On the design of a 3D optical interconnected memory system

Kang Wang¹, Huaxi Gu¹a), Yintang Yang², Ke Chen¹, and Kun Wang³

¹ State Key Laboratory of ISN, Xidian University, Xi’an, China
² School of Microelectronics, Xidian University, Xi’an, China
³ School of Computer Science and Technology, Xidian University, Xi’an China
a) hxgu@xidian.edu.cn

Abstract: In this letter, a multi-rank parallel accessing memory system based on 3D stacking and optical interconnection technologies is proposed. This proposed memory system can be a promising solution to future many-core system’s memory accessing bottleneck. Simulation results show that the proposed memory system yields the latency reduction of transactions and enhancement of bandwidth effectively. The worst case latency of 4 ranks and 8 ranks proposed memory system can be 1/3 and 1/5 that of the electronic bus-based ones. Bandwidth enhancement of 4 ranks and 8 ranks proposed memory system can be 3.5 times and 4 times higher than that of the traditional bus-based ones.

Keywords: optical interconnect, memory system, 3D stacking
Classification: Fiber optics, Microwave photonics, Optical interconnection, Photonic signal processing, Photonic integration and systems

References
[1] B. Jacob, S. W. Ng and D. Wang: Memory System: Cache, DRAM, Disk (Elsevier, Burlington, 2008) 1st ed. 413.
[2] P. V. Mejia, R. Amirtharajah, M. K. Farrens and V. Akella: ACM/IEEE NOCS Int. Symp. (2010) 215. DOI:10.1109/NOCS.2010.31
[3] Y. Xie: DATE (2013) 964. DOI:10.7873/DAT.2013.202
[4] S. Beamer, C. Sun, Y. J. Kwon, A. Joshi and C. Batten: SIGARCH Comput. Archit. News (2010) 129.
[5] D. Brunina, D. Liu and K. Bergman: IEEE J. Sel. Top. Quantum Electron. 19 (2013) 3700407. DOI:10.1109/JSTQE.2012.2224096
[6] P. Rosenfeld, E. Cooper-Balis and B. Jacob: IEEE Comput. Archit. Lett. 10 (2011) 16. DOI:10.1109/L-CA.2011.4

1 Introduction

In traditional electrical bus-based memory system, only can one rank communicate with memory controller through the electric bus at one time. Fig. 1(a) shows a traditional bus-based memory system. Memory modules working in lockstep are
linked to from a rank [1]. When the number of processors increasing, more processors will attempt to communicate with memory at one time. The bottleneck of bandwidth becomes more severe under this circumstance. When the memory storage capacity increasing, more ranks will be linked to this electric bus-based architecture, which induces higher accessing latency. Chip pin-bandwidth density cannot be increased dramatically in the future and one-layer design brings huge area consumption. The electrical bus-based memory system cannot fulfill future many-core computing system.

![Diagram](image)

**Fig. 1.** (a) The traditional bus-based one. (b) The proposed 3D optical interconnected one.

Future memory system should provide high bandwidth processor-memory communication [2]. 3D stacking and high-speed low-power CMOS-compatible photonic devices can provide a new way to build an optical interconnected memory system [3]. A redesigned DRAM memory system using monolithically integrated silicon-photonic technology is proposed [4]. But in [4], the usage of photonic devices deep into every DRAM array faces more technology challenges compared with adding an optical layer into a DRAM chip. A 4 × 4 optical crossbar network is investigated linking microprocessors with their associated main memory [5], this could connect processors with memory optically, but some links in the crossbar is wasted because one memory module will not send/receive data to/from another memory module inside a memory system.

In this letter, we design an optical interconnected layer and integrate it into a memory system. Dense wavelength-division multiplexing (DWDM) allows multiple wavelength to share the same waveguide. The enhancement in processor-memory bandwidth and reduction in memory accessing latency can be achieved by multi-rank parallel accessing property in this 3D stacking memory system.

## 2 Architecture

The proposed 3D stacking optical interconnected memory system consists of one optical layer and several memory layers. Fig. 1(b) shows an example of the proposed memory system’s 3D architecture. Memory modules in the same rank are placed at different memory layers vertically and linked with Opti. Receiver and Opti. Transmitter on the optical layer through through-silicon vias (TSVs).
On the optical layer, there are Opti. Receiver, Opti. Transmitter, waveguides and broadband micro-resonators (BMRs). Waveguides connect memory system with on-chip processors through off-chip fibers. Processors send optical memory transactions through one of these waveguides according to the wavelengths usage. That means if one wavelength in a certain waveguide is being used at the moment, the processor can use either another wavelength in the same waveguide or the same wavelength in another waveguide. The competition of accessing to the same rank is reduced. Opti. Receiver and Opti. Transmitter use a series of narrowband micro-resonators (NMRs) to demodulate/modulate optical memory transactions from/to waveguides. Control logics and buffers in Opti. Receiver and Opti. Transmitter ensure the communication between optical layer and memory layers is in order. As memory transactions (including data) are transmitted in a certain wavelength optically, deserialization in Opti. Receiver and serialization in Opti. Transmitter will be carried out during O/E and E/O process. A proposed n ranks memory system needs n Opti. Receivers, n Opti. Transmitters, n waveguides and n types \(2^n\) BMRs. These BMRs have different coupling wavelengths to serve the communication. If \(m\) wavelengths can be propagated in one waveguide, then each BMR should couple in \(m/n\) wavelengths. The coupling wavelengths of different types of BMRs are totally different. In order to realize the multi-rank parallel communication property, a BMRs mapping matrix is designed. The types of BMRs used on different waveguides for each rank’s communication can be mapped into the following \(n \times n\) matrix:

\[
\begin{pmatrix}
MR_1 & MR_2 & MR_3 & \cdots & MR_n \\
MR_n & MR_1 & MR_2 & \cdots & MR_{n-1} \\
MR_{n-1} & MR_n & MR_1 & \cdots & MR_{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
MR_2 & MR_3 & MR_4 & \cdots & MR_1
\end{pmatrix}
\]

(1)

\[MR_i : \Lambda_i = \{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{in}\} i = 1, 2, \ldots, n\]  
(2)

\[\Lambda_i \cap \Lambda_j = \emptyset (i \neq j; \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, n)\]  
(3)

In this matrix, columns mean different ranks and rows mean different waveguides. For example, element at the intersection of column 3 and row 3 is MR1, which means the type of BMRs used for rank 3’s communication on waveguide 3 is MR1. Fig. 2(a) shows an example of an optical layer which is used for 4 ranks communication. In this example, 4 groups of Opti. Receiver and Opti. Transmitter, 4 waveguides (labelled as W1, W2, W3 and W4) and 4 types 32 BMRs are used.

Memory modules are placed on memory layers. In order to utilize optical communication, a new memory module is designed. In a memory module, Receiver (R) and Transmitter (T) are the interfaces to optical layer’s Opti. Receiver and Opti. Transmitter. Transaction Control Logic (TCL) are placed inside the memory module. TCL receives the memory transaction, and translates the logic address into physical address and processes it. The process differs according to the type of transaction. Data Register and Driver (DRD) receives data from TCL or DRAM Array and sends acknowledge (ACK) to on-chip processors. Command Control
Logic (CCL) conforms data’s transmission in DRD and keeps the right sequence of data’s accessing into/out of memory array based on accessing time constrain. In case of storage mistakes, CCL sends extra commands such as precharge command or refresh command as well. Row Decoder and Column Decoder can get row and column address from CCL and locate the position in DRAM Array where data is stored. Fig. 2(b) shows the design of the memory module.

3 Communication method

It is assumed that memory transactions sent to memory system are optical signals. Write transactions are sent together with write data. The proposed architecture ensures wavelengths between different ranks are not mutual interference and they can be coupled into different ranks by BMRs independently.

A memory transaction sent to memory system through the waveguides will be coupled into a rank’s Opti. Receiver which will register the wavelength it used and convert it to electronic signal. When data with ACK or just an ACK needs to be sent back to processor, the registered wavelength will be used again in Opti. Transmitter. After being converted to electronic signal, memory transaction will be sent to memory module and processed depend on the type of transaction. There are two types of transactions: a read transaction and a write transaction with data.

During the rank processing time, logic address in a read transaction will be translated into physical address in TCL and then sent to CCL. CCL will schedule the accessing to DRAM Array according to the array current situation and timing constrains. Data and the ACK signal will be sent to Transmitter from DRD only when it gets the conformation from the CCL.
As for a write transaction with data, TCL will separate the data and the transaction additionally. The data will be sent to DRD, waiting for CCL’s permission. The transaction will be processed in TCL, CCL by order and finally generate the row and column address. Row Decoder and Column Decoder get the row address and column address. When permission is confirmed, data can be sent to DRAM Array. After the data is stored in DARM Array correctly, an ACK signal will be sent from DRD to Transmitter.

4 Simulation

In this section, the latency and bandwidth enhancement are simulated compared with the traditional bus-based structure. DRAMSim2 [6] is used in this work, which is a cycle accurate memory system simulator. This simulator can simulate the usage and release of resources on DRAM modules in a cycle-accurate basis.

In this simulation, the total memory storage is 4096 MB (64 bits width) built with 2 types of memory modules, which are 256 MB (16 bits width) module and 128 MB (16 bits width) module. Parameters of these two module are shown in Fig. 3(c). Row buffer policy is open page. The max number of open page is 4. In order to get a more realistic performance, we use trace file which has 695520 transactions from application to support this simulation. This trace file includes the sending cycle, the types and the logic address of a memory transaction. Fig. 3(a) shows the comparison of 4 ranks traditional memory system and the proposed memory system built with 256 MB (16 bits width) memory module. Fig. 3(b) shows the results of 8 ranks traditional memory system and the proposed memory system built with 128 MB (16 bits width) memory module.

Fig. 3. Traditional bus-based and proposed optical interconnected memory system. (a) 4 ranks comparisons. (b) 8 ranks comparisons.
The worst case latency of the proposed 4 ranks and 8 ranks memory system shown in Fig. 3(a) and Fig. 3(b) can be $1/3$ and $1/5$ that of the comparison traditional memory system respectively. It can be found in the simulation that when optical interconnection is used, large number of transactions can be executed and sent back to processors with low latency. Compared with the optical one, traditional bus-based structure shows a long tail feature, which means several transactions will be completed after a very long time. When the worst case comes, the long tail latency of traditional one can be several thousand cycles, which dramatically influences the performance of memory system. The average bandwidth results of two types of proposed optical interconnected memory systems are shown in Fig. 3(d). Bandwidth enhancement of a 4 ranks and 8 ranks optical interconnected memory system can be nearly 3.5 times and 4 times higher than that of the traditional ones. These simulations show that the property of multi-ranks parallel accessing brought by optical interconnection can enhance the bandwidth and reduce the latency of memory system significantly.

Acknowledgments

This work is supported by the National Science Foundation of China Grant No. 61472300 and 61334003, the Fundamental Research Funds for the Central Universities Grant No. K5051301003, Shenzhen Research Funding No. 20130401171935815, the 111 Project Grant No. B08038.