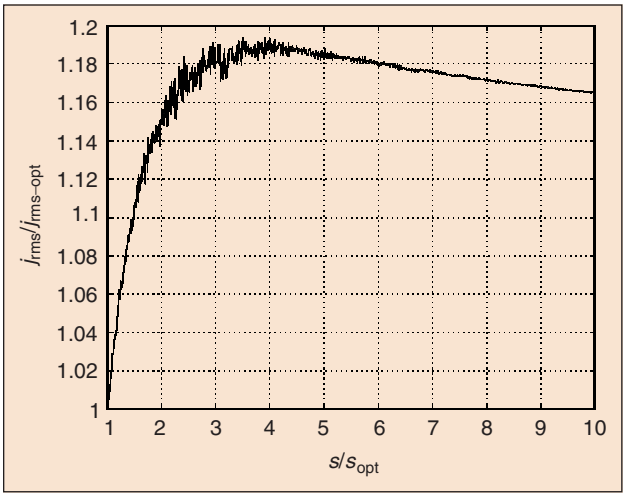
Reliability Issues In Vias

In this section we investigate the possible cases when the performance-based rms current density may potentially exceed the reliability rms current limit. We first examine the case of global interconnects when the buffer size is increased beyond s_{opt}. Figure 17 plots the rms current density as a function of buffer size driving an interconnect of length l_{opt} as obtained from SPICE simulations. We observe that j_{rms} saturates at a value of approximately $1.2 \times j_{\text{rms-opt}}$, which is well below the reliability limit. The rms current density in global interconnects becomes more than the reliability current density if the buffer size exceeds $3 \times s_{\text{opt}}$ and the interconnect length is less than 5 µ. This is highly unlikely in any practical design.

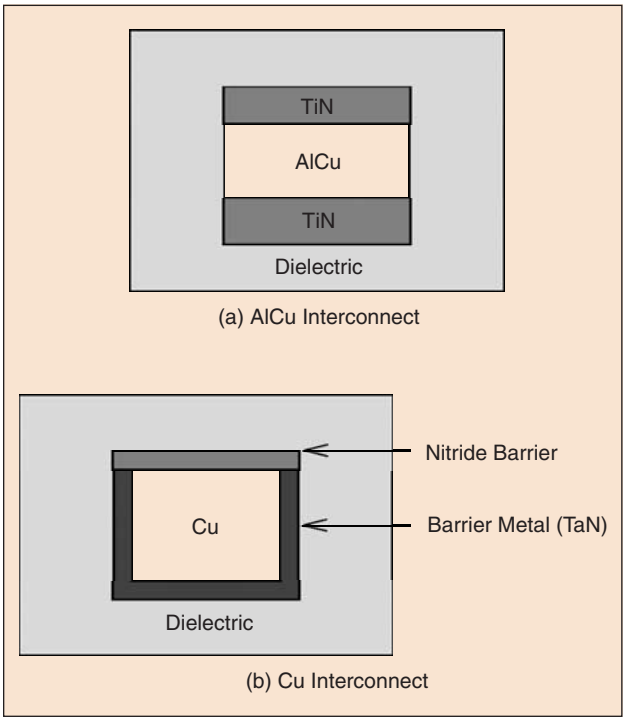
We now investigate whether rms current density in vias used in these large buffers can exceed the reliability limit. As shown in Fig. 2, in a technology with multiple levels of interconnect, the buffers are connected to the global lines by a series of vias and possibly short interconnects at intermediate layers of metal if the technology does not allow stack vias. The vias at the local and semiglobal layers are pitch-matched to the interconnects at these layers. However, they are carrying the same current as the global metal layers and therefore can potentially be a thermal reliability concern. For copper vias, Eq. (35) holds (with a different proportionality constant κ_{via}). Therefore an equation of the form of Eq. (36) can be derived for copper vias as well, which can be used to calculate the maximum allowed j_{rms} for the via. In our work, κ_{via} was determined using the measured value of temperature rise $(T_{\text{via}} - T_{\text{ref}})$ for a given current density (j_{rms}) [39]. These values are summarized in Table 5 for various ITRS technology nodes along with the minimum via area such that the current density in the via due to performance considerations is equal to the maximum allowed j_{rms} from thermal and reliability considerations. Also shown are the areas of minimum-sized vias at the local tier. It can be observed that for optimally buffered global lines, in order to meet the thermal reliability constraints, the area of the via at the local tier should be at least three times the area of the minimum sized via at that tier.

Thermal Effects Under High-Current Stress Conditions

As discussed earlier, a special case of thermally accelerated interconnect failure can also occur under ESD conditions. ESD is a



17. j_{rms} as a function buffer size for an interconnect length of l_{opt} for the 100-nm technology mode..



18. Schematic cross section of interconnects.

high-current (>1A) short-time scale (< 200 ns) phenomenon that can lead to catastrophic open-circuit failures and latent damage [10, 40, 41]. Also, IC failures due to ESD have significantly increased in importance as VLSI technologies continue to shrink towards the deep sub-micron regime [42]. Protection circuits are typically designed to reduce the impact of ESD events. Interconnects in these protection circuits and in the I/O circuitry, which are subjected to ESD stress, have to withstand very high-current stress conditions during such events. In older technology nodes the ESD robustness of ICs was mostly dominated by the failures in the protection devices. This was due to the fact that the performance of the ICs was not wire limited or limited by the number of I/O pins, which allowed extra flexibility in wire sizing and spacing to avoid any interconnect failures. However, for deep submicron technology nodes, interconnect failures can become a serious design and reliability concern. In fact, it has been shown that at the 0.25-μm technology node, interconnects become the dominant ESD failure mechanism [43]. Furthermore, the introduction of low-k materials can increase the susceptibility of interconnects to thermally accelerated failures during ESD events [9].

Self-heating under ESD type events are nonsteady-state events since they are much smaller in time scale compared with the thermal time constants of the interconnects, which are of the order of a few microseconds. Due to the nonsteady-state situation, the Joule heat generated in the wires does not have sufficient time to flow out through the entire underlying insulator stack. Banerjee et. al. formulated a model to account for the nonsteady-state Joule heating of interconnects [21, 19]. They proposed a dielectric sheath model, wherein only a thin sheath of the surrounding dielectric around the interconnect heats up along with the metal interconnect. Hence, interconnect temperature rise under such conditions ($t_{pulse} << \tau$, where τ is the thermal time constant of the interconnect) is independent of the total underlying insulator thickness or the thermal impedance, contrary to the temperature rise expected in interconnects under steady-state conditions as described earlier. The temperature rise of the interconnect under such conditions is given by

Table 5. Maximum allowed j_{rms} for a via and minimum via area for various technology nodes.

Tech.	Jrms (MA/cm^2)	Minimum Area (μ^2)	Via Area (Local Tier) (μ^2)
180	4.18494	0.091718	0.049087
130	4.14404	0.056753	0.026159
100	4.09105	0.028924	0.013789
70	3.95897	0.016835	0.067201
50	3.50195	0.0062626	0.0033183

$$\Delta T = \frac{E}{C_\theta} \tag{37}$$

where E is the pulse energy and C_θ is the effective thermal capacity of the interconnect that includes the thermal capacity of the metal and the surrounding dielectric sheath. The pulse energy is given by [19]

$$E \approx \frac{1}{2} I^2 \Delta t (R_0 + R_f) = \frac{1}{2} I^2 \Delta t R_0 (2 + \alpha_{AlCu} \Delta T) \tag{38}$$

where I is the magnitude of the constant current pulse, Δt is the pulse duration, R_0 and R_f are the initial ($t = 0$) and final ($t = \Delta t$) values of the line resistance, and α_{AlCu} is the temperature coefficient of resistance for AlCu. The time-dependent thermal capacity C_θ is given by

$$C_\theta(t) = \sum_i C_i^m + C_{\text{dielectric sheath}}(t) \tag{39}$$

where the summation includes all materials in the metal stack and

$$C_{\text{dielectric sheath}}(t) \propto \sqrt{a_d t}$$

where a_d is the thermal diffusivity in the surrounding dielectric. Substituting Eqs. (38) and (39) in Eq. (37), a *critical current*, I_{crit}, that causes a certain amount of temperature rise ΔT_{crit} can be calculated [19]

$$I_{crit} = \sqrt{\frac{2\left[\sum_i C_i^m + C_{\text{dielectric sheath}}(t)\right]}{\Delta t R_0 (2 + \alpha_{AlCu} \Delta T_{crit})}}. \tag{40}$$

The model presented for TiN/AlCu/TiN interconnects in [19] can be easily extended to analyze and design by damascene Cu interconnects (see Fig. 18). The models can be compared by applying

$$\frac{E_{crit,Cu}}{E_{crit,AlCu}} = \frac{\Delta T_{crit,Cu}}{\Delta T_{crit,AlCu}} \times \frac{C_{\theta,Cu}}{C_{\theta,AlCu}}.$$

We can show that

$$\frac{I^2_{crit,Cu}}{I^2_{crit,AlCu}} = \left[\frac{\rho_{0,Cu}}{\rho_{0,AlCu}}\right]\left[\frac{W_{AlCu} H_{AlCu}}{W_{Cu} H_{Cu}}\right] \times \left[\frac{2 + \alpha_{Cu} \Delta T_{crit,Cu}}{2 + \alpha_{AlCu} \Delta T_{crit,AlCu}}\right]\left[\frac{\Delta T_{crit,Cu}}{\Delta T_{crit,AlCu}}\right] \times \left[\frac{C_{\theta,Cu}}{C_{\theta,AlCu}}\right]$$

where $\rho_{0,Cu}$ is the effective resistivity of Cu as described above. If we assume that dimensions of AlCu and Cu interconnects are the same for a given pulse width and dielectric material, we can get

$$\frac{I^2_{\text{crit,Cu}}}{I^2_{\text{crit,AlCu}}} \approx \left[\frac{\rho_{0,\text{Cu}}}{\rho_{0,\text{AlCu}}}\right]\left[\frac{2+\alpha_{Cu}\Delta T_{\text{crit,Cu}}}{2+\alpha_{\text{AlCu}}\Delta T_{\text{crit,AlCu}}}\right]\times\left[\frac{\Delta T_{\text{crit,Cu}}}{\Delta T_{\text{crit,AlCu}}}\right].$$

ΔT_{crit} for open-circuit failure for Cu is expected to be higher than that of AlCu for a given insulator material since the melting point of Cu (~1100 °C) is larger than the melting point of AlCu (~660 °C). Therefore, for open-circuit failure conditions

$$I_{\text{crit,Cu}} > I_{\text{crit,AlCu}}.$$

In [19] it has been shown that the critical current density for causing open-circuit metal failure in AlCu interconnects is ~60 MA/cm^2, at which the critical temperature rise $\Delta T_{\text{crit}} = 1000$ °C. Also, interconnects can suffer latent damage under subcritical current pulses if the lines resolidify after melting, which has been shown to degrade the EM lifetime of AlCu interconnects [21, 22]. Equation (40) can also be used to avoid any latent reliability hazard by designing for a lower value of ΔT_{crit}. These interconnect design rules must be obeyed for high-current robustness.

Summary

Thermal effects in advanced high-performance interconnect systems arising due to self-heating under various circuit conditions, including ESD, has been examined in detail. A self-consistent approach that simultaneously comprehends self-heating and EM under both unipolar and bipolar stress conditions has been used to compute the maximum allowed rms current density and to compare it with performance-limited interconnect current density. This methodology is used to analyze the implications of interconnect technology (Cu, low-k, etc.) scaling using ITRS data on thermal effects and performance optimization for global-tier metal lines, wherein the effects of increasing interconnect (Cu) resistivity with decreasing line dimensions, the effect of a finite barrier metal thickness, and thermal coupling between wires have been included.

Using circuit-level simulations we have confirmed that the effective duty cycles for optimized global interconnects remain nearly invariant across metal layers and technologies. A value of 0.31 ± 0.01 was obtained for all cases. This analysis has helped provide a more realistic value of the duty cycle for analyzing thermal and EM effects using the self-consistent approach. Second, the implications of these thermal effects on the design (driver sizing) and optimization of the interconnect length between repeaters at the upper-level signal lines have been investigated. We have shown that the maximum allowed rms current density based on the self-consistent criteria is greater than the rms current density obtained from optimized driver and interconnect configurations (i.e., $j_{\text{rms-reliability}} > j_{\text{rms-performance}}$). Furthermore, we have shown that in real 3-D interconnect arrays the greater self-heating due to thermal coupling effects could further lower the $j_{\text{rms-reliability}}$ in advanced technology nodes. Hence, for DSM technologies that employ low-k dielectrics, thermal effects in interconnects need careful consideration.

The reliability implications for minimum-sized vias in optimally buffered signal nets have also been quantified. This analysis suggests that for the optimally buffered interconnects, while the current density in the line remains limited by the performance, the current density in the vias significantly exceeds the reliability-based limits, which has important implications for the physical design process flow and various optimization techniques. Finally, design guidelines for high-current robustness of interconnects used in ESD protection and I/O circuits have been presented.

Kaustav Banerjee is a research associate at the Center for Integrated Systems at Stanford University. He received the Ph.D. degree in electrical engineering and computer sciences from the University of California at Berkeley in 1999. His research interests include modeling and analysis of nanometer-scale circuit problems and their implications for computer-aided design of integrated circuits. He is also interested in emerging device and circuit interconnect architectures such as 3-D ICs, optical interconnects, and hybrid SET-FET designs. At Stanford, Dr. Banrjee co-mentors several doctoral students. He also co-advises doctoral students in the electrical engineering departments of the University of Southern California, Los Angeles, and the Swiss Federal Institute of Technology, Lausanne, Switzerland. In the past, he has held several summer research positions at the Semiconductor Process and Device Center of Texas Instruments Inc., Dallas, during 1993-1997. He has authored or co-authored over 50 research publications in archival journals and refereed international conferences and has presented numerous invited talks and tutorials. He is currently the Technical Program Chair of the International Symposium on Quality Electronic Design (ISQED) and also serves on the technical program committees of the International Symposium on Physical Design (ISPD), the EOS/ESD Syymposium, and the International Reliability Physics Symposium (IRPS). He was the recipient of a Best Paper Award at the 2001 Design Automation Conference (DAC). E-mail: kaustav@cis.stanford.edu.

Amit Mehrotra received a B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1994 and a master's and Ph.D. from the Department of Electrical Engineering and Computer Science of the University of California at Berkeley in 1996 and 1999, respectively. In August 1999 he joined the University of Illinois at Urbana-Champaign where he is currently an assistant professor with the Department of Electrical and Computer Engineering and a research assistant professor with the Illinois Center for Integrated Micro-Systems group at the Coordinated Science Laboratory. He is a member of the IEEE. He has co-authored over 20 research publications in archival journals and refereed international conferences and has presented numerous invited talks. He currently serves on the technical program committee of International Symposium on Quality Electronic Design (ISQED). He is the recipient of a Best Paper Award at the 1997 International Conference on Computer Design and 2001 Design Automation Conference (DAC). E-mail: amehrotr@mail.icims.csl.uiuc.edu.

References

[1] M.T. Bohr, "Interconnect scaling—The real limiter to high performance ULSI," in *Tech. Dig. Int. Electron Device Meeting*, 1995, pp. 241-244.

[2] W.J. Dally, "Interconnect limited VLSI architecture," in *Proc. Int. Interconnect Technology Conf.*, 1999, pp. 15-17.

[3] K. Banerjee, "Thermal effects in deep sub-micron VLSI interconnects and implications for reliability and performance," Ph.D. dissertation, University of California at Berkeley, 1999.

[4] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu, "On thermal effects in deep submicron VLSI interconnects," in *Proc. 36th ACM Design Automation Conf.*, 1999, pp. 885-891.

[5] K. Banerjee, A. Mehrotra, W. Hunter, K.C. Saraswat, K.E. Goodson, and S.S. Wong, "Qualitative projections of reliability and performance for low-k/Cu interconnect systems," in *Proc. 36th Annu. Int. Reliability Physics Symp.*, 2000, pp. 354-358.

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

[7] R.H. Dennard, F.H. Gaensslen, H. Yu, V.L. Rideout, E. Bassous, and A.R. LeBank, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 256-268, 1974.

[8] J. Ida, M. Yoshimaru, T. Usami, A. Ohtomo, K. Shimokawa, A. Kita, and M. Ino, "Reduction of wiring capacitance with new low dielectric SiOF interlayer film for high speed/low power sub-half micron CMOS," in *VLSI Technology Symp., Digest of Technical Papers*, 1994, pp. 59-60.

[9] 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," in *Tech. Dig. IEEE Int. Electron Devices Meeting*, 1996, pp. 65-68, 1996.

[10] C. Duvvury and A. Amerasekera, "ESD: A pervasive reliability concern for IC technologies," *Proc. IEEE*, vol. 81, pp. 690-702, May 1998.

[11] T. Green, "A review of EOS/ESD field failures in military equipment," in *Proc. 10th, EOS/ESD Symp.*, 1998, pp. 7-14.

[12] M.J. Pelgrom and E.C. Dijkmans, "A 3/5 V compatible I/O buffer," *IEEE J. Solid-State Circuits*, vol. 30, no. 7, pp. 823-825, 1995.

[13] *International Technology Roadmap for Semiconductors (ITRS)*, 1999.

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

[15] T. Sakurai and K. Tamaru, "Simple formulas for two- and three-dimensional capacitances," *IEEE Trans. Electron Devices*, vol. ED-30, no. 3, pp. 183-185, 1983.

[16] J. Rabaey, *Digital Integrated Circuits — A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

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

[18] B.K. Liew, N.W. Cheung, and C.Hu, "Projecting interconnect electromigration lifetime for arbitrary current waveforms," *IEEE Trans. Electron Devices*, vol. 37, pp. 1343-1350, 1990.

[19] 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.

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

[21] K. Banerjee, A. Amerasekera, and C. Hu, "Characterization of VLSI circuit interconnect heating and failure under ESD conditions," in *Proc. IEEE Int. Reliability Physics Symp.*, 1996, pp. 237-245.

[22] 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," in *Proc. 38th IEEE Annu. Int. Reliability Physics Symp.*, 2000, pp. 283-288.

[23] N.S. Nagaraj, F. Cano, H. Haznedar, and D. Young, "A practical approach to static signal electromigration analysis," in *Proc. 35th Design Automation Conf.*, 1998, pp. 572-577.

[24] 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.

[25] 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. Comp., Packag., and Manufact. Technol. A*, vol. 21, no. 3, pp. 1-6, 1998.

[26] J.C. Anderson, Ed., *The Use of Thin Films in Physical Investigations*. New York: Academic, 1966.

[27] J.R. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," *IEEE Trans. Electron Devices*, vol. 57, no. 9, pp. 1587-1594, 1969.

[28] C.-K. Hu, R. Rosenberg, and K.Y. Lee, "Electromigration path in Cu thin-film lines," *Applied Phys. Lett.*, vol. 74, no. 20, pp. 2945-2947, 1999.

[29] B.H. Jo and R.W. Vook, "In-situ ultra-high vacuum studies of electromigration in copper films," *Thin Solid Films*, vol. 262, no. 1-2, pp. 129-134, 1995.

[30] D.W. Malone and R.E. Hummel, "Electromigration in integrated circuits," *Critical Rev. Solid State and Materials Sci.,* vol. 22, no. 3, pp. 199-228, 1997.

[31] A. Chandrakasan and R.W. Brodersen, *Low Power Digital CMOS Design*. Norwell, MA: Kluwer, 1995.

[32] A.A. Bilotti, "Static temperature distribution in IC chips with isothermal heat sources," *IEEE Trans. Electron Devices*, vol. ED-21, pp. 217-226, 1974.

[33] 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.

[34] K. Banerjee and A. Mehrotra, "Coupled analysis of electromigration reliability and performance in ULSI signal nets," in *Int. Conf. Computer-Aided Design. Digest of Technical Papers*, 2001, to be published.

[35] H.A. Schafft, "Thermal analysis of electromigration test structures," *IEEE Trans. Electron Devices*, vol. ED-34, pp. 664-672, Mar. 1987.

[36] R.H.J.M. Otten, and R.K. Brayton, "Planning for performance," in *Proc. 1998 Design Automation Conf.*, pp. 122-127, 1998.

[37] F. Chen and D. Gardner, "Influence of line dimensions on the resistance of Copper interconnections," *IEEE Electron Device Lett.*, vol. 19, pp. 508-510, Dec. 1998.

[38] E.C. Jordan and K.G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[39] K. Banerjee, G. Wu, M. Igeta, A. Amerasekera, A. Majumdar, and C. Hu, "Investigation of self-heating phenomenon in small geometry vias using scanning joule expansion microscopy," in *Proc. IEEE Int. Reliability Physics Symp.*, 1999, pp. 297-302.

[40] C. Duvvury and A.Amerasekera, "State-of-the-art issues for technology and circuit design of ESD protection in CMOS ICs," *Semicond. Sci. Technol.*, no. 833-850, 1996.

[41] S.H. Voldman, "The state of the art of electrostatic discharge protection: physics, technology, circuits, design, simulation, and scaling," *IEEE J. Solid-State Circuits*, vol. 34, no. 9, pp. 1272-1282, 1999.

[42] A. Amerasekera and C. Duvvury, "The impact of technology scaling on ESD robustness and protection circuit design," in *Proc. EOS/ESD Symp.*, 1994, pp. 237-245.

[43] S.H. Voldman, "The impact of technology scaling on ESD robustness of aluminum and copper interconnects in advanced semiconductor technologies," *IEEE Trans. Comp., Packag., and Manufact. Technol. C*, vol. 21, no. 4, pp. 265-277, 1998. CD■