Dynamic thermal management for 3D multicore 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: Stacking core layers is emerging as an alternative for future high performance computing, but thermal problems have to be tackled first. When adaptive voltage scaling is adopted to hide the growing variation in the performance of cores, as a result, heat generation of each core varies. By exploiting the static thermal characteristics, the efficiency of dynamic thermal management can be improved. The proposed thermal management reduces the energy consumption by up to 30.02% compared with existing techniques, while keeping the ratio of temperature violations around 1%.

Keywords: 3D, multicore, dynamic thermal management, process variations

Classification: Integrated circuits

References

- [1] A. K. Singh, M. Shafique, A. Kumar and J. Henkel: Proc. Design Automation Conference (2013) 1.
- [2] P. D. Franzon, W. R. Davis, M. B. Steer, H. Hao, S. Lipa, S. Luniya, C. Mineo, J. Oh, A. Sule and T. Thorolfsson: Proc. Int. Symp. Signals, Syst. and Electronics (2007) 263.
- [3] S. T. 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. Huand, S. Velusamy, K. Sankaranarayanan and D. Tarjan: Proc. Int. Symp. Computer Architecture (2003) 2.
- [5] X. Zhou, J. Yand, Y. Xu, Y. Zhang and J. Zhao: IEEE Trans. Parallel Distrib. Syst. **21** [1] (2010) 60.
- [6] H. Hong, J. Lim, H. Lim and S. Kang: IEICE Electron. Express **10** (2013) 20130463.
- [7] A. K. Coskun, J. L. Ayala, D. Atienza, T. S. Rosing and Y. Leblebici: Proc. Design, Automation & Test in Europe Conference & Exhibition (2009) 1410.
- [8] T. Tobita and H. Kasahara: J. Sched. **5** [5] (2002) 379.

1 Introduction

Three dimensional (3D) integration technology has been emerging as a solution to the problems faced by traditional two-dimensional (2D) integration technology. In 3D ICs, the global wire length is considerably reduced by a factor of $\sqrt{k}$, where k is the number of stacked layers. The performance of multicore processors can be improved using 3D integration technology because the interconnections between the cores are significantly shortened. 3D multicore processors are accepted as an alternative for future high performance computing systems [1].

Although 3D stacking of core dies offers many advantages, some existing problems may be exacerbated. One of the most critical challenges is the thermal crisis. The power density per unit volume considerably increases compared to 2D technology, so the peak temperature and the temperature gradients may increase. Thermal crises degrade performance, raise cooling costs, and exacerbate reliability issues. Heat generation is primarily dependent on the power consumption so that task mapping and dynamic voltage scaling methodologies can be effective at mitigating thermal problems. This paper proposes a thermal management approach for 3D multicore processors according to the process variation by using task mapping and voltage scaling for cores. The thermal characteristics of cores that are statically obtained are exploited to make dynamic decisions. This improves both the energy efficiency and temperature management.

2 3D multicore architecture and thermal characteristics

We consider homogeneous 3D multicore processors that consist of identical layers. This assumption of homogeneous 3D ICs is reasonable since stacking identical layers can reduce both design efforts and manufacturing costs [2]. Eight cores are on a die and the die count varies from two to four. The granularity of voltage and frequency scaling is per-core in order to fully support energy efficiency. The on-chip communication is based on the network-on-chip (NoC). Mesh topology is the most popular 2D NoC topology due to its regularity, thus a stacked mesh topology NoC is assumed. The structure of the router of the stacked mesh topology is a straightforward extension of that of the 2D mesh topology: two physical ports, one for up and one for down, are added to support inter-layer data transmission.

Process variations lead to the asymmetry on the maximum frequency of cores. The transistor switching delay, t_d , is dependent on the transistor threshold voltage V_t and the effective channel length L_{eff} , which are two major process parameters affected by process variations. The varied delay can be estimated as

$$t_d \propto \frac{V_{dd}L_{eff}}{(V_{dd} - V_t)^\alpha}, \quad (1)$$

where α is typically 1.3 [3]. To equalize the performance of cores, the adaptive voltage scaling (AVS) is adopted: slow cores only use some of the higher supply voltages from the power delivery network while fast cores use lower supply voltages. Using Eq. (1), the appropriate supply voltage level

for the required frequency to hide the delay can be selected.

A die is divided into eight core tiles and the variations on their process parameters are modeled. The standard deviation of the parameters, σ_{total} , is modeled as

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

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

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

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 a map of maximum core frequencies in a 3D multicore processor of 16 cores. Layer 0 is the top layer, and layer 1 is contiguous to the heat sink. The maximum frequencies range from 370 MHz to 620 MHz, where σ_{total}/μ is 0.12.

Fig. 1. The frequency map of the 3D multicore processor of 16 cores 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 when AVS is adopted.

The thermal characteristics of cores are determined after AVS is completed. We modified HotSpot [4] to model 3D stacked dies for static temperature simulation. There are some fixed thermal characteristics in 3D ICs [5]: in the bottom layer it is easy to expel generated heat so it tends to be cooler than the upper layers. There is a strong correlation between vertically neighboring cores due to the wafer thinning. As the modified HotSpot takes the thermal characteristics into account, the results of the steady state temperature simulation provide a good evaluation for the 3D multicore processors. To extract the thermal characteristics of cores, every core runs at the same frequency without any idle time. The result of the steady state temperature simulation is shown in Fig. 1 (b); it corresponds

to the process variation profile in Fig. 1 (a). Although all cores run at the same frequency, the resulting temperatures differ as the supply voltages of the cores and the distance from the heat sink differ. These thermal characteristics should be statically analyzed and be considered while making dynamic decisions [6]. The core with the lowest steady state temperature is given the top priority. Since the analysis also reflects the thermal characteristics of 3D ICs, the priority does not simply match the power consumption profile of each core.

3 Thermal management for D multicore processors

A task graph is a directed acyclic graph that represents the workloads. 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 also given. Firstly, the frequency for each task should be selected under the timing constraints. Tasks are slowed down to eliminate slack by selecting one of the available frequencies. The critical path of the task graph is found, and then the available lowest frequency without timing violations is selected. The task graph is updated for all tasks with the selected frequency. If there are still tasks with slack, they are assigned a lower frequency. This is iterated until there is no slack in any remaining tasks.

Once the frequency scaling is completed, tasks are mapped to the cores. To select cores that minimize the total energy consumption for tasks, the static thermal characteristics of cores are used. For a given operating frequency, fast cores operate at a lower supply voltage than the others, so they consume less power. The key of task mapping for thermal management is to assign the tasks that should operate at higher frequency to fast cores. However, this may not minimize the total energy consumption by increasing the distance between tasks which communicate each other. The priority generated by the steady state temperature simulation is modified to take communication energy into account as

$$Pr(C_{l_c}) = Pr_{steady} + \alpha \sum_{s=1}^S CW_s \times D_s. \quad (4)$$

Pr_{steady} is the priority generated by the steady state temperature simulation. CW_s is the amount of communication between the c th core on layer l , C_{l_c} and the predecessor cores. D_s is the Manhattan distance between C_{l_c} and the predecessor cores. The weighting factor, α , is the empirical value that minimize the total energy consumption. Using the modified priority, the core with the lowest $Pr(C_{l_c})$ is selected first for the next mapping.

4 Experiments

The thermal management framework was built in C/C++ to estimate the energy consumption and temperature traces of the proposed thermal-aware (TA) technique and some existing techniques. The baseline is a random task mapping (Random). Round-Robin (RR) mapping was also evaluated; this reduces the chance of uneven task distribution over Random. A basic

dynamic thermal management technique (DTM) that maps tasks requiring higher frequency to the cores whose current temperature is the lowest was evaluated. The last existing technique is an adaptive-random mapping for 3D multicore processors using each core's thermal history (Adap3D) [7].

The number of cores on a die is eight and the layer count considered is two, three, and four. The core is modeled based on UltraSPARC T1 cores. 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 technology used is assumed to be 45 nm process. The process parameter variance due to process variation was obtained using Eq. (2) and Eq. (3), where σ_{total}/μ is 0.12, and φ is 0.5. We modeled 32 dies with various process variation profiles so that both die-to-die and within-die variations can be also taken into account. Therefore, 16 3D multicore processors of two layers, ten of three layers, and eight of four layers are modeled. The following results are the averaged values of the results from the multiple models.

We evaluated the energy consumption and temperature using task graphs. The three real task graphs; the robot control, sparse matrix solver, and SPEC95 fpppp kernel, were also used [8]. The number of tasks in the real task graphs is 88, 96 and 334, respectively. Twenty task graphs were randomly generated in the standard task graph format. In each randomly generated task graph, the number of tasks ranges from 300 to 500. The estimations are all based on the same task graphs in which the frequencies of tasks are scaled, thus ideally, the performance should be the same. No tasks are mapped to the cores in which temperature violations occur until their temperature stabilizes again, and this leads to the performance degradation. However, the degradation is not very significant. The maximum additional cycle by TA is 1.16% for the 3D multicore processor of four layers, and it becomes negligible as the layer count increases.

The energy savings are shown in Fig. 2; the results are normalized to that of the baseline, Random. Both DTM and Adap3D save more energy than RR, but the energy saving from TA is the most significant: the energy saving reaches up to 32.02% for the fpppp kernel. Since DTM and Adap3D focus only on peak temperature management, the asymmetry in the power dissipation of cores is not considered, thereby resulting in smaller energy savings.

The task graphs are run sequentially to obtain the temperature traces. Fig. 3 (a) illustrates the percentage of sampling points when the temperatures of one or more cores exceed the threshold temperature, 85°C. As the number of cores and the layer count increases, this percentage increases due to the difficulty of heat dissipation in stacked dies. RR reduces temperature violations over Random, but approximately 5% of temperature violations remain, which leads to performance degradation. DTM, Adap3D and TA result in similar efficiency in the management of peak temperature, while keeping the percentage of temperature violations around 1%. Although the percentage of thermal violations is the lowest using Adap3D, computation overheads are needed for calculating the probability of each core at every sampling cycle. The peak temperature according to layer count is shown in Fig. 3 (b). The results show that Random and RR cannot effectively manage peak temperatures. With DTM, Adap3D and TA, the peak temperature remains slightly above the threshold temperature for every layer count.

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

Fig. 3. Temperature managements; (a) percentage of sampling points of temperature violations, (b) peak temperatures.

5 Conclusion

When AVS is adopted to hide the variation in the performance of cores, consequently, the heat generation varies for cores in 3D multicore processors. These thermal characteristics may vary from die to die, and from core to core, but they are fixed after fabrication. By exploiting the static thermal characteristics, the efficiency of dynamic thermal managements can be improved. The proposed thermal management consumes up to 30.02% less energy than the existing technique while keeping the ratio of peak temperature violations around 1%.

Acknowledgments

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