Man Prakash Gupta
G. W. Woodruff
School of Mechanical Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332
e-mail: mp.gupta@gatech.edu

Minki Cho
e-mail: mcho8@gatech.edu

Saibal Mukhopadhyay
e-mail: saibal@ece.gatech.edu

School of Electrical and Computer Engineering,
Georgia Institute of Technology,
Atlanta GA, 30332

Satish Kumar
G. W. Woodruff
School of Mechanical Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332
e-mail: satish.kumar@me.gatech.edu

Thermal Investigation Into Power Multiplexing for Homogeneous Many-Core Processors

In this paper, a proactive thermal management technique called "power multiplexing" is explored for many-core processors. Power multiplexing involves redistribution of the locations of active cores at regular time intervals to obtain uniform thermal profile with low peak temperature. Three different migration policies namely random, cyclic, and global coolest replace have been employed for power multiplexing and their efficacy in reducing the peak temperature and thermal gradient on chip is investigated. For a given migration frequency, global coolest replace policy is found to be the most effective among the three policies considered as this policy provides 10 °C reduction in peak temperature and 20 °C reduction in maximum spatial temperature difference on a 256 core chip. Power configuration on the chip is characterized by a parameter called "proximity index" which emerges as an important parameter to represent the spatial power distribution on a chip. We also notice that the overall performance of the chip could be improved by 10% using global multiplexing. [DOI: 10.1115/1.4006012]

Keywords: many-core processors, thermal management, power multiplexing, core-hopping, migration policy

1 Introduction

In the field of computing, we have already witnessed the critical transition from single core processors to multicore processors in the last 5–6 years to address the growing demands of higher performance and faster computing. This trend has been so steep that the number of cores on a single die, that are commercially available in the market, has already reached up to six on desktop CPUs (∼Intel and AMD chips) [1,2]. Server and workstation processors have even higher number of cores per die. Graphic chips by NVIDIA and AMD already have hundreds of parallel processing units on a single die [3]. The current multicore architectures and programming models are suitable for 2–32 core processors but with the strong potential of parallel computing, the transition from multicore to many-core is imminent where the number of cores on a single chip is expected to reach in hundreds or even thousands per single processor die [4–6]. Such large-scale integration and very high power densities will bring a significant challenge of heat dissipation which is likely to act as first order constraint for many-core chip design. The traditional air-cooling devices begin to reach their flow and acoustic limits for very high power density (∼1.5 W/mm^2) apart from being highly inefficient from economic point of view when applied to many-core technology [7–10].

Spatiotemporal nonuniformity in the thermal field on chip due to uneven workload distribution among the cores is detrimental to both performance and reliability [11]. The leakage power increases exponentially with temperature resulting in higher power dissipation and cooling costs [12–14]. Proximity of high temperature zones (∼hotspots) on chip affects peak temperature and it needs to be optimized to improve thermal performance (rise in temperature per unit power) of multicore processors [15]. For a small feature size (∼15 nm node technology) chip, thermal coupling between the neighboring cores is highly pronounced and leads up to 65% temperature overhead [12]. Hence, spatially optimized power dissipation on chip becomes very important for many-core processors. A uniform on-chip temperature distribution and low peak temperature can be obtained by efficient heat redistribution techniques which in turn can improve energy-efficiency and coefficient of performance (∼compute/cooling power) [16,17]. This brings new opportunities for the dynamic thermal management (DTM) techniques and their role to address the new challenges of power dissipation issues for many-core processors becomes very critical. Many dynamic thermal management techniques have been explored such as clock gating, dynamic voltage and frequency scaling, and thread migration for single and multicore processors [11,18–23]. All these methods have power and performance overheads apart from the hardware and software implications.

The DTM techniques mentioned above are based on the "reactive" approach which rely on either reducing amount of dissipated energy or redistributing the energy over chip area (thread migration) only when the chip temperature rises above the stipulated temperature threshold. The temperature threshold could be set arbitrarily low to control the migration frequency in a traditional DTM technique. However, the peak temperature on chip may reach the threshold value frequently during moderate to high workload, and hence these techniques can have the adverse effects on the performance. The peak and average temperature on the chip can be lowered by a proper use of otherwise idle or underutilized cores, if the workload is redistributed among all the cores at regular intervals instead of waiting for the peak temperature to cross threshold value and then apply a DTM technique. This approach is defined as "proactive" thermal management. These proactive methods can be utilized as a supplementary approach to the reactive methods for effective thermal management of many-core processors. Our previous work suggested that significant reduction in peak temperature and higher thermal uniformity on a many-core chip can be achieved using power multiplexing technique [17,24]. To the best of our knowledge, no detailed investigation into power multiplexing has been performed to explore and understand the thermal physics behind the application of this technique for many-core processors. The insights from the thermal physics analysis could form the basis of designing effective migration policies.

In the present work, a detailed 3D thermal model of an electronic package and attached cooling devices has been developed

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 20, 2011; final manuscript received October 24, 2011; published online April 30, 2012. Assoc. Editor: Kenneth Goodson.

to explore the thermal response of a homogeneous 256-core processor chip, while investigating and comparing three power multiplexing policies: random, cyclic, and global coolest replace. The effect of migration frequency on peak temperature and thermal profile on chip, and limits of thermal performance using different policies have been investigated. Finally, performance improvement at different multiplexing frequencies is discussed, and an index is presented to characterize and understand the on-chip power configurations obtained during application of different migration policies.

2 Thermo-Fluidics System and 3D Modeling

In order to accurately analyze the effect of power multiplexing on on-chip thermal profile, a detailed 3D fluidics and thermal modeling of an electronic package and attached cooling system has been performed using finite volume method based commercial solver ANSYS FLUENT [3]. The computational domain is comprised of a flow tunnel, a heat sink, a heat spreader, the thermal interface material (TIM), a chip, and a substrate (Fig. 1). Approximately 450 K hexahedral cells are considered for the electronic package; grid independence tests with approximately 1000 K cells show less than 0.5 K change in chip temperature and verify that these cells are sufficient for further simulations. The properties of the various components of the system are listed in Table 1. The dimensions of chip are 12 mm × 0.5 mm × 12 mm (along x, y, and z directions) and the typical size of a grid cell inside chip is 0.375 mm × 0.1 mm × 0.375 mm. A predictive tile-type homogeneous 256-core processor is considered, where the cores are arranged in a 16 × 16 2D array and each core is assumed to have its own local cache operating at 3 GHz clock frequency. The total power dissipation on the chip is considered to be 128 W. The power dissipation value has been selected based on the prediction by International Technology Roadmap of Semiconductor for 16 nm node technology [25]. Our model considers 2 W of power dissipation in each active core which is reasonable for cores with 16 nm node technology running at 3 GHz. A detailed discussion of the core power estimation can be found in Ref. [24].

A uniform velocity profile at the inlet of the air flow tunnel is considered with constant velocity of 5 m/s. An outflow boundary condition is imposed at the outlet of the tunnel and no-slip boundary condition is imposed at all four walls of the tunnel and outer surfaces of the electronic package (Fig. 1(*a*)). The flow inside the tunnel is turbulent as Reynolds number based on the inlet flow rate and tunnel width is greater than 20,000. As accurate turbulent flow computations are not critical in the present study, we use Spalart–Allmaras turbulence model [26] which is a simple one-equation model and appropriate for applications involving wall-bounded flows and for avoiding fine meshing near the wall. We consider SIMPLE scheme for pressure–velocity coupling, implicit scheme for transient formulation, and second order upwind scheme for the discretization of all governing equations [27]. The characteristic timescale (τ) of the system, defined as the thermal diffusion time from the chip to ambient, is used to normalize temporal variables and parameters. We estimated that this timescale is approximately 0.1 s.

Table 1 Properties of the components of the system

Component	Material	ρ (kg/m^3)	cp (J/kg K)	k (W/mK)
Heat sink	Copper	8978	381	387.6
Heat spreader	Aluminum	2719	871	202.4
TIM	Grease	2550	700	4
Chip	Silicon	2330	712	141.2

3 Power Configurations and Multiplexing

Power configuration is defined as a particular distribution of power dissipating "active" cores on the chip. For a fixed number of active cores, the different permutations of the locations of active cores lead to different power configurations which in turn yield different temperature profiles and which have direct effect on the size, location, and the peak temperature of the hot spots on the chip. Power multiplexing offers an approach that can be utilized for the dynamic thermal management of many-core processors. This technique involves the change in power configuration of chip at the specified time intervals to spread the power envelope on the chip. The guiding rules which drive the change in power configuration are referred to as the migration policy. The specified time interval at which this migration takes place is termed as timeslice. A smaller timeslice corresponds to faster multiplexing. Timeslice is typically chosen such that it is smaller than the characteristic time scale (τ) of the system. Here, τ is defined as the thermal diffusion time from the chip to ambient. This criterion for the timeslice selection is based on the requirement that the 2D effects of power multiplexing need to be realized faster than the 3D thermal diffusion in order to get full advantage of multiplexing. It can be argued that power multiplexing will not be useful when the computational workload on the processor is 100%. But, typically only a fraction of the total number of cores is used for the compute work and rest are free of workload such that they can take up the workload from the busy cores during migration. The workload may differ from one core to the other even when all the cores are active. So, the power dissipated by different cores would not be the same. This scenario would again allow power multiplexing to play a useful role as the cores with high power dissipation can exchange the workload with the cores having low power dissipation. Thus, overall this approach is expected to play a useful role in all workload situations. For the most of our analysis, we have considered partial workload of 25% such that only 64 out of 256 cores are active (power dissipating) at a given time. Three migration policies explored in the present study are described next.

Random Migration Policy. According to this policy, an arbitrary set of cores is activated at each migration step, but the total

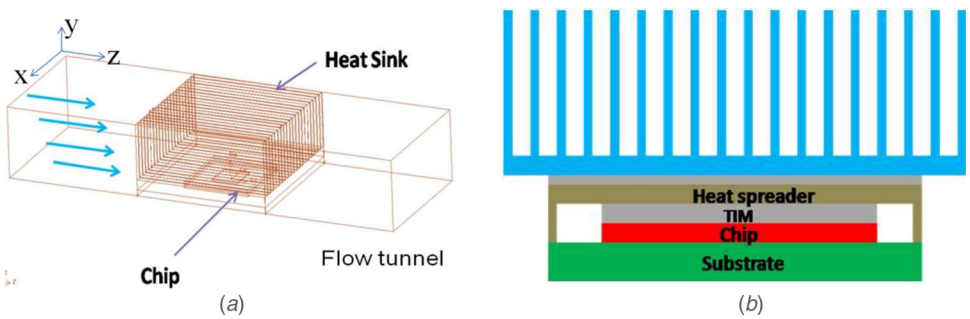

Fig. 1 (*a*) Flow tunnel with a heat sink and an electronic package used for the thermal modeling. (*b*) Schematic of the heat sink and electronic package of the multicore processor which includes heat spreader, TIM, chip, and substrate (view along the direction of inlet flow).

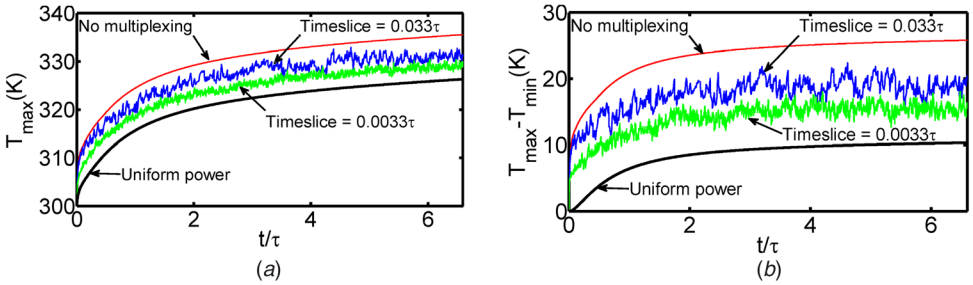

Fig. 2 Effect of timeslice variation on (*a*) peak temperature and (*b*) spatial temperature difference for random multiplexing. The top and bottom curves in both figures correspond to no multiplexing and uniform power, respectively. 25% cores are considered to be active with total power = 128 W.

number of active cores remains constant. This policy involves random redistribution of all active cores at each migration step. So, it might be the most difficult policy from the implementation perspective.

Cyclic Migration Policy. For the cyclic policy, the 256 cores on the chip are grouped into smaller blocks of 2×2 cores. The active cores are assigned in a checkerboard configuration and shifted in a circular fashion at each timeslice maintaining checkerboard configuration. This policy also requires redistribution of all cores at each migration step. However, the level of complexity is less than that of the random policy as migration of cores is predefined.

Global Coolest Replace Migration Policy. The basic working principle behind the global coolest replace policy (also referred as global policy) is to exchange the workload from the hottest cores to the coolest ones at each timeslice. The "global coolest replace" is a semiproactive policy as it requires information about the instantaneous temperature of all cores at each migration step. This policy may not require migration of all cores at each migration step.

4 Results and Discussion

4.1 Effect of Migration Policies

Random Policy. We study four cases to illustrate the impact of random multiplexing and timeslice variation on peak temperature and thermal profile: (i) fixed random power configuration, (ii) random multiplexing with timeslice $= 0.033\tau$, (iii) random multiplexing with timeslice $= 0.0033\tau$, and (iv) uniform power distribution. First three cases represent without multiplexing, slower multiplexing, and faster multiplexing, respectively. The fourth case is a reference for the other cases and it can help compare thermal effects of random power multiplexing for different timeslices. The timeslices 0.033τ and 0.0033τ correspond to 10,000 K and 1000 K clock cycles, respectively, considering 3 GHz as operating frequency of the chip.

Results indicate that the random power multiplexing reduces the peak temperature and brings more uniformity in the thermal profile of the chip. The peak temperature reduction and uniformity of thermal profile depend on the timeslice. Faster multiplexing accompanies with higher reductions in the peak temperature (T_{max}) (Fig. 2(*a*)) and the maximum spatial temperature difference ($T_{max} - T_{min}$) (Fig. 2(*b*)). The reductions in T_{max} and ($T_{max} - T_{min}$) at the chip reach toward 10 °C and 15 °C, respectively, as we apply extremely fast (i.e., very small timeslice) random multiplexing. However, it should be noted that the decrease in the timeslice adversely affects the overall chip performance due to migration overhead, and hence, there is a trade-off which requires significant attention while selecting suitable timeslice. A graphic comparison of the thermal profile on the chip at time instant, $t = 6.6\ \tau$, for the cases (i), (ii), and (iii) is shown in Fig. 3 which shows that faster the multiplexing higher the uniformity in thermal profile on the chip.

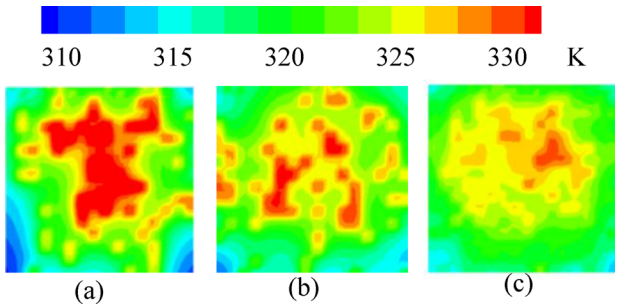

Fig. 3 Thermal profile on 256 core chip at time instant, $t/\tau = 6.6$, for (*a*) no multiplexing, (*b*) multiplexing with timeslice $= 0.033\tau$, and (*c*) multiplexing with timeslice $= 0.0033\tau$ using random core migration policy. 25% active cores with total power = 128 W.

Cyclic Policy. We again consider four cases: (i) fixed checkerboard power configuration, (ii) cyclic multiplexing with timeslice $= 0.033\tau$, (iii) cyclic multiplexing with timeslice $= 0.0033\tau$, and (iv) uniform power distribution. The relevance of these four cases is same as that explained for random policy.

Results indicate that the cyclic policy reduces the peak temperature (Fig. 4(*a*)) but the temperature reduction is only 3 °C even for the limiting case (uniform power distribution) of this policy. This small reduction can be attributed to the pre-existing checkerboard configuration. The spatial temperature difference across the chip is significantly lowered by the cyclic multiplexing (Fig. 4(*b*)). The maximum reduction in ($T_{max} - T_{min}$) is about 7 °C which is substantial compared to the peak temperature reduction. A graphic comparison of the effect of timeslice on multiplexing is shown in Fig. 5. For a fixed configuration, small hotspots tend to show up but thermal profile becomes more uniform as cyclic multiplexing becomes faster.

Global Coolest Replace Policy. Global policy is intrinsically different from the previous two policies in two ways. First, it takes decisions based on the instantaneous chip temperature. Second, fewer active cores are involved during the multiplexing at each timeslice. The two important parameters for this policy that may affect the thermal profile on chip are the timeslice and the number (N) of hot cores that are swapped with the equal number of cool cores. Please note that here the swapping of cores means swapping of workload on the respective cores.

In order to analyze the effect of timeslice we consider five cases here (i) fixed random power configuration, (ii) multiplexing with timeslice $= 0.33\tau$, (iii) multiplexing with timeslice $= 0.033\tau$, (iv) multiplexing with timeslice $= 0.0033\tau$, and (v) uniform power distribution on chip. First four cases represent no multiplexing, slow, medium, and fast multiplexing, respectively. The last case is just a reference case and unlike the previous two policies, it does not represent an extremely fast multiplexing for the global policy.

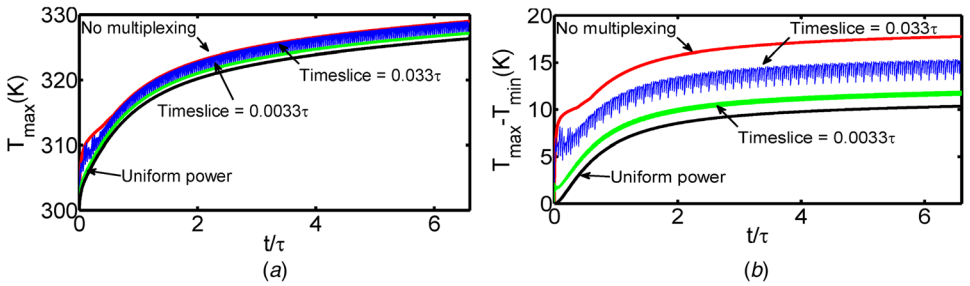

Fig. 4 Effect of timeslice variation on (*a*) peak temperature and (*b*) spatial temperature difference for cyclic multiplexing. 25% cores are considered to be active with total power = 128 W.

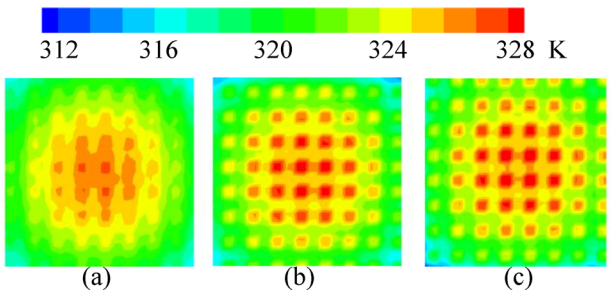

Fig. 5 Thermal profile on a chip for cyclic policy at $t = 6.6\tau$ for (*a*) timeslice = 0.0033τ, (*b*) timeslice = 0.033τ, and (*c*) no change in configuration. Higher spatial thermal uniformity can be observed for high frequency of multiplexing. 25% active cores with total power = 128 W.

Here, we keep $N = 1$ for all the three timeslices which means that only a pair of cores are involved during the multiplexing at each timeslice.

Global policy shows significant improvement in thermal profile even for the timeslices which are greater than the characteristic time (τ). This behavior underscores the intrinsic intelligent nature of the policy which is further substantiated by the fact that the peak temperature reduction of about 10 °C is observed for the slow multiplexing (timeslice = 0.33τ) at $t = 6.6\tau$ (Fig. 6(*a*)). The decrease in timeslice has very little impact on the peak temperature reduction after $t = 4\tau$.

A significant reduction (~20 °C) in maximum spatial temperature difference (T_{max}–T_{min}) is observed for fast multiplexing (Fig. 6(*b*)) which remains almost constant in time. This response of global multiplexing is very encouraging since peak temperature is reduced almost by the same amount as that obtained in uniform power distribution case even for slow multiplexing (timeslice = 0.33τ). Moreover, the reduction in maximum spatial temperature difference for the fast global multiplexing (timeslice = 0.0033τ) surpasses the reference case by almost 4 °C. A high degree of thermal uniformity (Fig. 7) can be achieved using global multiplexing even for a very slow multiplexing (timeslice = 0.33τ) and further decrease in timeslice to 0.033τ does not lead to much difference in the uniformity of the thermal profile on chip.

This behavior is very important from policy formation perspective as it suggests that mere uniform distribution of power may not yield the optimal results for thermal profile. By analyzing the power map at each migration step, we find that the global coolest policy ingeniously places the active cores away from the center of the chip such that it not only reduces peak temperature by a significant amount but also reduces thermal nonuniformity. The results heavily advocate the strength of global policy.

In the previous paragraphs, we discussed the effect of timeslice for $N = 1$, i.e., only one pair of cores (the hottest and the coolest) are involved during global multiplexing. We study the effect of variation in N considering two cases (i) $N = 1$, and (ii) $N = 4$ for timeslice of 0.33τ and 0.033τ. Intriguingly, changing N does not bring any significant change in the peak temperature variation for a given timeslice (Fig. 8(*a*)). However, high fluctuations are observed for slow multiplexing (timeslice = 0.33τ) especially for $N = 4$. A possible reason for this behavior could be the proximity of the coolest cores which are to be swapped with the hottest cores giving rise to higher peak temperature as the hottest cores lie close to each other even after the migration. In Fig. 8(*b*), we show the effect of N on variation of spatial temperature difference. Here, we also observe that the spatial temperature difference remains almost constant in time for both $N = 1$ and $N = 4$ for faster multiplexing (timeslice = 0.033τ). However, we notice greater fluctuations for slow multiplexing (timeslice = 0.33τ); the amplitude of the fluctuations increases as N is increased. It can be inferred from the results that $N = 1$ is sufficient for global policy to achieve favorable thermal profile.

4.2 Comparison of Policies. So far we have discussed the thermal performance of different policies and studied the effect of variation of different relevant parameters. The three policies discussed above have one common feature which is the simplicity of the guiding principles for core migrations. However, these policies significantly differ from the implementation and thermal performance perspectives. Each policy has its own advantages and

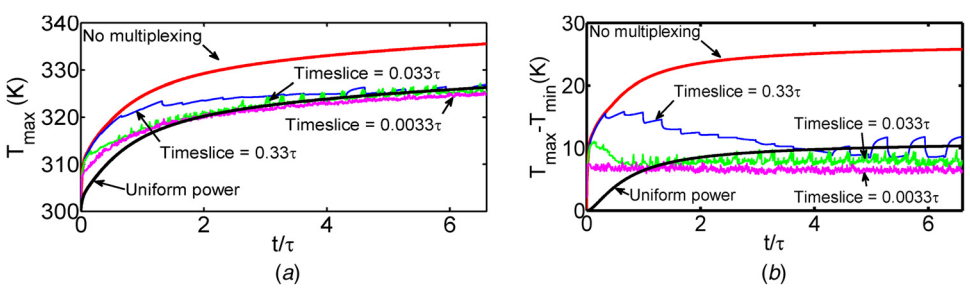

Fig. 6 Effect of timeslice variation on (*a*) peak temperature and (*b*) spatial temperature difference for global multiplexing. 25% cores are considered to be active with total power = 128 W.

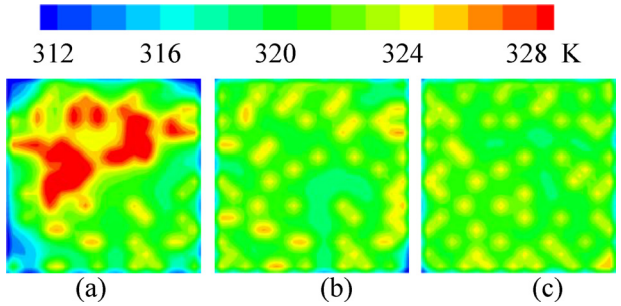

Fig. 7 Thermal profile on a chip for global coolest replace policy at $t/\tau = 6.6$ for (*a*) no multiplexing, (*b*) timeslice = 0.33τ, and (*c*) timeslice = 0.033τ. Significant improvement in the uniformity of the thermal profile can be observed from case (a) to case (b). 25% active cores with total power = 128 W.

limitations. We enlist a few categories under which these policies can be compared.

The Thermal Effect. Thermal performance of different policies is adjudged here by the degree of reductions in the peak temperature and the spatial temperature difference for a fixed timeslice = 0.033τ. Figure 9(*a*) shows peak temperature variation with time for different policies with few special cases. Cyclic policy shows better performance compared to random policy but global coolest replace policy shows much better thermal performance compared to the random and cyclic policies in terms of higher peak temperature reduction and better thermal uniformity on the chip (Fig. 9(*b*)). A graphic comparison of the thermal profile on chip can be seen in Fig. 10 for the three policies and it can be noticed that global policy easily outperforms the other policies.

Implementation. Random policy would require a simple random number generator to identify next set of cores at each migration step which is considerably an easy option from the implementation perspective. Cyclic policy rotates the arrays of active cores in cyclic fashion; this policy can also be implemented without much complexity since the migration sequence is predefined and simple. In contrast, global coolest policy requires the temperature measurement at multiple points on chip in real time as it needs a sorted list of cores based on the instantaneous temperature. Therefore, fast on-chip thermal sensors will be an important requirement for this policy.

Migration Traffic. Random and cyclic policies involve migration of all cores at each migration step. In random policy, this active core migration could be highly complex as power configuration changes randomly from one set of active cores to another. In cyclic case, this migration is more systematic but nonetheless all the cores are involved in the migration. Global coolest policy has great advantage in this aspect since migration of only a pair of cores may be sufficient to get best reduction in peak temperature and uniformity in temperature profile.

4.3 Proximity Analysis. We already discussed that the peak temperature may vary depending upon the power configuration. This means that each power configuration may have some characteristic which can be directly related to the peak temperature. We define one such characteristic named "proximity index" which can be calculated by adding up the relative distances between the active cores.

$$\text{Proximity index} = \sum_{i,j} 0.5 \times \left|\vec{r}_i - \vec{r}_j\right| \tag{1}$$

where $\vec{r}_i$ and $\vec{r}_j$ are the position vectors of the active cores such that $1 \le i,j \le 64$ as we consider 25% active cores. Proximity index represents the degree of proximity of active cores, i.e., higher proximity index corresponds to more sparsely located active cores. We observe that the peak temperature (T_{max}) decreases linearly with the proximity index under steady state conditions without multiplexing (Fig. 11). A band of T_{max} values has been observed for a given proximity index, which can be attributed to the finite size of the chip and also to the geometrical/thermal features of the 3D electronic package and the attached cooling solution. Despite these effects, the peak temperature is observed to be strongly correlated to the proximity index of the active cores which indicates that this index is an important metric to represent the thermal interaction of active cores.

The peak temperature behavior of the migration policies can also be understood using proximity index. The power configurations corresponding to the random policy cover the entire range of proximity index as any power configuration is probable due to the random re-arrangement of the active cores on the chip. All checkerboard configurations can be represented by a single point (solid ellipse in Fig. 11). These configurations have high proximity index and corresponds to low T_{max}. This is consistent with our analysis using cyclic multiplexing which shows that this policy does not yield much advantage in reducing T_{max} as initial checkerboard configurations have a good arrangement of active cores. The checkerboard configuration seems to be an ideal power configuration as active cores are placed alternatively in a uniform fashion, but we observe other power configurations obtained by global policy can lead to even lower T_{max} (see Fig. 11). The power configurations for the global policy (points in dashed ellipse) are obtained after employing global multiplexing for significantly long time. Global policy places the active cores in the checkerboard fashion near the edges, while keeping very few active cores near the center leading to an optimal power profile. This policy uses the current temperature distribution to decide the next power configuration and has the potential to include the effect of both geometrical and thermal properties of the 3D system and therefore lead to great improvement in the on-chip thermal profile. Peak temperature difference between the checkerboard and global configurations is about 2 °C as shown in Fig. 11. In transient case,

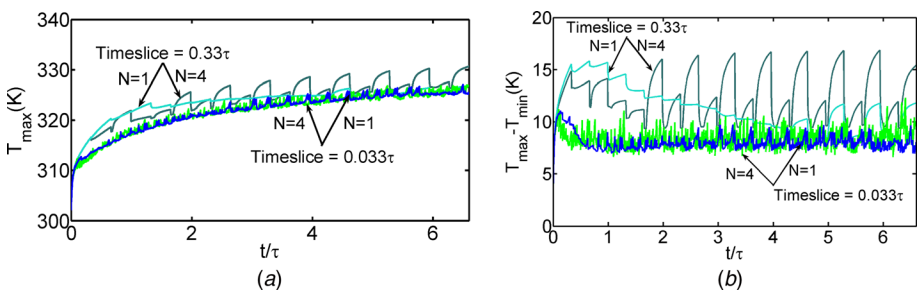

Fig. 8 Effect of variation in number of swapped cores (*N*) on (*a*) peak temperature and (*b*) spatial temperature difference for global multiplexing. 25% cores are considered to be active with total power = 128 W.

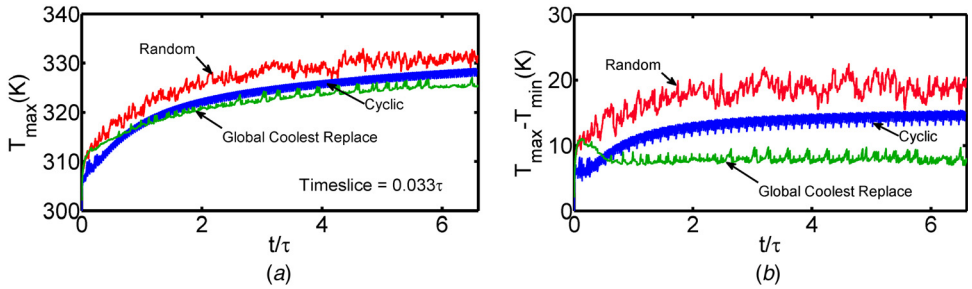

Fig. 9 Comparison of the effect of different migration policies on (*a*) peak temperature and (*b*) spatial temperature difference. Timeslice is kept as 0.033τ during power multiplexing. 25% cores are considered to be active with total power = 128 W.

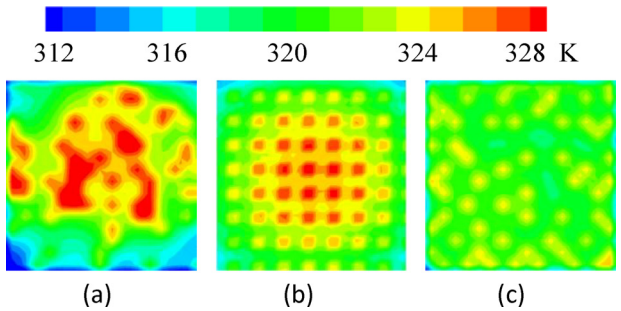

Fig. 10 Thermal profile on a chip at $t/\tau = 6.6$ for (a) random policy, (b) checkerboard, (c) global coolest replace. Timeslice is taken as 0.033τ for all cases. Very high spatial thermal uniformity can be seen for the global multiplexing. 25% active cores with total power = 128 W.

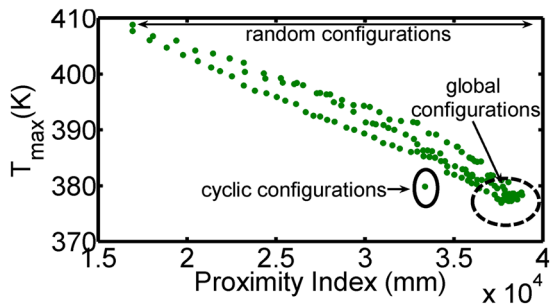

Fig. 11 Variation of peak temperature with proximity index under steady state conditions. Each point in the plot corresponds to a unique power configuration. 25% cores are considered to be active with total power = 128 W.

when power multiplexing is employed, the difference between the peak temperatures obtained by cyclic and global policies increases to 4 °C at $t = 6.6\tau$ for timeslice = 0.033τ; this difference will further increase with time during multiplexing. These results indicate that the effect of power multiplexing approach has two important components, (a) the first one is spatial which is related with the power configurations and (b) the second one is the transient which corresponds to the interaction of active cores with themselves and the rest of the 3D system due to the switching of power in time. Even though the proximity index turns out to be a relevant representation of the power configurations and hence the spatial effects, the spatial and transient effects are not decoupled to each other. The design of any effective power multiplexing needs to consider these spatial-temporal effects together indicating that other representative parameters similar to "proximity index" are probable [28–30].

4.4 Performance Overhead Analysis. As discussed earlier, power multiplexing improves the thermal profile on the chip which in turn allows operation in a higher frequency range. A higher clock frequency improves the performance for a constant number of execution cycles. However, multiplexing is also associated with additional time necessary for migration of the threads which negatively affects the performance. Therefore, in order to demonstrate the overall impact of power multiplexing on the performance, we combine the two effects and present a first order analysis. We define effective timeslice including migration time. We consider the migration time overhead as follows:

$$T_{\text{eff}} = T_{\text{timeslice}} + T_{\text{migrate}} \tag{2}$$

where T_{eff} is effective timeslice, $T_{\text{timeslice}}$ is originally defined timeslice and T_{migrate} is total time for power migration. As migration time increases, migration overhead increases and overall performance degrades more. On the other hand, a faster migration leads to reduced maximum chip temperature. It is well-known that a lower peak temperature results in higher operating speed of the logic circuits. For a many-core chip the operating frequency is determined by the slowest core. Hence, a lower peak temperature of the chip will result in a faster operating speed, i.e., higher clock frequency. A higher clock frequency has a positive impact on performance. Therefore, we define effective performance improvement (PI) for a given migration interval ($T_{\text{migration}}$) as

$$\text{PI} = \underbrace{\left(\frac{T_{\text{timeslice}}}{T_{\text{timeslice}} + T_{\text{migrate}}}\right)}_{T_{\text{norm}}} \times \underbrace{\frac{f_{\text{mod}}}{f_{\text{nom}}}}_{f_{\text{norm}}} \tag{3}$$

where f_{nom} is nominal operating frequency, f_{mod} is the modified operating frequency which reflects thermal field improvement, f_{norm} is the normalized operating frequency, and T_{norm} is the normalized timeslice with respect to effective timeslice. Note that T_{norm} represents the impact of migration on the number of cycles required to complete a given task, while f_{norm} represents the impact on frequency of each cycle. We have evaluated the impact of migration on the above mentioned performance improvement index. We have considered different migration intervals and estimated the peak temperature. The operating frequency of a ring-oscillator in 16 nm node was estimated at different temperatures using circuit simulation as described in Ref. [24] and used to compute f_{norm}. Next, we have used the experimental results presented in Ref. [31] to obtain a simple estimate of the number of cycles required to perform thread migration. This is next used to compute the T_{norm} considering different migration interval. It is expected that a faster migration will reduce T_{norm} but it will increase f_{norm}. Therefore, the combined effect shows interesting trend as presented in Fig. 12. Analysis shows that f_{norm} effect dominates for slow (0.33τ) and medium (0.033τ) power multiplexing which give 10% improvement in the overall performance. However, at very fast migration, although peak temperature reduces and f_{norm}

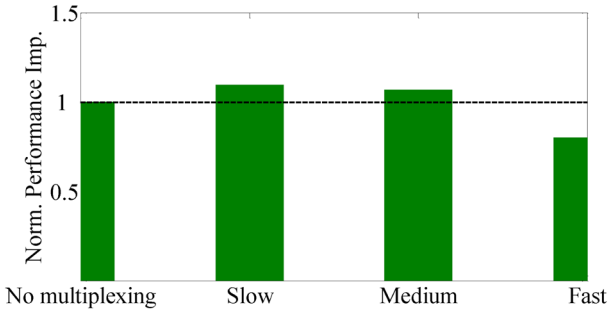

Fig. 12 Variation of normalized Performance Improvement with no multiplexing, slow (timeslice = 0.33τ), medium (timeslice = 0.033τ) and fast (timeslice = 0.0033τ) multiplexing

increases, the overall performance degrades due to high migration overhead.

5 Conclusion

In the present work, spatiotemporal power multiplexing has been analyzed as a prospective thermal management technique for many-core processors. The global coolest replace policy is found to be the most effective among the three policies discussed in the present work as the peak temperature reduction of 10 °C and the maximum spatial temperature difference ($T_{max} - T_{min}$) reduction of 20 °C is observed using global policy. This policy leads to the optimal power map required for the uniform thermal profile. A comparison between the three policies also suggests that the global policy may be more suitable from the implementation perspective as only a pair of cores is involved at each migration step during multiplexing. The proximity index is observed to be an important spatial parameter to characterize the power configurations on a chip. A simple performance analysis shows that overall 10% increase in performance of the chip can be achieved using power multiplexing. The current work may be considered as a first order analysis of migration policies as simple policies are applied in case of the homogeneous many-core processors. More evolved policies can be formulated to handle thermal management for heterogeneous many-core processors.

Acknowledgment

The authors acknowledge the support from National Science Foundation Grant ECCS-1028569. M. Cho and S. Mukhopadhyay would like to acknowledge Semiconductor Research Corporation (#2084.001), Intel Corp, and IBM Faculty Award for financial support.

Nomenclature

c_P = specific heat (J/kg K)
f = operating frequency of processor
k = thermal conductivity (W/m K)
N = number of swapped cores during global policy
$\vec{r}$ = position vector of active core
t = time (s)
T = temperature (K)

Greek Symbols

τ = characteristic time scale of the system (s)

Subscripts

max = maximum
min = minimum
mod = modified
nom = nominal
i, j = number index for active cores

References

[1] Intel, "Product Brief" http://download.intel.com/products/processor/ corei7EE/323307.pdf
[2] AMD, "Physical Cores V. Enhanced Threading Software: Performance Evaluation Whitepaper," http://www.amd.com/us/Documents/Cores_vs_Threads_Whitepaper.pdf
[3] http://www.ansys.com/products/fluid-dynamics/fluent
[4] Asanovic, K., Bodik, R., Catanzaro, B. C., Gebis, J. J., Husbands, P., Keutzer, K., Patterson, D. A., Plishker, W. L., Shalf, J., Williams, S. W., and Yelick, K. A., 2006, "The Landscape of Parallel Computing Research: A View from Berkeley," EECS Department, University of California, Berkeley, Technical Report No. UCB/EECS-2006-183.
[5] Garver, S.-L., and Crepps, B., 2009, "The New Era of Tera-Scale Computing," http://software.intel.com/en-us/articles/the-new-era-of-tera-scale-computing
[6] Yeh, D., Peh, L.-S., Borkar, S., Darringer, J., Agarwal, A., and Hwu, W.-M., 2008, "Thousand-Core Chips [Roundtable]," IEEE Des. Test Comput., **25**, pp. 272–278.
[7] Rodgers, P., Eveloy, V., and Pecht, M. G., 2005, "Limits of Air-Cooling: Status and Challenges," presented at IEEE Twenty First Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[8] Krishnan, S., Garimella, S. V., Chrysler, G. M., and Mahajan, R. V., 2007, "Towards a Thermal Moore's Law," IEEE Trans. Adv. Packag., **30**, pp. 462–474.
[9] Zhou, P., Hom, J., Upadhya, G., Goodson, K., and Munch, M., 2004, "Electro-Kinetic Microchannel Cooling System for Desktop Computers," presented at Twentieth Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[10] Wei, H., Stan, M. R., Gurumurthi, S., Ribando, R. J., and Skadron, K., 2010, "Interaction of Scaling Trends in Processor Architecture and Cooling," presented at 26th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[11] Mukherjee, R., and Memik, S. O., 2006, "Physical Aware Frequency Selection for Dynamic Thermal Management in Multi-Core Systems," presented at IEEE/ACM International Conference on Computer-Aided Design (ICCAD).
[12] Janicki, M., Collet, J. H., Louri, A., and Napieralski, A., 2010, "Hot Spots and Core-to-Core Thermal Coupling in Future Multi-Core Architectures," presented at 26th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[13] Srinivasan, J., Adve, S. V., Bose, P., and Rivers, J. A., 2004, "The Case for Lifetime Reliability-Aware Microprocessors," presented at 31st Annual International Symposium on Computer Architecture.
[14] Kursun, E., and Chen-Yong, C., 2009, "Temperature Variation Characterization and Thermal Management of Multicore Architectures," IEEE MICRO., **29**, pp. 116–126.
[15] Guoping, X., 2006, "Thermal Modeling of Multi-Core Processors," presented at The Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM).
[16] Wei, H., Skadron, K., Gurumurthi, S., Ribando, R. J., and Stan, M. R., 2010, "Exploring the Thermal Impact on Manycore Processor Performance," presented at 26th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[17] Gupta, M. P., Minki, C., Mukhopadhyay, S., and Kumar, S., 2010, "Thermal Mangament of Multicore Processors Using Power Multiplexing," presented at 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm).
[18] Rao, R., Vrudhula, S., and Chakrabarti, C., 2007, "Throughput of Multi-Core Processors Under Thermal Constraints," presented at ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED).
[19] Donald, J., and Martonosi, M., 2006, "Techniques for Multicore Thermal Management: Classification and New Exploration," presented at 33rd International Symposium on Computer Architecture (ISCA).
[20] Hanumaiah, V., Vrudhula, S., and Chatha, K. S., 2009, "Maximizing Performance of Thermally Constrained Multi-Core Processors by Dynamic Voltage and Frequency Control," presented at IEEE/ACM International Conference on Computer-Aided Design—Digest of Technical Papers (ICCAD).
[21] Chaparro, P., Gonzalez, J., Magklis, G., Cai, Q., and Gonzalez, A., 2007, "Understanding the Thermal Implications of Multi-Core Architectures," IEEE Trans. Parallel Distrib. Syst., **18**, pp. 1055–1065.
[22] Brooks, D., and Martonosi, M., 2001, "Dynamic Thermal Management for High-Performance Microprocessors," presented at The Seventh International Symposium on High-Performance Computer Architecture (HPCA).
[23] Zhigang, H., Buyuktosunoglu, A., Srinivasan, V., Zyuban, V., Jacobson, H., and Bose, P., 2004, "Microarchitectural Techniques for Power Gating of Execution Units," presented at International Symposium on Low Power Electronics and Design (ISLPED).
[24] Cho, M., Sathe, N., Gupta, M., Kumar, S., Yalamanchilli, S., and Mukhopadhyay, S., 2010, "Proactive Power Migration to Reduce Maximum Value and Spatiotemporal Non-Uniformity of On-Chip Temperature Distribution in Homogeneous Many-Core Processors," presented at 26th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM).
[25] International Technology Roadmap for Semiconductors (2008), http://www.itrs.net/
[26] Spalart, P., and Allmaras, S., 1992, "A One-Equation Turbulence Model for Aerodynamic Flows," American Institute of Aeronautics and Astronautics, Technical Report No. AIAA-92-0439.

[27] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publishing Corporation, Washington, DC/McGraw-Hill, New York.
[28] Mallows, C. L., 1972, "A Note on Asymptotic Joint Normality," Ann. Math. Statist., **43**, pp. 508–515.
[29] Czado, C., and Munk, A., 1998, "Assessing the Similarity of Distributions—Finite Sample Performance of the Empirical Mallows Distance," J. Statist. Comput. Simul., **60**, pp. 319–346.
[30] Rubner, Y., Tomasi, C., and Guibas, L. J., 1998, "A Metric for Distributions With Applications to Image Databases," presented at Sixth International Conference on Computer Vision.
[31] Jeonghwan, C., Chen-Yong, C., Franke, H., Hamann, H., Weger, A., and Bose, P., 2007, "Thermal-Aware Task Scheduling at the System Software Level," presented at ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED).