Three-Dimensional Chip-Multiprocessor Run-Time Thermal Management

Changyun Zhu, *Student Member, IEEE*, Zhenyu Gu, *Student Member, IEEE*, Li Shang, *Member, IEEE*, Robert P. Dick, *Member, IEEE*, and Russ Joseph, *Member, IEEE*

***Abstract*—Three-dimensional integration has the potential to improve the communication latency and integration density of chip-level multiprocessors (CMPs). However, the stacked high-power density layers of 3-D CMPs increase the importance and difficulty of thermal management. In this paper, we investigate the 3-D CMP run-time thermal management problem and describe efficient management techniques. This paper makes the following main contributions: 1) It identifies and describes the critical concepts required for optimal thermal management, namely the methods by which heterogeneity in both workload power characteristics and processor core thermal characteristics should be exploited; and 2) it proposes an efficient proactive continuously engaged hardware and operating system thermal management technique governed by optimal thermal management polices. The proposed technique is evaluated using multiprogrammed and multithreaded benchmarks in an integrated power, performance, and temperature full-system simulation environment. We find that proactive power-thermal budgeting allows a 30% improvement in instruction throughput compared to a proactive thermal management approach that bases decisions only upon local information. The software components of the proposed thermal management technique have been implemented in the Linux 2.6.8 kernel. This source code will be publicly released. The analysis and technique developed in this paper provide a general solution for future 3-D and 2-D CMPs.**

***Index Terms*—Chip-multiprocessor, thermal management, 3-D integration.**

I. Introduction

CONTINUED increases in integration density, and achieving higher application performance without corresponding increases in processor frequency, are now primary goals for microprocessor designers. As a result, microprocessor design is rapidly moving toward highly scalable chip-multiprocessor (CMP) architectures. Today's mainstream microprocessors are multicore [1]–[6]. The trend for future CMPs is to increase the number of on-chip cores: 80-core prototypes have recently been demonstrated by Intel [7].

Performance scalability is a major challenge in CMP design. Using the mainstream 2-D planar CMOS fabrication process, on-chip interconnect shows poor scalability in both performance and power consumption [8]. Three-dimensional integration has the potential to overcome the limitations of 2-D technology [9]–[12]. By stacking multiple device layers connected through interdie vias, 3-D integration increases logic integration density significantly and reduces on-chip wire length, particularly for global and semiglobal wires. This has motivated computer architects to evaluate 3-D technology for CMP architecture design [10], [13]–[15]. However, none of these papers describes a thermal management solution appropriate for 3-D CMPs.

Thermal issues are a large and growing concern for CMPs [16]–[19]. Increasing chip power consumption and temperature affect circuit reliability (via negative bias temperature instability, electromigration, time-dependent dielectric breakdown, thermal cycling, etc.), power and energy consumption (via increased leakage power), and system cost (via increased cooling and packaging cost). The use of 3-D integration magnifies power dissipation problems [10], [20]–[22]. Chip cross-sectional power density increases linearly with the number of vertically stacked active circuit layers. Three-dimensional integration holds promise but without solutions to the thermal problems it brings, 3-D CMPs will be impractical.

Run-time thermal management techniques, such as dynamic voltage and frequency scaling, clock throttling, execution unit toggling, and workload migration, have been proposed for 2-D high-performance microprocessors [16]–[19], [23], [24]. Using these techniques, cooling solutions and packages need not be designed for worst case power consumption scenarios. Cooling cost can thereby be significantly reduced. Past work, however, cannot effectively optimize the performance–temperature tradeoff in 3-D CMPs for the following reasons.

First, the thermal management techniques deployed in current microprocessors and operating systems (OSs) are primarily used to handle rare worst case processor power consumption events and eliminate thermal emergencies. Although they can potentially introduce significant performance overhead, they are rarely invoked. In contrast, the higher power densities of future 3-D (and some 2-D) CMPs will frequently require operation at or near thermal limits. Already, processors contain reactive techniques to permit the use of reduced-cost packaging and cooling configurations that are not capable of handling maximum power dissipation. Today's laptops frequently invoke thermal management mechanisms that drastically reduce performance, even under normal operating conditions [25]. Power

Manuscript received February 12, 2008; revised March 31, 2008. This work was supported in part by the NSERC under Discovery Grant 388694-01, in part by the National Science Foundation under Awards CNS-0347941, CNS-0702761, and CNS-0720691, and in part by the Semiconductor Research Corporation under Award 2007-HJ-1593. This paper was recommended by Associate Editor D. Atienza.

C. Zhu is with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada.

Z. Gu is with Synopsys, Inc., Sunnyvale, CA 94086 USA.

L. Shang is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80305 USA (e-mail: li.shang@colorado.edu).

R. P. Dick and R. Joseph are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA.

Digital Object Identifier 10.1109/TCAD.2008.925793

should be viewed as a limited resource and processor cores should spend carefully budgeted amounts. Thermal management should be used to proactively and continuously optimize CMP performance and temperature, instead of merely reacting to emergencies.

Second, 3-D CMPs have heterogeneous power and thermal characteristics. On-chip processor cores have different cooling efficiencies. For instance, cores in the layers closer to the heat sink have higher cooling efficiencies than those farther from the heat sink. Processor cores farther from the heat sink will have higher temperatures than their neighbors nearer the heat sink, even when their power consumptions are lower. Intercore thermal correlation is heterogeneous. The thermal correlation between vertically aligned processor cores is stronger than that between processor cores within the same layer. The power and thermal heterogeneity of 3-D CMP poses unique challenges for run-time thermal management. Achieving optimal 3-D CMP performance under a temperature constraint requires careful system-wide control of each processor core's performance and power consumption. Local control alone is insufficient.

In this paper, we develop the analytical framework necessary to determine the thermal impact of every core in a 3-D CMP upon every other core. This framework yields guidelines for near-optimal thermal management. The guidelines are embodied in a proactive global power-thermal budgeting algorithm, performance counterbased workload monitor, and distributed thermal control techniques, which we have implemented in version 2.8.6 of the Linux kernel; this code will be publicly released. The resulting 3-D CMP thermal management solution, which we call ThermOS, is evaluated using detailed full-system simulation with M5 [26]. We have integrated power modeling and thermal analysis tools within the simulator, allowing unified architectural/power/thermal simulation of arbitrary single-threaded and multithreaded applications and the Linux OS. Our results for a wide range of multiprogrammed and multithreaded applications indicate that, given a peak temperature constraint, ThermOS improves CMP throughput by an average of 29.84% when compared to state-of-the-art proactive distributed thermal management. This improvement is primarily due to the power-thermal budgeting guidelines used by ThermOS.

II. Related Work

This section summarizes the current status of 3-D integration in microprocessor design, surveys related work in microprocessor thermal management, and indicates the special thermal management challenges 3-D CMPs will bring.

Several 3-D fabrication technologies have been proposed and developed [9], [11], [12]. Topol *et al.* [9] review the 3-D fabrication process and design techniques developed at IBM. Tezzaron [12] and Samsung [11] developed 3-D fabrication technologies, and Intel is planning to use 3-D integration in the Terascale project [7].

Three-dimensional integration increases the importance of, and complicates, thermal management. The 2-D heat flux density through the heat sink increases roughly linearly with the number of stacked wafers. As a result, unless per-layer power densities are greatly reduced, 3-D CMPs will often operate near their thermal limits. Today's 2-D CMPs already operate at or near their thermal limits, and rely on reactive management techniques to maintain thermal safety.

In addition to increasing the importance of thermal management, 3-D integration complicates thermal management policy design. In contrast with 2-D CMPs, the temperatures of some pairs of 3-D CMP processor cores, e.g., vertically adjacent cores, are highly correlated. Moreover, in 2-D CMPs, processor cores have similar thermal resistances to the ambient and high thermal resistances to other cores. In 3-D CMPs, core resistance to ambient and thermal interaction are heterogeneous. For example, heat generated in cores farther from the heat sink must flow through more layers of silicon.

We next survey work in microprocessor thermal management. Initially, thermal control strategies were seen as an infrequently engaged final resorts. However, due to increasing transistor densities and limitations in cooling technology, in the future, thermal control will be constantly engaged. ThermOS was developed for this emerging thermal management paradigm.

Black *et al.* [10] evaluated the performance improvement yielded by stacking memory and logic layers. Healy *et al.* [27] proposed a microarchitecture-level floorplanning algorithm that works for both 2-D and 3-D ICs. Kgil *et al.* [14] proposed an architecture in which processing core layers are vertically integrated with main memory consisting of multiple DRAM dies, permitting performance and power consumption improvements compared to 2-D designs. Li *et al.* [13] proposed a 3-D topology that combines the benefits of network-on-chip, and 3-D technology to reduce L2 cache latencies. Tsai *et al.* [28] explored cache implementation in 3-D technologies.

Thermal issues are critical for 3-D integration. Puttaswamy and Loh [20] evaluated the thermal impact of 3-D integration on high-performance microprocessors. They also proposed a family of techniques that reduce 3-D power density and assign more power to the die closet to the heat sink [21]. These approaches are principally applied at design time. Skadron *et al.* [19] described a compact thermal analysis technique that has been extended to support 3-D integration. Loi *et al.* [29] studied processor and memory behavior under temperature constraints for 3-D technology. Link and Vijaykrishnan [22] examine thermal effects in 3-D technologies.

Brooks and Martonosi presented one of the first evaluations of dynamic thermal management (DTM) [18]. In essence, DTM allows microprocessor designers to constrain the average-case, instead of worst case, power profile. They instead allow run-time mechanisms to detect and resolve potential thermal emergencies. This yields better overall performance than pessimistically designing systems based on the worst case power profile. Li *et al.* [30] examined the impact of several design constraints, including thermal effect, on CMP architecture design. Sun *et al.* [31] proposed a temperature-aware synthesis technique for 3-D CMPs, but do not consider run-time OS management.

Migration strategies can improve the use of multicore processors by distributing heat generation more uniformly across the chip. Heo *et al.* [32] proposed reducing peak power density by moving computation to another physical location.

Powell *et al.* [23] explored the benefit of OS thermal management for simultaneous multi-threading (SMT) and CMPs. They proposed the Heat and Run strategy, in which the OS coschedules and migrates SMT threads to maximize resource utilization before a thermal emergency arises and then migrates computation to an idle core. Kumar *et al.* [33] examined hardware–software thermal management that uses hardware performance counters to characterize thermal behavior and kernel support to schedule tasks. They evaluated their mechanism on a real system with SMT support and find significant benefits from considering system-level effects which cannot be accounted for with pure hardware techniques. We also take advantage of kernel scheduling and performance counters but also consider multicore management. Recent work by Park *et al.* [34] examined energy-performance tradeoffs in multithreaded applications.

This paper is most closely related to Donald's and Martonosi's [17] research on CMP thermal management using distributed control-theoretic core management and a global controller that guides migration. Both their thermal management technique and ThermOS are continuously engaged thermal management techniques. However, existing proactive thermal management techniques are not appropriate for CMPs with heterogeneous thermal environments, such as 3-D CMPs. Global guidance and power-thermal budgeting are particularly beneficial for 3-D CMPs. By matching core cooling characteristics, application features, and voltage levels, we can improve performance by limiting throttling and migration. We are the first to examine the impact of thermal heterogeneity on thermal management of 3-D architectures. We evaluate our proposed policies in a full system simulator. This experimental setup accounts for the overhead of DTM in the OS, including migration costs and context switches.

III. Heat Flow in 3-D CMPs

This section uses examples to explain the special thermal characteristics of 3-D CMPs and develop a mathematical model that will be used to derive the thermal management policies described in Section IV and validated in Section VI.

A. Introduction to Thermal Modeling

Heat conduction within CMP chip and package can be modeled using Fourier heat flow analysis, which has been the standard method used by industry and academia for circuit-level and architecture-level IC chip–package thermal analysis during the past few decades [19], [35]–[37]. This method is analogous to Georg Simon Ohm's method of modeling electrical current.[1] Using Fourier heat flow analysis, heat flow is analogous to electrical current, and temperature is analogous to voltage. The CMP is virtually partitioned into numerous discrete blocks, as shown in Fig. 1. The thermal conductance of each block is a linear function of the conductivity of its material and its cross-sectional area divided by length; it is analogous to electrical conductance. Blocks also have heat capacities that are analogous to electrical capacitance. Therefore, an instantaneous change in heat generation results in a gradual change in temperature. As a result, the temperature profile of a CMP is essentially its power profile after applying a complicated RC filter. We will deal with this effect in detail in Section III-C. For a thermal model to be accurate, each block must be so small that the temperature within it is uniform. A fine-grained, and thus more accurate, model was used to validate ThermOS. However, for the sake of explanation, this section will describe the coarse-grained model shown in Fig. 1, in which each core is represented with a single thermal model element.

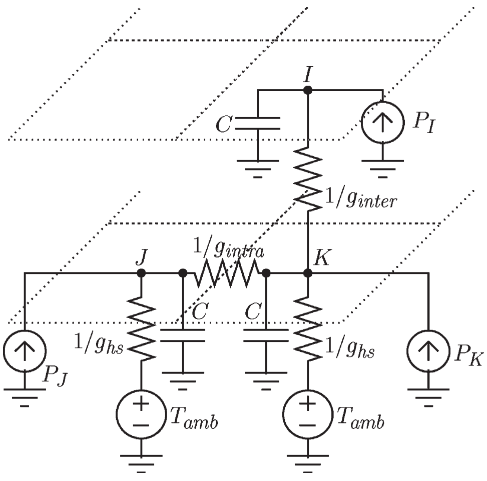


Fig. 1. Interlayer and intralayer thermal heterogeneity and dominance in 3-D CMPs.

In 3-D CMPs fabricated from multiple stacked wafers, the thermal environment varies from layer to layer. Moreover, the intralayer and interlayer thermal relationships among CMP cores are heterogeneous. The rest of this section explains the impact of this heterogeneity on heat flow and builds the theoretical foundations for developing near-optimal 3-D CMP thermal management policies. This understanding is essential for proper thermal management of 3-D CMPs but no prior work is based on it.

Homogeneous Intralayer Characteristics: Fig. 1 shows a simplified heat conduction model for a pair of adjacent CMP cores on the same layer (J and K) and a pair of adjacent CMP cores on different layers (I and K) of a 3-D CMP. As shown in this figure, since the heat dissipation paths of Cores I and K are nearly identical, the thermal conductances of these two cores are nearly equal. In other words, processor cores within the same layer have similar cooling efficiencies.

Heterogeneous Interlayer Characteristics: In contrast to cores on the same layer, Cores I and K have different conductances to the ambient: $g_{\mathrm{hs}} = 0.82$ W/K for Core K and $1/(1/g_{\mathrm{hs}} + 1/g_{\mathrm{inter}}) = 0.73$ W/K for Core I.[2] In addition, the

[1] In fact, Ohm borrowed this model from Fourier, and it was initially proposed to model heat flow.

[2] The thermal conductance values in this section are derived using a thermal analysis package developed by the coauthors [37], which constructs a fine-grained 3-D CMP thermal model based on the material properties and physical structure of the chip–package configuration described in Section V-A2, Tables III and IV. For the sake of explanation, coarse-grained thermal model with compact equations are used in this section to simplify the explanation of fundamental 3-D CMP thermal properties.

steady state temperature of Core I is always higher than that of Core K, even if Core I has a lower power consumption. The following equations formalize this effect, which we refer to as *thermal dominance*. Neglecting the limited intralayer heat flow

$$T_K = T_{\mathrm{amb}} + (P_K + P_I)/g_{\mathrm{hs}} \quad (1)$$

$$\begin{aligned} T_I &= T_K + P_I/g_{\mathrm{inter}} \\ &= T_{\mathrm{amb}} + (P_K + P_I)/g_{\mathrm{hs}} + P_I/g_{\mathrm{inter}} \end{aligned} \quad (2)$$

where T_K and T_I are the temperatures of Cores K and I, T_{amb} is the ambient temperature, P_K and P_I are the power consumptions of Cores K and I, g_{hs} is the thermal conductance from Core K to the ambient through the cooling solution, and g_{inter} is the interlayer thermal conductance between Cores I and K. In addition to Core I thermally dominating Core K, it also has a higher total resistance to the ambient, i.e., it has a lower *cooling efficiency*. As a result, a unit of power consumption on Core I will have at least as great an impact on temperature as a unit of power consumption on Core J or K.

Thermal Coupling: The thermal conductance between J and K (g_{intra}) is approximately 0.41 W/K. Heat can flow between Cores J and K. As a result, the power consumption of one can influence the temperature of the other. However, this thermal coupling is relatively minor compared to that between vertically aligned cores. The thermal conductance between Cores I and K (g_{inter}) is approximately 6.67 W/K, almost $16 \times g_{\mathrm{intra}}$. The large interface area between Cores I and K results in a high thermal conductance, despite the interposed high thermal resistivity (but thin, and therefore low resistance) 10-μm polyimide bonding layer.

Summary and Open Questions: At this point, we can draw some qualitative conclusions. The temperatures of vertically aligned cores are highly correlated, relative to the temperatures of horizontally adjacent cores. Cores farther from the heat sink have higher temperatures than their neighbors closer to the heat sink. In addition, the temperature impact of a unit of power dissipation will be at least as high for Core I as for Cores J and K, due to their differing thermal conductances to the ambient. However, a few questions remain.

1) How can we use this knowledge of thermal environment heterogeneity to guide the development of a CMP thermal management algorithm?
2) What is the impact of the power consumption of each core upon all other cores in the system?

We will now introduce a general analytical framework that answers these questions.

B. Three-Dimensional CMP Heat Flow Analytical Framework

In this section, we formulate the problem of determining the impact of a unit change in power consumption for any given processor core upon the temperatures of all other cores. This formulation provides the theoretical foundation for determining the principals of near-optimal thermal management. We can represent the thermal characteristics of a 3-D CMP using the following notation, which follows naturally from the heat conduction analysis ideas discussed in Section III-A:

$$\mathbf{C}\frac{dT(t)}{dt} + \mathbf{A}T(t) = Pu(t). \quad (3)$$

In this equation, given a system of N thermal elements, $\mathbf{C}$ is a an $N \times N$ matrix with thermal element heat capacities along the diagonal and zeros elsewhere, T is a length N thermal element temperature vector, t is time, $\mathbf{A}$ is an $N \times N$ matrix containing the thermal conductances of adjacent elements at the corresponding row–column intersections and zeros elsewhere, P is a length N thermal element power vector, and $u(t)$ is a step function that changes from 0 to 1 at time t. In addition, matrix $\mathbf{A} = \mathbf{L}^{\mathrm{T}}\mathbf{K}\mathbf{L}$, where $\mathbf{L}$ is a Laplacian matrix and $\mathbf{K}$ is a diagonal matrix containing the thermal conductances of adjacent thermal elements. Given an IC chip–package partition with N connected thermal elements plus a ground element that models the ambient temperature, matrix $\mathbf{A}$ is full rank or nonsingular [38]. The impact of the $\mathbf{C}dT(t)/dt$ term will be explained in detail in Section III-C. In order to ease explanation, neglect $\mathbf{C}$, then solve (3) for T as follows:

$$T = P\mathbf{A}^{-1}. \quad (4)$$

This leads to an interesting observation: $\mathbf{A}^{-1}$ gives the thermal impact of unit changes in power consumption. It is conventionally referred to as the thermal resistance matrix [39], but it would be better to view it as a thermal impact matrix. In order to determine the thermal impact of one core's power consumption on another core's temperature, we need only consider the value in the corresponding row–column intersection in $\mathbf{A}^{-1}$. Let us assume that Core I is currently the hottest in the CMP. ζ_{ij} is the thermal impact coefficient for thermal i due to j. This value indicates the change in the temperature for element i as a consequence of a unit change in power consumption for element j. To determine the impact of power consumed in Cores J and K upon Core I's temperature, we need only consider the thermal impact coefficients in row I in $\mathbf{A}^{-1}$, i.e., $[\zeta_{I,I}, \zeta_{I,J}, \zeta_{I,K}]$. Thus

$$T_I = P_I \times \zeta_{I,I} + P_J \times \zeta_{I,J} + P_K \times \zeta_{I,K}. \quad (5)$$

The thermal impact matrix will be used extensively in Section IV to develop thermal management guidelines. It also gives us a new view of thermal heterogeneity in 3-D CMPs. For a representative stacked-wafer 3-D CMP design, the ζ value for vertically adjacent cores is 1.22 K/W and the ζ value for laterally adjacent cores is 0.39 K/W, yielding a thermal impact ratio of 3.12 for the two cases.

C. Power Model, Dynamic Thermal Analysis, and Modeling Granularity

In the previous sections, we made a number of simplifying assumptions about the thermal environment in order to ease explanation. Our actual analysis and thermal management implementation relaxes many of these assumptions for greater accuracy. We now expound on our thermal model.

In order to determine thermal profile, the power profile must first be known. We model both dynamic power consumption and leakage power consumption [40]. Dependence on voltage, switching activity, capacitance, and temperature are considered. These equations are used together with a Wattch-based EV6 power model [41] to determine the power consumption distribution among architectural units. The power distributions of real multiprogrammed and multithreaded workloads on CMPs may be spatially and temporally heterogeneous. The proposed modeling approach allows us to capture the impact of workload heterogeneity on power and thermal profiles.

As explained in Section III-B, the thermal analysis of real ICs must consider heat capacity $(\mathbf{C})$ as well as thermal conductance, i.e., transient analysis is necessary. The thermal analysis infrastructure we use in architectural–thermal simulation captures these effects using a frequency-domain moment matching analysis technique. Our online thermal management technique continuously adjusts its behavior based on thermal sensor readings. Prior sections assumed that each CMP core is represented by a single thermal element to simplify explanation. In reality, our analysis infrastructure is capable of dividing each CMP core into numerous 3-D thermal elements to permit accurate temperature estimation.

Heat capacity plays a role in thermal modeling and management. Considering transient effects complicates the power and thermal analysis infrastructure. Fortunately, heat capacity limits the rate of temperature change, i.e., the maximum temperature change of a CMP core in a given time interval is limited by the RC thermal time constant of the core and the maximum power consumption change. Although we used a thermal analysis infrastructure that considers transient thermal effects in detail, the proposed thermal management technique is designed to react to transient thermal effects by periodically adapting its behavior based on temperatures measured with thermal sensors or estimated using run-time thermal models.

IV. Three-Dimensional CMP Thermal Management

In this section, we investigate the 3-D CMP run-time thermal management problem and propose efficient management techniques. Given a 3-D CMP with N on-chip processor cores, our goal is to maximize the CMP throughput under run-time thermal constraints. CMP throughput is defined as the total number of instructions executed by the CMP per second

$$\text{CMP\_IPS} = \sum_{i=0}^{N-1} \text{IPC}_i \times f_i \tag{6}$$

where IPC_i and f_i are the run-time instructions per cycle and frequency of Core i.

Run-time thermal safety requires that

$$\forall_{i=0}^{N-1}\, T_i \leq T_{\text{MAX}} \tag{7}$$

i.e., the temperature of each processor core cannot exceed the maximum safe temperature: T_{MAX}.

In the following sections, we analyze the thermal management problem for 3-D CMPs and determine the policies necessary for performance optimization under temperature constraints. This paper will be used to guide the development of our run-time thermal management techniques.

A. *Conditions Required for Optimal 3-D CMP Thermal Management and Derivations of Resulting Policy Guidelines*

This section derives performance optimization guidelines. The central theme is to optimize the performance of CMP cores under a constraint on peak temperature during workload assignment and power-thermal budgeting.

Observation: To maximize CMP throughput, processor cores should operate at different voltages and frequencies due to heterogeneous processor core thermal characteristics and heterogeneous run-time workloads.

As described in Section III-A, processor cores in a 3-D CMP are thermally correlated. The temperature of each Core i, is affected by the power consumptions of all cores, as follows:

$$T_i = \sum_{j=0}^{N-1} \zeta_{i,j} \times p_j \leq T_{\text{MAX}} \tag{8}$$

where T_i is the temperature of processor Core i; $\zeta_{i,j}$, $\{i,j\} \in [0, N-1]$ is an intercore thermal impact coefficient, which indicates the impact of a unit power consumption of Core j on the temperature of Core i; p_j is Core j's power consumption; and N is the number of processor cores of the CMP.

We would like to guide migration of tasks among cores, and budget power to cores, in order to optimize CMP throughput under a temperature constraint. To facilitate developing the necessary guidelines, we introduce the concept of thermal impact per performance gain TIP

$$\text{TIP}_{i,j}^{f} = \frac{dT_i}{df_j}, \quad \text{TIP}_{i,j}^{\text{IPC}} = \frac{dT_i}{d\text{IPC}_j}. \tag{9}$$

$\text{TIP}_{i,j}$ indicates the thermal impact on processor Core i due to the increase in Core j's performance, by either increasing its frequency and voltage, and/or assign a high IPC job to this core. Intuitively, TIP is the thermal cost per unit increase in processor core performance. It can be viewed as the inverse of a core's thermal efficiency. Subject to a temperature bound, maximizing CMP performance thus requires that all the processor cores achieve the same thermal impact per performance improvement on the maximum-temperature core, that is

$$\text{TIP}_{i,0}^{f,\text{IPC}} \equiv \text{TIP}_{i,1}^{f,\text{IPC}} \equiv \cdots \equiv \text{TIP}_{i,N-1}^{f,\text{IPC}}. \tag{10}$$

Note that the impact on T_i due to the power consumption of core j is $\zeta_{i,j} P_j$. Given that dynamic power consumption, $P_j = \xi_j V_j^2 f_j$ (where V_j and f_j are the supply voltage and frequency of Core j), $V_j \propto f_j^{\beta}$, and $\beta \approx 1$ [42]; ξ_j is Core j's run-time switching activity multiplied the capacitance of the switched nodes (which is approximately linearly proportional to the IPC of the job running in Core j), then

$$\zeta_{i,0} f_0^{2\beta+1} \equiv \zeta_{i,1} f_1^{2\beta+1} \equiv \cdots \equiv \zeta_{i,N-1} f_{N-1}^{2\beta+1}$$
$$\zeta_{i,0}\xi_0 f_0^{2\beta} \equiv \zeta_{i,1}\xi_1 f_1^{2\beta} \equiv \cdots \equiv \zeta_{i,N-1}\xi_{N-1} f_{N-1}^{2\beta}. \qquad (11)$$

This result indicates that processor cores with heterogeneous power and thermal characteristics, i.e., different power-thermal impact coefficients $\zeta_{i,j}$ running jobs with different IPCs should be clocked at different frequencies. A similar conclusion can be drawn when both dynamic and leakage power variants are considered.

As shown in Section III-A, the interlayer and intralayer thermal characteristics of 3-D CMPs show distinct differences. This leads to different thermal management policies for interlayer and intralayer processor cores. In the following sections, we determine the conditions required for optimal 3-D CMP thermal management and derive the resulting policy guidelines.

1) Interlayer Power-Thermal Budgeting and Workload Assignment: Interlayer processor cores have heterogeneous thermal characteristics. In addition, vertically aligned cores have strongly correlated temperatures. We now derive heterogeneity-aware guidelines for power-thermal budgeting and workload assignment among vertically aligned cores.

Guideline I: To maximize CMP throughput, the thermal efficiencies of vertically aligned processor cores should be optimized under the thermal constraint, i.e., the voltage and frequency assignment among vertically aligned processor cores should follow (8)–(11).

As shown in Section III-A, among each group of vertically aligned processor cores, the Core i farthest from the heat sink is thermally dominant, i.e., it has the highest temperature and also the lowest cooling efficiency. Therefore, given the thermal constraint for processor Core i, i.e., $T_i \le T_{\mathrm{MAX}}$, the performance-optimal voltage and frequency setup produced by (8)–(11) also guarantees the thermal safety for other vertically aligned processor cores. In other words, (8)–(11) provide the performance-optimal power-thermal budget policy for vertically aligned processor cores. Considering Cores I and K in Fig. 1

$$\zeta_I (= 1/g_{\mathrm{inter}} + 1/g_{\mathrm{hs}}) > \zeta_K (= 1/g_{\mathrm{hs}}),$$
$$\text{and} \quad T_I (= \zeta_I \times P_I + \zeta_K \times P_K) > T_K (= \zeta_K \times P_I + \zeta_K \times P_K).$$

Equations (8)–(11) yield $f_I/f_K = ((\mathrm{IPC}_K \times \zeta_K)/(\mathrm{IPC}_I \times \zeta_I))^{1/2\beta}$. Given homogeneous workload assignment, i.e., $\mathrm{IPC}_K \equiv \mathrm{IPC}_K$, this implies that $f_K > f_I$, i.e., to optimize CMP throughput, the processor core with higher cooling efficiency should be clocked at a higher frequency.

Guideline II: Given jobs with different IPCs, the maximal CMP throughput can only be achieved by maximizing the IPC heterogeneity during workload distribution. To maximize throughput, jobs with higher IPCs should be assigned to cores with higher thermal efficiencies.

This guideline indicates how to distribute run-time workload among vertically aligned processor cores. We will again use Fig. 1 to illustrate the reason for this guideline. Given a temperature constraint T_{MAX} and an arbitrary workload assignment with Core I's IPC equal to IPC_I and Core K's IPC equal to IPC_K, (8)–(11) yield the following performance-optimal power and thermal budget assignment under the given workload distribution:

$$f_I = f_K \times \left(\frac{\mathrm{IPC}_K \times \zeta_K}{\mathrm{IPC}_I \times \zeta_I}\right)^{\frac{1}{2\beta}} \qquad (12)$$

$$f_K = \left(\frac{T_{\mathrm{MAX}}}{\zeta_K \times \mathrm{IPC}_K \left(1 + \left(\frac{\zeta_K \times \mathrm{IPC}_K}{\zeta_I \times \mathrm{IPC}_I}\right)^{\frac{1}{2\beta}}\right)}\right)^{\frac{1}{2\beta+1}}. \qquad (13)$$

Next, we switch the workload between Core I and Core K, (8)–(11) yield the following performance-optimal power and thermal budget assignment for the new distribution:

$$f'_I = f'_K \times \left(\frac{\mathrm{IPC}_I \times \zeta_K}{\mathrm{IPC}_K \times \zeta_I}\right)^{\frac{1}{2\beta}} \qquad (14)$$

$$f'_K = \left(\frac{T_{\mathrm{MAX}}}{\zeta_K \times \mathrm{IPC}_I \left(1 + \left(\frac{\zeta_K \times \mathrm{IPC}_I}{\zeta_I \times \mathrm{IPC}_K}\right)^{\frac{1}{2\beta}}\right)}\right)^{\frac{1}{2\beta+1}}. \qquad (15)$$

Then, simple calculation can show that difference in the CMP throughput between these two workload distributions

$$(\mathrm{IPC}_I \times f_I + \mathrm{IPC}_K \times f_K) - (\mathrm{IPC}_K \times f'_I + \mathrm{IPC}_I \times f'_K)$$
$$\ge 0 \iff \mathrm{IPC}_I \le \mathrm{IPC}_K. \qquad (16)$$

In other words, assigning jobs with higher IPCs to cores with higher thermal efficiencies yields higher overall throughput under the same temperature constraint.

2) Intralayer Power-Thermal Budgeting: Intralayer cores have mostly homogeneous thermal characteristics with almost identical cooling efficiencies (see Section III-A), i.e., $\zeta_{i,i} \approx \zeta_{j,j}$, when Core i and Core j are in the same layer. In addition, the intercore thermal impact is significantly lower than the self-power-thermal impact of each core, i.e., $\zeta_{i,i} \gg \zeta_{i,j}$, when $i \ne j$. We derive the following policies for intralayer power-thermal budgeting and workload assignment.

Guideline III: To maximize aggregate CMP frequency or instruction throughput, power-thermal budget and workload should be balanced among intralayer processor cores.

Consider two intralayer processor cores J and K with $\zeta_{J,J} \equiv \zeta_{K,K} \gg \zeta_{J,K} \equiv \zeta_{K,J}$. The temperature of each core depends mainly on its own power consumption, i.e., $T_J \approx \zeta_{J,J} \times P_J$ and $T_K \approx \zeta_{K,K} \times P_K$ (steady state). Given thermal constraint T_J, $T_K \le T_{\mathrm{MAX}}$, performance optimization yields $P_J \equiv P_K$ and $\mathrm{TIP}_J \equiv \mathrm{TIP}_K$, i.e., both cores should be clocked at the same frequency and execute workload with the same IPC. This guideline can also be motivated as follows. Assume both cores are assigned the same voltage V, frequency f, and workload (ξ and IPC). Therefore, $T_J \equiv T_K$. Next, by adjusting the workload assignment, we increase the IPCs of the jobs assigned to one core and decrease the IPCs of the jobs assigned to another

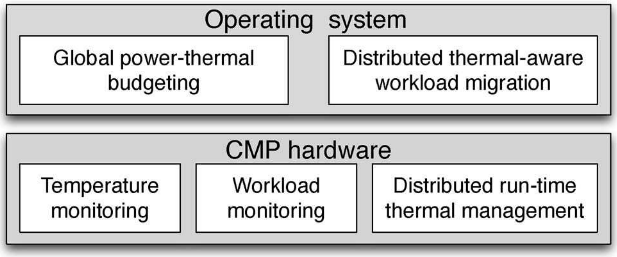

Operating system
Global power-thermal budgeting
Distributed thermal-aware workload migration
CMP hardware
Temperature monitoring
Workload monitoring
Distributed run-time thermal management

Fig. 2. ThermOS: 3-D CMP run-time thermal management.

core. Since $\zeta_{J,J}$, $\zeta_{K,K} \gg \zeta_{J,K}$, $\zeta_{J,K}$, the temperature of one of the cores increases and the peak temperature of these two cores increases. As a result, frequency reduction and performance degradation are required to meet temperature constraints.

B. ThermOS: 3-D CMP Thermal Management

Based on the thermal management guidelines developed in Section IV-A, we have developed ThermOS, a unified hardware and OS thermal management solution to maximize thermally safe for 3-D CMP performance. As shown in Fig. 2 and Table I, ThermOS consists of hardware-based temperature–workload monitoring and distributed run-time thermal management built into a 3-D CMP microarchitecture, as well as a temperature-aware Linux kernel equipped for global power-thermal budgeting and distributed temperature-aware workload migration. ThermOS is a proactive continuously engaged solution designed to handle 3-D CMP power-thermal heterogeneities, distribute run-time workload, and manage the limited power-thermal budget to optimize performance under temperature constraint. Our ThermOS is built upon the Linux 2.6.8 kernel. The 2.6 kernel is presently the latest stable version. It has an $\mathcal{O}$ (1) time complexity scheduler. Our temperature-aware scheduling algorithm maintains the same time complexity. Table I summarizes the proposed offline, run-time, and hardware management techniques.

1) Temperature Monitoring: ThermOS gathers CMP temperature profiles at run-time, which are used to guide temperature-aware workload migration as well as power-thermal budgeting. Either thermal sensors or online thermal analysis may be used for online temperature monitoring. Thermal sensors have been widely used in high-performance microprocessors [1], [43]. Efficient software-based online thermal analysis techniques have also been developed [19].

2) Workload Monitoring: In addition to CMP thermal profile, ThermOS gathers run-time performance and power characteristics to guide job migration as well as power-thermal budgeting. A processor core's activity factor is a function of the capacitances of its functional units and the corresponding run-time activity factors resulting from its workload. Most modern processors provide hardware performance counters for monitoring specific events [1], [44]. These performance counters can be used to inform accurate and efficient regression-based run-time performance and power models [33], [45]. ThermOS uses this technique for linear regression estimation of run-time processor core activity factors. The model was developed offline and integrated with the OS. During execution, each processor core's hardware performance counter values are gathered periodically when triggered by OS timer interrupts (every 1 ms in Linux 2.6.8 kernel). These performance counter values are used for run-time workload activity and IPC estimation.

3) Distributed Thermal-Aware Workload Migration: ThermOS contains a distributed online workload migration technique to support performance optimization. The proposed technique follows the guidelines derived in Section IV-A and carefully handles 3-D CMP interlayer thermal heterogeneity and run-time workload heterogeneity. ThermOS uses a distributed approach that swaps jobs with high IPCs to processor cores with higher thermal efficiencies.

Consider two vertically adjacent processor cores: Core I and Core K. Assume Core K has higher cooling efficiency than Core I. To optimize instruction throughput, ThermOS compares the jobs stored in each processor core's job queue. It first identifies the lowest IPC job (IPCMIN_K) on core K and the highest IPC job (IPCMAX_I) on Core I. If $\mathrm{IPCMIN}_K < \mathrm{IPCMAX}_I$, ThermOS swaps the corresponding jobs. Intralayer thermal heterogeneity and thermal correlation are small. Therefore, ThermOS balances the intralayer IPC distribution to optimize instruction throughput. Average IPCs of jobs on horizontally adjacent cores are compared. If appropriate, they are swapped to further balance the distribution. The proposed distributed thermal-aware workload migration technique has been integrated within the default Linux kernel workload balancing policy. In the current implementation, workload migration occurs every 20 ms.

4) Global Power-Thermal Budgeting: ThermOS dynamically adjusts the power-thermal budgets of processor cores to optimize 3-D CMP performance. Following the guidelines in Section IV-A, ThermOS balances the power-thermal budget assignment among processor cores in the same layer. Equations (8)–(11) are used to guide interlayer power-thermal budgeting. The leakage-temperature dependence introduces temperature variables on both sides of (10). Solving this equation requires numerical iteration and detailed chip-package thermal analysis, which are computationally intensive. To minimize run-time overhead, we have developed a hybrid offline/online budgeting technique.

Given the switching activity (or IPC) range of the workload, the optimal voltage and frequency settings for vertically aligned processor cores are precomputed. The offline component of the budgeting algorithm is iterative. During each iteration, based on the IPC and the switching activity of each processor core, (8)–(11) are used to determine the optimal processor core power-thermal budgets. Thermal analysis is then used to estimate the 3-D CMP thermal profile and update the leakage power profile estimate. This process iterates until the chip-package thermal profile converges, subject to feedback from temperature-dependent leakage power consumption. The final voltage and frequency configurations are stored in a lookup table for efficient use during online power-thermal budgeting. Given that the number of processor layers is L and the number of activity factor settings is n, the lookup table has n^L entries. Increasing n, i.e., the resolution of the activity factor index, improves performance but increases storage overhead, as demonstrated in Section VI-D2. In ThermOS, run-time

TABLE I
ThermOS Implementation

Offline computation			Given the activity factor range of on-chip processor core, derive the look-up table, which contains the optimal voltages and frequencies yielded by Equations 8–11.
Online	OS	rebalance_tick()	Invoke cluster_opt() and group_opt() at the beginning of each workload migration time interval (every 20 ms).
		cluster_opt()	Conduct inter-layer migration according to Guideline II.
		group_opt()	Conduct intra-layer migration according to Guideline III.
		scheduler_tick()	1) Monitor the activity factors of run-time processes using hardware performance counters. 2) Determine the global power–thermal budgeting using run-time table lookup.
	Hardware	Local DVFS	Proactive distributed DVFS based on global guidance and local variation.
		Local clock	Reactive distributed clock throttling to guarantee thermal safety.

TABLE II
DVFS and Clock Throttling Comparison

	Area overhead	Response	Performance impact
DVFS	High	Slow	Low
Clock throttling	Low	Fast	High

power-thermal budgeting is implemented in the Linux kernel and invoked periodically. Periods ranging from 1 to 100 ms are currently supported.

5) Distributed Run-Time Thermal Management: ThermOS uses distributed run-time thermal management to honor the power and thermal budgets described in Section IV-B4 and adhere to a temperature constraint. Periodically, each processor core adjusts its voltage and frequency based on its assigned power-thermal budget. However, transient variations may not be immediately detected by the OS. In order to honor the temperature constraint, ThermOS uses local dynamic voltage and frequency scaling (DVFS) and clock throttling to react to transient variation with lower latency than global power-thermal budgeting. Table II compares these two widely used power management techniques. DVFS has high area overhead, mainly due to complex power supply circuitry and the need of off-chip capacitors and inductors for each independent voltage domain. It also has a higher response latency than clock throttling. For modern high-performance microprocessors equipped with DVFS, the voltage transition rate is in the range of 10 mV/μs [46]. Clock throttling, on the other hand, has low area overhead and low latency. However, DVFS has less performance impact per unit power reduction than clock throttling, thanks to the superlinear dependence of power on voltage. Note that most modern high-performance processors already support DVFS. We are proposing to use this existing DVFS hardware to the best effect. In ThermOS, local DVFS continuously tracks temperature changes and clock throttling is used as a final defense to guarantee thermal safety.

V. Experimental Setup

This section describes the experimental setup used to evaluate the proposed 3-D CMP DTM techniques. We describe our simulation and OS infrastructure, 3-D chip and package models, and benchmark suites.

A. Infrastructure

Performance and temperature estimation for 3-D CMP architectures is challenging. Estimating spatial and temporal thermal profiles requires time-varying power profiles. This, in turn, requires timing and power analysis. To accurately estimate the run-time characteristics of 3-D CMPs, we developed a full-system out-of-order multiprocessor simulation environment with integrated processor performance, power, and thermal models.

1) Full-System Simulation Setup: We use the M5 Full System Simulator [26]. M5 provides a detailed cycle-accurate out-of-order simulation mode and a faster functional simulation mode. We use a combination of full-system checkpoints and the functional simulation mode to boot the system and fast-forward past the initialization portion of our benchmarks. We then switch to detailed simulation mode to evaluate thermal and performance characteristics.

We added a Wattch-based EV6 power model to M5 [41], scaled to a 90-nm process. Our cache power model is based on CACTI [47]. Static power consumption was estimated using an area-based temperature-sensitive leakage model [48]. A 3-D frequency-domain dynamic thermal analysis package was used [37]. Each active layer was modeled using numerous thermal elements.

2) Processor Architecture: There are two ways to stack device layers: face-to-face and face-to-back. For designs with more than two layers, face-to-back bonding decreases worst case interwafer via delay. We evaluate a three-layer front-to-back CMP structure. As shown in Fig. 3, there are eight Alpha 21264 microprocessor cores in the top two layers. Each layer contains four microprocessor cores. Layers are connected with polyimide glue. There is 50 μm of thermal grease between the heat sink and die. Parameters for thermal grease and interface material follow Samson *et al.* [49].

Each processor core has 32 KB L1 data cache and 64 KB L1 instruction cache. There is a 16-MB-shared L2 cache on Layer 2 and 1024 MB of main memory. A 90-nm technology is modeled. Details can be found in the Tables III and IV.

We have accounted for interlayer vias in the thermal model in the following way. The via density in a region follows $\rho_{\mathrm{via}} = nA_{\mathrm{via}}/(wh)$ where n is the number of vias in the region, A_{via} is the cross-sectional area of each via, w is the width of the region, and h is the height of the region. The relationship between via density and effective vertical thermal conductivity follows:

$$K_{\mathrm{eff}} = \rho_{\mathrm{via}} K_{\mathrm{via}} + (1 - \rho_{\mathrm{via}}) K_{\mathrm{layer}} \qquad (17)$$

where K_{via} is the thermal conductivity of the via material and K_{layer} is thermal conductivity of the region without any

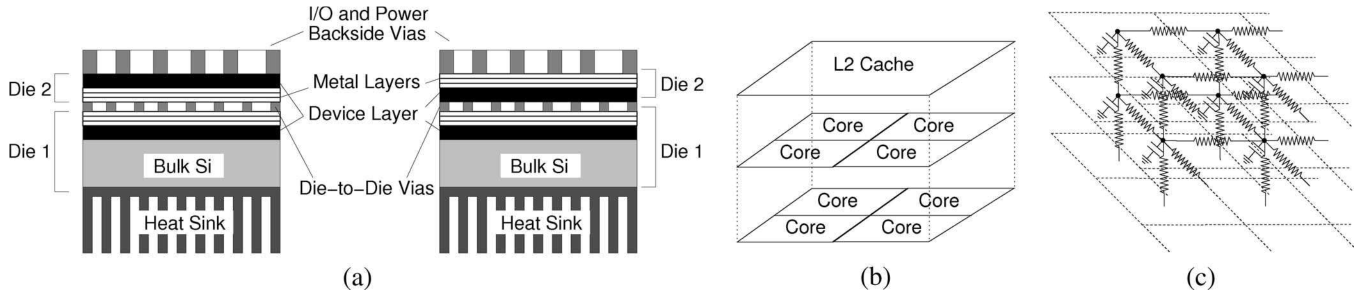


Fig. 3. (a) (Left) Comparison of face-to-face and (right) face-to-back configurations for two stacked dies, (b) 3-D three stacked die floorplan used in this paper, and (c) 3-D CMP chip-package thermal modeling.

TABLE III
DESIGN PARAMETERS FOR ALPHA 21264

Alpha 21264 Configuration (90 nm)	
Die size	$4.56\times4.56\,\text{mm}^2$
Frequency and Voltage	2 GHz, 1.2 V
Instruction Queue	64 entries
Functional Units	4IXU, 2FPU, 1BPU
Physical Registers	80 GPR, 72 FPR
Branch Predictor	1 K local, 4 K global
Memory Hierarchy	
L1 DCache/core	32 KB, 2-way, 64 B blocks, 3 cycle lat.
L1 ICache/core	64 KB, 2-way, 64 B blocks, 1 cycle lat.
Shared L2 Cache	16 MB, 8-way LRU, 64 B blocks, 25 cycle lat.

TABLE IV
THREE-DIMENSIONAL PACKAGE SETUP

Layer	Thermal cond. (W/mK)	Heat cap. ($\text{J/m}^3\text{K}$)	Depth (μm)
Eff. Active Layer (Silicon)	160.11	1.66×10^6	50
Eff. Interface Layer (Polyimide)	6.83	3.99×10^6	10
Heatsink (Cu)	400	3.55×10^6	6,900
Thermal Grease [49]	3–5 (5 used)	4×10^6*	50

* From configuration used in HotSpot [19].

vias. Here, the via is assumed to be copper with a thermal conductivity of 400 W/mK. A typical via size is $15\times15\ \mu$m.

For the Alpha 21264, there are 587 package pins (389 die pins). Interconnect vias use 0.64% of the core area. This results in the effective bulk silicon layer and interface layer thermal conductivities reported in Table IV. There are three types of heat sinks: extruded, folded-fin, and integrated vapor-chamber. In this paper, we assume an extruded copper heat sink with a thermal conductivity of 400 W/mK [50].

3) OS: The ThermOS run-time thermal management algorithms are implemented within the Linux 2.6.8 kernel. We made two main changes to the kernel.

1) *Performance-counterbased power modeling:* We enable OS-level power estimation using performance counters. Hardware event counters of the sort typical for modern processors were added to M5. A regression-based power model was added to the OS [45].
2) *Power-thermal budgeting, task migration, and thermal management:* The proposed power-thermal budgeting and temperature-aware task migration techniques were implemented in the Linux kernel. We modified M5 to support kernel control of DVFS and clock throttling temperature monitoring through privileged machine registers.

B. Benchmark Suites

Multithreaded and multiprogrammed benchmarks from SPEC2000, Media Bench, ALPBench [51], and SPLASH2 [52] are used. Phansalkar *et al.* [53] did a detailed analysis of SPEC2000 and found that it can be divided into different groups based on several benchmark-specific metrics. In order to build a complete set of test cases for our proposed techniques, we selected two benchmark-specific metrics: IPC and expected temperature variation. Although the absolute values of these metrics depend on microarchitectural characteristics, their relative differences in a set of benchmarks are mostly microarchitecture independent.

1) *IPC:* IPC is approximately linearly related to power consumption, which has a strong influence on temperature.
2) *Expected temperature variation:* The main goal of the proposed 3-D CMP thermal management technique is to maximize performance subject to a temperature constraint. In order to evaluate it, we have selected a set of benchmarks with a wide range of spatial and temporal thermal characteristics.

Based on these metrics, the benchmarks were analyzed, yielding the results in Table V. Dynamic power traces were gathered during 500 ms to determine average power consumption, the temporal average of peak temperature, and the maximum peak temperature variation.

We created 17 test setups (see Table VI). Ten of these were for multiprogrammed benchmarks. Each contains mixes of benchmarks with high and low temperature variation and IPC. Each test setup contains two SPEC or Media benchmarks. For multithreaded benchmarks, seven test setups are created. Each test setup contains one ALPBench or SPLASH2 benchmark with two parallel threads. During experiments, each run contains eight copies of each test setup, i.e., 16 processes/threads in total with two processes or threads per core on average.

VI. EXPERIMENTAL RESULTS

This section evaluates ThermOS, the proposed run-time thermal management solution for 3-D CMPs.

A. Comparison of Thermos With Alternatives

In this section, we first contrast ThermOS with solutions used in existing processors. Then, we provide a detailed

TABLE V
BENCHMARK CHARACTERISTICS

Group	Name	Avg. IPC	Avg. Pow. (W)	Max. T	Max. δT
SPEC High IPC	gcc	3.36	14.67	64.88	0.20
	applu	3.13	14.37	65.64	0.12
	gzip	2.78	13.34	63.49	0.34
	mgrid	2.58	13.66	61.84	0.31
SPEC Low IPC	twolf	1.58	11.33	64.30	0.19
	parser	1.55	10.41	60.70	0.28
	vpr	1.47	10.63	60.43	0.29
	mcf	1.25	10.91	63.79	0.25
Media High IPC	gsmenc	3.10	13.50	63.38	0.09
	jpegdec	2.72	13.42	65.89	0.13
Media Low IPC	g721enc	1.94	11.91	61.39	0.08
Multithreaded (two threads)	MPGenc	2.95	14.34	68.78	0.20
	Sphinx3	1.13	9.93	61.68	0.02
	cholesky	2.83	14.27	70.57	0.32
	lu	2.26	12.10	66.97	0.08
	radix	0.84	5.81	57.17	0.28
	water-nsquared	1.85	11.99	65.32	0.12
	water-spatial	1.74	10.57	62.35	0.08

TABLE VI
BENCHMARK SUITES

Multiprogrammed test setups

Group	Filename	Clusters	Benchmarks
SPEC	hv-hipc	High T var., high IPC	gzip, mgrid
	lv-hipc	Low T var., high IPC	applu, gcc
	hv-lipc	High T var., low IPC	parser, vpr
	lv-lipc	Low T var., low IPC	twolf, mcf
	hv-mipc1	High T var., mixed IPC	gzip, parser
	hv-mipc2	High T var., mixed IPC	mgrid, vpr
	lv-mipc1	Low T var., mixed IPC	applu, mcf
	lv-mipc2	Low T var., mixed IPC	gcc, twolf
Media	media-hipc	High IPC	jpegdec, gsmenc
	media-mipc	Mixed IPC	gsmenc, g721enc

Multithreaded test setups

MPGenc, sphinx3, cholesky, lu, radix, water-nsquared, water-spatial

quantitative comparison with a state-of-the-art continuously engaged thermal management technique. The following experiments use 85 °C as a predefined thermal constraint.

Most thermal management techniques used in practice react to emergencies instead of being continuously engaged. They detect dangerously high temperatures and reduce power consumption, generally via hardware clock throttling. Such solutions are adequate when temperatures approach their limits only very rarely. However, high-power densities and constraints on cooling costs require proactive thermal management. Some researchers have moved in this direction.

Donald and Martonosi [17] proposed a distributed continuously engaged thermal management technique for 2-D CMPs. Their approach is based on closed-loop control theory and continuously adjusts the voltage and frequency of each processor core to maintain safe temperatures. Each core has its own controller, and the controllers act independently, without knowledge of the conditions of other cores. This permits significantly better performance than reactive approaches because DVFS can generally reduce power consumption by the same amount as clock throttling with a smaller performance penalty. In fact, their results indicate that, compared with a stop-go-based thermal control policy, distributed DVFS improves throughput by 2.5×. However, independent local control has limitations. The power consumed in one processor can impact the temperatures of other processors in nonuniform ways. As a result, continuously engaged global control can permit better performance than continuously engaged local control. This is particularly true for 3-D architectures, in which the power consumption of a particular processor core has great impact on the temperature of vertically aligned cores and relatively less impact on other cores.

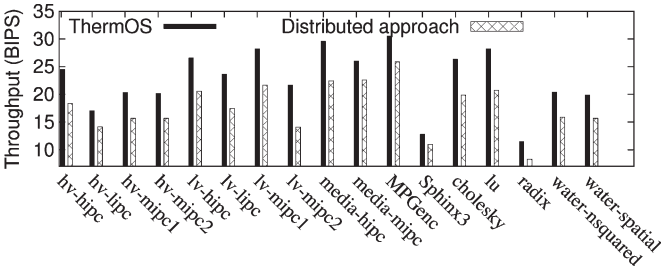


Fig. 4. Comparison of ThermOS and distributed approach [17].

ThermOS uses continuously engaged, distributed global/local control to maximize performance given a temperature bound. It supports both 3-D and 2-D architectures. It has two primary differences with state-of-the-art temperature control techniques. First, it uses global power budgeting that takes into account the thermal interaction between processor cores. Second, it directs temperature-aware workload migration of threads among processor cores.

Fig. 4 shows 3-D CMP run-time instruction throughput (BIPS: billion instructions per second), achieved by ThermOS and Donald's and Martonosi's approach. Compared to the distributed local approach, ThermOS improves instruction throughput by 29.84% on average (ranging from 15.22% to 53.79%). This can be explained as follows. In 3-D CMPs, the strong thermal correlation among interlayer vertically aligned processor cores has significant impact on the temperature of the processor layer farthest from the heat sink. Using the proposed power-thermal budgeting and thermal-aware workload migration techniques, ThermOS determines appropriate power budgets for each group of vertically aligned processor cores. In addition, it uses DVFS to optimize the power-thermal efficiency of each processor core. Together, these techniques maximize overall throughput. Donald's and Martonosi's work, on the other hand, is a distributed processor-local technique. Using this technique, each processor core regulates its power and performance to ensure local thermal safety without considering the thermal impact on neighboring cores. As a result, vertically aligned processor cores are unable to collaboratively share the power and thermal budget, which can reduce CMP performance. In other words, when a distributed local management technique is used, power consumption on processor cores near the heat sink can push processor cores farther from the heat sink to their thermal limits.

B. *Efficiency Impact of Guaranteeing Thermal Safety*

In this section, we establish an upper bound on performance by evaluating a thermal management technique with

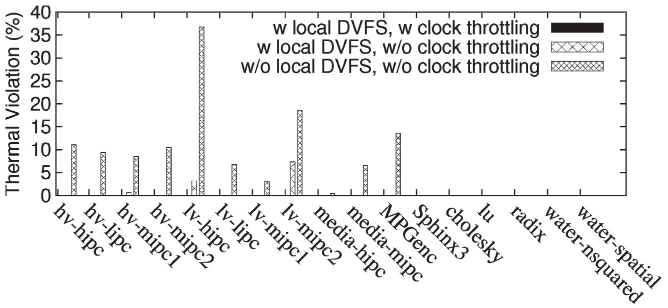


Fig. 5. Reduction in temperature constraint violations due to local DVFS and elimination of temperature constraint violations due to clock throttling.

near-optimal performance, but vulnerability to temperature constraint violations due to transient changes in workload. We then show that there is only a small performance reduction resulting from the additional management techniques ThermOS uses to guarantee thermal safety is small.

ThermOS uses the temperature-aware workload migration and global power-thermal budgeting guidelines derived in Section IV-A. These techniques can potentially offer near-optimal run-time performance subject to a temperature constraint. However, they do not immediately react to transient workload variation occurring in individual processor cores, which may cause run-time temperature constraint violations. ThermOS uses distributed run-time thermal management techniques to guarantee thermal safety, i.e., local DVFS and clock throttling dynamically adjust the voltage and frequency of each processor core to eliminate thermal emergencies. Compared to DVFS, clock throttling is more responsive but degrades performance more for the same thermal improvement. Therefore, in ThermOS, DVFS is continuously engaged and clock throttling is invoked only when local DVFS cannot guarantee thermal safety. These techniques, however, may cause the run-time operations of the processor cores to deviate from the guidelines derived in Section IV-A. Straying from these guidelines has the potential to reduce performance.

Fig. 5 shows the levels of thermal safety achieved by various control techniques. As shown in this figure, when distributed control is disabled, the voltage and frequency of each processor core are solely controlled by global power-thermal budgeting, which does not consider the temporal workload variation within each processor core. This local workload variation can cause significant run-time power variation, and therefore temperature constraint violations. Local DVFS can adapt to rapid workload variation occurring within each processor core and adjust voltage and frequency accordingly, thereby reducing run-time thermal emergencies. When clock throttling is also enabled, processor thermal emergencies are completely eliminated (see Fig. 5).

To further illustrate the effectiveness of the distributed run-time control techniques, Fig. 6 shows the run-time thermal profiles of eight processor cores when running the lv-mipc2 benchmark, with and without local clock throttling. Processors 0–3 are adjacent to the heat sink and processors 4–7 are farther from it. Local DVFS balances CMP thermal profile, and run-time temperature constraint violations (exceeding 85 °C, a predefined thermal threshold used in this experiment) occur only rarely. When both local DVFS and clock throttling are enabled, the temperature constraint is never violated.

Fig. 7 shows that the performance penalty introduced by the distributed control techniques required to guarantee thermal safety is low. To help quantify the performance impact, we normalize the CMP throughput to the value achieved by global power-thermal budgeting and then evaluate the CMP throughput with local DVFS only with both local DVFS and clock throttling. These results indicate that local DVFS degrades instruction throughput by 0.55% on average. Since local DVFS is capable of eliminating most run-time thermal emergencies, clock throttling is rarely invoked. As shown in these figures, enabling both local DVFS and clock throttling results in performance penalties of only 0.60% on average for instruction throughput. In summary, the proposed distributed run-time thermal control technique achieves thermal safety with little performance impact.

C. Robustness to Changes in 3-D Integration

In order to show the robustness of ThermOS to variation in 3-D integration style, we evaluated the performance improvement when used for CMPs using front-to-back and front-to-front wafer integration (see Section V-A). We simulated the proposed technique and Donald's and Martonosi's distributed local approach [17] for both integration styles using all benchmark mixes shown in Table VI. The average CMP instruction throughput improvement was 29.84% for front-to-back integration and 23.77% for front-to-front integration. For all combination of benchmarks and packages, the instruction throughput improvements were greater than 7%. We can conclude that ThermOS permits substantial improvements in performance over Donald's and Martonosi's distributed local technique for different 3-D integration styles.

D. Scalability Analysis of Thermos

ThermOS uses distributed temperature-aware workload migration, global power-thermal budgeting, and distributed run-time thermal control techniques to optimize 3-D CMP throughput and guarantee thermal safety. In contrast with purely local distributed techniques, run-time power-thermal budgeting is global. This might raise concerns about the scalability of ThermOS when used on many-core 3-D CMPs. In this section, we evaluate the scalability of the proposed global power-thermal budgeting technique.

1) Performance Impact: ThermOS periodically decides power-thermal budgets for processor cores. This involves interlayer and intralayer assignment. Run-time interlayer assignment uses efficient table lookup. Intralayer assignment uses an efficient homogeneous assignment policy, i.e., processor cores within the same layer are assigned the same power-thermal budgets. In the current setup, i.e., an eight-core 3-D CMP with a 1-ms global guidance interval, detailed simulation shows that the overall run-time overhead introduced by global power-thermal budgeting is only 0.22%.

The run-time overhead of global power-thermal budgeting is linearly proportional to the run-time global guidance/budgeting

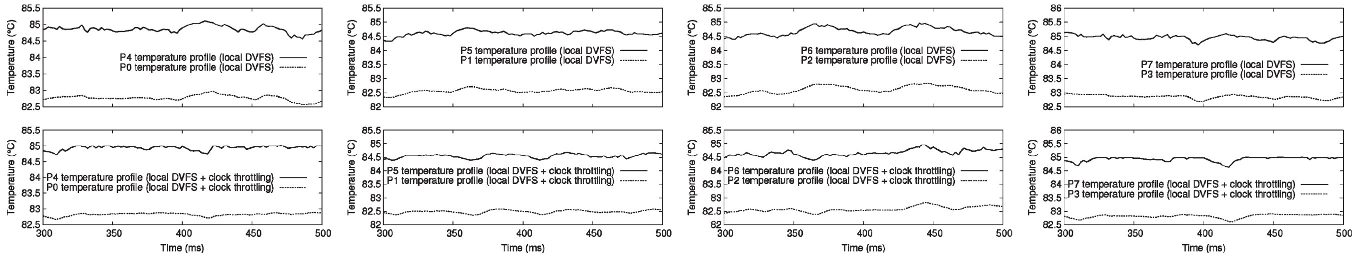


Fig. 6. Temporal temperature variation for eight processor cores (P0–P7) running lv-mipc2 using local DVFS (top) without and (bottom) with clock throttling.

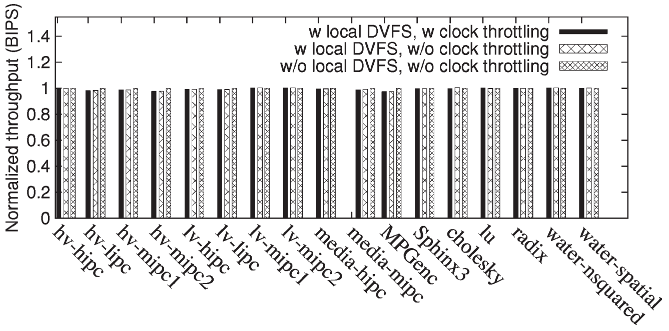


Fig. 7. Negligible CMP instruction throughput reduction resulting from local DVFS and clock throttling.

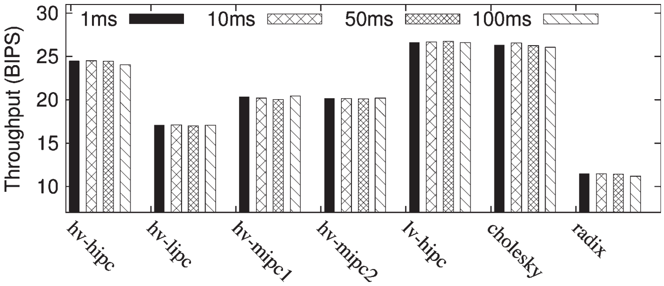


Fig. 8. Impact of global guidance interval.

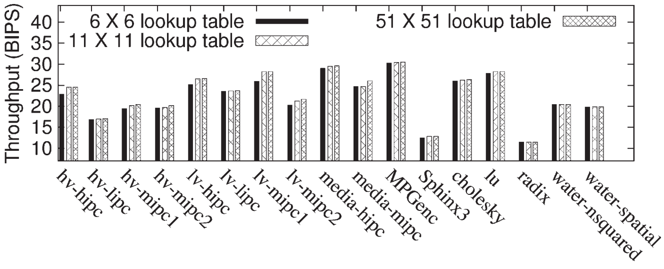


Fig. 9. Impact of lookup table size.

interval. In general, shorter global guidance intervals can more accurately track run-time workload variation but may introduce more run-time overhead and communication contention when aggregating data from different CMP cores. It might therefore be useful to reduce this overhead by increasing the global guidance interval.

In the current setup, a 1-ms guidance interval is used. This is frequent enough to allow adjustments in global power-thermal budget before temporal workload variation can produce large temperature changes, i.e., a higher frequency is unnecessary. To evaluate the impact of increasing global guidance interval on system performance, we run all six benchmarks with high workload variation from Table VI. One low-variation benchmark (lv-hipc) is also included for the sake of comparison. The results are shown in Fig. 8. They indicate that, for guidance intervals up to and including 100 ms, ThermOS maintains nearly identical performance. Only hv-hipc, cholesky, and radix experience noticeable performance degradation, due to their high temporal workload variation. However, changing the global guidance interval from 1 to 100 ms only reduces CMP instruction throughput by 1.81%, 1.06%, and 2.61% for hv-hipc, cholesky, and radix, respectively. We conclude that even if it were necessary to reduce global guidance interval by two orders of magnitude in order to maintain low global power-thermal budgeting run-time overhead in many-core 3-D CMPs, there would be little reduction in thermally safe performance.

2) Storage Impact: As described in Section IV-B4, ThermOS uses an offline iterative budgeting algorithm to precompute some power-thermal budgeting decisions, which are stored using a lookup table in the main memory for efficient run-time usage. This lookup table has n^L entries. Each entry requires 4-B storage. L is the number of processor layers. It is expected that the number of processor layers in 3-D CMPs will be limited. n is the number of activity factor settings, which affects the power-thermal budgeting resolution. Higher resolution improves the accuracy of the run-time power-thermal budgeting decisions, but also increases the storage requirements for the table. In the current setup, we use a 2-D lookup table with 51 × 51 entries (10.4 KB) which provides sufficient resolution for accurate power-thermal budgeting.

It might be useful to decrease lookup table resolution for many-core systems in order to limit storage overhead. We evaluated the impact of decreasing lookup table resolution on thermally safe CMP performance by running all benchmark mixes using 51 × 51, 11 × 11, and 6 × 6 tables. As shown in Fig. 9, compared to the 51 × 51 lookup table, the 11 × 11 lookup table setting reduces the memory usage from 10 404 to 484 B, with average CMP instruction throughput reductions of 0.75%. When the table is reduced to 6 × 6 entries, memory usage decreases to 144 B, with average CMP instruction throughput reductions of 2.87%. We conclude that ThermOS requires little storage, and that its performance degrades slowly with reduced lookup table size.

E. Interaction With 3-D CMP Floorplan Optimization

This experiment evaluates ThermOS for 3-D CMPs with different floorplans. CMP thermal profile is strongly influenced by on-die power distribution. In 3-D CMPs, interlayer vertically

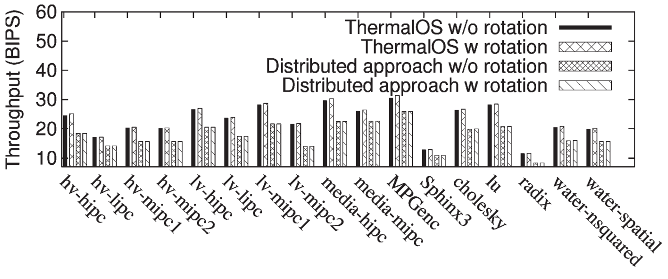


Fig. 10. Impact of floorplan rotation.

aligned processor cores have strong thermal correlation. If all cores have identical floorplans, functional units with high power densities are vertically aligned, potentially creating local thermal hotspots. Intelligent interlayer floorplan arrangement can potentially balance interlayer power profile and minimize chip peak temperature. Using the three-layer 3-D CMP setup with processor core layers and one L2 cache layer, detailed thermal analysis shows that, by rotating the floorplan of top-layer processor cores by 180°, chip power profile is more balanced, intracore local hotspots are minimized, and chip peak temperature is reduced by 1.99 °C on average and 4.24 °C maximum among the multiprogramming and multithreading benchmarks. Fig. 10 compares ThermOS and the baseline distributed technique, with and without floorplan rotation. It shows that both run-time techniques can leverage the temperature reduction offered by floorplan rotation and achieve higher throughput under the same temperature constraint. In addition, ThermOS consistently outperforms the distributed technique by 31.45% and 29.84% on average with and without floorplan rotation, respectively.

VII. Conclusion

Three-dimensional integration has the potential to significantly improve performance and integration density. However, it will also increase power density, thereby increasing the importance of using continuously engaged thermal management techniques. It will also increase the heterogeneity in thermal interaction among processor cores. This requires careful consideration during thermal management policy design.

We have developed a mathematical formulation for optimizing workload assignment, power-thermal budgeting, and voltage mode selection for 3-D CMP thermal management. This formulation has been used to develop a continuously engaged hardware–software thermal management solution for 3-D CMPs. The proposed solution has been implemented within the Linux kernel and evaluated using full-system 3-D CMP and OS simulation. Our strategy outperforms a state-of-the-art proactive thermal management technique that does not make use of power-thermal budgeting.

References

[1] R. Kalla, B. Sinharoy, and J. M. Tendler, "IBM Power5 chip: A dual-core multithreaded processor," *IEEE Micro*, vol. 24, no. 2, pp. 40–47, Mar./Apr. 2004.

[2] P. Kongetira, K. Aingaran, and K. Olukotun, "Niagara: A 32-way multithreaded SPARC processor," *IEEE Micro*, vol. 25, no. 2, pp. 21–29, Mar./Apr. 2005.

[3] *AMD Multi-Core White Paper*. [Online]. Available: http://www.amd.com

[4] *Intel Multi-Core Processor Architecture*. [Online]. Available: http://www.intel.com

[5] M. B. Taylor *et al.*, "Evaluation of the raw microprocessor: An exposed-wire-delay architecture for ILP and streams," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2004, pp. 2–13.

[6] K. Sankaralingam *et al.*, "Exploiting ILP, TLP, and DLP using polymorphism in the TRIPS architecture," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2003, pp. 422–433.

[7] S. Vangal *et al.*, "An 80-tile 1.28 TFLOPS networks-on-chip in 65 nm CMOS," in *Proc. Int. Solid-State Circuits Conf.*, Feb. 2007, pp. 98–100.

[8] V. Agarwal *et al.*, "Clock rate vs. IPC: The end of the road for conventional microarchitectures," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2000, pp. 276–283.

[9] A. W. Topol *et al.*, "Three-dimensional integrated circuits," *IBM J. Res. Develop.*, vol. 50, no. 4/5, pp. 491–506, 2006.

[10] B. Black *et al.*, "Die stacking (3D) microarchitecture," in *Proc. Int. Symp. Microarchitecture*, Dec. 2006, pp. 469–479.

[11] Samsung. [Online]. Available: http://www.samsung.com/

[12] Tezzaron. [Online]. Available: http://www.tezzaron.com/technology/FaStack.htm

[13] F. Li *et al.*, "Design and management of 3D chip multiprocessors using network-in-memory," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2006, pp. 130–141.

[14] T. Kgil *et al.*, "PicoServer: Using 3D stacking technology to enable a compact energy efficient chip multiprocessor," in *Proc. Int. Conf. Architectural Support Program. Lang. Operating Syst.*, Oct. 2006, pp. 117–128.

[15] J. Kim *et al.*, "A novel dimensionally-decomposed router for on-chip communication in 3D architectures," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2007, pp. 138–149.

[16] Y. Li *et al.*, "Performance, energy, and thermal considerations for SMT and CMP architectures," in *Proc. Int. Symp. Comput. Archit.*, Feb. 2005, pp. 71–82.

[17] J. Donald and M. Martonosi, "Techniques for multicore thermal management: Classification and new exploration," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2006, pp. 78–88.

[18] D. Brooks and M. Martonosi, "Dynamic thermal management for high-performance microprocessors," in *Proc. Int. Symp. High-Perform. Comput. Archit.*, Jan. 2001, pp. 171–182.

[19] K. Skadron *et al.*, "Temperature-aware microarchitecture," in *Proc. Int. Symp. Comput. Archit.*, Jun. 2003, pp. 2–13.

[20] K. Puttaswamy and G. H. Loh, "Thermal analysis of a 3D die-stacked high-performance microprocessor," in *Proc. Great Lakes Symp. VLSI*, May 2006, pp. 19–24.

[21] K. Puttaswamy and G. H. Loh, "Thermal herding: Microarchitecture techniques for controlling hotspots in high-performance 3D-integrated processors," in *Proc. Int. Symp. High-Perform. Comput. Archit.*, Feb. 2007, pp. 193–204.

[22] G. M. Link and N. Vijaykrishnan, "Thermal trends in emerging technologies," in *Proc. Int. Symp. Quality Electron. Des.*, Mar. 2006, pp. 625–632.

[23] M. D. Powell, M. Gomaa, and T. N. Vijaykumar, "Heat-and-run: Leveraging SMT and CMP to manage power density through the operating system," in *Proc. Int. Conf. Architectural Support Program. Lang. Operating Syst.*, Nov. 2004, pp. 260–270.

[24] J. McGregor, "x86 power and thermal management," *Microprocess. Rep.*, vol. 18, no. 12, pp. 1–8, Dec. 2004.

[25] A. Mallik *et al.*, "PICSEL: Measuring user-perceived performance to control dynamic frequency scaling," in *Proc. Int. Conf. Architectural Support Program. Lang. Operating Syst.*, Mar. 2008, pp. 70–79.

[26] N. L. Binkert *et al.*, "The M5 simulator: Modeling networked systems," in *Proc. Int. Symp. Microarchitecture*, 2006, vol. 26, pp. 52–60. no. 4.

[27] M. Healy *et al.*, "Multi-objective microarchitectural floorplanning for 2D and 3D ICS," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 1, pp. 38–52, Jan. 2007.

[28] Y. Tsai *et al.*, "Three-dimensional cache design exploration using 3DCacti," in *Proc. Int. Conf. Comput. Des.*, Oct. 2005, pp. 519–524.

[29] G. L. Loi *et al.*, "A thermally-aware performance analysis of vertically integrated (3-D) processor-memory hierarchy," in *Proc. Des. Autom. Conf.*, Jul. 2006, pp. 991–996.

[30] Y. Li *et al.*, "CMP design space exploration subject to physical constraints," in *Proc. Int. Symp. High-Perform. Comput. Archit.*, Feb. 2006, pp. 17–28.

[31] C. Sun, L. Shang, and R. P. Dick, "Three-dimensional multi-processor system-on-chip thermal optimization," in *Proc. Int. Conf. Hardware/Software Codes. Syst. Synth.*, Oct. 2007, pp. 117–122.

[32] S. Heo, K. Barr, and K. Asanovic, “Reducing power density through activity migration,” in *Proc. Int. Symp. Low Power Electron. Des.*, Aug. 2003, pp. 217–222.

[33] A. Kumar *et al.*, “HybDTM: A coordinated hardware–software approach for dynamic thermal management,” in *Proc. Des. Autom. Conf.*, Jul. 2006, pp. 548–553.

[34] S. Park *et al.*, “Managing energy-performance tradeoffs for multithreaded applications,” in *Proc. Int. Conf. Meas. Modeling Comput. Syst.*, Jun. 2007, pp. 169–180.

[35] *COMSOL Multiphysics*, COMSOL, Inc. [Online]. Available: http://www.comsol.com/products/multiphysics

[36] ANSYS. [Online]. Available: http://www.ansys.com

[37] Y. Yang *et al.*, “Adaptive multi-domain thermal modeling and analysis for integrated circuit synthesis and design,” in *Proc. Int. Conf. Comput.-Aided Des.*, Nov. 2006, pp. 575–582.

[38] U. Miekkala, “Graph properties for splitting with grounded Laplacian matrices,” *BIT Numer. Math.*, vol. 33, no. 3, pp. 485–495, Sep. 1993.

[39] Y.-K. Cheng *et al.*, *Electrothermal Analysis of VLSI Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[40] Y. Zhang *et al.*, “HotLeakage: A temperature-aware model of subthreshold and gate leakage for architects,” Univ. Virginia, Charlottesville, VA, Tech. Rep. CS-2003-05, May 2003.

[41] D. Brooks, V. Tiwari, and M. Martonosi, “Wattch: A framework for architectural-level power analysis and optimizations,” in *Proc. Int. Symp. Comput. Archit.*, Jun. 2000, pp. 83–94.

[42] K. A. Bowman *et al.*, “A physical alpha-power law MOSFET model,” *IEEE J. Solid-State Circuits*, vol. 34, no. 10, pp. 1410–1414, Oct. 1999.

[43] D. Pham *et al.*, “The design and implementation of a first-generation CELL processor,” in *Proc. Int. Solid-State Circuits Conf.*, Feb. 2007, pp. 49–52.

[44] R. Sprunt, “Pentium 4 performance-monitoring features,” *IEEE Micro*, vol. 22, no. 4, pp. 72–82, Jul./Aug. 2002.

[45] C. Isci and M. Martonosi, “Runtime power monitoring in high-end processors: Methodology and empirical data,” in *Proc. Int. Symp. Microarchitecture*, Dec. 2003, pp. 93–104.

[46] C. Isci *et al.*, “An analysis of efficient multi-core global power management policies: Maximizing performance for a given power budget,” in *Proc. Int. Symp. Microarchitecture*, Dec. 2006, pp. 347–358.

[47] D. Tarjan, S. Thoziyoor, and N. P. Jouppi, “CACTI 4.0,” HP Laboratories, Palo Alto, CA, Tech. Rep. HPL-2006-86, Jun. 2006.

[48] J. Srinivasan *et al.*, “Exploiting structural duplication for lifetime reliability enhancement,” in *Proc. Int. Symp. Comput. Archit.*, Jun. 2005, pp. 520–531.

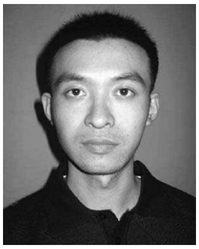
[49] E. C. Samson *et al.*, “Interface material selection and a thermal management technique in second-generation platforms built on Intel Centrino mobile technology,” *Intel Technol. J.*, vol. 9, no. 1, pp. 75–86, Feb. 2005.

[50] R. Viswanath *et al.*, “Thermal performance challenges from silicon to systems,” *Intel Technol. J.*, vol. 4, no. 3, pp. 1–16, Aug. 2000.

[51] M.-L. Li *et al.*, “The ALPBench benchmark suite for complex multimedia applications,” in *Proc. Int. Symp. Workload Characterization*, Oct. 2005, pp. 34–35.

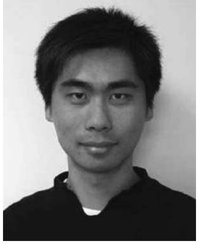
[52] *SPLASH2 Website*. [Online]. Available: http://www-flash.stanford.edu/apps/SPLASH/

[53] A. Phansalkar *et al.*, “Measuring program similarity: Experiments with SPEC CPU benchmark suites,” in *Proc. Int. Symp. Perform. Anal. Syst. Software*, Mar. 2005, pp. 10–20.



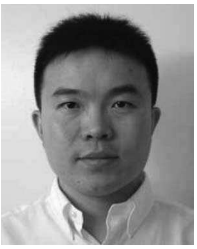
Changyun Zhu (S’06) received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, in 2002 and 2005, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Queen’s University, Kingston, ON, Canada.

His research interests include computer-aided design of integrated circuits, reliability modeling and optimization, and design for nanotechnologies.



Zhenyu Gu (S’04) received the B.S. and M.S. degrees from Fudan University, Shanghai, China, in 2000 and 2003, respectively, and the Ph.D. degree from Northwestern University, Evanston, IL, in 2007.

He has published in the areas of low-power design, behavioral synthesis, and thermal analysis of integrated circuits. He is currently with Synopsys, Inc., Sunnyvale, CA.



Li Shang (S’99–M’04) received the B.E. degree (with honors) from Tsinghua University, Beijing, China, and the Ph.D. degree from Princeton University, Princeton, NJ.

He is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Colorado, Boulder. Before that, he was with the Department of Electrical and Computer Engineering, Queen’s University. He has published in the areas of design automation for embedded systems, design for nanotechnologies, distributed computing, and computer architecture, particularly in thermal/reliability modeling, analysis, and optimization.

Dr. Shang was nominated for the Best Paper Award at DAC 2007 for his work on hybrid SET/CMOS reconfigurable architecture. His work on thermal-aware incremental design flow was nominated for the Best Paper Award at ASP-DAC 2006. His work on temperature-aware on-chip network has been selected for publication in MICRO Top Picks 2006. He also won the Best Paper Award at PDCS 2002. He is currently serving as an Associate Editor of IEEE TRANSACTIONS ON VLSI SYSTEMS and serves on the technical program committees of several design automation conferences. He won his department’s Best Teaching Award in 2006. He is the Walter Light Scholar.

Robert P. Dick (S’95–M’02) received the B.S. degree from Clarkson University, Potsdam, NY, and the Ph.D. degree from Princeton University, Princeton, NJ.

He is an Assistant Professor of electrical engineering and computer science with Northwestern University, Evanston, IL. He worked as a Visiting Professor with the Department of Electronic Engineering, Tsinghua University, Beijing, China, and as a Visiting Researcher with NEC Laboratories America, Inc. He has published on numerous topics within computer engineering.

Dr. Dick received an National Science Foundation CAREER award and won his department’s Best Teacher of the Year award in 2004. His technology won a Computerworld Horizon Award, and his paper was selected by DATE as one of the 30 most influential in the past ten years in 2007. He served as an International Conference on Hardware/Software Codesign and System Synthesis technical program subcommittee chair and serves on several embedded systems and computer-aided design/very large scale integration technical program committees. He is an Associate Editor of IEEE TRANSACTIONS ON VLSI SYSTEMS.

Russ Joseph (S’00–M’05) received the B.S. degree in electrical and computer engineering with an additional major in computer science from Carnegie Mellon University, Pittsburgh, PA, in 1999, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2004.

He is an Assistant Professor of electrical engineering and computer science with Northwestern University, Evanston, IL. His research focuses on power and reliability issues in high-performance computer architecture.

Dr. Joseph received an National Science Foundation CAREER Award.