

On-chip Bus Thermal Analysis and Optimization

Feng Wang, Yuan Xie, N. Vijaykrishnan, and M. J. Irwin
The Pennsylvania State University, University Park, PA 16802, USA
{fenwang, yuanxie, vijay, mji}@cse.psu.edu

Abstract

As technology scales, increasing clock rates, decreasing interconnect pitch, and the introduction of low-k dielectrics have made self-heating of the global interconnects an important issue in VLSI design. In this paper, we study the self-heating of on-chip buses and show that the thermal impact due to self-heating of on-chip buses increases as technology scales, thus motivating the need of finding solutions to mitigate this effect. Based on the theoretical analysis, we propose an irredundant bus encoding scheme for on-chip buses to tackle the thermal issue. Simulation results show that our encoding scheme is very efficient to reduce the on-chip bus temperature rise over substrate temperature, with much less overhead compared to other low power encoding schemes.

1. Introduction

On-chip bus power consumption has become an important part of the overall system power consumption. For example, a recent research on the power breakdown of a commercial chip ARM946ES has demonstrated that the on-chip bus power consumption is comparable to other primary sources of power consumption, such as embedded processor and caches [1]. Power dissipation directly transfers into heat dissipation and causes interconnect temperature to rise. Interconnect thermal effect is further exacerbated as technology scales, because of increasing clock rates, decreasing interconnect pitch, and the introduction of low-k dielectric materials (which have low thermal conductivity [2]). It has been shown that the interconnect temperature can be as high as 90 °C [14].

High temperature has a dramatic negative impact on interconnect performance and reliability [2][3]. For example, interconnect (Elmore) delay increases approximately 5% for every 10°C temperature increase, and Electromigration (EM) is significantly accelerated at high temperature, reducing the interconnect reliability [2]. Therefore, it is very important to minimize the interconnect temperature.

Bus encoding techniques [9-12] have been proposed to reduce the power consumption. However, in general, many low-power encoding techniques have insufficient impact on chip temperature because they do not directly address the *spatial and temporal* behavior of the operating temperature.

In this paper, we first characterize the thermal impact of the on-chip buses based on the bus energy and thermal models, and then propose an irredundant bus encoding scheme to spread the switching activity among all bus lines, which can efficiently reduce the transient peak temperature of on-chip buses. The spreading encoding is very efficient and can be combined with

existing low power encoding techniques to further reduce the bus temperature.

The rest of this paper is organized as follows: Section 2 reviews related work; section 3 describes the power and thermal model of on-chip buses; section 4 describes the thermal impact of system bus when technology scales; section 5 shows theoretical analysis to mitigate the thermal impact and gives details of our bus encoding scheme to minimize bus temperature and experimental results are presented in section 6; section 7 concludes the paper.

2. Related work

Many techniques have been proposed to reduce the power consumption of buses. The spatial locality and temporal locality of the *address buses* have been exploited to reduce power consumption by reducing the switching activities on the bus. For example, The *Bus-Invert Code* [11] toggles the polarity of the signals according to the Hamming distance (the number of differing bits) between two consecutive address values by using an additional line on the bus. The T0 code [10] exploits the property that the address lines are typically in incremental mode. TOCAC [12] is an irredundant bus encoding method that combines an adaptive codebook with an extension of T0 code [10], eliminating the redundant bit. However, these techniques to reduce address bus toggling cannot be applied to the *instruction bus*, since bits transferred on an instruction bus are highly irregular. Recently Petrov and Orailoglu [9] proposed a low power encoding framework for embedded processor instruction buses, using efficient instruction transformation to minimize the bit transitions on the instruction bus lines.

Although thermal effects in global interconnect including P/G network and clock networks have been extensively studied [3], the on-chip bus has not gained enough attention. Chiang *et al.* [4] first proposed the analytical models to characterize the thermal effects due to Joule heating in high performance Cu/Low-k interconnects. Based on this thermal model, Sundaresan *et al.* [7] developed a thermal model for on-chip buses. A bus energy model was proposed by Sotiriadis [6], which take into account the inter-wire coupling effects.

3. Energy and thermal model for on-chip bus

In this section, we first present the energy model and thermal model, which are improved over previous work [4][5][6][7] and used in our research to evaluate the effectiveness of our encoding scheme.

3.1 Energy model for on-chip bus

Sotiriadis [6] proposed an elaborate bus energy model for deep submicron technology. The energy drawn from power supply consists of two parts: 1) the energy stored in the capacitances in

the repeater and bus, 2) the energy transformed into heat. The energy transformed into heat drawn from power supply by the i th driver during a transition, E_{heat0} , is given by [6]

$$E_{heat0} = E - \Delta E_c \quad (1)$$

where, E is the total energy drawn from the power supply

$$E = V_i^f e_i^T C^t (V^f - V^i) \quad (2)$$

and ΔE_c is the difference of energy stored on the capacitances after and before switching:

$$\Delta E_c = \frac{1}{2} V_i^f e_i^T C^t V^f - \frac{1}{2} V_i^i e_i^T C^t V^i \quad (3)$$

V_i^f is the final voltage of wire i after switching, V_i^i is the initial voltage of wire i before switching, V^f and V^i are the voltage vectors of all the wires after switching and before switching, and e_i represents a vector, which has one at the i th position and zeros at the other positions. In equation (2), C^t is a matrix of the capacitances:

$$C^t = \begin{bmatrix} c_{1,1} + c_{1,2} & -c_{1,2} & 0 & 0 & 0 \\ -c_{1,2} & c_{2,2} + c_{1,2} + c_{2,3} & -c_{2,3} & 0 & 0 \\ 0 & -c_{2,3} & c_{3,3} + c_{2,3} + c_{3,4} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & -c_{31,32} \\ 0 & 0 & 0 & -c_{31,32} & c_{31,32} + c_{32,32} \end{bmatrix},$$

where $c_{i,i}$ is the total capacitance between wire i and ground, and it includes the capacitance of the driver and the receiver; $c_{i,j}$ is the coupling capacitance between two wires i and j .

The energy model presented above only considers the heat (E_{heat0}) generated when the bus line is charged. However, interconnect joule heating causes temperature to rise because of current flow through the metal, and happens regardless of the direction of the current flow. Therefore, our bus energy model takes into account the heat generated while the bus line is discharged. The energy transformed to heat during bus line discharging can be defined as E_{heat1} :

$$E_{heat1} = \Delta \tilde{E}_c \quad (4)$$

where $\Delta \tilde{E}_c = -\frac{1}{2} \tilde{V}_i^f e_i^T C^t \tilde{V}^f + \frac{1}{2} \tilde{V}_i^i e_i^T C^t \tilde{V}^i$ is the difference of the energy stored in capacitance before and after the transition. Therefore, the total energy dissipated as heat on the driver and bus line can be estimated as the summation of E_{heat1} and E_{heat0} .

To estimate the interconnect temperature, it is necessary to identify the heat dissipation on the bus line. In practice, long interconnects, such as system buses, are divided into small segments; repeaters (inverters) are inserted in order to minimize the propagation delay. The optimal length of interconnects at which to insert the repeaters and the optimal size of repeater are given by [15].

$$l_{opt} = \sqrt{\frac{0.7r_0(c_0 + c_p)}{0.4rc}} \quad (5)$$

$$s_{opt} = \sqrt{\frac{r_0c}{rc_0}} \quad (6)$$

where r_0 , c_0 and c_p are the effective resistance, input capacitance, and output capacitance of the minimum sized inverter respectively, and r and c are interconnect resistance and capacitance per unit length.

As the total energy transformed into heat is determined, we calculate the self-heating power in interconnects. The average power transformed into heat is given as

$$I_{avg}^2 (R_{repeater} + R_{wire}) = P_{heat} = \frac{E_{heat1} + E_{heat0}}{T} \quad (7)$$

where I_{avg} is the average current flow through the interconnects, $R_{repeater}$ is the effective resistance of the repeater, R_{wire} is the wire resistance, P_{heat} is the power dissipated as heat, and T is the period of bus transition from 1 to 0 and from 0 to 1.

Having the s_{opt} and l_{opt} , the power dissipated as heat on interconnects ($I_{avg}^2 R_{wire}$) can be calculated as

$$\frac{l_{opt} r (E_{heat1} + E_{heat0})}{(l_{opt} r + \frac{r_0}{s_{opt}}) T}, \text{ based on equation (7).}$$

3.2 Thermal model for on-chip bus

Chiang *et al.* [4] first proposed a fast electro-thermal simulation methodology to characterize the thermal effects due to Joule heating in high performance interconnects. The parallel thermal coupling between wires and the temperature distribution along interconnects (due to via effects) have been modeled in [4]. Two assumptions have been made in that model. First, the four sidewalls and top surface of the chip, containing the interconnect area, are assumed to be thermally isolated. The only heat dissipation path is through the underlying layers, which is a valid assumption [2]. Second, the variation of thermal conductivity along interconnects is ignored. Having these assumptions and the thermal-electrical analogy, a distributed thermal circuit model can be developed. The equivalent thermal RC network of 32 bit bus wires [7] are shown as Fig.1.

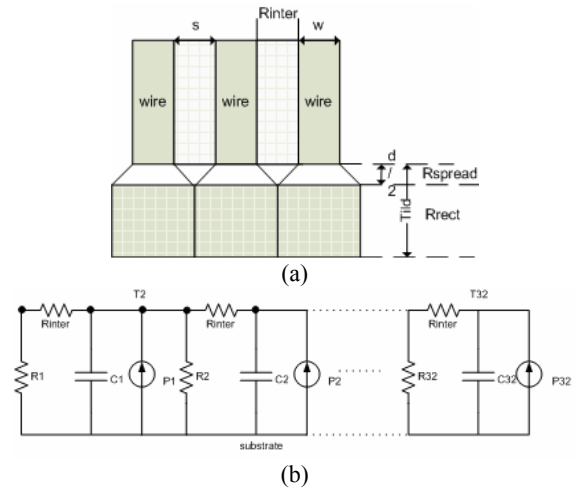


Fig. 1. (a) Geometry used for calculating R_{spread} and R_{rect} (space between any two wires is shared for heat dissipation [4]). (b) Equivalent thermal RC network for 32 bit bus [7].

In the thermal RC network, temperature difference between two nodes corresponds to a voltage difference and the heat transfer rate corresponds to current. Similar to Kirchoff's current law, at every

node, the sum of all heat flowing into a node must equal zero. So, we have the differential equations without the via effect terms [7]:

For edge wires

$$P_1 = C_1 \times \frac{\partial T_1}{\partial t} + \frac{T_1 - T_0}{R_1} + \frac{T_1 - T_2}{R_{inter}} \quad (8)$$

$$P_{32} = C_{32} \times \frac{\partial T_{32}}{\partial t} + \frac{T_{32} - T_0}{R_{32}} + \frac{T_{32} - T_{31}}{R_{inter}} \quad (9)$$

For the wires in the middle

$$P_i = C_i \times \frac{\partial T_i}{\partial t} + \frac{T_i - T_0}{R_i} + \frac{(2T_i - T_{i-1} - T_{i+1})}{R_{inter}} \quad (10)$$

where T_i is the temperature of wire i , P_i is the instantaneous power dissipated as heat in the wire i , and T_0 is the substrate temperature. C_i is thermal capacitance per unit length. R_i is the thermal resistance per unit length of wire along the heat transfer path downward. For R_i , the heat transfers downward as well as spreads laterally [4]. Hence, we have $R_i = R_{spread} + R_{rect}$

$$R_i = \frac{\ln\left(\frac{w_i + s_i}{w_i}\right)}{2K_{ild}} + \frac{t_{ild} - 0.5s_i}{K_{ild}(w_i + s_i)} \quad (11)$$

where K_{ild} is thermal conductivity of the dielectrics, R_{spread} and R_{rect} is the thermal resistances per unit length of the total cross sectional area through which heat conduction takes place laterally and vertically as shown in Fig. 1(a). The lateral thermal resistance, R_{inter} , is used to account for the parallel thermal coupling effect between the wires [4].

According to Chiang's model, heat can flow through the vias within the range of thermal characteristic length [4]; while the via effect is diminished beyond thermal characteristic length. From the previous section, the optimal length l_{opt} of repeater is determined (we estimate the optimal length of repeater for 32nm technology nodes using scaling), and the distance between the vias is at least 10 times larger than thermal characteristic length for each technology generation in this study. As we are investigating the peak temperature along wire due to the joule heating, consequently, the via effect has no impact on the temperature of the center of the wire.

4. Thermal impact of system bus with technology scaling

To capture the actual effect of the technology scaling on interconnects, we perform thermal analysis based on the real data, not maximum current density as done in [4][17][18]. Using a worst case current density for thermal analysis may lead to excessive conservation in thermal aware design. For example, the thermal coupling effects heavily depend on the behavior of the program.

Technology parameters in ITRS 2004 edition are used in our analysis and shown in Table 1. The values of thermal conductivity of low-k materials are taken from [17]. The interconnect capacitances are estimated using Berkeley prediction model [16]. Detailed instruction address bus traces are collected for two SPEC2000 benchmark programs using Simplescalar [8]. For each benchmark program, 100 million instruction addresses were generated. The energy consumption of each 10k cycles of each individual line is estimated using the energy model in section 3.

The temperature rise on each wire is estimated by solving the equivalent thermal RC network presented in Section 3.2. Note that the assumption that we made is the substrate temperature is 70°C .

Table 1. Technology and device parameters for various technology nodes based on the ITRS 2004 edition.

Parameter	Technology nodes			
	90nm	65nm	45nm	32nm
Wire width (nm)	205	145	102.5	70
Space (nm)	205	145	102.5	70
Height of wire (nm)	430.5	319	235.75	168
Height of ILD (nm)	389.5	290	215.25	154
Effective dielectric constant	3.3	2.7	2.3	2.3
Kild(W/m K)	0.19	0.12	0.07	0.07
Clock (MHz)	4171	9285	15079	22980
Supply voltage(V)	1.2	1.1	1	0.9

Fig. 2 shows the thermal impact increases with the technology scaling.

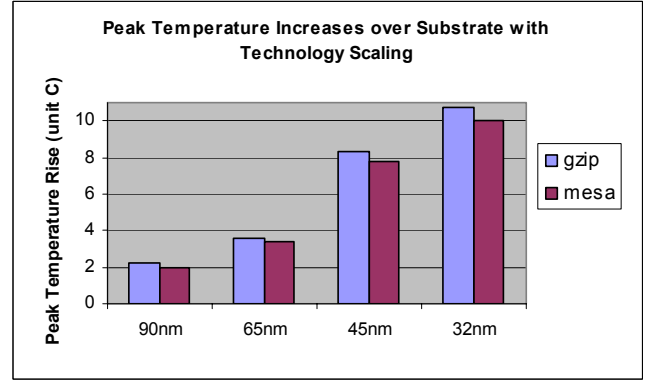


Fig. 2. Thermal impact increases with the technology scaling (peak temperature increase in wires as compared to substrate temperature).

5. Thermal optimization using bus encoding

In this section, we first present a theoretical analysis, which shows that the peak temperature is minimized when the switching activities spread evenly among all bus lines, given that the total switching activities are fixed. Based on the analysis, we present a simple thermal spreading encoding scheme and its implementation.

5.1 Theoretical analysis of bus thermal optimization

A straightforward approach to solve the problem is to reduce the overall switching activities of the system bus. In this sub-section, we approach this problem by finding the optimal power consumption distribution among the on-chip bus to achieve lowest peak temperature over the system bus.

In general, a system bus may consist of several segments of parallel lines. A simple case of one segment of parallel lines with 2 repeaters at its ends is examined here and the results can be scaled for the entire bus.

Problem formulation: Given that the total power consumption across the system bus, P_{total} , is only a function of time t ,

$$P_{total} = \sum_{i=1}^{Buswidth} P_i, \text{ where } P_i \text{ is the power consumption on bus}$$

wire i , and $Buswidth$ is the number of the bus wires, find an optimal distribution of the power consumption among the bus wires, which causes least peak temperature rise over the bus.

The following lemma helps to identify this optimal distribution.

Lemma: When the power consumption on each bus wire,

$P_i = \frac{P_{total}}{Buswidth}$, the minimum peak temperature of the bus can be obtained, given that P_{total} is only a function of time t .

Proof:

Summing the power of all the bus lines using equation (8), (9) and (10) in section 3.2, we get

$$\sum_{i=1}^{Buswidth} P_i = C \times \frac{\partial \sum_{i=1}^{Buswidth} T_i}{\partial t} + \frac{\sum_{i=1}^{Buswidth} T_i - T_0 * Buswidth}{R} \quad (12)$$

Therefore, we have

$$P_{total} = C \times \frac{\partial T_{total}}{\partial t} + \frac{T_{total} - T_0 * Buswidth}{R} \quad (13)$$

where,

$$P_{total} = \sum_{i=1}^{Buswidth} P_i, T_{total} = \sum_{i=1}^{Buswidth} T_i.$$

T_{total} is a function of P_{total} and time t , and is fixed at any instance given that P_{total} is only a function of time t .

We first prove that **the peak temperature across the bus is at least $\frac{T_{total}}{Buswidth}$ at any instance.**

We prove this statement using *proof by contradiction*. Assuming the peak temperature of bus is less than $\frac{T_{total}}{Buswidth}$, then, we have

T_i for each bus lines is less than $\frac{T_{total}}{Buswidth}$. So the summation of the temperature value of all the bus lines:

$$\sum_{i=1}^{Buswidth} T_i < \left(\frac{T_{total}}{Buswidth} \right) * Buswidth = T_{total},$$

which is a contradiction. Thus the minimum value of peak temperature over the bus is not less than $\frac{T_{total}}{Buswidth}$ at any instance. In other words, if the peak temperature is equal to $\frac{T_{total}}{Buswidth}$, then the peak temperature is the minimum.

We then prove that **only when the temperature of each bus line T_i equals to $\frac{T_{total}}{Buswidth}$, the minimum peak temperature on the bus lines can be obtained.**

We prove this statement using *proof by contradiction*. Assuming that the lowest peak temperature can be obtained even when there

is a bus lines i such that T_i is not equal to $\frac{T_{total}}{Buswidth}$. Since the

lowest peak temperature is $\frac{T_{total}}{Buswidth}$, so T_i should be less than

$\frac{T_{total}}{Buswidth}$. Then, $\sum_{j=1}^{Buswidth} T_j$ is less than T_{total} , which is a contradiction.

Substituting $T_i = \frac{T_{total}}{Buswidth}$ into equations (8), (9) and (10), we

get all the $P_i = \frac{P_{total}}{Buswidth}$. Thus balancing the power consumption on the bus lines minimizes the peak temperature across all bus lines.

The conclusion from the above analysis is that, given fixed total switching activities on a bus, the peak temperature will be minimized when the switching activities spread evenly among all bus lines.

5.2 Thermal spreading encoding method

Based on the analysis in subsection 5.1, we propose an efficient thermal spreading encoding scheme to reduce the peak temperature on bus. The encoding scheme has the following advantages over previous low power encoding techniques: low area overhead, low power overhead, and low complexity.

Spreading coding distributes the switching activities across all bus lines by using a very simple and effective method with little overhead. The basic concept is to migrate the switching activities among all bus lines, by rotating the bus line position at a certain period. The implementation of the spreading encoding is very simple, which consists of a shift register and an NxN crossbar logic (N is the bus width). For example, the encoder for a 32-bit bus is shown in the Fig. 3. The decoder is just a reverse version of the encoder. To estimate overhead, we designed and implemented the encoder in the TSMC 90nm process technology. The design is 4151 μm^2 and the power consumption is only 85.69 μW . To compare against low power bus encoding schemes, we also implemented one of the most effective low power encoding schemes T0CAC, the encoder for T0CAC has a much larger area (9243 μm^2) and power consumption (41.10 mW). In terms of performance, our spreading decoder is about 3 times faster than the T0CAC encoder that operates at 1.4 GHz.

The more frequently rotation is performed, the more evenly switching activities are distributed. However, the power overhead of the bus encoder and decoder would also increase with the increase in rotation frequency. Thus a trade-off has to be made. Our experimental results have shown that rotating the bus every 100 clock cycles evenly distributes the switching activities across the bus with far less overhead compared to existing low power bus encoding techniques, which require the encoder and decoder operate at the same frequency of the system bus.

Since the spreading encoding doesn't change the total switching activities on the bus (it only migrates the activities among bus lines), it can be combined with other low power bus encoding techniques to have further temperature reduction: First reduce the total switching activities via low power encoding and then distribute the activities among all the bus lines via spreading encoding. Fig. 4 shows one example of the complete block diagram of the proposed address bus encoder. First block is the simple low power bus encoder with little overhead (for example, T0 or Bus-invert code), which purely aims to reduce the switching

activities on the address lines. The second block is used to distribute these switching activities evenly among the system bus lines.

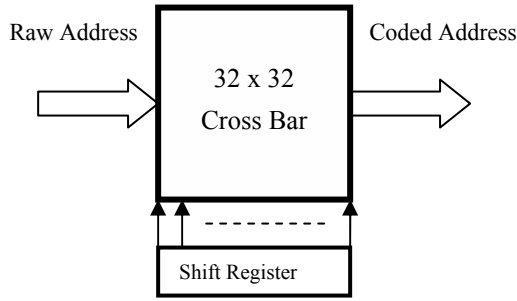


Fig. 3. Block diagram of spreading encoder.

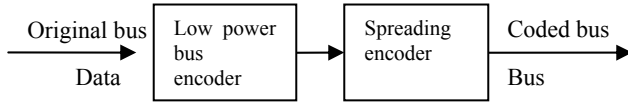


Fig. 4. Block diagram of the proposed address bus encoder.

6. Experimental results

To evaluate the effectiveness of the proposed encoding methods, we collected instruction address bus traces as well as instruction bus traces using the SimpleScalar [8] simulator for eight SPEC 2000 benchmarks. Each trace consists of 100 million instructions. We demonstrate our results on both instruction address bus and instruction bus.

6.1 Instruction address bus

First we show the normalized switching activities of all the bus lines before and after the encoding for benchmark *mesa* in Fig. 5 (the switching activities are normalized to the maximum switching activities among all the address buses). It shows that, before encoding, the instruction address bus switches mainly on the least significant bits (bit 4 switches most, because the instruction length is 64 bit and therefore instruction address increases by 8 bytes). After spreading encoding, the switching activities are effectively distributed across the address bus lines.

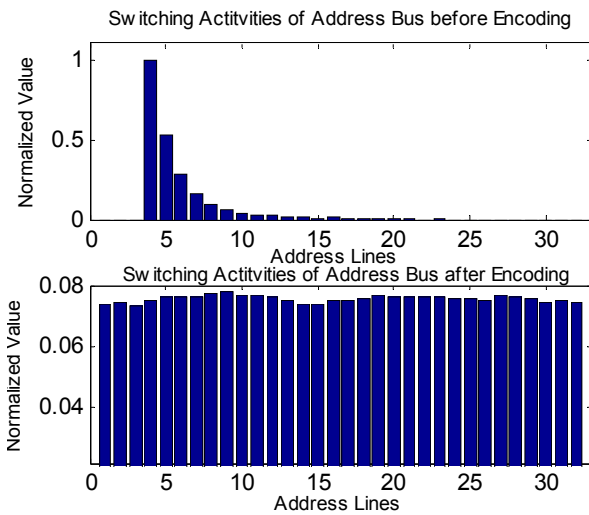


Fig. 5. Switching activities distribution of address bus.

Fig. 6 shows the average power consumption on the bus lines. It shows that, because the switching activities are spread among the bus lines, after encoding the total power consumption is quite evenly distributed among all bus lines except the boundary lines.

Although the two boundary bus lines, line 1 and line 32, have approximately the same switching activities as the middle ones, their coupling effects is much less than that of middle buses and it results in less power dissipation on the boundary bus lines.

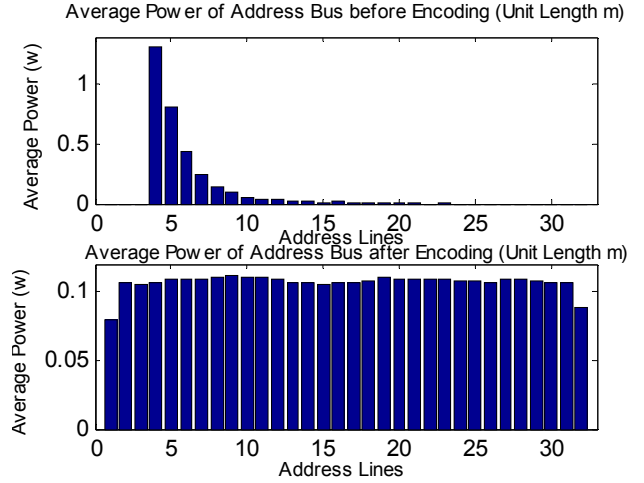


Fig. 6. Power distribution of address bus.

The effectiveness of our encoding in reducing the temperature is demonstrated in Table 2 and Fig. 7.

Table 2. Peak temperature (C) on instruction address bus for gzip benchmark.

Technology nodes	Before encoding	After encoding	Reduction
32nm	80.7	72.2	8.5
45nm	78.3	71.8	6.5
65nm	73.6	70.8	2.8
90nm	72.2	70.4	1.8

From Table 2, it can be observed that spread coding technique will be quite effective in future technology nodes, when the temperature rise due to self-heating becomes more pronounced because of higher current density, worse coupling effects, and lower k materials. Consequently, at 45 nm and 32nm technology node, the average peak temperature is reduced about 6 degrees and 8 degrees respectively using our technique and even for current 90nm technology node, the temperature decrease is 1.8 (which is not small considering that the interconnect temperature is only 2.2 degrees higher than the substrate temperature of 70 degrees in the unencoded case).

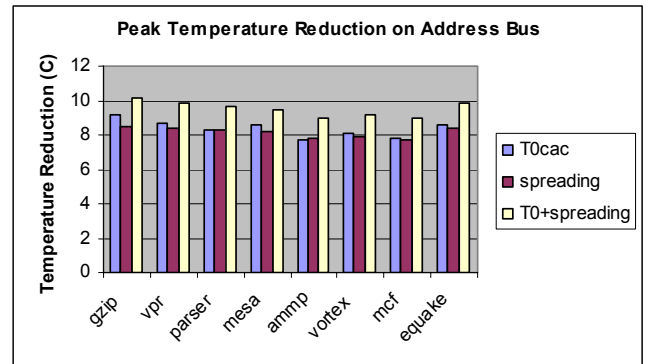


Fig. 7. Peak temperature reduction for low power bus and spreading encoding methods for 32nm technology.

As shown in Fig. 7, temperature reduction of our spreading coding technique is comparable with that of T0CAC, which is a

very effective low power encoding scheme but has much more overhead than spreading coding. Since our spreading encoding scheme doesn't change the total power consumption, combining with less efficient and simpler low power encoding scheme, such as T0 coding, it can further reduce the peak temperature. Fig. 7 gives our results of the T0-spreading coding scheme.

6.2 Instruction bus

As shown in Fig. 5, instruction *address* bus has regular switching patterns, which was taken advantage by the T0CAC encoding scheme. In this section, we demonstrate that our spreading scheme works for buses that have irregular switching patterns as well, such as the instruction bus itself.

Fig. 8 shows the normalized average switching activities on the instruction bus for benchmark *mesa* before and after the spreading encoding.

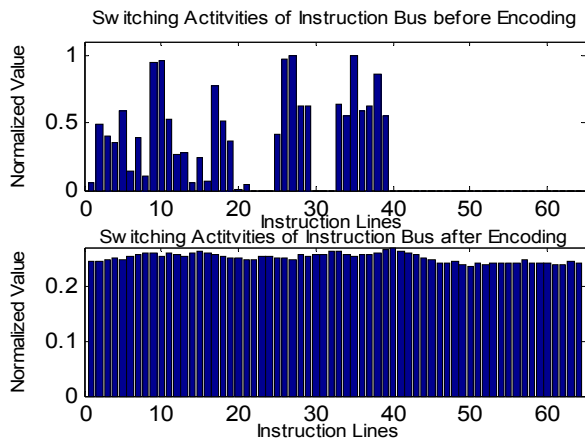


Fig. 8. Switching activities on instruction bus before and after spreading encoding (normalized to the peak value of switching activities).

Fig. 9 compares our encoding scheme and the low power bus encoding scheme Bus-Inversion [11]. We can see that the low power encoding scheme achieves very little temperature reduction while our spreading scheme is still very effective. Although bus inverting saves the overall bus power, it might increase the peak temperature due to the thermal coupling effects. (Note that T0CAC is inefficient for instruction bus due to the irregular switching patterns).

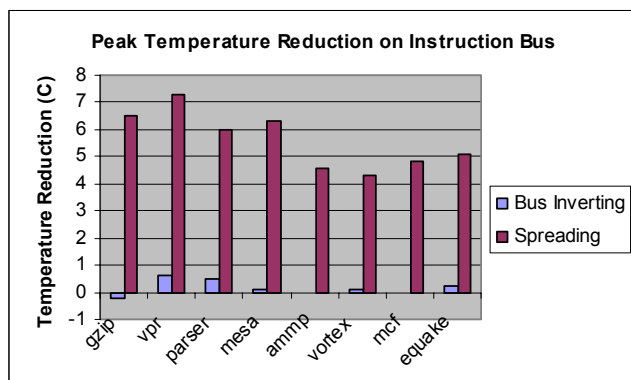


Fig. 9. The peak temperature reduction on the instruction bus using spreading encoding and the bus inversion coding at 32nm technology node.

7. Conclusions

In this paper, we characterize the thermal impact due to self-heating of the on-chip buses based on improved bus energy and thermal models, and propose an irredundant bus encoding scheme, which can spread the switching activity among all bus lines and efficiently reduce the transient peak temperature of on-chip buses. The spreading encoding is very efficient and can be combined with existing low power encoding techniques to further reduce the bus temperature.

8. References

- [1] K. Lahiri and A. Raghunathan, "Power analysis of system-level on-chip communication architectures," *International Symposium on Hardware/Software Codesign and System Synthesis*, September 2004.
- [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*, pp. 158–164, 2001.
- [3] A. H. Ajami, K. Banerjee, A. Mehrotra and M. Pedram, "Analysis of IR-Drop Scaling with Implications for Deep Submicron P/G Network Designs," *IEEE International Symposium on Quality Electronic Design (ISQED)*, San Jose, CA, pp. 35-40, March 2003.
- [4] T.-Y. Chiang, K. Banerjee, and K. Saraswat, "Compact Modeling and SPICE-based Simulation for Electrothermal Analysis of Multilevel ULSI Interconnects," In *Proceedings of the International Conference on Computer-Aided Design*, pp. 165–172, 2001.
- [5] 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.
- [6] P. Sotiriadis and A. Chandrakasan, "A Bus Energy Model for Deep Submicron Technology," *IEEE Transactions on VLSI Systems*, pp. 341–350, June 2002.
- [7] K. Sundaresan and N. R. Mahapatra "Accurate Energy Dissipation and Thermal Modeling for Nanometer-Scale Buses," *Proceedings of the 11th Int'l Symposium on High-Performance Computer Architecture*, 2005
- [8] <http://www.simplescalar.com/>.
- [9] P. Petrov and A. Orailoglu, "Low-Power Instruction Bus Encoding for Embedded Processors," *IEEE Trans. on VLSI*, pp. 812-826, August 2004.
- [10] L. Benini, G. De Micheli, E. Macii, D. Sciuto and C. Silvano, "Asymptotic Zero-Transition Activity Encoding for Address Buses in Low-Power Microprocessor-Based Systems," *IEEE 7th Great Lakes Symposium on VLSI*, pp. 77-82, March 1997.
- [11] M.R. Stan and W.P. Burleson, "Bus-Invert Coding for Low Power I/O," *IEEE Transactions on VLSI*, March 1995.
- [12] S. Komatsu and M. Fujita, "Irredundant Address Bus Encoding Techniques based on Adaptive Codebooks for Low Power," *International Symposium on Low Power Design*, pp. 9-14, Aug 2001,
- [13] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware microarchitecture," *Proc. ISCA-30*, pp. 2–13, June 2003.
- [14] Li Shang, Li-Shiuan Peh, Amit Kumar, and Niraj K. Jha, "Thermal Modeling, Characterization and Management of On-Chip Networks," *Proceedings of the 37th International Symposium on Microarchitecture (MICRO)*, 2004
- [15] R. H.J.M. Otten and R.K. Brayton, "Planning for performance," In *Proceedings of the Annual ACM/IEEE Design Automation Conference*, pp. 122-127, 1998.
- [16] <http://www-device.eecs.berkeley.edu/~ptm>.
- [17] S. Im and K. Banerjee, "Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs," In *Proceedings of the IEDM*, pp. 727–730, 2000.
- [18] Kaustav Banerjee, Massoud Pedram, and Amir H. Ajami, "Analysis and Optimization of Thermal Issues in High-Performance VLSI," In *ACM/SIGDA International Symposium on Physical Design (ISPD)*, pp. 230–237, April, 2001.