Three-Dimensional Multiprocessor System-on-Chip Thermal Optimization

Chong Sun, Li Shang
ECE Department
Queen's University
Kingston, ON K7L 3N6, Canada
(4cz1@qlink., li.shang@) queensu.ca

Robert P. Dick
EECS Department
Northwestern University
Evanston, IL 60208, U.S.A.
dickrp@northwestern.edu

ABSTRACT

3D stacked wafer integration has the potential to improve multiprocessor system-on-chip (MPSoC) integration density, performance, and power efficiency. However, the power density of 3D MPSoCs increases with the number of active layers, resulting in high chip temperatures. This can reduce system reliability, reduce performance, and increase cooling cost. Thermal optimization for 3D MPSoCs imposes numerous challenges. It is difficult to manage assignment and scheduling of heterogeneous workloads to maintain thermal safety. In addition, the thermal characteristics of 3D MPSoCs differ from those of 2D MPSoCs because each stacked layer has a different thermal resistance to the ambient and vertically-adjacent processors have strong temperature correlation.

We propose a 3D MPSoC thermal optimization algorithm that conducts task assignment, scheduling, and voltage scaling. A power balancing algorithm is initially used to distribute tasks among cores and active layers. Detailed thermal analysis is used to guide a hotspot mitigation algorithm that incrementally reduces the peak MPSoC temperature by appropriately adjusting task execution times and voltage levels. The proposed algorithm considers leakage power consumption and adapts to inter-layer thermal heterogeneity. Performance evaluation on a set of multiprogrammed and multithreaded benchmarks indicates that the proposed techniques can optimize 3D MPSoC power consumption, power profile, and chip peak temperature.

Categories and Subject Descriptors: C.1.4 [Processor Architectures]: Parallel Architectures; C.5.4 [Computer System Implementation]: VLSI Systems

General Terms: Design, Algorithms, Performance

1. INTRODUCTION

Multiprocessor system-on-chips (MPSoCs) are now widely used in application-specific systems and high-performance computing. They offer performance, design and implementation complexity, power consumption, and thermal benefits over massively superscalar uniprocessor architectures. Their use, and scales, will increase dramatically in the coming years. According to Tony Massimini, chief of technology at semiconductor research and consulting company Semico Research, 16-core processors will be common within the next four years [1]. Intel plans to deliver processors that have dozens or hundreds of cores during the next decade [2].

Increasing functionality and performance requirements, combined with the increasing impact of global interconnect delay, will push designers toward more aggressive technologies in order to increase integration density and decrease communication delay. Using the mainstream two-dimensional (2D) planar CMOS fabrication process, on-chip interconnect shows poor scalability in both performance and power consumption. Three-dimensional (3D) fabrication technology can improve integration density and interconnect delay by using vertical interconnects [3, 4, 5, 6, 7]. Major vendors plan to start shipping 3D MPSoCs incorporating through-silicon-vias within a year [8].

Temperature control is one of the key challenges for 3D MPSoC design. Increasing chip power consumption and temperature affect circuit performance, reliability, cooling cost, packaging cost, and power consumption. 3D integration results in power density increasing linearly in the number of vertically-stacked active layers, complicating thermal-aware design relative to 2D MPSoCs. Thermal issues that can sometimes be safely ignored in 2D planar processes (e.g., self-heating, thermal runaway, and temperature-induced performance degradation) become increasingly prominent in 3D MPSoCs. 3D integration holds promise for MPSoCs. However, before it is practical new solutions are needed for the thermal problems it brings.

Thermal optimization of 3D MPSoCs is complex. An MPSoC must support numerous concurrent tasks. Different tasks exhibit distinct power and thermal characteristics. Managing heterogeneous workloads to optimize performance and temperature is challenging. In addition, the thermal characteristics of 3D MPSoCs differ from those of 2D MPSoCs. In 3D MPSoCs, each active layer has a different thermal resistance to the ambient. Moreover, since the thickness of active layers in 3D MPSoCs is within the range of tens of microns, inter-layer temperature correlation is much higher than intra-layer temperature correlation, i.e., heat flows easily between vertically-adjacent processor cores while lateral core-to-core heat flow is limited. This heterogeneity complicates thermal optimization of 3D MPSoCs.

2. PAST WORK AND CONTRIBUTIONS

Our work draws upon research in MPSoC synthesis and thermal-aware integrated circuit (IC) design.

Given an embedded system specification, MPSoC synthesis [9] is the process of determining the set of processor cores to use, the assignment of computational tasks and communication events to processor cores and interconnect, and the schedule of tasks and communication events. Some work also considers and integrates MPSoC physical design and voltage control. Most prior work on MPSoC synthesis attempts to minimize or constrain system cost and execution time [10, 11, 12, 13], or in some cases energy. Schmitz, Al-Hashimi, and Eles developed a task mapping and scheduling algorithm for energy minimization in distributed embedded systems [14]. Mishra et al. describe a technique dynamic and static power management techniques for multiprocessor real-time systems [15]. Hu et al. present a method of using voltage islands in SoC designs that minimizes power consumption, area overhead, and number of voltage islands [16].

This work was supported in part by NSERC under Discovery Grant #388694-01, in part by the NSF under award CNS-0347941, and in part by the SRC under awards 2007-HJ-1593 and 2007-TJ-1589.

CODES+ISSS'07, September 30–October 3, 2007, Salzburg, Austria.

Thermal analysis and thermal-aware IC design are becoming increasingly important. Various thermal modeling and optimization techniques have been proposed. Skadron et al. develop HotSpot, a compact thermal modeling technique for steady-state and dynamic IC thermal analysis [17]. Li et al. propose a steady-state IC thermal model using multigrid iterative method [18]. Yang et al. developed ISAC, a multi-domain chip-package thermal analysis method, using spatially and temporally adaptive techniques to speedup steady-state, time-domain, and frequency domain thermal analysis [19]. Our uses an enhanced version of ISAC for thermal analysis.

Mukherjee, Ogrenci Memik, and Memik developed a temperature-aware algorithm for resource allocation and binding in high-level synthesis [20]. Gu et al. propose a thermal-aware unified physical-level and high-level synthesis system [21]. Paci et al. indicate that thermal optimization is unnecessary in low-power 2D MPSoCs [22]. However, this conclusion does not generalize to high-performance 3D MPSoCs, for which power densities are higher and the thermal resistances to ambient are larger. Xie and Hung propose a thermal-aware task allocation and scheduling algorithm to minimize IC peak temperature [23].

Thermal optimization of 3D ICs has focused on physical design. Cong et al. propose a thermal-aware 3D floorplanning algorithm [24]. Hung et al. develop a thermal-aware floorplanner which considers the interconnect power consumption [3]. Goplen and Sapatnekar propose a thermal-aware placement solution for 3D ICs [25].

In this work, we propose a 3D MPSoC temperature optimization algorithm. Task assignment, scheduling, and voltage scaling are conducted to optimize 3D MPSoC peak temperature under functionality and timing constraints. Peak temperature is interesting because it limits maximum operating frequency, influences the cost of cooling solutions, and impacts reliability. This work has the following main contributions:

1. We investigate the impact of the heterogeneous characteristics of 3D IC thermal properties and the heterogeneous workload power characteristics on 3D MPSoC thermal optimization. Our study provides general guidance for 3D MPSoC thermal optimization.

2. We propose and evaluate an iterative system-level thermal optimization algorithm for 3D MPSoCs that conducts thermal-aware task assignment, scheduling, and voltage scaling. This algorithm first balances spatial power density and then uses feedback from thermal analysis of a detailed 3D MPSoC thermal model to guide an iterative hotspot mitigation algorithm that minimizes peak temperature. Our results indicate that although power minimization and power density balancing serve as useful starting points for thermal optimization, feedback from a detailed thermal model permits significantly lower peak temperatures. To the best of our knowledge, this is the first system-level thermal optimization algorithm for 3D MPSoCs.

3. 3D MPSOC THERMAL OPTIMIZATION

This section defines the 3D MPSoC thermal optimization problem, gives an overview 3D-Wave, the proposed optimization flow, and explains the optimization algorithms in detail.

3.1 Problem Analysis

In this article, we propose a solution to the following problem. Given

1. A multi-layer 3D chip-level multiprocessor composed of numerous processing elements;

2. A geometrical thermal model consisting of the chip and package heat capacities and thermal conductivities; and

3. A real-time workload consisting of a set of periodic directed acyclic graphs of data-dependent tasks, each of which has an execution time and power consumption at a predefined peak operating voltage and frequency,

determine an assignment of tasks to processor cores, a schedule of tasks, and independent dynamic voltage and frequency scaling

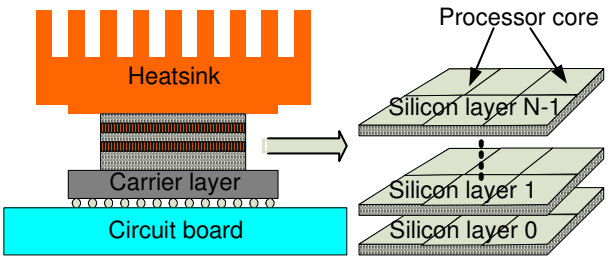

Figure 1: 3D MPSoC chip package structure.

(DVFS) schedules for each processor core in order to optimize the 3D MPSoC peak temperature under the following operation and timing constraints:

1. All tasks must finish execution before their deadlines;

2. For each inter-task precedence constraint, the start time of the child must be later than the finish time of the parent;

3. The execution intervals assigned to the same processor core may not overlap; and

4. The voltage of each processor core must always be within the valid operating range.

The optimization variables for this problem are task execution times, task start times, the processor cores to which all tasks are assigned, and operating voltages for all tasks.

Next, we discuss 3D MPSoC thermal properties and run-time workload characteristics. We then explain the challenges of the thermal optimization problem.

3D MPSoC thermal characteristics: Figure 1 illustrates 3D MPSoC chip and package design. An MPSoC chip contains multiple vertically-stacked silicon layers. Each silicon layer contains processor cores and memory modules. One side of the MPSoC chip is connected to a carrier layer, is attached to the circuit board. The other side of the chip is attached to the cooling solution. The primary heat dissipation path is from the silicon layers through the cooling solution to the ambient. Therefore, for each silicon layer i, the thermal resistance to the ambient can be estimated using the following equation.

$$R_{i,\text{ambient}} = \sum_{j=1}^{i} R_{j,j-1} + R_{0,\text{cooling}} + R_{\text{cooling, ambient}} \qquad (1)$$

where $R_{j,j-1}$ is the thermal resistance between silicon layer j and $j-1$; $R_{0,\text{cooling}}$ is the thermal resistance between silicon layer 0 and the cooling solution; and $R_{\text{cooling, ambient}}$ is the thermal resistance from cooling solution to the ambient. Equation 1 implies the thermal heterogeneity for 3D MPSoCs. Processor cores in the silicon layers closer to cooling solution have higher thermal efficiencies, i.e., lower thermal resistances to the ambient.

Within a 3D MPSoC, inter-processor thermal correlation is heterogeneous. Inter-layer thermal correlation is significant. Among vertically-adjacent processors, each processor's power consumption has direct impact on its neighbors' temperatures. On the other hand, since the thickness of each silicon layer in 3D MPSoC is within the range of tens of microns, lateral heat flow among neighboring processors within the same silicon layer is negligible, i.e., intra-layer thermal correlation is weak.

Workload characteristics: A 3D MPSoC supports a large quantity of tasks with distinct performance and power characteristics. For each task with execution time exe^i, task start time $start^i$, and task deadline $deadline^i$, the slack-execution time ratio, $\frac{deadline^i - start^i - exe^i}{exe^i}$, characterizes the maximum allowed performance slowdown and corresponding potential power and temperature reduction of this task. In addition, different tasks are result in different run-time switching activities, directly affecting 3D MPSoC power consumption.

In summary, heterogeneous thermal characteristics and run-time workload complicate 3D MPSoC thermal optimization. To mini-

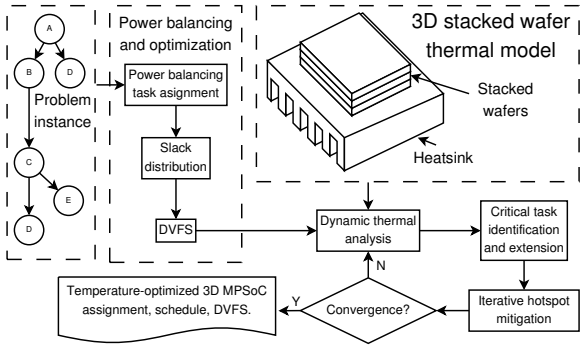

Figure 2: Overview of 3D MPSoC thermal optimization flow.

mize 3D MPSoC peak temperature under a performance constraint, one must consider thermal and workload heterogeneity throughout the 3D MPSoC synthesis flow, including task assignment, scheduling, slack distribution, as well as voltage and frequency scaling.

3.2 3D-Wave Overview

This section gives an overview of the proposed 3D MPSoC thermal optimization flow, which we call 3D-Wave. As shown in Figure 2, an energy-aware task assignment algorithm conducts workload distribution and power balancing among cores (Section 3.4.1). Timing slack distribution and voltage and frequency scaling then minimize 3D MPSoC average power consumption (Section 3.4.2). An iterative optimization loop is then entered (Section 3.5). Within this loop, dynamic thermal analysis is used to determine the 3D MPSoC spatial and temporal thermal profile. These data are used to locate the tasks responsible for the highest temperature and reduce their supply voltages and frequencies in order to permit reduced temporal power density, thereby reducing the peak temperature. The schedules and operating voltages of other tasks are again adapted using the iterative hotspot mitigation algorithm. This process repeats until convergence.

3.3 Modeling

Our 3D MPSoC power model estimates both dynamic power and leakage power consumption. Leakage power is dependent on operating voltage and temperature. We use the piecewise linear technique proposed by Liu et al. to model leakage power and temperature dependency [26].

Given an IC chip-package design with N discrete elements, the thermal analysis problem can be described as follows:

$$\mathbf{C}T(t)' + \mathbf{A}T(t) = Pu(t) \tag{2}$$

where the thermal capacitance matrix, $\mathbf{C}$, is an $[N \times N]$ diagonal matrix; the thermal conductance matrix, $\mathbf{A}$, is an $[N \times N]$ sparse matrix; $T(t)$ and $P(t)$ are $[N \times 1]$ temperature and power vectors; and $u(t)$ is the time step function. In this work, we consider temporal changes in thermal profile, i.e., we use dynamic thermal analysis in which the effects of heat capacity are explicitly modeled. For either the dynamic or steady-state versions of the problem, although direct solutions are theoretically possible, the computational expense is too high for use on high-resolution thermal models. In this work, we use ISAC for fine-grained dynamic modeling [19]. This algorithm is designed for accurate thermal analysis that is sufficiently fast for use in the inner loop of synthesis algorithms.

3.4 Constructive Power Balancing and Minimization Algorithm

This section describes a constructive algorithm for peak temperature minimization in 3D MPSoCs. The solution produced by this algorithm can be further optimized using the iterative hotspot mitigation algorithm described in Section 3.5.

3.4.1 Power Balancing Task Assignment

3D-Wave optimizes peak temperature by balancing temporal and spatial power density in 3D MPSoCs. First, it uses constructive task assignment to balance the average power and temperature distribution among processors. As shown in Algorithm 1, tasks are prioritized in the order of earliest finish time (EFT). In steps 4–7, the total energy consumed by tasks assigned to each processor is calculated. Processors in different layers have different power to thermal impact ratios. For example, processors closer to the heatsink can safely execute tasks with higher power consumptions than those farther from the heatsink. Step 5 evaluates the thermal impact of assigning to each processor core, p_T. It calculates the average power consumption of each processor p_p weighted by the average thermal resistance of the processor to the ambient r_p. Task s is then tentatively assigned to the processor with the minimal thermal impact. During task assignment, a new precedence constraint is added from the latest task on the target processor to the new task in order to prevent temporal overlap. The result of Algorithm 1 is a power distribution among processors that is appropriately weighted for temperature minimization. This initial distribution will be further improved by later optimization stages.

Algorithm 1 Power balancing task assignment

1: Compute EFT for each task
2: Prioritize the sequence **S** of the tasks in order of EFT
3: **for** all task $s \in \mathbf{S}$ **do**
4: Calculate $p_T = p_p \times r_p$, the thermal impact of processor p
5: $p_{Tmin} = \text{argmin}_{p \in \mathbf{T}}\, p_T$
6: Tentatively assign task s to p_{Tmin}, the processor with min. thermal impact
7: Add a precedence constraint from the latest task on p to s
8: **end for**

3.4.2 Slack Distribution and Voltage Scaling

Task slack distribution is used to permit voltage scaling, which permits temporal power density and peak temperature minimization. To address 3D MPSoC workload and heat flow heterogeneity, 3D-Wave uses the same weighted average power consumption as an objective function during slack distribution. Note that the results produced by this constructive algorithm will later be improved using the iterative hotspot mitigation algorithm described in Section 3.5.

Given a set of sequential tasks, the thermal resistance weighted slack distribution problem is equivalent to deciding the execution time of each task such that the thermal resistance weighted energy consumption is minimized under a hard constraint on path execution time. We shall use the following variables and constants: D is the bound on path execution time; p is the set of all tasks on the path; d_i is the task's execution time; v_i is the task's voltage; V_t is the threshold voltage constant; K is an execution time constant; E is the total path energy consumption; e_i is the thermal resistance weighted energy required for a task; C_i is the switched capacitance constant of a task's processor core; R_i is the thermal resistance from the task's processor core to the ambient; and α is the alpha power law constant [27].

$$d_i = \frac{Kv_i}{(v_i - V_t)^{\alpha}} \text{ subject to the constraint } D \geq \sum_{i \in p} d_i \tag{3}$$

However, V_t is small and a very low value of v will generally imply an unacceptable path delay that will be prevented by the constraint in Equation 3. Therefore, we may assume V_t is small, thus

$$d_i \simeq \frac{Kv_i}{v_i^{\alpha}} \quad \text{and} \quad v_i = \left(\frac{d_i}{K}\right)^{\frac{1}{1-\alpha}} \tag{4}$$

$$e_i = R_i C_i v_i^2 = R_i C_i \left(\frac{d_i}{K}\right)^{\frac{2}{1-\alpha}} \text{ and} \tag{5}$$

$$E = \sum_{i \in p} R_i C_i \left(\frac{d_i}{K_i}\right)^{\frac{2}{1-\alpha}} \text{ then } \min_{\substack{\forall i \in p \\ v_i}} \quad \sum_{i \in p} R_i C_i \left(\frac{d_i}{K_i}\right)^{\frac{2}{1-\alpha}} \tag{6}$$

Algorithm 2 Slack distribution

```
1: Compute all task slacks
2: Group task slacks into same-slack paths, P
3: Sort paths in order of increasing slack
4: while p in P do
5:    Assign slack to tasks in p according to Equations 3 and 7
6:    Recompute all task slacks
7: end while
```

Note that a decrease in v_i implies a decrease in e_i, which implies an increase in d_i. Therefore, for minimal E, $D = \sum_{i \in p} d_i$. Consider the delay and energy trade-off for an arbitrary pair of tasks:

$$e_{12} = \frac{R_1 C_1}{K_1^{\frac{2}{1-\alpha}}} (d_1)^{\frac{2}{1-\alpha}} + \frac{R_2 C_2}{K_2^{\frac{2}{1-\alpha}}} (d_{12} - d_1)^{\frac{2}{1-\alpha}} \quad (7)$$

We assign task durations that honor Equation 7 and the constraint in Equation 3. By granting slack to each task in the path such that its time is proportional to its time share, we produce an a favorable initial dynamic voltage and frequency scaling schedule. However, this schedule is not perfect because assigning non-minimal task time shares eventually reduces scheduling flexibility. Limitations on task start times may influence the earliest start times and latest finish times of tasks on other paths as a result of precedence constraints and resource contention. In order to avoid deadline violations, slack distribution is conducted on task paths in order of increasing path slack. A modified depth-first search for generating paths is conducted on a graph in which each vertex is a task labeled with its slack and each edge is a data dependency. Vertex children are visited in increasing order of slack, thereby guaranteeing that vertices on multiple paths will be included in minimal-slack paths.

As shown in Algorithm 2, starting from the minimal-slack path, slack is distributed to tasks according to Equations 3 and 7. After slack sharing is completed for a given path, the slacks of all nodes are recomputed and slack distribution proceeds for the next path. The voltage of each task is then scaled to the lowest value permitting completion within its assigned time interval. Although this algorithm is not guaranteed to produce minimal-energy solution, it provides a starting point for the iterative optimization algorithm described in Section 3.5.

3.5 Iterative Hotspot Mitigation Algorithm

This section describes an iterative hotspot mitigation algorithm to further improve the peak temperature resulting from the task assignment, scheduling, and DVFS solution produced by the constructive power balancing and minimization algorithm. This algorithm iteratively detects and eliminates temporal hotspots to optimize 3D MPSoC peak temperature.

As shown in Algorithm 3, within each iteration 3D MPSoC peak temperature is first computed using dynamic thermal analysis (step 2). Tasks within the peak temperature region are identified and recorded in a critical task set (step 3). For each selected task, this algorithm determines the potential reduction in peak temperature resulting from adjusting the slack distribution between this task and its neighboring tasks. The slack time of the task is increased by δt and the slack time of its parent and child tasks are each decreased by $\delta t/2$. Dynamic thermal analysis is used to estimate the peak temperature reduction (step 10), which is then recorded with the corresponding slack time adjustment setup (step 11 and 12). Next, the initial slack distribution is restored and the algorithm evaluates the impact of increasing the slack of other candidate tasks (step 13). After evaluating all the tasks primarily responsible for the temporal hotspot, this algorithm selects the task whose expansion decreased the peak temperature the most (step 15). It then updates the slack time adjustment coefficient δt (explained below) and conducts dynamic thermal analysis to identify next thermal hotspot. This process continues till no peak temperature improvement.

Several algorithm design decisions merit further discussion. First, we describe the method of identifying tasks responsible for temporal hotspots. In a 3D MPSoC, tasks executed concurrently on vertically-adjacent processor cores are thermally correlated. When a temporal hotspot occurs, tasks executed on both the hotspot processor and vertically-adjacent processors should be considered. Second, the slack time adjustment coefficient, δt, should be carefully selected. A large value should be used initially to speed peak temperature reduction. However, the value should decrease as temporal thermal variation decreases to permit convergence. Our analysis shows that a simple multiplicative adaption policy, i.e., reducing δt by 10% every 100 iterations, provides both good runtime efficiency and stability. Third, dynamic thermal analysis is used to guide the proposed iterative optimization flow. Even though thermal analysis increases computational complexity, it can accurately locate temporal hotspots and determine the complex thermal implications of changes to task execution times and voltages. As shown in Section 4.2, this algorithm consistently produces higher-quality solutions than a solution based on task power consumption, alone.

Algorithm 3 Iterative hotspot mitigation

```
1: while 3D MPSoC peak temperature can be reduced do
2:    Compute peak temperature
3:    Find critical task set S within the peak temperature region
4:    for each task s ∈ S do
5:       Prolong s' execution time by δt
6:       for each task s', where s' is s's parent or child task do
7:          Reduce s''s execution time by δt/2
8:          Slack time validation
9:       end for
10:      Recompute peak temperature
11:      Record local slack adjustment
12:      Record new peak temperature
13:      Restore initial slack distribution
14:   end for
15:   Apply the slack adjustment with lowest peak temperature
16:   Adjust δt
17: end while
```

4. EXPERIMENTAL RESULTS

In this section, we present experimental results for 3D-Wave, the proposed 3D MPSoC thermal optimization algorithm. Section 4.1 describes the experimental setup and the benchmarks used to evaluate 3D-Wave. The algorithm consists of a constructive power balancing and minimization algorithm and a thermal analysis driven iterative hotspot mitigation algorithm. Section 4.2 gives a detailed characterization of each optimization stage. These data indicate that 3D-Wave produces solutions with low average power consumption and then balances spatial and temporal power variation in order to minimize 3D MPSoC peak temperature.

The experiments were conducted on AMD 64 X2 Linux workstations with 2 GB of RAM. All the optimization runs require less than 400 s of CPU time.

4.1 Experimental Setup

This section describes the 3D MPSoC chip package setup and the benchmarks used to evaluate 3D-Wave.

We consider a two-layer front-to-back eight-core 3D MPSoC architecture. Each silicon layer contains four Alpha 21264 microprocessor cores. Each processor core has a size of 4.56 mm×4.56 mm. The thickness of the top silicon layer is 50 µm. The thickness of the bottom silicon layer is 0.5 mm; this layer is thicker in order to provide mechanical support. There is a 10 µm polyimide glue layer between silicon layers. We model a forced-air cooling solution. The 3D MPSoC chip is attached to a copper heat sink through a 50 µm thermal grease interface layer. We use a detailed thermal analysis algorithm to evaluate heat flow through this stacked wafer thermal model [19]. These analysis results provide guidance to 3D-Wave during optimization.

The proposed 3D MPSoC thermal optimization algorithm is evaluated using testing traces generated from 22 multiprogrammed and multithreaded benchmarks. Table 1 shows the benchmarks we used from MediaBench, ALPBench, and SPEC2000 benchmark suits. The M5 multi-processor full-system simulation environment [28] with integrated dynamic and leakage power models is used to gather the execution and power consumption traces of these benchmarks and convert them into 22 task graphs. The traces are produced by dividing the execution of processes into short discrete

Table 1: Members Benchmark Suites

SPEC2000	applu, bzip2, gap, gcc, gzip, lucas, mcf, mgrid, parser, perlbmk, twolf
MediaBench	adpcmdec, g721enc, g721dec, gsmenc, gsmdec, jpegenc, jpegdec
Alpbench	mpgenc, mpgdec, sphinx3, tachyon

time intervals corresponding to tasks and assigning each such task a power consumption based on processor power simulation. Using these 22 task graphs, ten benchmarks were constructed. Each of these is composed of 11 of the 22 task graphs.

4.2 Performance Evaluation

3D MPSoC peak temperature is loosely related to average power consumption; minimizing average power consumption is a useful first step in thermal optimization. However, it is also necessary to consider detailed spatial and temporal power density in order to optimize peak temperature. 3D-Wave uses a two-stage optimization flow. A constructive power balancing and minimization algorithm (PBMCA) combined with a thermal analysis driven iterative hotspot mitigation algorithm (IHM) is used to minimize chip power consumption and optimize thermal profile. To evaluate 3D-Wave, we characterize the power and thermal impact of PBMCA and PBMCA+IHM (3D-Wave) separately. The thermal optimization of 3D-Wave is guided by dynamic thermal analysis. To determine whether power consumption can be used as a computationally-efficient estimate of temperature, we also consider IHM-P, which uses the iterative hotspot mitigation algorithm guided by peak power consumption instead of peak temperature. More specifically, during each iteration, the task with the highest power consumption instead of the highest temperature is chosen for voltage reduction.

Table 2 shows the 3D MPSoC peak temperature and power characteristics produced by PBMCA, IHM-P, and 3D-Wave. We will explain the implications of the data in this table in the following sections.

4.2.1 Power Optimization

3D MPSoC peak temperature is roughly related to chip average power consumption. If power density variation is neglected, MPSoC temperature is linearly proportional to average power consumption. PBMCA uses path-based slack distribution as well as voltage and frequency scaling to minimize chip power consumption. As shown in Table 2, compared to initial workload average power consumption (column 2), PBMCA can effectively reduce chip average power consumption (column 3). Among these ten benchmarks, PBMCA can reduce chip average power consumption by 23.4% on average and 29.3% at most.

4.2.2 Spatial Thermal Profile Balancing

The temperature at any time and position in a 3D MPSoC is strongly influenced by spatial and temporal variations in power density. PBMCA uses energy-aware task assignment to balance 3D MPSoC spatial power profile. In Table 2, the column labeled "Var." shows the maximum difference among the average power consumption of the eight processor cores. PBMCA can balance 3D MPSoC spatial power profile and constrain the maximum inter-core average power variation to 11.2%.

4.2.3 Temporal Thermal Profile Balancing

In Table 2, the three columns labeled "Temp." show the peak 3D MPSoC temperatures produced by PBMCA, IHM-P, and 3D-Wave. Note the corresponding average power consumptions in the columns labeled "Power". Although these three algorithms produce similar power consumptions, considering temporal variations in power density permits further temperature reduction. IHM-P balances temporal power profile. 3D-Wave uses dynamic thermal analysis to guide the iterative hotspot mitigation algorithm. It is more effective because its optimization moves are guided by accurate dynamic temperature estimates instead of task power consumptions, which permit only rough estimates of temperature.

Table 2: Results for PBMCA, IHM-P, and 3D-Wave

Benchmark	PBMCA				IHM-P		3D-Wave	
Num.	Power (W)	Power (W)	Var. (%)	Temp. (K)	Power (W)	Temp. (K)	Power (W)	Temp. (K)
1	72.9	57.3	9.9	448.9	55.3	367.6	55.7	364.7
2	83.3	62.4	9.6	391.0	61.0	374.4	60.0	363.6
3	86.4	61.1	8.9	393.4	60.9	372.6	60.3	363.6
4	70.4	58.9	9.8	388.5	56.9	382.3	58.4	365.8
5	72.4	54.0	8.8	377.9	53.5	368.0	53.8	364.4
6	82.4	62.3	8.7	389.1	60.5	372.4	61.0	366.2
7	74.2	59.3	7.6	384.0	58.4	375.5	59.0	368.7
8	96.6	72.6	7.6	395.9	64.0	378.8	70.0	376.7
9	72.3	56.1	11.2	391.6	54.3	369.0	55.4	364.7
10	80.6	60.3	11.0	382.2	59.3	368.9	59.9	365.4

Figure 3 shows the 3D MPSoC run-time power consumption and peak thermal profile produced by PBMCA, IHM-P, and 3D-Wave. Due to space limitations, Figure 3 only includes the results of three benchmarks (Benchmark 1, 6, and 10). The other benchmarks show a similar trend. These figures demonstrate the impact of temporal power variation on 3D MPSoC peak temperature and the effectiveness of the temporal power balancing and peak temperature mitigation algorithm used by 3D-Wave. Compared to PBMCA, 3D-Wave can reduce 3D MPSoC peak temperature by 27.9 °C on average.

Figure 4 demonstrates the iterative thermal optimization process of the dynamic thermal analysis driven iterative hotspot mitigation (IHM) algorithm used in 3D-Wave. For comparison, this figure also shows the power-driven iterative optimization alternative called IHM-P. We make the following observations. First, IHM is effective in detecting and eliminating temporal thermal hotspots and optimizing 3D MPSoC peak temperature. Second, Figure 4 demonstrates that the temperature reduction process of IHM is not monotonic. To eliminate local hotspots, IHM adjusts the slack distribution, voltage, and frequency assignments of the tasks responsible for the hotspot as well as its immediate neighbors. Slack reduction in neighboring tasks may introduce new hotspots. However, the iterative slack mitigation process then mitigates these newly-introduced hotspots until convergence, i.e., a condition in which any average power consumption reduction of the tasks responsible for the highest MPSoC temperature results in the introduction of an equally-high temperature elsewhere.

A comparison of IHM-P and 3D-Wave demonstrates the need for thermal analysis during hotspot mitigation instead of merely considering task power consumption. 3D-Wave permits a lower MPSoC peak temperature than IHM-P.

5. CONCLUSIONS

High chip power density and temperature complicate 3D MPSoC design. In this paper, we have described 3D-Wave, a thermal optimization algorithm for 3D MPSoC design. 3D-Wave uses a two-stage optimization flow consisting of (1) a constructive power balancing and minimization algorithm and (2) a thermal analysis driven iterative hotspot mitigation algorithm. Experimental results indicate that 3D-Wave can reduces 3D MPSoC peak temperature below that of similar techniques that optimize power, alone. To the best of our knowledge, this is the first system-level thermal optimization algorithm for 3D MPSoCs.

6. ACKNOWLEDGMENTS

We would like to acknowledge Zhenyu Gu at Northwestern University and Changyun Zhu at Queen's University for helpful suggestions and the extraction of power traces that were used for the benchmarks in Section 4.

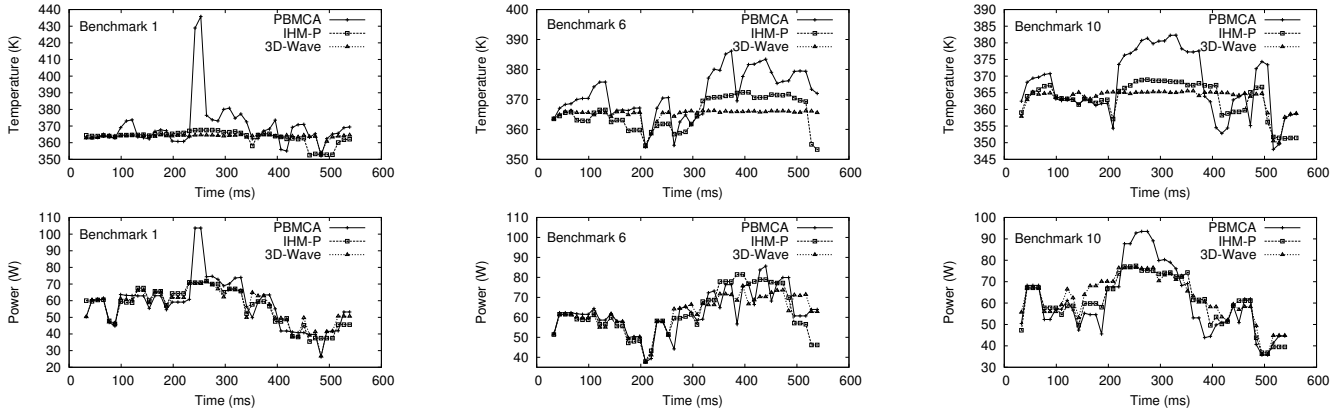

Figure 3: Comparison of different optimization heuristics.

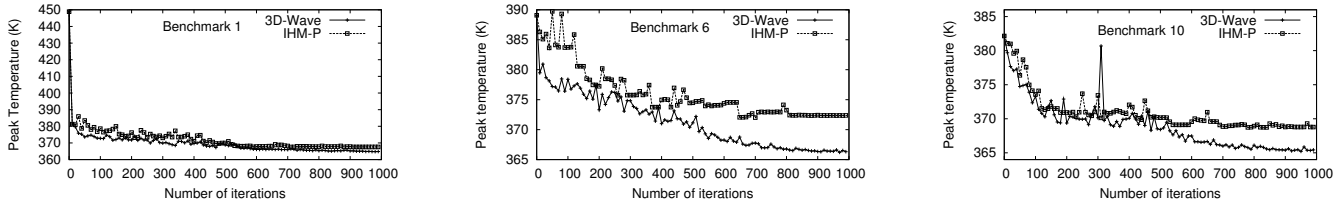

Figure 4: Comparison of 3D-Wave and iterative hotspot mitigation (IHM-P) temperature minimization.

7. REFERENCES

[1] E. Milchman, "Intel dual-core FAQ," *Wired News*, July 2006.

[2] S. Y. Borkar, et al., "Platform 2015: Intel processor and platform evolution for the next decade," Intel Corporation, Tech. Rep., Mar. 2005.

[3] W.-L. Hung, et al., "Interconnect and thermal-aware floorplanning for 3D microprocessors," in *Proc. Int. Symp. Quality of Electronic Design*, Mar. 2006, pp. 98–104.

[4] A. W. Topol, et al., "Three-dimensional integrated circuits," *IBM J. Research and Development*, vol. 4, 2006.

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

[6] Samsung, http://www.samsung.com/PressCenter/PressRelease/PressRelease.asp?seq=20060413_0000246668.

[7] Tezzaron, http://www.tezzaron.com/technology/FaStack.htm.

[8] W. M. Bulkeley, "IBM touts breakthrough in 3-d chips," *The Wall Street J.*, Apr. 2007.

[9] A. Jerraya, H. Tenhunen, and W. Wolf, "Multiprocessor systems-on-chips," *IEEE Computer*, vol. 38, no. 7, pp. 36–40, July 2005.

[10] D. Lyonnard, et al., "Automatic generation of application-specific architectures for hetereogeneous multiprocessor system-on-chip," in *Proc. Design Automation Conf.*, June 2001, pp. 518–523.

[11] G. Qu and M. Potkonjak, "System synthesis of synchronous multimedia applications," *ACM Trans. Embedded Computing Systems*, pp. 74–97, Feb. 2003.

[12] T. Givargis, F. Vahid, and J. Henkel, "System-level exploration for Pareto-optimal configurations in parameterized systems-on-a-chip," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2001, pp. 25–30.

[13] C. Lee and S. Ha, "Hardware-software cosynthesis of multitask MPSoCs with real-time constraints," in *Proc. Int. Conf. ASIC*, Oct. 2005, pp. 919–924.

[14] M. T. Schmitz, B. M. Al-Hashimi, and P. Eles, "Energy-efficient mapping and scheduling for DVS enabled distributed embedded systems," in *Proc. Design, Automation & Test in Europe Conf.*, Feb. 2002, pp. 514–521.

[15] A. Mishra and P. Banerjee, "An algorithm-based error detection scheme for the multigrid method," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 52, no. 9, pp. 1089–1099, Sept. 2003.

[16] J. Hu, et al., "Architecting voltage islands in core-based system-on-a-chip designs," in *Proc. Int. Symp. Low Power Electronics & Design*, Aug. 2004, pp. 180–185.

[17] K. Skadron, et al., "Temperature-aware microarchitecture," in *Proc. Int. Symp. Computer Architecture*, June 2003, pp. 2–13.

[18] P. Li, et al., "Efficient full-chip thermal modeling and analysis," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2004, pp. 319–326.

[19] Y. Yang, et al., "ISAC: Integrated Space and Time Adaptive Chip-Package Thermal Analysis," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Jan. 2007.

[20] R. Mukherjee, S. O. Memik, and G. Memik, "Temperature-aware resource allocation and binding in high-level synthesis," in *Proc. Design Automation Conf.*, June 2005.

[21] Z. P. Gu, et al., "TAPHS: Thermal-Aware Unified Physical-Level and High-Level Synthesis," in *Proc. Asia & South Pacific Design Automation Conf.*, Jan. 2006, pp. 879–885.

[22] G. Paci, et al., "Exploring "temperature-aware design" in low-power MPSoCs," in *Proc. Design, Automation & Test in Europe Conf.*, Mar. 2006.

[23] Y. Xie and W.-L. Hung, "Temperature-aware task allocation and scheduling for embedded multiprocessor systems-on-chip (MPSoC) design," *J. VLSI Signal Processing*, vol. 45, no. 3, pp. 177–189, Dec. 2006.

[24] J. Cong, J. Wei, and Y. Zhang, "A thermal-driven floorplanning algorithm for 3D ICs," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2004, pp. 306–313.

[25] B. Goplen and S. Sapatnekar, "Efficient thermal placement of standard cells in 3D ICs using a force directed approach," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2003, pp. 86–89.

[26] Y. Liu, et al., "Accurate Temperature-Dependent Integrated Circuit Leakage Power Estimation is Easy," in *Proc. Design, Automation & Test in Europe Conf.*, Mar. 2007, pp. 204–209.

[27] K. A. Bowman, et al., "A physical alpha-power law MOSFET model," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1410–1414, Oct. 1999.

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