

Thermal-aware dynamic voltage frequency scaling for many-core processors under process variations

Hyejeong Hong, Jaeil Lim, Hyunul Lim, and Sungho Kang^{a)}

*School of Electrical and Electronic Engineering, Yonsei University,
50 Yonsei-ro, Seodaemun-gu, Seoul 120–749, Korea*

a) shkang@yonsei.ac.kr

Abstract: Process variations yield the asymmetry on core performance in many-core processors. Adaptive voltage scaling can hide the variations, but that results in the different thermal characteristics of cores. By using the thermal characteristics, the efficiency of energy optimization and temperature management can be improved. Experiments showed that the proposed dynamic voltage frequency scaling consumes up to 25.2% less energy than the existing thermal management technique while remaining the ratio of peak temperature violations under 1%.

Keywords: many-core, dynamic voltage frequency scaling, thermal-aware

Classification: Integrated circuits

References

- [1] V. Hanumaiah and S. Vrudhula: IEEE Trans. Comput. **61** [10] (2012) 1484.
- [2] M. Elgebaly and M. Sachdev: IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **15** [5] (2007) 560.
- [3] S. R. Sarangi, B. Greskamp, R. Teodorescu, J. Nakano, A. Tiwari and J. Torrellas: IEEE Trans. Semicond. Manuf. **21** [1] (2008) 3.
- [4] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan and D. Tarjan: Proc. Int. Symp. Computer Architecture (2003) 2.
- [5] T. Tobita and H. Kasahara: J. Scheduling **5** [5] (2002) 379.
- [6] M. Gomaa, M. D. Powell and T. N. Vijaykumar: ACM SIGOPS Operating Systems Review **38** [5] (2004) 260.

1 Introduction

The performance improvement of microprocessors has resulted in the increased power density thereby worsening thermal problems. The increase in temperature may degrade performance, raise cooling costs, and exacerbate reliability issues, such as negative bias temperature instability and hot carrier injection. A lot of studies have been conducted to mitigate the thermal stresses in processors, and most of them are combined with traditional power management techniques, such as dynamic voltage

frequency scaling (DVFS). The heat generation is mainly dependent on the power consumption so that, DVFS can also mitigate thermal problems [1].

Process variations become severe as technology scales down. They raise the inter-die and intra-die asymmetry on performance: the former is die-to-die (D2D), and the latter is within-die (WID) process variation. We consider homogeneous many-core processors, whose cores are identical. The performance of each many-core die may vary because of D2D variation. The performance of cores in a many-core processor may also vary due to WID process variations, which is resulted in core-to-core (C2C) process variation. Adaptive voltage scaling (AVS) is adopted to hide the variations, by scaling up the supply voltage V_{dd} of the core whose maximum frequency is lower than the others [2]. Consequently, the power consumption of the cores also varies and that affects the thermal characteristics of each core. Exploiting the thermal characteristics that are obtained statically can improve both the energy efficiency and temperature management.

2 Thermal-aware dynamic voltage frequency scaling

We assume homogeneous many-core processors. The required performance for each core is predefined, and AVS is assumed. Each core has its own supply voltage setting to support the predefined frequency setting, according to the level of process variations. Cores are prioritized through the static temperature analysis, and the priority is utilized in task mapping and dynamic voltage scaling.

2.1 Process variations and AVS

Process variation is a combination of random and systematic components. Two major process parameters which affected by process variations are the transistor threshold voltage V_t and the effective channel length L_{eff} . The standard deviation of the parameters, σ_{total} is modeled as

$$\sigma_{total} = \sqrt{\sigma_{rand}^2 + \sigma_{sys}^2}, \quad (1)$$

where σ_{rand} and σ_{sys} represents the random and systematic components respectively and they are assumed to equally impact on the total variation. The random process variation is caused by random doping effects. The systematic process variation is caused by the lack of accuracy in manufacturing processes, so it appears as spatial similarity. It can be modeled using a multivariate normal distribution with a spherical spatial correlation structure. This model was shown to match empirical data [3]. The correlation function of V_t , $\rho(r)$, is given as

$$\rho(r) = \begin{cases} 1 - \frac{3r}{2\phi} + \frac{r^3}{2\phi^3}, & r \leq \phi, \\ 0, & r > \phi \end{cases} \quad (2)$$

where r is the distance and ϕ is the predefined finite distance. The systematic variation of L_{eff} is considered to be half of that of V_t [3].

Fig. 1(a) illustrates the map of maximum core frequencies in a 16×16 many-core processor: a small square represents a core tile. The maximum frequencies range from 360 MHz to 620 MHz, as a result of process variations. There is a tendency that cores having similar maximum

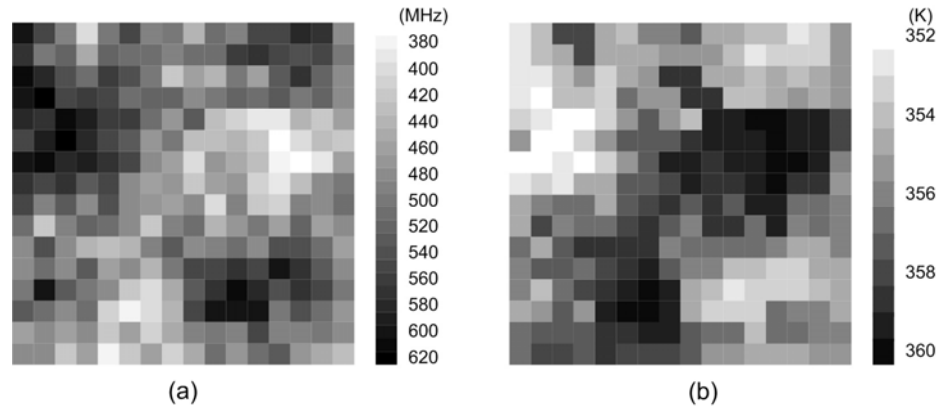


Fig. 1. The frequency map of 16×16 many-core and the corresponding temperature map. (a) The map of maximum frequency of cores varied due to process variations where σ_{total}/μ is 0.12. (b) The map of steady state temperature where AVS is adopted.

frequencies are contiguous due to the systematic components. To equalize the performance of cores, the supply voltage setting is selected for each core: slow cores only use some of higher supply voltages among all the supply voltages that the power delivery network supports, and fast cores use lower supply voltages. The transistor switching delay, t_d , is dependent on V_t and L_{eff} , which vary due to process variations. The varied delay can be estimated as

$$t_d \propto \frac{V_{dd}L_{eff}}{(V_{dd} - V_t)^\alpha} \quad (3)$$

where α is typically 1.3 [3]. Using Eq. (3), the appropriate supply voltage level for a specific frequency to hide the delay can be selected.

Once AVS is completed, the thermal characteristics of cores can be known. HotSpot [4] is used for the steady state temperature simulations. We assume that every core runs at the maximum frequency without any idle time. Although the frequency is the same, the supply voltage of each core varies due to AVS. Therefore, the power consumption of each core differs, consequently the heat generation differs. Fig. 1(b) shows the result of the steady state temperature simulation, corresponding to the process variation profile in Fig. 1(a). Although all cores run at the same frequency, the resulted temperatures differ. Therefore, these thermal characteristics should be statically analyzed and be considered while making dynamic decisions. The core with the lowest steady state temperature is given the top priority. Since the analysis also reflects the thermal correlation between neighboring cores, the priority does not perfectly match to the power consumption profile of cores.

2.2 Dynamic frequency scaling and voltage selection

A task graph is a directed acyclic graph that represents the workloads to be executed on the many-core processor. The vertices represent the tasks, and the edges represent the communication between the connected vertices. A computation time is defined for each vertex, and the timing constraints are given. Prior to mapping tasks on cores and selecting a proper supply

voltage level, the frequency for each task should be selected under the timing constraints. Tasks are slowed down to eliminate their slacks by selecting one of the available frequencies. First, the critical path of the task graph is found then the frequency is selected not to violate the timing constraint but to reduce the slack of the critical path as much as it can. The task graph is updated for all tasks to have the selected frequency. If there are still tasks with slacks, the lower frequency is assigned to the tasks. This is iterated until there are no remaining tasks having slacks.

Once the frequency scaling is completed, tasks are mapped to the cores. The task scheduling and mapping is based on as-soon-as-possible (ASAP) approach, so the tasks on the critical path are dealt with first. To select cores which minimize the total power consumption for tasks, the static thermal characteristics of cores are utilized. Fast cores operate at lower supply voltage than others to support a certain operating frequency, so they consume less power. Therefore, assigning the tasks which should operate at higher frequency to fast cores reduces the total energy consumption without performance loss.

It is rare that the temperature of any core exceeds the threshold, as cores whose power consumption is larger than others take care of tasks at lower frequencies. The basic thermal throttling technique is also adopted to prevent the rare cases. If the temperature of a core exceeds the threshold, no tasks are mapped on the cores until the temperature goes below the threshold.

3 Experiments

The numbers of cores considered are 64 (8×8), 144 (12×12), and 256 (16×16). The core is modeled based on the cores of UltraSPARC T1 core, which is relatively small and simple. The voltage levels used vary from 0.8 V to 1.1 V in 0.1 V increments, and the maximum frequency is 500 MHz.

The process parameters varied due to process variations were obtained using Eq. (1) and Eq. (2), where σ_{total}/μ is 0.12, and φ is 0.5. We modeled eight process variation profiles for each many-core platforms so that D2D variations can be also taken into account. The following results are the averaged values of the results from the eight profiles.

We have evaluated the energy consumption and temperature using task graphs. The three real task graphs, the robot control program, sparse matrix solver, and SPEC95 fpppp kernel, were also used [5]. The numbers of tasks in the real task graphs are 88, 96 and 334, respectively. Twenty task graphs were randomly generated in the standard task graph format. The number of tasks in the randomly generated task graphs range from 300 to 500.

The frequency scaling and voltage selection framework was built in C/C++ to estimate energy consumption and temperature traces. The proposed technique assigns timing-critical tasks on fast cores thereby reducing the total energy consumption. To clarify the advantage, we also estimated the energy consumptions by two other techniques: a random task assignment (Random) and a dynamic thermal management technique (DTM) which assigns tasks requiring higher frequency to the cores whose current temperature is the lowest [6]. Fig. 2 shows the comparison of the energy consumptions: the results are normalized to the proposed technique

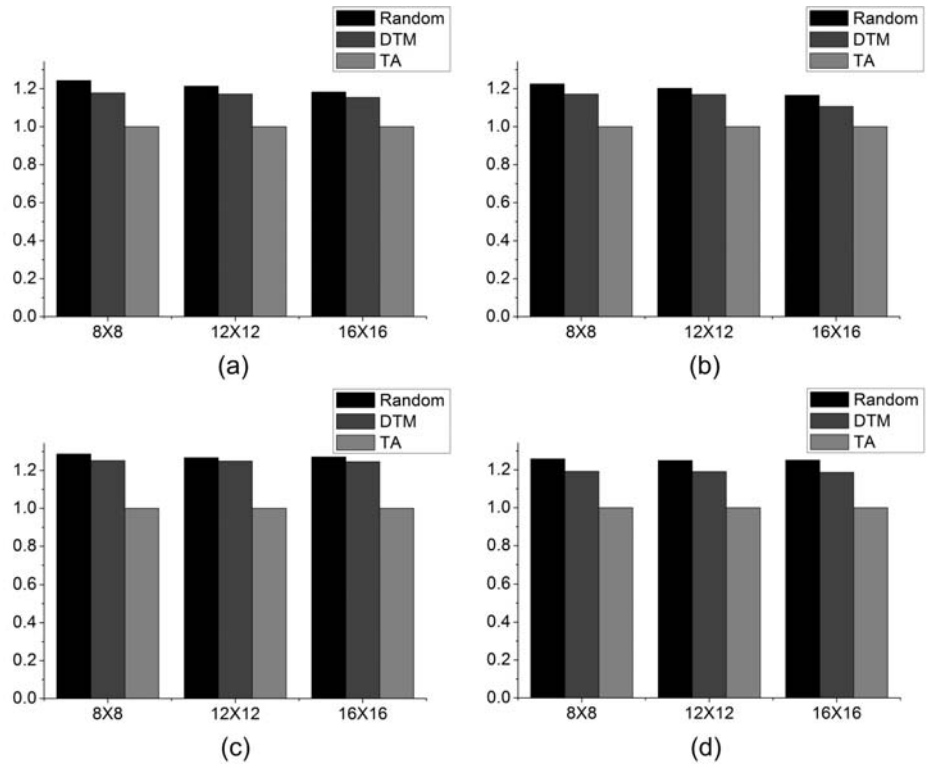


Fig. 2. Energy consumptions with different task graphs; (a) robot control program, (b) sparse matrix solver, (c) fpppp-kernel, and (d) average of 20 randomly generated task graphs.

(TA).

DTM consumes slightly less energy than Random, but the energy consumption is higher than that of TA: the energy saving reaches up to 25.2% for fpppp kernel. DTM focuses on the peak temperature management so the asymmetry on the power dissipation of cores is not considered. The cores that no tasks are mapped are powered off, and the number of the cores becomes larger as the number of integrated cores in a many-core processor becomes larger. Thus, the energy saving becomes less significant as the number of cores becomes larger.

The task graphs are run sequentially to observe the temperature traces. Fig. 3 illustrates the percentage of sampling points that the temperatures of one or more cores exceed the threshold temperature, 85°C. Although the number of cores is the smallest, the percentages are the highest in 8×8

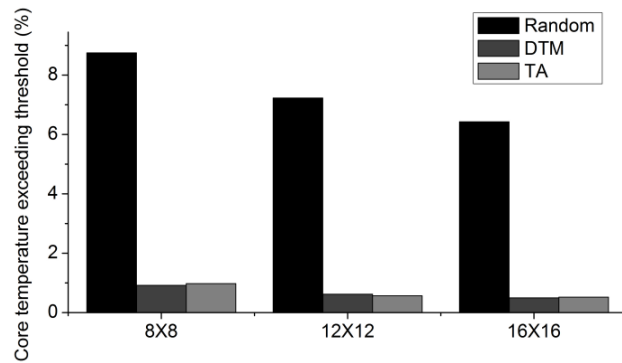


Fig. 3. Percentage of the sampling points that core temperature exceeds the threshold temperature.

many-core platforms since almost every core always works. DTM and TA result in the similar efficiency in the management of peak temperature, while keeping the percentage of temperature violations under 1%.

The estimations of Random, DTM, and TA are based on the same task graphs which the frequencies of tasks are scaled, thus ideally, the performance is the same. No tasks are mapped to the cores which temperature violations occur until their temperature stabilizes again, which results in the performance degradation. However, the degradation is not very significant; the additional cycle by Random is 2.42%, and it becomes ignorable as the number of cores increases.

4 Conclusion

AVS is adopted to hide the variations in the performance of cores. That results in the difference in the heat generation of cores in many-core processors. These thermal characteristics may vary from die to die, and from core to core, but they are fixed after fabrication. Exploiting the static thermal characteristics, the efficiency of dynamic thermal managements can be improved. The proposed thermal-aware dynamic voltage frequency scaling consumes up to 25.2% less energy than the existing thermal management technique while remaining the ratio of peak temperature violations under 1%.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A1A03006255).