Thermal-aware 3D Multi-core Processor Design using Core and Level-2 Cache Placement

Dong Oh Son, Hong Jun Choi, Hyung Gyu Jeon and Cheol Hong Kim*

School of Electronics and Computer Engineering, Chonnam National University, Gwangju, 500-757 Korea
{sdo1127, chj6083, hggodman1108, cheolhong}@gmail.com

Abstract

As integration densities continue to increase, interconnection has become one of the most important factors in determining the performance of multi-core processors. Recently, in order to reduce the delay due to interconnection, many studies have focused on the 3D multi-core processors. Compared to 2D multi-core architecture, 3D multi-core architecture gets decreased interconnection delay and lower power consumption owing to reduced wire length. Despite the benefits mentioned above, 3D design cannot be practical because it causes serious thermal problems in multi-core processors due to high power density. In this paper, we analyze temperature behavior of 3D multi-core processors according to various placement of core and level-2 cache. According to our simulation results, the floorplan where the core is stacked adjacently to the level-2 cache can reduce the temperature by 22% with 4-layers, and by 13% with 2-layers on the average, compared to the floorplan where the core is stacked adjacently to the core.

Keywords: *3D Architecture; Floorplan; Multi-core Processor; Thermal Problem*

1. Introduction

Even though continuing advances in semiconductor technology have allowed dramatic performance improvement of processors, it causes increased interconnection delay which has become one of the major constraints in improving the performance of multi-core processors. One of the most effective alternative to solve the interconnection delay problem in multi-core processors is 3D integration technology [1]. When designing 3D multi-core processors, multiple cores are stacked vertically and each core on different layers are connected by direct vertical through-silicon vias (TSVs) [2]. Therefore, the 3D architecture can reduce the interconnection delay of multi-core processors by reducing total wire length significantly compared to the 2D architecture. The 3D technology can have the benefits of performance improvement and reduced power consumption [3]. However, as a negative consequence, temperature of the 3D architecture has increased dramatically due to high power density caused by stacking multiple layers vertically. Unfortunately, high temperature in the processor causes increased cooling costs, negative impact on the reliability, and performance degradation [5]. Consequently, 3D technique cannot be practical without proper solutions for high temperature.

Existing mechanical cooling schemes using heat insulating board or cooling fan are being limited due to the increased cooling costs. To mitigate the thermal problems with minimizing cooling costs, a lot of high-performance microprocessors use thermal management technique,

* Corresponding Author
This paper is an extended version of [4]

such as Dynamic Thermal Management (DTM). Unfortunately, DTM degrades the performance while it can mitigate the thermal problems in the processor [6-7].

In this work, we propose the effective methods to reduce the temperature in the 3D multi-core processor. Then, the proposed thermal-aware schemes are analyzed in detail. The rest of this paper is organized as follows. Section 2 explains proposed floorplan schemes and Section 3 provides our experimental methodology and results. Finally, Section 4 concludes this paper.

2. Floorplan Schemes

Heat transfer from adjacent functional units is another important factor that affects the temperature distribution of a chip. It depends on the functional unit adjacency determined by the floorplan of the microprocessor. Therefore, floorplan scheme has been focused on the solutions for reducing the peak temperature in the processor with minimal performance degradation.

In previous floorplan studies [8-12], the hottest functional unit in core is replaced with cool functional unit to control the temperature. However, in this paper, proposed floorplan schemes change the forms of vertical placement of the core and the level-2 cache while the location of functional units in the core is fixed. Modifying the floorplan of the core affects the datapath of 3D multi-core processors because the distance between functional units is determined by the floorplan. However, in this work, we don't consider the overhead due to the datapath since the location of functional units is not changed.

In order to analyze the temperature behavior according to the number of layers, proposed floorplan schemes are categorized into two types: One is 2-layer 3D quad-core processor and the other is 4-layer 3D quad-core processor.

A. 4 -layer Architecture

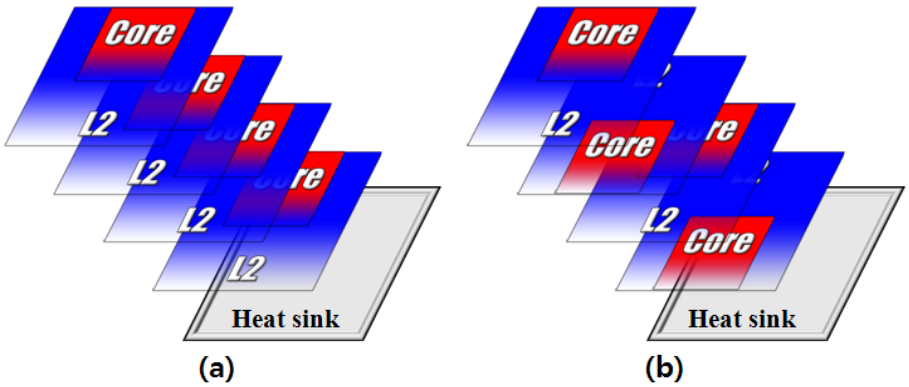


Figure 1. Floorplans of 4-layer Architecture

Figure 1 shows the two floorplans of 4-layer architecture. Figure 1(a) represents the baseline floorplan of 4-layer architecture. We use the Appha21364 floorplan for each core and stacked four cores vertically to configure the quad-core processor. In Figure 1(b), core and L2 cache are relocated to reduce the temperature without increasing area. Compared to the baseline, core and L2 cache placement of adjacent layer is inverted, as shown in Figure 1(a) and Figure 1(b).

B. 2 -layer Architecture

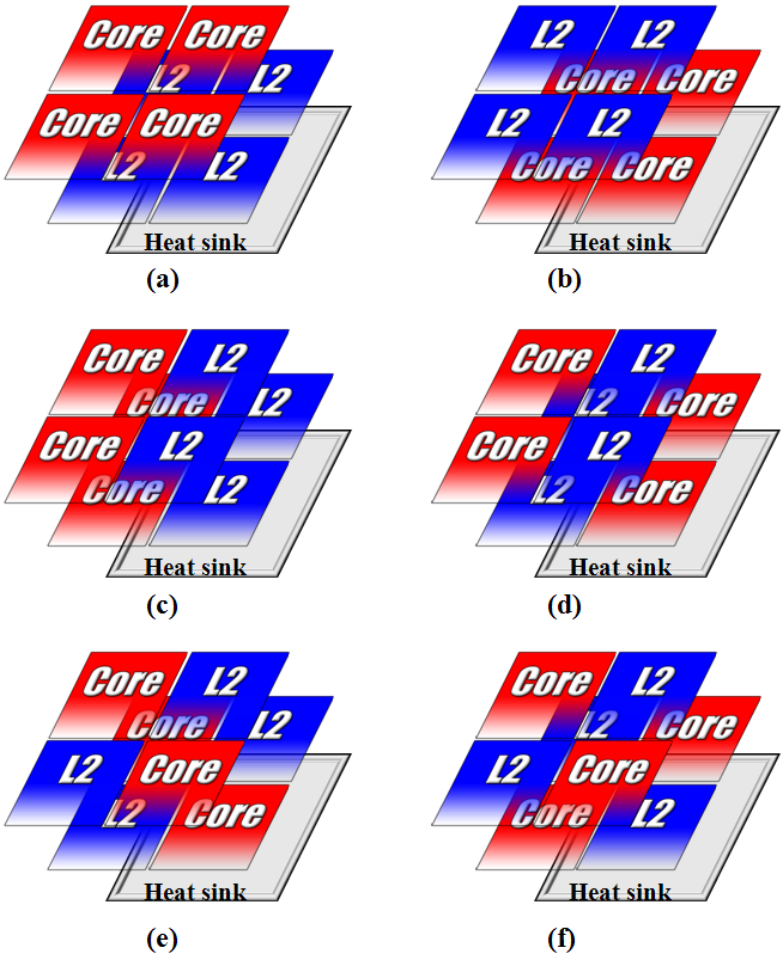


Figure 2. Floorplans of 2-layer Architecture

Figure 2 shows six floorplans of 2-layer 3D multi-core processors. In this work, we implement a lot of experiments to find efficient floorplans for reducing the temperature and describe only six floorplan schemes owing to limited page. In 2-layer architecture, we assume that core size is same to L2 cache size. Placement of the core and L2 cache in the proposed floorplan schemes can be summarized as follows:

Name : bottom-layer / upper-layer (left part – right part)

2-layer(a) : all L2 caches / all cores

2-layer(b) : all cores / all L2 caches

2-layer(c) : two cores & L2 caches / two cores & L2 caches

2-layer(d) : two L2 caches & cores / two cores & L2 caches

2-layer(e) :core-L2-L2-core / core-L2-L2-core

2-layer(f) : L2-core-core-L2 / core-L2-L2-core

3. Experiments

Temperature evaluation of 3D multi-core processors is very challenging because the heat flow is unpredictable. In this section, we briefly describe the experimental environments and thermal characteristics according to proposed floorplan schemes.

A. Experimental Environment

To measure the temperature behavior of each layer, we use HotSpot Version 5.0 [13] which is capable of configuration for 3D multi-core processors. In order to have precise experiments, the configuration and the layer parameters in HotSpot are obtained from the material properties in CRC handbook [14]. In addition, Wattch [15] is integrated to obtain the detailed power trace. In our experiments, baseline processor is the Alpha 21264(Ev6) [16]. Table 1 shows the main processor and memory hierarchy parameters used in the simulation.

Table 1. System Parameters

parameters	Value
Functional units	4 integer ALUs, 4 FP ALUs, 1 integer multiplier/divider, 1 FP multiplier/divider
L1 I-Cache	32KB, 4-way, 32byte lines, 1 cycle latency
L1 D-Cache	32KB, 4-way, 32byte lines, 1 cycle latency
L2 Cache	512KB, 8-way, 32byte lines, 12 cycle latency

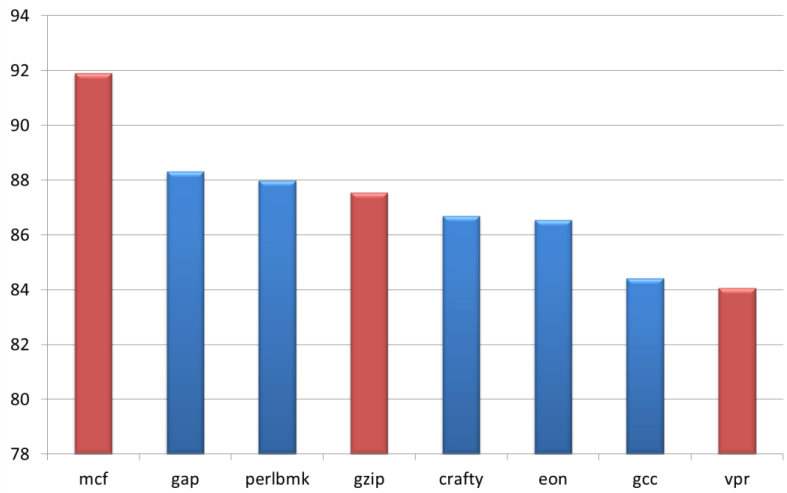
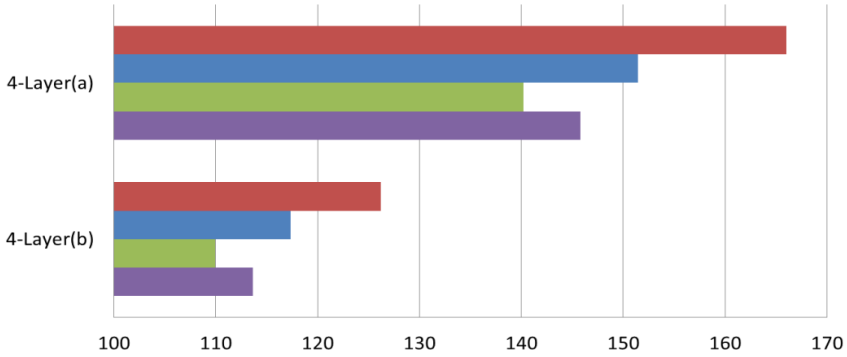


Figure 3. Peak Temperature

Figure 3 shows the peak temperature of applications from SPEC2000 benchmarks[17]. We select three application(mcf, gzip, vpr) because the chosen applications show diverse thermal characteristics(high temperature, medium temperature, low temperature).

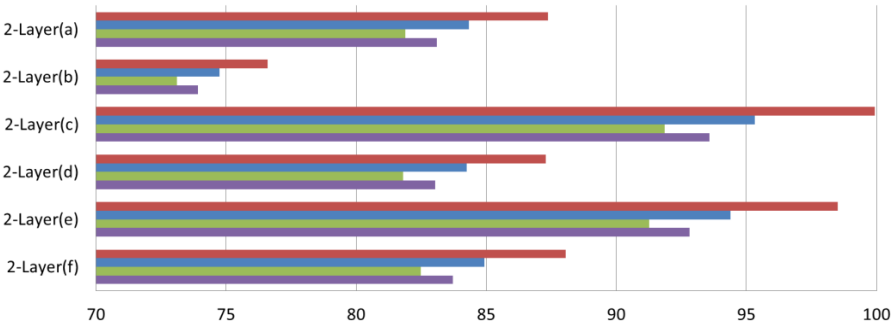
B. Experimental Results



	4-Layer(b)	4-Layer(a)
mcf	126.24	166.01
gzip	117.33	151.46
vpr	109.92	140.19
Avg	113.63	145.83

Figure 4. Peak Temperature Comparison of 4-layer Architecture

Figure 4 shows the peak temperature of 3D quad-core processors according to each floorplan scheme. In the graph, 4-layer(a) and 4-layer(b) represent the baseline 4-layer floorplan and proposed 4-layer floorplan, respectively. We assume that each core executes same application. For example, mcf means that mcf is executed on all cores. As shown in the graph, the temperature of 4-layer(a) shows serious thermal problems. Contrary to that, 4-layer(b) reduces the peak temperature compared to the 4-layer(a) significantly. According to our simulation results, 4-layer(b) decreases the temperature by 34.72℃ on average (mcf:39.77℃, gzip:34.13℃, vpr:30.27℃) compared to 4-layer(a).



	2-Layer(f)	2-Layer(e)	2-Layer(d)	2-Layer(c)	2-Layer(b)	2-Layer(a)
mcf	88.06	98.51	87.29	99.93	76.6	87.38
gzip	84.93	94.38	84.25	95.32	74.75	84.33
vpr	82.49	91.27	81.81	91.85	73.11	81.89
Avg	83.71	92.825	83.03	93.59	73.93	83.11

Figure 5. Peak Temperature Comparison of 2-layer Architecture

Figure 5 shows the peak temperature according to various floorplans of 2-layer architecture. Compared to 2-layer(c) which shows the highest temperature, 2-layer(b) decreases the temperature by 19.66℃ on average (mcf:23.33℃, gzip:20.57℃, vpr:18.74℃).

4. Conclusion

In this paper, we investigated how the floorplan schemes handle the thermal problems in 3D multi-core processors. We proposed thermal-aware foorplan schemes for 2-layer and 4-layer 3D architecture. The proposed floorplan schemes change the vertical placement of the core and the level-2 cache. According to our analysis, the floorplan where the core is stacked adjacently to the level-2 cache reduces the temperature of 3D multi-core processors significantly compared to the floorplan where the core is stacked adjacently to the core. We expect that this work presents the guideline for thermal-aware 3D multi-core processors.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012R1A1B4003492) and MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)(NIPA-2012-H0301-12-3005)

References

[1] C. Zhu, Z. Gu, L. Shang, R. P. Dick and R. Joseph, "Three-dimensional chip-multiprocessor run-time thermal management", Proc. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **(2008)** August, pp. 1479-1492.
[2] S. W. Yoon, D. W. Yang, J. H. Koo, M. Padmanathan and F. Carson, "3D TSV processes and its assembly/Packaging technology", Proc. the IEEE International Conference on 3D System Integration, **(2009)** September, pp. 1-5.
[3] J. W. Joyner, P. Zarkesh-Ha and J. D. Meindl, "A Stochastic Global Net-length Distribution for a Three-Dimensional System on Chip (3D-SoC)", Proc. the 14th IEEE International ASIC/SOC Conference, **(2001)** September, pp. 147-151.
[4] D. O. Son, H. J. Choi, H. G. Jeon and C. H. Kim, "Temperature Analysis of the 3D Multi-core Processor according to the Placement of Core and Level-2 Cache", Proc. International Conference on SMA, **(2012)** August, pp. 31.
[5] A. K. Coskun, J. L. Ayala, D. Atienza, T. S. Rosing and Y. Leblebici, "Dynamic Thermal Management in 3D Multicore Architectures", Proc. Design, Automation & Test in Europe Conference & Exhibition, **(2009)** April, pp. 1410-1415.
[6] D. Brooks and M. Martonosi, "Dynamic Thermal Management for High-performance Microprocessors", Proc. of the 7th International Symposium on High-Performance Computer Architecture, **(2001)** January, pp. 171-182.
[7] A. K. Coskun, J. L. Ayala, D. Atienza, T. S. Rosing and Y. Leblebici, "Dynamic Thermal Management in 3D Multicore Architectures", Proc. Design, Automation & Test in Europe Conference & Exhibition, **(2005)** April, pp. 1410-1415.
[8] E. Wong and S. K. Lim, "3D Floorplanning with Thermal Vias", Proc. the conference on Design, Automation and Test in Europe, **(2006)** March, pp. 878-883.
[9] K. Puttaswamy and G. H. Loh, "Thermal Analysis of a 3D Die Stacked High Performance Microprocessor", Proc. ACM Great Lakes Symposium on VLSI, **(2006),** pp. 19-24.
[10] A. K Coskun, A. B Kahng and T. S Rosing, "Temperature- and Cost-Aware Design of 3D Multiprocessor Architectures", Proc. 12th Euromicro conference on Digital System Design, Architectures, Schemes and Tools, **(2009)** August, pp. 183-190.
[11] W. L. Hung, G. M Link, Y. Xie, N. Vijaykrishnan and M. J. Irwin, "Interconnect and Thermal-aware Floorplanning for 3D Microprocessors", Proc. the 7th ISQED, **(2006)** March, pp. 98-104.
[12] D. O. Son, J. W. Ahn, J. H. Park, J. M. Kim and C. H. Kim, "Analysis on the Temperature of 3D Multi-core Processors according to Vertical Placement of Core and L2 Cache", Journal of The Korea Society of Computer and Information, vol. 16, no. 6, **(2011)** June, pp. 1-10.
[13] W. Huang, M. R. Stan, K. Skadron, K. Sankaranarayanan and S. Ghosh, "HotSpot: A Compact Thermal Modeling Method for CMOS VLSI Systems", Journal of IEEE Transactions on VLSI Systems, vol. 14, **(2006)**.

[14]CRC Press, "CRC Handbook of Chemistry", http://www.hbcpnetbase.com/.
[15]D. Brooks, V. Tiwari and M. Martonosi, "Wattch: A Framework for Architectural-level Power Analysis and Optimizations", Proc. the 27th Annual International Symposium on Computer Architecture, **(2000)** June, pp. 83-94.
[16]R. E. Kessler, "The Alpha 21364 microprocessor", Journal of IEEE MICRO, vol. 19, no. 2, **(1996)**, pp. 24-36.
[17]J. L. Henning, "SPEC CPU2000: measuring cpu performance in the new millennium", Journal of IEEE Computer, vol. 33, no. 7, **(2000),** pp. 28-35.

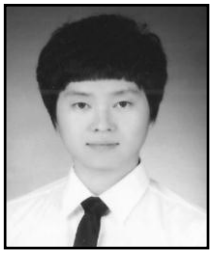
Authors

Dong Oh Son

Ph.D. student, Chonnam National University, School of Electronics and Computer Engineering, Korea.

Hong Jun Choi

Ph.D. student, Chonnam National University, School of Electronics and Computer Engineering, Korea.



Hyung Gyu Jeon

Master Student, Chonnam National University, School of Electronics and Computer Engineering, Korea.

Cheol Hong Kim

Professor, Chonnam National University, School of Electronics and Computer Engineering, Korea.