A new processor design based on 3D cache

Lei YI[1,a] , Guangbao SHAN[2,b], Song LIU[3,c], Chengmin XIE[4,d]

[1] Xi'an Microelectronics Technology of Institute, Xi'an, 710029, China

[2]Xi'an Microelectronics Technology of Institute, Xi'an, 710029, China

[3]Xi'an Microelectronics Technology of Institute, Xi'an, 710029, China

[4]Xi'an Microelectronics Technology of Institute, Xi'an, 710029, China

[a]1311970283@qq.com , [b]18092060235@189.cn , [c]song.liu771@hotmail.com, [d]hglnew@sina.com

Keywords: Three-dimensional integration technology, TSV, 3D cache, 3D processor

Abstract. The interconnection is becoming one of main concerns in current and future microprocessor designs from both performance and consumption. Three-dimensional integration technology, with its capability to shorten the wire length, is a promising method to solve issues related the interconnection. In this paper, we propose a new processor architecture based on 3D cache. We integrate 3D cache with the processor which reduces the global interconnection, power consumption and improves access speed. In addition, we simulate the performance of the 3D processor and 3D cache at different node using 3D Cacti tools. Comparing with 2D, the results show power consumption of the memory system is reduced by about 50%, access time and cycle time of the processor increase 18.57% and 21.41%, respectively.

Introduction

As technology scaling comes into the deep sub-micron era, interconnect has emerged as the major source of delay and power consumption, in particularly high-density interconnect layout design [1-2]. 3D integration technology, using TSV (Through-Silicon Via) to transfer signal, is a promising solution for overcoming obstacle in technology scaling [3-4], thereby offering an opportunity to improve circuit performance, especially for processor [5-6].

Despite the merits mentioned above, there are a few works focusing on 3D processor architecture exploration stacking the memory and the processor logic module. In [7], the processor-DRAM-stacked is investigated and turns out that 3D integration technology can effectively reduce the inter-module interconnect length and consumption. Paul Reed et al [8] has studied the 3D integrated memory-processor and analyzed the sense amplifier. In [9], the author has studied a 3D stacked register file with cache in high-performance microprocessor architecture. However, most of these works simply consider the 3D memory and the processor to be different level or only stack entire memory with another memory to increase cache capacity, at the same time, there are few works focusing on overall performance of the 3D processor.

In this paper, we explore a high-performance processor architecture based 3D on-chip cache. We vertically stack 3D on-chip cache with logic module of processor into a single chip using TSV technology. Therefore, we make a analysis about the delay, power consumption and overhead footprint of 3D integrated processor using 3D Cacti tool.

The rest of the paper is organized as follows. Section 2 introduces basic design principles of the 3D on-chip cache processor and makes a description of the 3D cache design structure. Section 3, investigating 3D Cacti, presents and discusses experimental results. We make a conclusion in the last section.

Processor architecture based on the 3D cache

3D processor architecture

Figure 1 shows the basic structure of a conventional 3D stacked processor [10]. The processor is divided into memory layers (die #2) and logic unit layer (die #1), which are stacked by vertical interconnect, as shown in fig 1. The structure could reduce the processor footprint, the global interconnect length, system power consumption and the critical path delay by vertically stacking multiple strata.

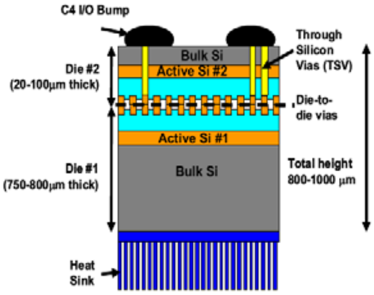

Figure 1 . The basic structure of a conventional 3D stacked processor [10]

We proposed a novel processor architecture which stacks the 3D on-chip cache and logic module into a single chip. Compared with the traditional processor, we use the 3D cache to substitute original cache and connect the bottom layer of 3D cache with the logic module by common interconnect. In 3D on-chip cache, we constitute directly vertical interconnect in different cache layers which shortens the global interconnection length and furthest optimizes processor interconnection. 3D on-chip cache further decreases the size of cache memory module, speeds up information transmission and lowers latency and power consumption by dividing the arrays of on-chip cache memory. In addition, the structure stacked the 3D on-chip cache with processor ulteriorly refines the overall footprint and reduces the global interconnect length, delays and costs.

3D cache structure design

Figure 2.a shows the 2D cache structure. It is partitioned along the bit line by 3D technology.

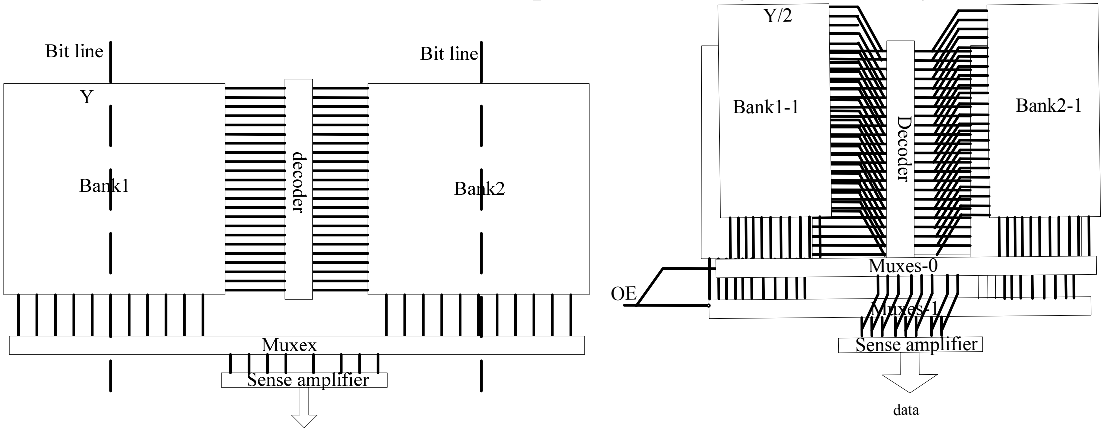

Figure 2. (a) The 2D cache structure. (b)The 3D cache structure

Figure 2.b shows 3D cache structure. As can be seen by fig 2.a, 2D cache is divided into two equal-sized sub-arrays, and stacking equal-sized sub-arrays constitutes 3D cache which decreases cache internal and external interconnect length. (For instance, the width of memory arrays is assumed to be Y in fig 2.a, so the distance that decoders access the furthest memory units becomes Y/2 in fig 2.b). Therefore, all addresses and data lines are routed locally on each sub-array and other layers are connected with by TSVs. The sense amplifiers are placed lower stratum, sharing with the upper strata with vertical interconnect. In addition, splitting along bit lines reduces the number of multiplexers (MUXES), for example, when splitting along two bit lines, the number of muxes become a quarter of the original. In order to ensure the stacked structure outputs data bit widths to be consistency, the number of muxes should be increased based on the division strategy. At the same time, there is the enable signal (OE) added to multiplexer to select conduction of each cache layers.

The 3D structure has huge impacts on interconnects reduced between cache and logic module or among different cache strata, which improves access time and lower power consumption. In this 3D cache structure, the length of word lines and the number of transistors connected with word lines are decreased to half. Therefore, the stacked structure reduces load resistance of word lines which is half of original on a single plane. 3D cache selects different strata with enable signal (OE), reducing overall power consumption of processors. At the same time, 3D technology reduces processor footprint, costs and improves processor performance. We will make a detail analysis in the section 3.

Analysis and Simulation

In this section, we simulate all relevant metrics, i.e, access time, cycle time, power and delay of 16, 32 and 64 caches, considering different division ways at 90 or 180 nm technology node based 3D Cacti.

Figure 4 shows the performance of 2D and 3D 16K, 32K, 64K at 90nm node. The results are that the delay of 16K, 32K, 64K cache is decreased by 32%, 27% and 23% respectively, and the power consumption of 16K, 32K, 64K cache is reduced by 45% respectively. Our analyses indicate that the capacity of cache is different and the reduced degree of delay, power consumption is also different at the same node.

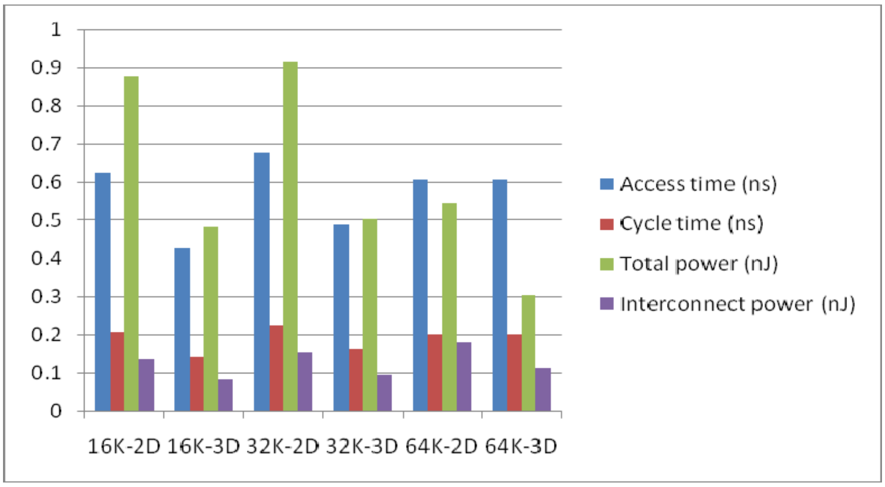

Figure 4 shows 16K, 32K, 64K divided case at 90nm node.

Figure 5 shows performance of 64K stacked cache compared with the 2D cache at different node. The results show in table 1 for the improvement of the access time, the cycle time, the power consumption of 3D cache. We observe the improvement performance of different aspects of cache exists difference at different node. It is expressed in table 1 that the access time and cycle time are improved by 18.57%, 21.41%(90nm), and 38.26%,39.64% (180nm), the power consumption is improved by 43.47%(90nm) and 37.41% (180nm), and the power of interconnect is reduced by 30.48% (90nm) and 40.82% (180nm). These are expected, since due to 3D stacked of the cache, the footprint is reduced proportionally to the number of dies, and the wires are shorter. The exceptions are explained by the longer wires presents indisproportionate sub-array layout because of TSVs.

Table 1. the improvement of the access time, the cycle time, the power consumption of 3D cache, combined with 2D cache.

	access time (ns)	Cycle time (ns)	Totalpower consumption (nJ)	Interconnect power consumption (nJ)	MAXpower consumptionof block(nJ)
90nm	18.57%	21.41%	43.47%	30.48%	25.79%
180nm	38.26%	39.64%	37.41%	40.82%	15.54%

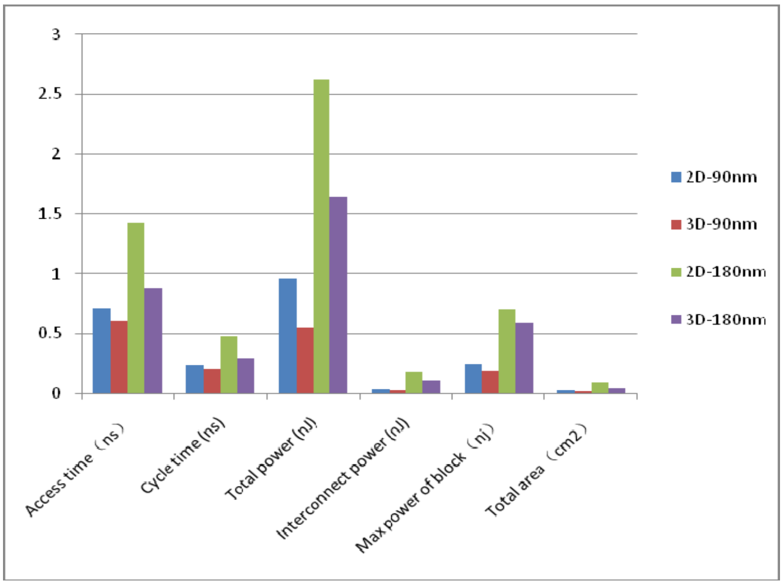

Figure 5 shows performance of 64K stacked cache compared with the 2D cache at different node.

Based on the above article, the 3D cache, making a rapid communication between 3D cache and logic model of processor, improves speed and latency of the processor. It is 3D cache that reduces the footprint of cache and processor and shortens the global interconnect among logic models and cache, further improving processor performance. 3D cache reduces normal operating power of cache and overall power consumption, latency and global interconnection of 3D processor.

Conclusion

In this work we propose a 3D processor architecture which uses a 3D on-chip cache to replace the traditional 2D cache. This paper addresses the issue of on-chip cache design in 3D processor integrated structures. It reduces the area of processor, improves performance of 3D processor such as access time, delay, power consumption. As a result, access time and cycle time of the processor increase 18.57% and 21.41%, respectively, and the area efficiency is brought to above 25%.

References

[1] Puttaswamy K, Loh G H. Implementing register files for high-performance microprocessors in a die-stacked (3D) technology, Emerging VLSI Technologies and Architectures, 2006. IEEE Computer Society Annual Symposium on. IEEE, 2006: 6 pp.

[2] Puttaswamy K, Loh G H. Implementing caches in a 3D technology for high performance processors, Computer Design: VLSI in Computers and Processors, 2005. ICCD 2005. Proceedings. 2005 IEEE International Conference on. IEEE, 2005: 525-532.

[3] Pavildis V, Friedman E. 3D integrated circuit design, Page 93-94, Burlington: Morgan Kaufmann Publishers, 2009.

[4] Xie Y, Cong J, Sapatnekar S S. Three-dimensional integrated circuit design. Springer, 2010.

[5] Black B, Nelson D W, Webb C, et al. 3D processing technology and its impact on iA32 microprocessors[C]//Computer Design: VLSI in Computers and Processors, 2004. ICCD 2004. Proceedings. IEEE International Conference on. IEEE, 2004: 316-318.

[6] Loh G H. 3D-stacked memory architectures for multi-core processors, ACM SIGARCH computer architecture news. IEEE Computer Society, 2008, 36(3): 453-464.

[7] Chen S S, Hsu C K, Shih H C, et al. Processor and DRAM integration by TSV-based 3-D stacking for power-aware SOCs, Design Automation Conference (ASP-DAC), 2013 18th Asia and

South Pacific. IEEE, 2013: 429-434.

[8] Reed P, Yeung G, Black B. Design aspects of a microprocessor data cache using 3D die interconnect technology, Integrated Circuit Design and Technology, 2005. ICICDT 2005. 2005 International Conference on. IEEE, 2005: 15-18.

[9] Puttaswamy K, Loh G H. Dynamic instruction schedulers in a 3-dimensional integration technology, Proceedings of the 16th ACM Great Lakes symposium on VLSI. ACM, 2006: 153-158.

[10] Borkar S. 3D integration for energy efficient system design, Proceedings of the 48th Design Automation Conference. ACM, 2011: 214-219.