# Accurate Energy Dissipation and Thermal Modeling for Nanometer-Scale Buses*

Krishnan Sundaresan and Nihar R. Mahapatra
Department of Electrical & Computer Engineering
Michigan State University, East Lansing, MI, U.S.A.
{sundare2, nrm}@egr.msu.edu

## Abstract

*With technology scaling, power dissipation and localized heating in global and semi-global bus wires are becoming increasingly important, and this necessitates the development of accurate models to explore these effects during design stage, with simulators and using realistic workloads. In this work, we present a unified nanometer-scale bus energy dissipation and thermal model that helps designers monitor energy dissipation and temperature rise in individual wires during dynamic simulation. In addition to self capacitance, our model incorporates the effects of capacitive coupling between adjacent as well as non-adjacent pairs of wires on switching energy, effect of repeater insertion, and the effect of lateral heat transfer between adjacent wires, all of which have been ignored in earlier models. Next, using our integrated model in a first-of-its-kind study, we study energy dissipation and thermal characteristics of instruction and data address buses with traces obtained from standard SPEC CPU2000 benchmarks and using technology parameters for various nanometer technology nodes from the ITRS road-map. We also evaluate the effect of some well-known low-power bus design schemes on bus line energy dissipation.*

## 1. Introduction

As fabrication technologies scale down and low-permittivity (low-K) inter-metal and inter-layer dielectrics are introduced to reduce RC delay, dynamic power, and cross talk in global signal bus lines/wires, study of thermal effects are becoming more and more important. This is more so because, due to the increase in clock frequencies, the time available for cooling reduces. Energy dissipation and temperature rise in global signal lines are aggravated due to the following reasons also: (1) increasing use of repeaters in long signal lines to reduce delay leads to higher energy dissipation [12]; (2) a steady increase in the number of metal layers, and particularly the number of global metal layers, also increases overall energy dissipation; (3) increased capacitive coupling between global signal lines, as a result of smaller inter-wire spacing, increases energy dissipated during a switching event; (4) low-K dielectrics that are used to reduce cross talk effects in high-aspect-ratio global lines have inherently poor thermal conductivities leading to lesser heat being dissipated from the wire into the surrounding medium; and (5) long via separations in upper metal layers also contribute to higher average wire temperatures (vias are normally better thermal conductors than surrounding low-K dielectrics) [6]. All of the above effects are expected to worsen as wire aspect ratios continue to increase and inter-wire spacings decrease in global wires for future technologies.

* This research was supported by U.S. National Science Foundation grant # 0102830.

### 1.1. Scope and contributions of this work

By virtue of their carrying smaller currents than power supply lines, bus line energy dissipation and thermal characteristics of signal lines (both clock and data lines) have not been the subject of serious study. But this will need to change as activity rates become higher and clock frequencies increase as technology scales down. High switching activity rates also mean that the fluctuating line currents drawn by the driving circuitry can influence inductive voltage drop in power supply lines, particularly since the long global signal lines present a high load capacitance. In this work, we develop a model for activity-dependent bus line energy dissipation and temperature rise, and apply it to study address buses; however, our energy and thermal model can be used to study energy and thermal characteristics of any bus (address, data, or instruction) routed in the upper metal layers.

The nature of capacitive coupling between bus lines, which dominates bus line energy dissipation and ultimately wire temperature, depends highly on the type of informa-



IEEE COMPUTER SOCIETY

tion being carried on the bus. Also, it is known that utilization of buses, in general, varies dynamically as programs execute. Typically, transmissions on buses are interspersed with unequal periods of idling during which no power is dissipated in the bus lines (assuming they hold the last value that was transmitted). This idle period presents opportunities for cooling in the bus lines. Thus, the temperature profile of a bus line is both time- and information-dependent. Hence, interconnect thermal models that estimate temperature and reliability based on the assumption that all bus lines carry the maximum RMS current density (worst-case scenario) [5, 2], and models that use switching activity factors to estimate average self-heating power to determine temperature rise [8], may result in inaccuracies when estimating wire temperature. This may, in turn, lead to incorrect interconnect lifetime prediction, since dynamic heating and cooling effects are not taken into account. Also, designers will be forced to allow higher-than-required safety margins and, as a result, the system will incur higher packaging costs. Hence, thermal effects in buses are best studied using dynamic simulation on real-world bus traces; in this work, we present a methodology to do this.

Briefly, the contributions of this work are the following.

- We present a model to estimate bus line energy dissipation that can be used in a trace-driven setup or in a power/performance simulator. Existing bus energy models like [16] only estimate bus energy dissipation considering the bus as a whole, not in each bus line; our model is capable of estimating energy dissipated in each bus line. As we shall see later, this is necessary to model dynamic temperature effects in buses. Our bus model is also more accurate because it considers the effect of both adjacent and non-adjacent neighbor capacitive coupling on switching energy in addition to energy dissipated in the self capacitance (capacitance to ground). In addition, we also model the effect of repeaters, which increase the self capacitance and energy dissipation in bus lines. Although inductive coupling is becoming important in nanometer technologies, we do not consider its effects on energy dissipation in this work because, for long global signal lines (length $>$ 10 mm) in current ITRS technologies, the interconnect (modeled as a lumped RLC circuit) is over-damped, i.e., its energy dissipation can be approximated by an RC model without significant error [10].
- Using our bus line energy dissipation model, we study the effectiveness of some existing low-power bus encoding techniques when used for address bus encoding. To our knowledge, no previous work has studied these bus encoding techniques using realistic address traffic from SPEC CPU2000 benchmark programs; most of them have used random traffic patterns that do not behave like real-world address streams do. Our work is also the first to study whether technology scaling will have any impact on the need for address bus encoding in future technologies. In this context also, we use realistic technology parameters from the ITRS road-map for current and future nanometer technology nodes.
- Finally, we present a thermal model and a methodology to estimate temperature rise in each wire of a global signal bus with dynamic simulation. Our model incorporates the effect of inter-layer heat transfer (heat conduction from lower metal layers through the inter-layer dielectric) and intra-layer heat transfer between adjacent bus lines through the inter-metal dielectric. Our model assumes a constant substrate temperature and models only the temperature rise in the bus lines due to internal heat generation in the wire (arising from current flow) and conductivity of the surrounding dielectric. Simulations using our model estimate that, even for the current 130 nm technology node, switching activities in wires of an address bus can cause wire temperatures in global signal buses to rise by about 30 degrees from the ambient during execution of a SPEC CPU2000 benchmark. This temperature rise, when combined with dynamic substrate temperature changes reported by [15] can result in highly irregular variations in the temperature profile that can cause performance degradation due to changes in RC delay of wires (as a result of temperature-dependent resistivity) and/or decrease in electromigration reliability.

The organization of the rest of the paper is as follows. Sec. 2 briefly reviews related work. Next, in Sec. 3, we describe our energy dissipation for global signal lines. Then, we present our thermal model in Sec. 4. Then, in Sec. 5, we present simulations and results obtained using our model on address bus traces. Finally, we conclude in Sec. 6.

## 2. Related Work

Some methods for high-level and architectural-level interconnect power analysis have been proposed [17, 19]. Earlier modeling methods estimated bus energies based on switching activities (self transitions) only [19], whereas recent models consider adjacent inter-wire capacitances also for energy calculations [17]. Thermal effects in interconnects and their implications for performance, current density, and reliability have been studied in [3, 1]. Recently, interconnect thermal models have been proposed in [6, 8]. But these models either perform a worst-case analysis using maximum current metrics suitable only for power supply lines [6] or consider average switching activities [8]. Such approaches are not suitable for analyzing signal lines since: (1) signal lines carry much less current densities than


1530-0897/05 $20.00 © 2005 IEEE

IEEE COMPUTER SOCIETY

power supply lines and (2) activities of bus lines and hence their energy dissipation and thermal characteristics are tied to actual traffic patterns (with intermittent idling) carried on the bus.

## 3. Bus Line Energy Dissipation Model

In this section, we develop our bus line energy dissipation model that calculates energy dissipated as a result of a switching (both self and coupling) transition. This energy model is then used to determine change in wire temperature that occurs due to the combined effect of self-heating in the wire and heat conduction into the surrounding medium. Values for wire geometry (wire width, spacing, etc.) and technology and equivalent circuit parameters, like capacitance and resistance of lines and repeaters that we used for various nanometer-scale technologies are listed in Table 1. Note that wire spacing is assumed to be equal to the wire width per ITRS [14].

The energy drawn from the supply rails by the driving gates of a bus line is dissipated as $I^2R$ losses in the bus line as well in those capacitively coupled to it. This results in temperature rise in the coupled wires due to self-heating effect. Existing bus energy models like [16] only provide expressions for total energy dissipated in the bus. From the thermal design point of view, the energy dissipated in each bus line is important since it helps determine the temperature rise in each individual wire separately. This can be calculated as described below.

### 3.1. Energy dissipated in line self capacitance

First, we define $V_i = V_i^{fin} - V_i^{in}$, i.e., the difference between the final and initial voltages on line $i$. Note that $V_i^{in}$ and $V_i^{fin}$ can take either of two values: 0 or $V_{dd}$. Thus, $V_i = V_{dd}$ implies that the self capacitance of line $i$ charges due to a rising transition ($0 \rightarrow 1$) and $V_i = -V_{dd}$ means that the self capacitance is discharged due to a falling transition ($1 \rightarrow 0$). For each transition, energy that is dissipated in wire $i$ due to charging or discharging of the self capacitance of the wire can be calculated as: $E_i^s = 0.5 \times (c_{line} + C_{rep}) \cdot V_i^2$, where $c_{line}$ is the self capacitance of a wire and $C_{rep}$ is the total capacitance of the repeaters on the line. The energy $E_i^s$ is called *self energy* since it involves only the self or line capacitance (including the contribution of repeaters).

**3.1.1. Capacitance due to repeaters.** Repeaters are used to reduce delays in long global signal lines such as address buses. The capacitance of each repeater is proportional to its size. Assuming that the size of a repeater is $h$ times the size of a minimum-sized inverter which is technology-dependent, and $k$ is the number of repeaters needed to achieve optimum delay on the interconnect, the total capacitance contributed by all repeaters on the interconnect is: $C_{rep} = h \cdot k \cdot C_0$. The optimal size of repeaters and the number of such repeaters that need to be used for minimum delay are as follows:

$$h = \sqrt{\frac{R_0 \cdot C_{int}}{C_0 \cdot R_{int}}} \quad (1)$$

$$k = \sqrt{\frac{0.4(R_{int} \cdot C_{int})}{0.7(C_0 \cdot R_0)}}, \quad (2)$$

where $C_{int} = c_{wire} + 2 \cdot c_{inter}$ is the capacitance of a wire and $R_{int}$ and $R_0$ are the resistances of the wire and a minimum-sized repeater, respectively [12]. It can be seen that, effectively, $C_{rep} = 0.75 \times C_{int}$.

### 3.2. Energy dissipated in inter-wire capacitance

The second component of energy dissipation is the *coupling energy*, which is influenced by charging, discharging, or toggling of the coupling capacitance $c_{i,j}$ between two lines $i$ and $j$. A *coupling charge* transition occurring between two lines is one of: $00 \rightarrow 01$, $00 \rightarrow 10$, $11 \rightarrow 01$, or $11 \rightarrow 10$; a *coupling discharge* transition is one of: $01 \rightarrow 00$, $01 \rightarrow 11$, $10 \rightarrow 00$, or $10 \rightarrow 11$; and a *coupling toggle* transition is either $01 \rightarrow 10$ or $10 \rightarrow 01$. The coupling energy dissipated in lines $i$ and $j$ for each of the above three cases can be calculated as described next.

If $V_i, V_j \neq 0$ and $V_i = -V_j$, then the coupling capacitance toggles, i.e., one terminal discharges to ground and the other charges to $V_{dd}$ or vice versa, which results in the effective coupling capacitance doubling per the Miller effect [13]. The coupling energy dissipated in line $i$ is obtained as: $E_i^c = 0.5 \times c_{i,j}(V_i^2 - V_i \cdot V_j), i \neq j$. Note that $c_{i,i+1}$ is the same as $c_{inter}$ whose values are given in Table 1. Similarly, the coupling energy dissipated in line $j$ is obtained using: $E_j^c = 0.5 \times c_{i,j}(V_j^2 - V_i \cdot V_j)$. If $V_i = 0$ or $V_j = 0$ and $V_i + V_j = V_{dd}$, then the coupling capacitance is charged due to the resulting transition. If $V_i = 0$ or $V_j = 0$ and $V_i + V_j = -V_{dd}$, then the coupling capacitance discharges due to the resulting transition. The energy for these cases for each line can also be calculated using the expressions above. It can be seen that the toggle case dissipates an equal amount of energy in both coupled lines but the charge and discharge transitions result in energy dissipation in only one of the lines.

The total energy dissipated in a bus line is the sum of the self energy and coupling energies obtained as above.

**3.2.1. Capacitance extraction.** The ITRS road-map predicts values for self and adjacent-wire coupling capacitance only for current and future technology nodes. Hence, to estimate the coupling capacitance between all pairs of


1530-0897/05 $20.00 © 2005 IEEE
IEEE COMPUTER SOCIETY

| Parameter | Technology node | | | |
|---|---|---|---|---|
| | 130 nm | 90 nm | 65 nm | 45 nm |
| Number of metal layers | 8 | 9 | 10 | 10 |
| Wire width, $w_i$ (nm) | 335 | 230 | 145 | 103 |
| Wire thickness, $t_i$ (nm) | 670 | 482 | 319 | 236 |
| Height of inter-layer dielectric, $t_{ild}$ (nm) | 724 | 498 | 329 | 243 |
| Relative permittivity of dielectric, $\varepsilon_r$ | 3.3 | 2.8 | 2.5 | 2.1 |
| Thermal conductivity of dielectric, $k_{ild}$ (W/mK) | 0.6 | 0.19 | 0.12 | 0.07 |
| Clock frequency, $f_{clk}$ (GHz) | 1.68 | 3.99 | 6.73 | 11.51 |
| Supply voltage, $V_{dd}$ (V) | 1.1 | 1.0 | 0.7 | 0.6 |
| Maximum current density in a wire, $j_{max}$ (MA/cm$^2$) | 0.96 | 1.5 | 2.1 | 2.7 |
| Self capacitance of wire, $c_{line}$ (pF/m) | 44.06 | 32.77 | 25.07 | 19.05 |
| Coupling capacitance of wire, $c_{inter}$ (pF/m) | 91.72 | 76.84 | 68.42 | 58.12 |
| Resistance of wire, $r_{wire}$ (kΩ/m) | 98.02 | 198.45 | 475.62 | 905.05 |

**Table 1. Wire geometry and equivalent circuit parameters for topmost layer interconnect: The top six values are from ITRS-2001 [14]. The values for $c_{line}$ and $c_{inter}$ were obtained from FastCap [11] simulations and the value for $r_{wire}$ was calculated using the formula $r_{wire} = \rho \cdot l/(w_i \cdot t_i)$.**

wires (adjacent as well as non-adjacent wire-pairs) that we need to use in our energy model, we employed the publicly available three-dimensional capacitance extraction program called FastCap [11]. Using the wire geometry parameter from ITRS (see Table 1 for values) to model a co-planar global bus layout, similar to the one shown in Fig. 1(a), we obtained a complete capacitance matrix for the wires of a 32-bit bus. Fig. 1(b) shows the percentage distribution of capacitances for various technologies. From the figure, we observe that, for current 130 nm and 90 nm technologies, non-adjacent coupling capacitances are non-negligible (contribute ≈10%) while, even in a future 45 nm node, non-adjacent capacitances account for about 8% of the total capacitance.

### 3.3. Energy dissipation due to non-adjacent coupling capacitances

The influence of non-adjacent neighbor coupling effects on bus line energy dissipation and wire temperature can be illustrated with a simple example of a 5-wire bus like the one shown in Fig. 1(a). Consider transitions on the bus lines as follows: ↑↑↓↑↑—this represents the relative thermal worst-case since most of the energy dissipation is concentrated in the center line and hence the temperature distribution across the wires is non-uniform. Next, consider transitions occurring as: ↓↑↓↑↓—this is the worst-case for total energy dissipated, but all the lines are at the same temperature since energy dissipated in each line is approximately same. The representation ↑ indicates that, in the current cycle, the line charges to $V_{dd}$ from its previous ground state and ↓ indicates that the line discharges in the current cycle from $V_{dd}$ held in the previous cycle. With capacitance values we extracted for 130 nm technology and using our method to calculate the coupling energy dissipation in each line, we found that the energy dissipation is underestimated by up to 6.6% for the middle wire of a 32-bit address bus when non-adjacent neighbor capacitances are neglected. Also, we found that, although the non-adjacent capacitance values are decreasing with technology scaling, this energy estimation error remains more or less constant in future technologies also. Non-adjacent coupling capacitances are especially important to consider because many current bus encoding techniques try to reduce self energies of bus lines and coupling energies between a pair of adjacent bus lines and this may increase the relative contribution of energy dissipated in transitions involving non-adjacent coupling capacitances. Also, previous work has not considered the effect on non-adjacent coupling capacitances and its influence on energy; ours is the first to do so.

## 4. Thermal model

### 4.1. Equivalent thermal network for a bus

After the energies dissipated as heat in the individual bus lines have been determined, the temperature rise in each wire can be estimated by considering an equivalent thermal-RC network like the one shown in Fig. 2(a) for a 5-wire bus. In the analogy between thermal and electrical quantities, the temperature difference between two nodes corresponds to a voltage difference and the heat transfer rate corresponds to current. By equating the amount of heat flowing into a node in the thermal equivalent circuit to the amount of heat flow-


1530-0897/05 $20.00 © 2005 IEEE

IEEE COMPUTER SOCIETY

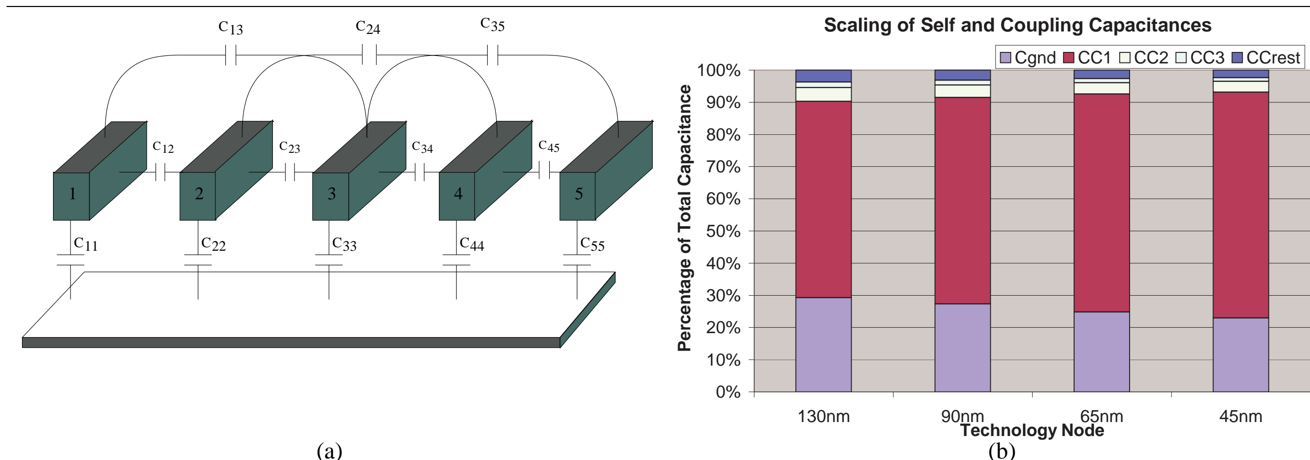


**Figure 1. Capacitance extraction: (a) Wire geometry used to extract capacitances for a 5-wire bus. (b) Distribution of capacitance values for a wire, extracted using the FastCap tool[11]. Cgnd = self capacitance of the wire; CC1= coupling capacitance between the wire and its adjacent neighbor; CC2 = coupling capacitance between the wire and a non-adjacent neighbor with 1 wire between them; CC3 = coupling capacitance between the wire and a non-adjacent neighbor with 2 wires between them; CCrest = coupling capacitance between wire-pairs with 3 or more wires between them. For current and near-future ITRS technology nodes (up to 45 nm), non-adjacent coupling capacitance is non-negligible (contributes ≈8–10%).**

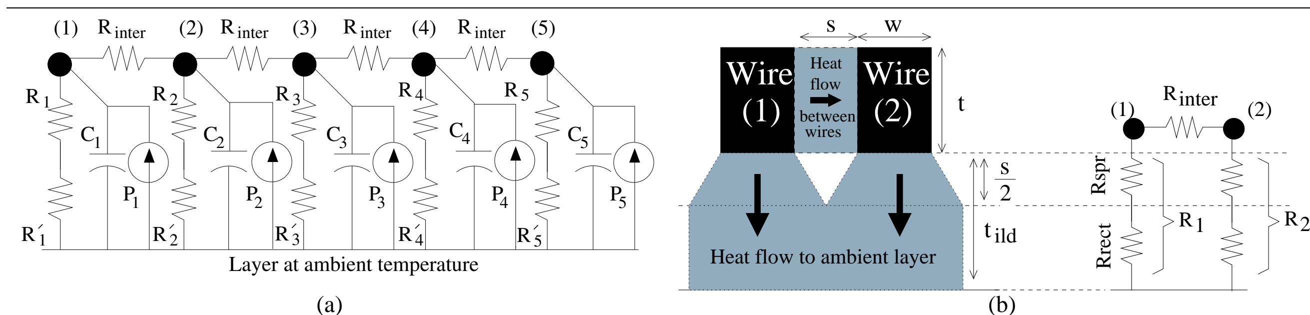


**Figure 2. Thermal model: (a) Complete equivalent thermal-RC network for a 5-wire bus. (b) Geometry for calculating equivalent thermal resistances for a wire based on [5]. The arrows represent heat flow between the conductors. The equivalent thermal-resistance network is shown on the right.**

ing out (analogous to Kirchoff's current law in electrical circuits), the following differential equations can be written.

For edge wires:

$$P_i = \mathcal{C}_i \cdot \frac{d\theta_i}{dt} + \frac{(\theta_i - \theta_0)}{\mathcal{R}_i} + \frac{(\theta_i - \theta_{i+1})}{\mathcal{R}_{inter}} \quad (3)$$

and for middle wires:

$$P_i = \mathcal{C}_i \cdot \frac{d\theta_i}{dt} + \frac{(\theta_i - \theta_0)}{\mathcal{R}_i} + \frac{(2\theta_i - \theta_{i-1} - \theta_{i+1})}{\mathcal{R}_{inter}}, \quad (4)$$

where $P_i$ is the instantaneous power dissipated in the $i^{th}$ wire and $\theta_0$ is the ambient temperature (45 C or 318.15 K).

In the above equations, $\mathcal{C}_i$, the *thermal capacitance per unit-length* of a wire is: $\mathcal{C}_i = \mathcal{C}_s \cdot (t_i \cdot w_i)$, where $\mathcal{C}_s$ is the specific heat per unit volume of the metal, and $w_i$ and $t_i$ are wire dimensions as shown in Fig. 2(b). $\mathcal{R}_i$ is the *thermal resistance per-unit length* of the wire along the heat transfer path as shown in the Fig. 2(b) and it can be calculated from the following expression using wire geometry and thermal conductivity $k_{ild}$ of the inter-layer dielectric (ILD) as described in [5]:

$$\mathcal{R}_i = \mathcal{R}_{spr} + \mathcal{R}_{rect} \quad (5)$$


1530-0897/05 $20.00 © 2005 IEEE
IEEE COMPUTER SOCIETY

$$\mathcal{R}_i = \frac{\ln(\frac{w_i+s_i}{w_i})}{2 \cdot k_{ild}} + \frac{t_{ild} - 0.5s_i}{k_{ild}(w_i + s_i)}. \quad (6)$$

The above expression is the sum of two terms: the first is the spreading resistance $\mathcal{R}_{spr}$, due to the spreading of heat from the bottom face of the wire in a trapezoidal manner, and the second is the thermal resistance due to rectangular heat flow ($\mathcal{R}_{rect}$) as depicted in Fig. 2(b). Equations 3 and 4 can be solved to find $\theta_i$, the wire temperature.

**4.1.1. Lateral thermal coupling between wires.** The lateral heat transfer between adjacent wires can be a significant amount due to the large exposed sidewall area in high aspect-ratio global lines and due to the difference in activity rates of the neighboring lines (which creates a temperature gradient and hence lateral heat flow). In our model, this effect is captured with a *lateral inter-wire thermal resistance per-unit length* whose value depends on wire geometry parameters, as shown in Fig. 2(a), and the inter-metal dielectric (IMD) thermal conductivity $k_{imd}$ given by the expression: $\mathcal{R}_{inter} = s_m/(k_{imd} \cdot t_i)$. Previous work on interconnect thermal modeling did not consider the effect of inter-wire heat transfer [8]; our model incorporates this for better accuracy.

**4.1.2. Heat transfer from lower layers.** The thermal model is not complete without considering the temperature rise in the global signal line due to heat transfer from underlying layers. This is needed because, in current C4/CBGA packages, a secondary heat transfer path exists from the substrate through the interconnect layers—heat flows from the bottom-most to the top-most interconnect layer—and finally flows through C4 bumps, ceramic substrate, CBGA joints, and the printed circuit board to the ambient air [8]. As a first order approximation, we attribute a constant temperature increase due to this effect to each global wire, assuming that wires in lower layers carry current at their maximum density $j_{max}$. A coverage factor $\alpha_i = 0.5$, that represents the probability a wire is thermally coupled to another wire in a lower layer, is also assumed [5]. Thus, the temperature correction for the global metal lines due to inter-layer heat transfer from lower layers can be calculated as follows:

$$\Delta\theta = \sum_{i=1}^{N} \frac{t_{ild,i}}{k_{ild,i} s_i \alpha_i} \cdot [\sum_{j=i}^{N-1} (j_{max})^2 \rho_j \alpha_j t_j], \quad (7)$$

where $\rho_j$ is the resistivity of the metal line (Copper). This is depicted as the resistance $R'_i$ in the network shown in Fig. 2(a).

# 5. Experiments and results

In this section, we present results for simulations using our bus-line energy dissipation and discuss the implications of our results.

## 5.1. Methodology for simulations

We collected address bus traces for processor to level-one cache address buses (separate instruction and data address buses) using the cachesim5 analyzer in SHADE5 [7] which simulates a SPARC-V9 processor. The memory system we modeled had a two-level cache hierarchy (separate I-L1 and D-L1 caches each: 16KB, 4-way set assoc., with block size = 32 bytes, and write-through, and a unified L2 cache: 256KB, 4-way set assoc., with block-size = 64 bytes and write-back) and main memory. We randomly chose eight benchmarks—*eon, crafty, twolf, and mcf* for integer programs and *applu, swim, art, and ammp* for floating-point programs—from the SPEC CPU2000 benchmark suite for our simulations. These programs were compiled as statically-linked, -O3 optimized, 32-bit V8plusa executables, i.e., they run on a 64-bit SPARC-V9 processor but have a 32-bit virtual address space. Note that the version of SHADE we used allows simulation of these benchmarks and traces only the lower order 32-bits of the SPARC-V9's 64-bit address bus for V8plusa executables. The higher order 32 bits of the address bus do not change during simulation. For studying energy dissipation and temperature characteristics for address buses, bus traces were collected for 300 million committed instructions after skipping the initial 500 million instructions for each benchmark; these results are reported in Sec. 5.3. For evaluating address bus encoding schemes (Sec. 5.2) , we collected a shorter trace corresponding to a run of 20 million instructions after the initial warm-up phase of 500 million.

## 5.2. Energy dissipation in address bus lines

The total energy dissipated in all the bus lines for unencoded and encoded address transmission obtained using our energy dissipation model is shown in Fig. 3. To study the effectiveness of low-power bus encoding techniques under technology scaling, we implemented schemes like bus-invert [18], odd/even bus-invert [20], and coupling-driven bus-invert [9] on address traffic and determined the energy dissipation using our model. To our knowledge, this is the first study to report energy dissipation results for actual SPEC benchmarks that represent real-world programs; most previous studies including the ones cited above reported energies for random traffic patterns.

Briefly, the bus-invert (BI), odd/even bus-invert (OEBI), and coupling-driven bus-invert (CBI) encoding schemes


IEEE COMPUTER SOCIETY

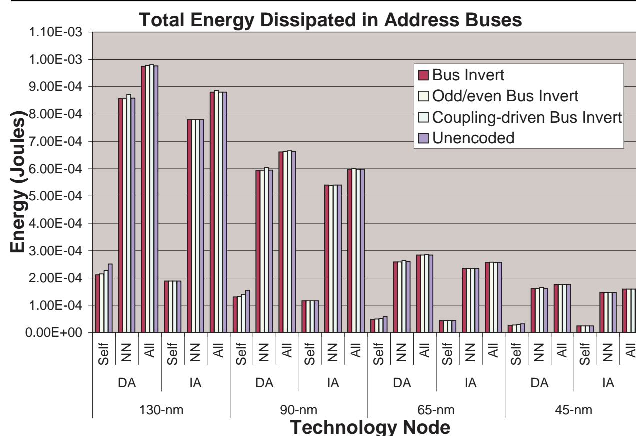


**Figure 3. Total energy dissipated in a 32-bit address bus for various ITRS technology nodes: 'Self' denotes self energy, 'NN' denotes coupling energy when only the nearest neighbor capacitance is considered, and 'All' denotes the energy when all pairs of coupling capacitances (adjacent and non-adjacent neighbor pairs) are considered.**

work as follows. The BI scheme examines the number of bits that are different between the current input pattern and the pattern transmitted on the bus in the last cycle (Hamming distance). If this number is greater than half the bus-width, all the bits of the current input pattern are inverted and a separate *invert* line is held high. Else, the value is transmitted in original form and the invert line is held low. In OEBI, even and odd bit positions are encoded (with bus inversion) separately and two invert lines are used to indicate one of four modes of transmission (00 – none of the bits are inverted, 01 – even bits are inverted, 10 – odd bits are inverted, and 11 – all bits are inverted). By evaluating the number of coupling transitions for all the four modes, the mode that will result in the lowest coupling activity is chosen and data is transmitted on the bus in that form. In a similar manner, the CBI encoding technique examines all pairs of adjacent bits and counts the number of coupling transitions. The current input bit pattern is inverted only if the coupling effect of the inverted pattern is less than that of the original pattern. Since odd and even lines are not handled separately, this scheme requires only one extra invert line.

**5.2.1. Results and discussion.** While bus-invert encoding reduces the number of transitions on the self capacitance of bus lines, OEBI and CBI schemes are designed to reduce transitions on the coupling capacitance between adjacent bus lines. In our simulations, the BI and CBI were implemented with one extra invert line as the MSB and the latter was implemented with two extra invert lines (LSB as the odd-invert line and MSB as the even-invert line). From Fig. 3, we find that all three encoding schemes reduce self energies with BI being the best, but the reduction is more observable in case of data address buses and in current technologies. For instruction address buses, the added complexity of encoding schemes seem to yield no benefits since these schemes perform no better than unencoded traffic. Our analysis of instruction address trace found that the number of bit transitions between consecutive cycles to be very low to cause inversion in any of these encoding schemes.

For coupling energies in data address buses, the OEBI scheme reduces energy dissipation only by a very small amount (less than 0.1% in most cases) but only when non-adjacent coupling effects are neglected. Also in many cases, CBI results in worse energies than the unencoded stream. When non-adjacent coupling capacitances are taken into account, both schemes perform marginally worse and coupling energies increase since these schemes were not designed to reduce transitions on non-adjacent neighbor capacitances. Here also, the energy-efficiency of OEBI and CBI seem to be only as good as BI. This is possible because both OEBI and CBI have modes in which all the bits are inverted (just like BI) and our analysis found that this mode occurred most of the time for our address traces. Also, addresses streams are bound to exhibit a high number of toggle transitions in lower-order bit pairs, it is also highly likely that their higher order portions may all change at the same time, particularly if different regions of the memory space are being accessed. In this case, the OEBI and CBI schemes invert all bits and hence result in a similar energy dissipation as BI. Based on our results, we can conclude that bus-inversion based encoding schemes do not work well for realistic address streams where the number of bits that transition between consecutive cycles is low.

### 5.3. Methodology for temperature simulations

Address traces collected for the benchmarks *eon* and *swim* were used in this analysis. Instantaneous energy and temperature profiles are shown for only one benchmark in each set—*eon*, an integer program, and *swim*, a floating-point program for 300 million cycles each. We found that other benchmarks also exhibited similar behavior as these benchmarks; hence these are not reported. An interval of 100K cycles was used to compute the instantaneous energy and power, and a fourth-order Runge-Kutta method was used to solve the differential equations for the thermal-RC network (Eqs. 3 and 4) to obtain the temperatures at the end of the interval. The instantaneous energies and temperatures are shown in Fig. 4(a)-(b) for *eon* and Fig. 4(c)-(d) for *swim*. Plots are shown for data and instruction address buses separately in each case. Initial temperature for all wires was


1530-0897/05 $20.00 © 2005 **IEEE**
IEEE COMPUTER SOCIETY

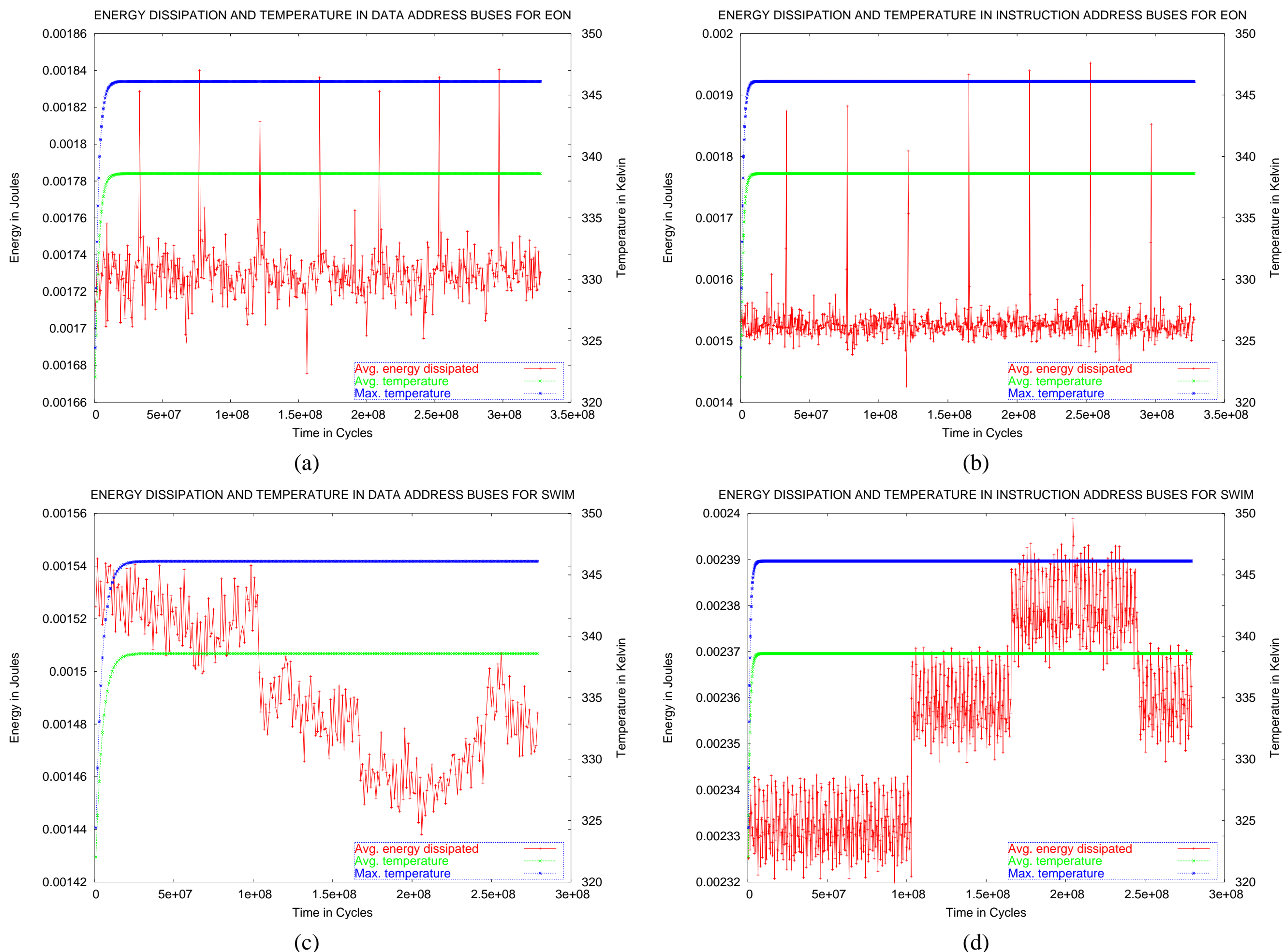


**Figure 4. Average bus line energy dissipation and temperature for 130 nm global buses as a result of switching activity: (a) Data address bus and (b) Instruction address bus for integer benchmark *eon*. (c) Data address bus and (b) Instruction address bus for floating-point benchmark *swim*. It can be observed that average temperature of data and instruction address buses reaches a steady state value of about 338 K (an increase of 20 about K from ambient).**

assumed to be 318.15 K.

**5.3.1. Results and discussion.** First, data address buses, on the average, were observed to dissipate higher energies and consequently run at slightly higher temperatures than instruction address buses for all benchmark programs. However, the temperature rise in instruction address buses were observed to be higher, presumably due to the higher activity in the latter. For the small sample of 12 million cycles we observed that, the hottest wire in a data address bus increased in temperature by 0.0003 K and the temperature for the hottest wire in an instruction address bus increased by about 0.0005 K for both *eon* and *swim*. However, this temperature rise is not insignificant for the time interval we considered; 12 million cycles correspond to a running time of only 6 ms in a processor clocked at 2 GHz. For execution of an entire SPEC benchmark program (mean running time of a SPEC benchmark program is about 288 billion cycles [4]), the temperature rise may be substantial. Note that collecting address traces and simulating the thermal-RC network for a complete run of a benchmark would have required huge computational and storage resources. Even if an execution-driven simulation is performed, it would have a significant time overhead. To obtain long-term temperature rise characteristics, we estimated wire temperatures using our model on a trace of 300 million cycles; these results are shown in Fig. 4. For this longer simulation run, we found that wire average temperatures saturated at about 338 K for both data and instruction buses. This is to be expected be-


IEEE COMPUTER SOCIETY

cause, the thermal network attains steady-state after some time and stays in this state as long as energy is being dissipated in the bus lines. We also observed that as shown in Fig. 5 that small intermittent periods of idling in the buses (about one million cycles, in this case) have no appreciable effect on reducing the temperature of buses.

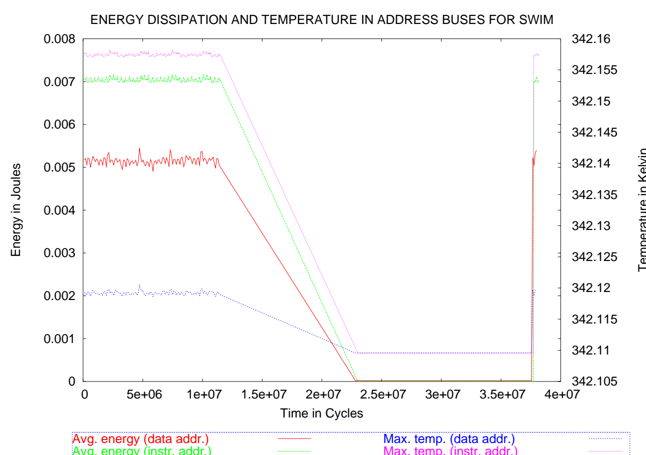


**Figure 5. Effect of intermittent idling of buses on temperature: This plot shows that intermittent periods of idling, when bus line energy dissipation drops dramatically (for about one million cycles in this case), has no appreciable cooling effect on the bus lines.**

As mentioned earlier, our models assume a constant substrate temperature. It has been reported in [15] that temperature variations of 10 degrees can occur in the substrate during various stages of benchmark execution. Due to this effect, interconnect temperatures may further increase from our projections.

Second, we observed that temporal fluctuations in energy dissipated in instruction address bus lines are more compared to those in data address lines although, as mentioned earlier, the amount of energy dissipated is lower. Instruction addresses buses which are issued typically every cycle can be expected to have a higher activity compared to data address buses whose activity depends on the frequency of load/store instructions for which data addresses are issued. Thus, in a given interval, there are fewer idle cycles in the instruction bus leading to more temporal variations.

The impact of these results, especially the highly fluctuating energy profile in instruction address buses that was observed, is the following. Since the energy dissipation of the wire roughly represents the square of the time-varying current flowing in the wire, fluctuations in the energy mean that a highly varying load is being placed on the power supply network by the driving circuits through which the currents flowing in the wires are drawn. This varying load can cause inductive voltage drops or $L\frac{di}{dt}$ noise. This motivates the need to smoothen temporal variations in energy dissipation of wires with appropriate techniques

## 6. Conclusions and Future Work

In this work, we presented a unified nanometer-scale bus energy dissipation and thermal model that can help designers monitor energy dissipation and temperature rise in individual wires during trace driven or power/performance simulation. In addition to self capacitance, our model incorporates the effects of adjacent and non-adjacent neighbor capacitive coupling on bus energy dissipation, effect of repeater insertion, and the effect of lateral heat transfer between adjacent wires. Unlike existing models which provide estimates for total bus energy, our model can estimate energy dissipated in each bus line; this feature helps to estimate wire temperatures also. Using this integrated model in a first-of-its-kind study, we studied energy and thermal characteristics of 32-bit instruction and data address buses with traces obtained from standard SPEC CPU2000 benchmarks. We found that encoding addresses to achieve low power address buses does not result in any significant benefits due to the characteristics of address traces. Based on simulations using our model, we found that average wire temperatures in unencoded data and instruction address buses can rise by about 20 degrees for SPEC CPU2000 benchmark programs. This temperature rise is solely due to heat generation as a result of currents flowing in the wire during bit switching between consecutive cycles. Changes in substrate temperature may cause other short-time effects in the temperature profile which we did not explore in this work. We also observed that instruction address buses, due to their high activity rates and fluctuating energy dissipation profiles, are susceptible to higher thermal stresses and electromigration failure and may also be potential contributors to inductive voltage drops in power supply grids.

## References


[1] K. Banerjee. Trends for ULSI Interconnections and Their Implications for Thermal, Reliability and Performance Issues (Invited Paper). In *Proceedings of the Seventh International Dielectrics and Conductors for ULSI Multilevel Interconnection Conference*, pages 38–50, Mar. 2001.

[2] K. Banerjee and A. Mehrotra. Coupled Analysis of Electromigration Reliability and Performance in ULSI Signal Nets. In *Proceedings of the International Conference on Computer-Aided Design*, pages 158–164, 2001.

[3] K. Banerjee and A. Mehrotra. Global Interconnect Warming. *IEEE Circuits and Devices*, pages 16–32, Sept. 2001.





IEEE COMPUTER SOCIETY

[4] J. Cantin and M. Hill. Cache Performance for SPEC CPU2000 Benchmarks. URL: http://www.cs.wisc.edu/multifacet/misc/spec2000cache-data/, 2003.

[5] T.-Y. Chiang, K. Banerjee, and K. Saraswat. Compact Modeling and SPICE-based Simulation for Electrothermal Analysis of Multilevel ULSI Inteconnects. In *Proceedings of the International Conference on Computer-Aided Design*, pages 165–172, 2001.

[6] T.-Y. Chiang and K. Saraswat. Closed-form Analytical Thermal Model for Accurate Temperature Estimation of Multilevel ULSI Interconnects. In *2003 Symposium on VLSI Circuits Digest of Papers*, pages 275–279, 2003.

[7] B. Cmelik and D. Keppel. SHADE: A Fast Instruction-Set Simulator for Execution Profiling. *ACM SIGMETRICS Performance Evaluation Review*, 22(1):128–137, May 1994.

[8] W. Huang, M. Stan, K. Skadron, K. Sankaranarayanan, S. Ghosh, and S. Velusamy. Compact Thermal Modeling for Temperature-Aware Design. In *Proceedings of the Annual ACM/IEEE Design Automation Conference*, June 2004.

[9] K. Kim, K. Back, N. Shanbhag, C. Liu, and S. Kang. Coupling-driven signal encoding scheme for low-power interface design. In *Proceedings of the International Conference on Computer-Aided Design*, pages 318–321, Nov. 2000.

[10] M. Mui, K. Banerjee, and A. Mehrotra. A Global Interconnect Optimization Scheme for Nanometer Scale VLSI with Implications for Latency, Bandwidth, and Power Dissipation. *IEEE Transactions on Electron Devices*, 51(3):195–203, Feb. 2004.

[11] K. Nabors, S. Kim, J. White, and S. Senturia. Fast Capacitance Extraction of General Three-Dimensional Structures. In *Proceedings of International Conference on Computer Design*, pages 479–484, 1991.

[12] A. Naeemi, R. Venkatesan, and J. D. Meindl. Optimal Global Interconnects for GSI. *IEEE Transactions on Electron Devices*, 50(4):980–987, Apr. 2003.

[13] J. Rabaey, A. Chandrakasan, and B. Nikolic. *Digital Integrated Circuits, Second edition*. Prentice-Hall, Dec. 2002.

[14] Semiconductor Industry Association. International Technology Roadmap for Semiconductors (ITRS), 2001 Edition, 2001.

[15] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, , and D. Tarjan. Temperature-Aware Microarchitecture. In *Proceedings of the 30th International Symposium on Computer Architecture*, June 2003.

[16] P. Sotiriadis and A. Chandrakasan. Bus Energy Minimization by Transition Pattern Coding (TPC) Using a Detailed Deep Sub-micron Bus Model. In *Proceedings of the International Conference on Computer-Aided Design*, 2001.

[17] P. Sotiriadis and A. Chandrakasan. A Bus Energy Model for Deep Submicron Technology. *IEEE Transactions on VLSI Systems*, 10(3):341–350, June 2002.

[18] M. Stan and W. Burleson. Bus-Invert Coding for Low-Power I/O. *IEEE Transactions on VLSI Systems*, 3:49–58, Mar. 1995.

[19] W. Ye, N. Vijaykrishnan, M. Kandemir, and M.J.Irwin. The Design and Use of SimplePower: A Cycle-Accurate Energy Estimation Tool. In *Proceedings of the Annual ACM/IEEE Design Automation Conference*, pages 340–345, June 2000.

[20] Y. Zhang, J. Lach, K. Skadron, and M. Stan. Odd/even Bus Invert with Two-Phase Transfer for Buses with Coupling. In *Proceedings of International Symposium on Low Power Electronics and Design*, pages 80–83, Aug. 2002.




1530-0897/05 $20.00 © 2005 **IEEE**

IEEE COMPUTER SOCIETY