A 3D SoC Design for H.264 Application With On-Chip DRAM Stacking

Tao Zhang[1], Kui Wang[2], Yi Feng[3], Yan Chen[2],
Qun Li[2], Bing Shao[2], Jing Xie[1], Xiaodi Song[2],
Lian Duan[1], Yuan Xie[1], Xu Cheng[3], and Youn-Long Lin[4]

[1]Department of Computer Science and Engineering, The Pennsylvania State University, PA, USA
[2]Peking University Unity Microsystems Technology Co. Ltd, Beijing, China
[3]Department of Computer Science, Peking University, Beijing, China
[4]Computer Science Department, National Tsing Hua University, Hsinchu, Taiwan

***Abstract*—Three-dimensional (3D) on-chip memory stacking has been proposed as a promising solution to the "memory wall" challenge with the benefits of low access latency, high data bandwidth, and low power consumption. The stacked memory tiers leverage through-silicon-vias (TSVs) to communicate with logic tiers, and thus dramatically reduce the access latency and improve the data bandwidth without the constraint of I/O pin count. To demonstrate the feasibility of 3D memory stacking, this paper introduces a 3D System-on-Chip (SoC) for H.264 applications that can make use of multiple memory channels offered by 3D integration. Two logic tiers are stacked together with each having an area of 2.5$\times$5.0mm^2, with a 3-layer 8-channel 3D DRAM stacked on the top. The design flow for this 3D SoC is also presented. The prototype chip has been fabricated with GlobalFoundries' 130nm low-power process and Tezzaron's 3D TSV technology. The 3D implementation shows that the 3D ICs can alleviate the pressure from I/O pin count and allow parallel memory accesses through multiple channels.**

I. INTRODUCTION

In recent years, 3D IC technology has been proposed as a valuable solution to continue the Moore's Law. With very dense TSVs, 3D IC can offer the following benefits: (1) low interconnect latency; (2) high data bandwidth; (3) low power consumption and (4) heterogenous design [1]. Specifically, 3D IC allows chip designers to stack memory layers on top of logic layers to reduce the memory access latency and improve the memory bandwidth. Moreover, moving the off-chip memory inside the chip can eliminate the limitation of I/O pin count on area, power, and cost, which can further improve the memory performance by integrating many memory controllers and/or channels on a single chip. Although 3D IC has great potentials, currently it still has many problems ranging from architecture immaturity to manufacture difficulty that are waiting for better solutions, from both industry and academia.

First of all, computer architects should rethink about the system architecture so that they can fully exploit benefits from TSV and 3D stacking. For example, the memory hierarchy and on-chip interconnects may have to be re-designed carefully to leverage the benefits 3D can offer. The lack of 3D EDA tools may prevent designers from adopting the emerging 3D technology [2]. Very few commercial EDA tools can fully support the 3D IC design. The use of conventional 2D EDA tools can not fully take advantage of the novel technology. Finally, both testability and cost are challenging problems for 3D IC. After stacking heterogeneous devices, the traditional homogeneous testing methodology may not be applicable any more [3]. In addition, the cost implication of adopting 3D IC is not clear yet. How to reduce the cost of either manufacture or testing has been one of the hottest topics recently [4].

To demonstrate the feasibility of 3D ICs, a System-on-Chip(SoC) design for H.264 application using 3D DRAM memory stacking is presented in this work. An H.264 encoder is deployed to process the stream data and a USB controller is used for image capture and wireless data transmission. Additionally, to fully leverage the benefit of on-chip 3D DRAM stacking, a dedicated memory controller is designed with simplified DDR protocol. Several optimizations, such as parallel access policy and TSV clustering, have been proposed to make better use of on-chip DRAM and 3D stacking.

II. RELATED WORKS

The traditional Double-Data-Rate (DDR) SDRAM has been proposed for many years as off-chip main memory, and DDR3 has become the mainstream product in this family [5]. Correspondingly, sophisticated DDR controllers have also been proposed to access DRAM with performance and power efficiency. In spite of growing memory size and lower access latency, the off-chip memory still suffers from the memory bandwidth limitation. The single channel between on-chip controller and off-chip memory is a severe bottleneck of memory bandwidth in a chip multi-processor (CMP) system. Even though DDR3 can support triple-channel access, the progress of memory bandwidth is still relatively slow due to the constraint of I/O pin count, which limits the number of memory channels to further increase. As one solution, System-in-Package (SiP) technology is currently widely used in mobile electronic devices (e.g., smartphone, PDA) [6]. SiP allows many memory chips to be stacked and encapsulated into one package so that the user can have ultra high memory capacity. Unfortunately, as SiP usually utilizes wire bonding for the inter-chip connectivity, it still has the limited memory bandwidth due to the long inter-chip wire length and few memory channels.

Different from SiP, the inter-layer distance in TSV-based 3D memory stacking (4μm~8μm [7]) is much shorter than the bonding wire in SiP (65μm~222μm [8]). As a result, the TSV-based 3D memory stacking can achieve better memory performance and power efficiency. Therefore, in recent years, significant researches have been done in this area to speed up the entire memory hierarchy via TSVs. Loh proposed several aggressive DRAM organizations that are stacked on a multi-core processer [9]. Both pseudo-3D DRAM and true-3D DRAM arcihtectures are exploited to take advantage of die-stacking. Saito et al. implemented a 3D SoC, in which bunches of SRAM slices are stacked on the logic chip [10]. The SRAM is reconfigurable so that the memory space can be reallocated to each core due to different demands from various SoC architectures. Woo et al. reorganized the memory hierarchy by implementing a wider memory bus with plenty of TSVs [11]. As a result, as many as 64 memory accesses can be processed in parallel between the L2 cache and the DRAM. To our best knowledge, however, most of prior work fall into software simulations rather than hardware implementations even though the 3D DRAM has been silicon proven by the industry. In addition to 3D technology, there are also some studies on increasing the number of memory controllers/channels to improve memory bandwidth in 2D platform. Kim et al. tried to improve the memory performance with multiple memory controllers [12]. Four memory controllers are introduced to optimize the memory scheduling scheme and increase the bandwidth with little coordination between each other. In spite of performance enhancement, using multiple memory controllers incurs heavy pressures on I/O pin count as well as large power consumption for 2D design.

In contrast to all of previous works, this project tries to implement a 3D SoC by integrating a 3D DRAM into the chip and thus demonstrate the feasibility of 3D on-chip DRAM stacking. The rest of paper is organized as follows. Section III introduces the overall architecture of the 3D SoC chip, which contains a two-tier SoC logic and a three-tier 3D DRAM memory. Section IV presents the logic design by discussing the 3D memory controller design and the logic partition scheme. Section V depicts the physical design flow, where certain steps, such as clock tree synthesis (CTS), power delivery, and TSV bonding technology, are discussed in detail. Section VI summarizes the 3D SoC chip prototyping using GlobalFoundries' 130nm technology with Tezzaron's TSV bonding technology, followed by Section VII to conclude this paper.

III. THE 3D SOC ARCHITECTURE

In this section, the overall 3D SoC architecture with on-chip DRAM stacking is presented first, followed by the functional schematic of the 3D SoC. The characteristics of 3D DRAM are introduced as the design consideration of the DRAM controller.

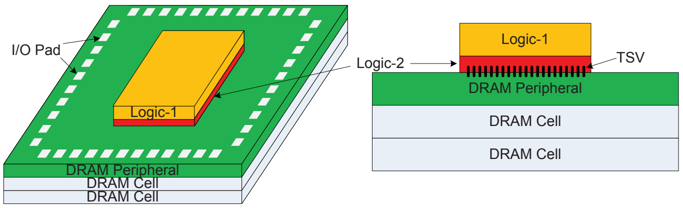


Fig. 1. 3D DRAM Stacking

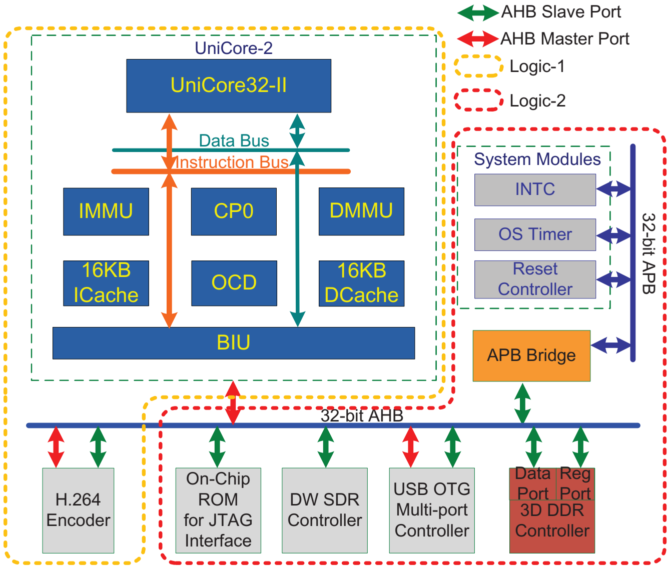


Fig. 2. Schematic View of 3D SoC

A. 3D DRAM Stacking

As shown in Fig. 1, our 3D chip is composed of two logic tiers and three DRAM tiers. Tier Logic-1 is on the top and Logic-2 lies in the middle. Both logic tiers are sized by $2.5 \times 5.0mm^2$ while the DRAM tiers are $12.3 \times 21.8mm^2$. Logic-2 is much thinner than other tiers since most of silicon substrate of this tier is burnished to expose TSVs. Note that all of the I/O pads are on the back surface of DRAM tier, which mandates TSVs to account for data and power delivery between DRAM and logic tiers as well.

B. 3D SoC

To support real-time H.264 applications, a 3D SoC is proposed as shown in Fig. 2. AMBA AHB is employed as the system bus [13]. An H.264 encoder, a USB On-The-Go controller, and a RISC processor UniCore-2 are also integrated into this SoC [14] [15] [16]. A dedicated 3D DDR controller is developed to communicate with the 3D DRAM and the design details will be discussed in Section IV.A. Additionally, a JTAG interface is used to load the initial instructions into memory for the system bootup.

C. 3D DRAM

The 3D DRAM used in this work is the state-of-the-art product from Tezzaron® [17]. To achieve both high performance and high cell density, the DRAM chip is separated into one peripheral tier and multiple cell tiers, where each tier can be optimized separately by different technologies. The total capacity of this 3D DRAM is 2Gb with 1Gb on each cell tier. Eight data channels with separate write/read ports (128b) are used to attain high memory bandwidth. By using the simplified DDR protocol, every data channel allows 256-bit data transfer in a single memory cycle. A prominent feature of these channels is that each channel is fully independent of the others, which means all channels can be accessed in parallel at different frequencies. In addition, a MailBox channel is used to initialize the DRAM when the system is powered on.

One data channel simply hooks to one bank[1] so that the bank and the data channel are interchangeable in this paper. Similar to the conventional DRAM technology, the 3D DRAM also has five main commands known as *Precharge*, *Refresh*, *Row Addressing*, *Column Read* and *Column Write*. Tezzaron has simplified the precharge and refresh command by providing the dedicated control signals for DRAM controller. The 3D DRAM can run as fast as 1GHz with 64ms refresh rate. Table I lists the basic parameters of this 3D DRAM. As shown in the table, up to 294 pins are used in each data channel so that the DDR controller should have more than 2,300 pins in total. Without 3D on-chip memory stacking, it's impossible to afford so many I/O pins in traditional off-chip DDR design due to the pin count constraints.

TABLE I
PARAMETER OF 3D DRAM CHIP

# of Tiers	3
Total Capacity	2G bits (256MB)
Clock Frequency	1GHz (Max.)
Refresh Mode	Automatic
Refresh Rate	64ms
# of Data Channel	8
Data Width Per Channel	128 bits
# of Pins Per Channel	294
Burst Length	4 or 8

IV. FRONT-END DESIGN

The comprehensive design flow used in this work is presented in Fig. 3. Analogous to the conventional 2D flow, we simply categorize it into frond-end flow and back-end flow. In this section, we focus on the front-end design while next section is devoted to the back-end design.

A. 3D DRAM Controller

1) Basic Structure: To take advantage of the on-chip DRAM, a custom DDR controller is implemented and integrated into the 3D SoC in the front-end design. As shown in Fig.4, the 3D DRAM controller mainly consists of three functional blocks: AHB wrapper, asynchronous FIFO, and DRAM wrapper. Both AHB wrapper and DRAM wrapper are controlled by finite state machines (FSM) respectively. The state transition in AHB (DDR) wrapper is triggered by AHB (DDR) clock. Note that a second DDR FSM is replicated to enable the parallel access policy that is detailed in next subsection. The address FIFO stores the starting address

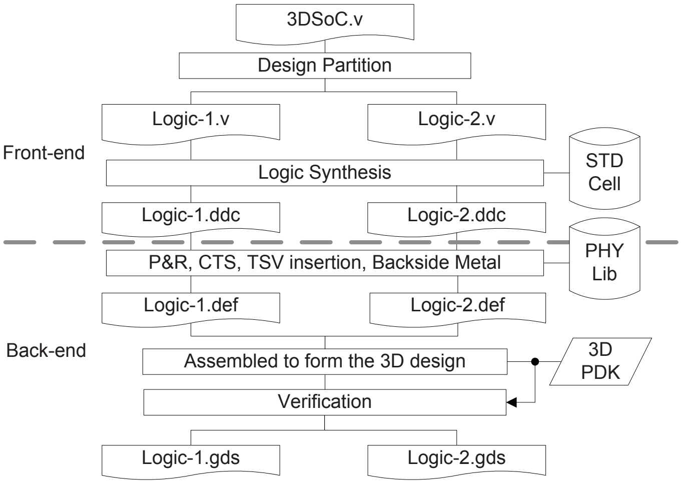

3DSoC.v
Design Partition
Logic-1.v
Logic-2.v
Front-end
Logic Synthesis
STD Cell
Logic-1.ddc
Logic-2.ddc
PHY Lib
P&R, CTS, TSV insertion, Backside Metal
Logic-1.def
Logic-2.def
Back-end
Assembled to form the 3D design
3D PDK
Verification
Logic-1.gds
Logic-2.gds

Fig. 3. 3D SoC Design Flow

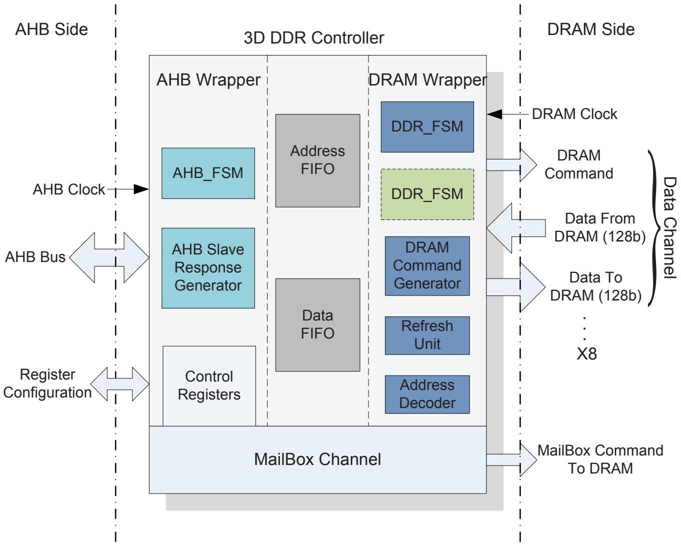

AHB Side
3D DDR Controller
DRAM Side
AHB Wrapper
DRAM Wrapper
DDR_FSM
DRAM Clock
Address FIFO
AHB_FSM
DRAM Command
AHB Clock
DDR_FSM
Data From DRAM (128b)
Data Channel
AHB Bus
AHB Slave Response Generator
DRAM Command Generator
Data FIFO
Data To DRAM (128b)
Refresh Unit
X8
Register Configuration
Control Registers
Address Decoder
MailBox Channel
MailBox Command To DRAM

Fig. 4. Block Diagram of 3D DRAM Controller

and the related transaction information. The data FIFO is composed of write FIFO and read FIFO. We double the write FIFO size to support the WAW parallel access. The simple First-Come-First-Served (FCFS) scheduling scheme is used so that the transaction on the top of FIFO will be fetched into DDR wrapper once it is available.

2) Parallel Access Policy: To make better use of on-chip memory stacking and multiple memory channels, a parallel access policy is developed in DDR controller. The two DDR FSMs can control two memory channels so that two memory accesses can be processed in parallel through each channel.We classify two sequential AHB transactions as: Read-After-Read (RAR), Read-After-Write (RAW), Write-After-Write (WAW), and Write-After-Read (WAR). Based on our understanding, except for WAR, we have the opportunity to optimize other three patterns by allowing the second outstanding transaction to be detected and served without any stall, if they have different bank access requests. For example, Fig.5 shows the case of RAR, where the second read operation follows the

---

[1] In Tezzaron's DRAM technology, a "bank" is equivalent to a "rank" in the conventional DRAM.

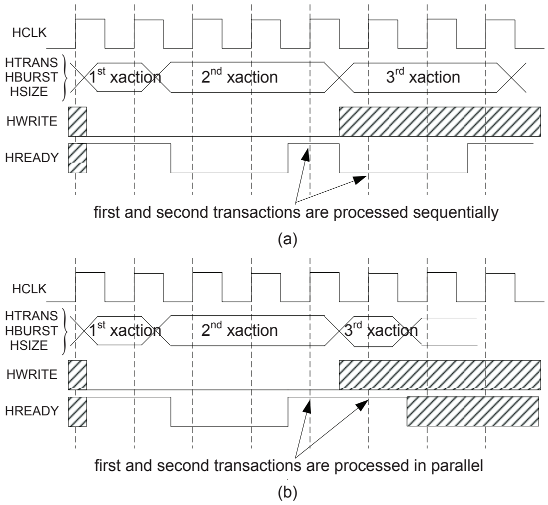

HCLK
HTRANS
HBURST
HSIZE
1st xaction
2nd xaction
3rd xaction
HWRITE
HREADY
first and second transactions are processed sequentially
(a)
HCLK
HTRANS
HBURST
HSIZE
1st xaction
2nd xaction
3rd xaction
HWRITE
HREADY
first and second transactions are processed in parallel
(b)

Fig. 5. Parallel Accessing Policy

first. As shown in Fig.5.(a), without parallel access, the DRAM controller can only process each transaction in sequence even if the second read requests the data from a different bank. As a result, the second read needs to stall for extra cycles before it can be served. In contrast, by enabling parallel access, the second transaction can be processed immediately, which can remove the redundant latency as shown in Fig.5.(b). Sometimes, the first access may suffer a row miss while the second one enjoys a row hit. In this case, considering the in-order property of AHB protocol, the DDR controller should hold the second result until the first access finishes.

B. Logic Partition

1) SoC Partition: The whole SoC architecture is partitioned into two logic tiers. Currently, partitioning is primarily based on the power and area budget of each tier. According to power and area evaluation results, UniCore-II and H.264 encoder, which consume more power and have larger area, are placed on Logic-1, and the rest of components including the DRAM controller are grouped into Logic-2 (shown in Fig. 2). In this way, the hotter tier gets closer to the heat sink, which help mitigate the thermal issue in 3D ICs. After partitioning, the logic synthesis is applied to each tier to generate the gate-level netlists by Synopsys™ Design Compiler® [18].

2) SRAM Partition: In addition to the SoC partition, it is necessary to slice large buffers and/or caches that are built by SRAM into smaller pieces, due to TSV density requirement. From Table. II, the TSV density is $250\mu m \times 250\mu m$, which means at least one TSV should be implanted in such a square area, even if it does nothing (the so-called dummy TSV). Unfortunately, sometimes the area of local buffer may be larger than the density requirement so that it's impossible to place the whole buffer into the tier without the overlapping with TSV. To solve this problem, as shown in Fig. 6, we simply split the single SRAM into mulitple smaller slices, and put them into the blank area.

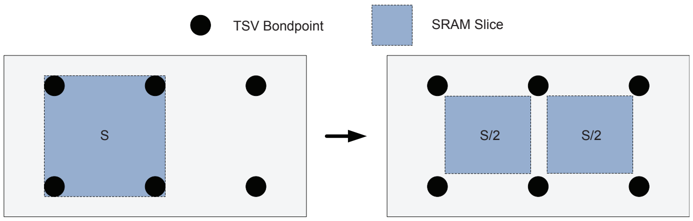

● TSV Bondpoint
SRAM Slice
S
S/2
S/2

Fig. 6. SRAM Partition

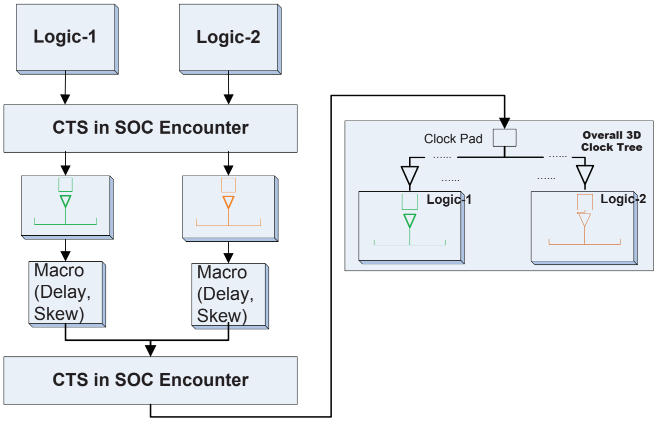

Logic-1
Logic-2
CTS in SOC Encounter
Macro (Delay, Skew)
Macro (Delay, Skew)
CTS in SOC Encounter
Clock Pad
Overall 3D Clock Tree
Logic-1
Logic-2

Fig. 7. Divide-and-Conquer Strategy on CTS

V. PHYSICAL DESIGN

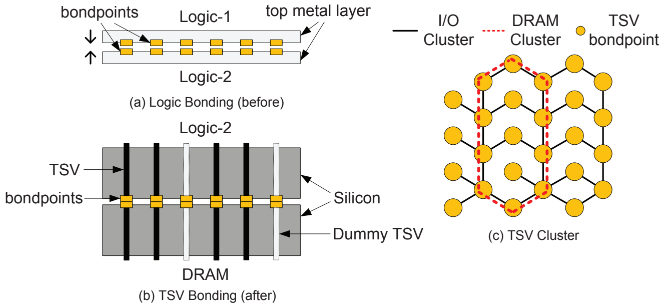
In this section, we concentrate on the physical design flow for 3D DRAM stacking. A 3D PDK developed by North Carolina State University (NCSU) is utilized to help us conceive the back-end flow [19]. As shown in Fig. 3, Divide-and-Conquer methodology is applied to the back-end flow. The two netlists are imported into SoC Encounter to run P&R, CTS, and TSV/Backside Metal Insertion, respectively. As soon as both tiers complete all these steps, the layouts of two tiers are reassembled for the 3D DRC/LVS checking. The GDS files are generated by Cadence™ Virtuoso® as the final outputs of back-end design flow [20].

A. Divide and Conquer Methodology

Divide-and-Conquer methodology is used by applying traditional 2D physical design methods to each tier. The floorplanning, placement, wire routing as well as clock tree generation are all done within a single tier. As an example, Fig. 7 illustrates the CTS flow. Both Logic-1 and Logic-2 conducts the conventional CTS by SoC Encounter. In the figure, the source clock from I/O pad is firstly delivered onto Logic-2 and further propagates to Logic-1 through the logic bondpoint covered in next paragraph.

B. 3D Bonding

Two bonding techniques are involved in this project. The connectivity of Logic-1 and Logic-2 is realized with face-to-face microbump bonding technology, while TSV and backside metal allow for back-to-back bonding between Logic-2 and DRAM.



(a) Logic Bonding (before)

(b) TSV Bonding (after)

(c) TSV Cluster

Fig. 8. 3D Bonding Technologies

1) Logic Bonding: Two logic tiers are stacked face-to-face as shown in Fig. 8.(a). The top metal layer is predefined and no modification is allowed to guarantee the alignment. Additionally, we take the shared-bus design into account to make it much easier to combine two tiers. The shared bus reveals an advantage in the bonding due to the characteristics of standardization and centralization. Without the shared bus, inter-tier communication is most likely distributed rather than centralized since the correponding functional components may scatter over the entire chip. The distributed inter-tier communication incurs more efforts on logic bonding to achieve the accurate alignment and thus introduces more complexity and potential failures. In contrast, shared bus offers centralized inter-tier connection, which can simplify the bonding process. In this work, AHB is placed in the center of Logic-2 by allocating two sets of AHB master interfaces for bonding (Fig. 10). Meanwhile, both AHB interfaces of Unicore-II and H.264 encoder are correspondingly placed in the center of Logic-1 so that they can adhere to Logic-2 directly.

2) TSV Bonding: TSVs are inserted between Logic-2 and DRAM as signal carriers. As presented in Fig. 8.(b), all of TSVs must be capped with backside metal bondpoints on one end. The backside metal bondpoint then connects to the bondpoint on DRAM with back-to-back bonding. Bunches of dummy TSVs are also inserted into Logic-2 to meet the requirement of TSV density. In addition, to achieve high reliability, multiple TSVs are aggregated to form a TSV cluster for signal delivery. Two types of TSV clusters are employed to deliver data as well as power. *DRAM Cluster* is utilized for dedicated DRAM traffics, while *I/O Cluster* is used to connect to the I/O pads on the surface of DRAM. Each DRAM cluster contains 10 TSVs and each I/O cluster contains 26 TSVs (Fig. 8.(c)). Table II lists the physical characteristics and the total number of effective TSVs (excluding dummy TSV) used in this work.

TABLE II
TSV PARAMETERS

Diameter	1.2μm
Pitch	4μm
Depth	6μm
Density	$250\times250\mu$m^2
# per DRAM Cluster	10
# per I/O Cluster	26
Total Number	16,924

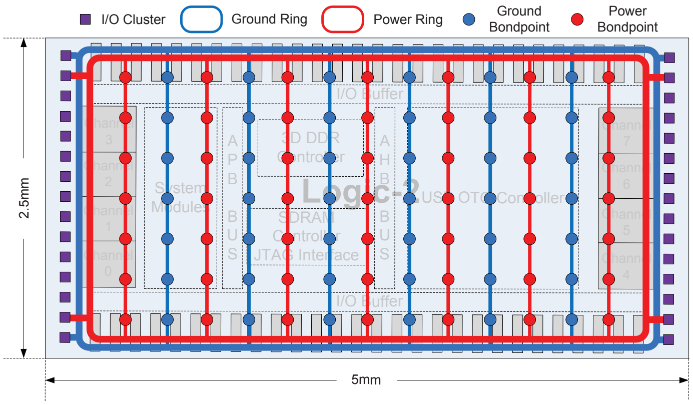


Fig. 9. Power Delivery Network on Logic-2

C. Power Delivery

Like the data signals, all power/ground signals are delivered from power I/O to tier Logic-2 first through I/O clusters. Fig. 9 illustrates the power delivery network on Logic-2. As shown in the figure, regular power/ground rings are generated along the edges of logic tiers. To deliver power to Logic-1 layer, multiple power/ground rows are implemented on both Logic-1 and Logic-2. One power/ground row is composed of an array of bonding points in both top metal layers. The interleaved power/ground rows cover the whole chip to meet the density requirement of the top vias.

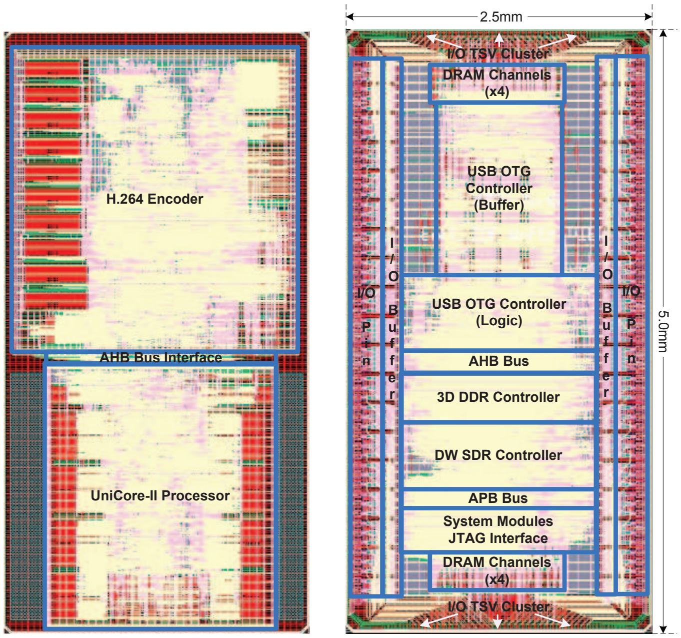
VI. THE 3D SOC CHIP PROTOTYPING

The prototype chip has been fabricated with GlobalFoundries' 130nm low-power process together with Tezzaron's TSV bonding technology. Table. III summarizes the design result. The supply voltage of this chip is 1.5V. Fig. 10 shows the layout of logic tiers. Stacking on the 3D DRAM chip, the DRAM controller can communicate with DRAM via eight data channels. In Fig.10.(b), eight DRAM channels are divided into two groups and placed on the top and at the bottom of tier Logic-2, respectively. Half of them, however, are used in practice due to the area limitation and routing complexity. The lack of PLL results in a single clock to drive the chip. Considering the quality of the input clock, the frequency of the whole SoC is set at 60MHz, which is sufficient to run desired multimedia applications.

TABLE III
SYSTEM DESIGN SUMMARY

Area	6.2mm^2 (Logic-1) 7.3mm^2 (Logic-2)
Power	101mW (Logic-1) 70mW (Logic-2)
Clock Freq.	60MHz
Temperature	45°C
Supply Voltage	1.5V
# of Data Pad	61
# of P/G Pad	128
# of DRAM Channels	4

In addition, we evaluate the timing, area, and power of DRAM controller by Design Compiler. Table. IV shows the simulation result. The DRAM controller consumes $0.78mm^2$ and can run at 133MHz, which can be translated to 4.25GB/s



(a) Logic-1 (b) Logic-2

Fig. 10. Layout of Two Logic Tiers

bandwidth without any optimization. With the parallel access policy, the channel utilization is improved and the peak bandwidth can be doubled to 8.5GB/s. Obviously, the bandwidth can be further improved if certain optimizations, such as replacing AHB with AXI and implementing multiple memory controllers, can be adopted to further leverage channel independence.

TABLE IV
DRAM CONTROLLER SIMULATION RESULT

Area	0.78mm^2
DDR Clock Freq.	133MHz
Power	12.57mW
Data Bandwidth	4.25GB/s (w/o Parallel Access) 8.5GB/s (with Parallel Access)

VII. CONCLUSION

In this paper, we demonstrate the feasibility of 3D memory stacking by building a 3D SoC for multimedia applications that can leverage the high memory bandwidth offered by 3D integration. The DRAM memory stacking mitigates the I/O pin count limitation so that it is possible to support as many as 8 independent channels. To take advantage of multiple channels, we develop a memory controller with parallel access policy that allows two access requests to be processed in parallel through two channels. In addition, both front-end and back-end design flow are discussed to exhibit the differences from conventional 2D flow. The chip has been fabricated with GlobalFoundries 130nm process and Tezzaron's 3D TSV technology.

ACKNOWLEDGEMENTS

This project is supported in part by NSF grant 0903432, 0643902 and 0702617, as well as SRC grants. The chip fabrication is supported by DARPA and Tezzaron. Special thanks to Gretchen Patti from Tezzaron for providing the design guidance. Also, thank the NCSU team for the support on design kits. Finally, we also thank other colleagues involved in this 3D IC MPW run for their constructive suggestions.

REFERENCES

[1] Y. Xie, J. Cong, and S. Sapatnekar, *Three-Dimensional Integrated Circuit Design: EDA, Design and Microarchitecture*. Springer, 2010.

[2] C. Chiang and S. Sinha, "The Road To 3D EDA Tool Readiness," in *Proceedings of the Asia and South Pacific Design Automation Conference*, January 2009, pp. 429–436.

[3] H.-H. S. Lee and K. Chakrabarty, "Test Challenges for 3D Integrated Circuits," *IEEE Design and Test*, vol. 26, no. 5, pp. 26–35, 2009.

[4] X. Dong and Y. Xie, "System-Level Cost Analysis and Design Exploration for Three-Dimensional Integrated Circuits (3D ICs)," in *Proceedings of the Asia and South Pacific Design Automation Conference*, January 2009, pp. 234–241.

[5] JEDEC Solic State Technology Association, "JEDEC Standard: DDR3 SDRAM Specification," September 2009.

[6] A. Fontanelli, "System-in-Package Technology: Opportunities and Challenges," *International Symposium on Quality Electronic Design*, vol. 0, pp. 589–593, 2008.

[7] International Technology Roadmap for Semiconductors, 2009.

[8] S. Babinetz, "Wire Bonding Solutions for 3-D Stacked Die Packages," http://www.kns.com/library/EME_Stacked_Die_Babinetz.pdf.

[9] G. H. Loh, "3D-Stacked Memory Architectures for Multi-core Processors," in *Proceedings of the International Symposium on Computer Architecture*, June 2008, pp. 453–464.

[10] H. Saito, M. Nakajima, T. Okamoto, Y. Yamada *et al.*, "A Chip-Stacked Memory for On-Chip SRAM-Rich SoCs and Processors," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 1, January 2010.

[11] D. H. Woo, N. H. Seong, D. L. Lewis, and H.-H. S. Lee, "An Optimized 3D-Stacked Memory Architecture by Exploiting Excessive, High-Density TSV Bandwidth," in *Proceedings of the International Conference on High Performance Computer Architecture*, January 2010.

[12] Y. Kim, D. Han, O. Mutlu, and M. Harchol-Balter, "ATLAS: A Scalable and High-performance Scheduling Algorithm for Multiple Memory Controllers," in *Proceedings of the International Conference on High Performance Computer Architecture*, January 2010.

[13] ARM, "AMBA Specification," May 1999.

[14] J.-W. Chen, C.-Y. Kao, and Y.-L. Lin, "Introduction to H.264 Advanced Video Coding," in *Proceedings of the Asia and South Pacific Design Automation Conference*, January 2006.

[15] "Introduction to USB On-The-Go," http://www.usb.org/developers/onthego/USB_OTG_Intro.pdf.

[16] X. Cheng, X. Wang, J. Lu, J. Yi *et al.*, "Research Progress of UniCore CPUs and PKUnity SoCs," *Journal of Computer Science and Technology*, vol. 25, no. 2, pp. 200–213, March 2010.

[17] Tezzaron, http://www.tezzaron.com.

[18] Synopsys, http://www.synopsys.com.

[19] P. D. Franzon, R. W. Davis, M. B. Steer, S. Lipa *et al.*, "Design and CAD for 3D Integrated Circuits," in *Proceedings of Design Automation Conference*, June 2008, pp. 668–673.

[20] Cadence, http://www.cadence.com.

[21] JEDEC Solic State Technology Association, "JEDEC Standard: DDR2 SDRAM Specification," April 2008.

[22] P. Garrou, C. Bower, and P. Ramm, *3D Integration: Technology and Applications*. Wiley-VCH, 2008.

[23] Tezzaron, "MPW Design Guide," 2010.

[24] G. H. Loh, Y. Xie, and B. Black, "Processor Design in 3D Die-Stacking Technologies," *Journal of IEEE Micro*, vol. 27, no. 3, pp. 31–48, 2007.

[25] T. Thorolfsson, K. Gonsalves, and P. D. Franzon, "Design Automation for a 3DIC FFT Processor for Synthetic Aperture Radar: A Case Study," in *Proceedings of Design Automation Conference*, July 2009, pp. 51–56.

[26] Tezzaron, "Preliminary Specification for 8-port Memory," 2010.

[27] Y. Pan and T. Zhang, "Improving VLIW Processor Performance Using Three-Dimensional (3D) DRAM Stacking," in *Proceedings of IEEE International Conference on Application-specific Systems, Architectures and Processors*, July 2009, pp. 38–45.

[28] S. Vangal, J. Howard, G. Ruhl, S. Dighe *et al.*, "An 80-Tile 1.28TFLOPS Network-on-Chip in 65nm CMOS," in *IEEE International Solid-State Circuits Conference*, Febuary 2007, pp. 98–589.