An Energy-Efficient Optically Connected Memory Module for Hybrid Packet- and Circuit-Switched Optical Networks

Daniel Brunina, *Student Member, IEEE*, Dawei Liu, *Student Member, IEEE*, and Keren Bergman, *Fellow, IEEE*

***Abstract*—Demands on future data centers and high-performance computing systems will require processor-memory interconnects with greater performance and flexibility than can be provided by existing electronic interconnects. The bandwidth density and bit-rate transparency offered by optical systems are uniquely suited to address these challenges facing memory interconnects. We, thus, investigate a hybrid packet- and circuit-switched optical interconnection network linking microprocessors with their associated main memory, which can simultaneously reduce memory access latency and improve energy efficiency performance. This novel hybrid approach allows low-bandwidth memory control data and small memory transactions to be efficiently transmitted as wavelength-striped optical packets, while long bursts of memory accesses are optically circuit switched. In this study, we experimentally demonstrate an optically connected memory system in which a microprocessor accesses multiple 80-Gb/s memory modules all-optically across a hybrid packet- and circuit- switched optical network. Error-free communication between the microprocessor and main memory is confirmed (bit-error rates less than 10^{-12}) with the optical network providing low-memory access latencies. The overall memory system reduces energy consumption by 28%.**

***Index Terms*—Circuit switching, optical communications, optics in computing, packet switching, photonic switching systems.**

I. Introduction

NEXT-generation large-scale high-performance computing (HPC) and data center systems will require microprocessors to support unprecedented off-chip bandwidths to memory, with low-access latencies and interconnect power dissipation. However, today's electronic interconnects face performance challenges with low bandwidth densities, as well as distance- and data-date-dependent energy dissipation. As a result, large-scale systems have experienced an exponentially growing performance gap between the computational performance of microprocessors and the performance of off-chip main memory systems [1]. This communications bottleneck will undoubtedly limit the overall system performance and scalability of future large-scale systems. Due to the tradeoffs among the communication bandwidth, latency, and energy efficiency requirements, high-performance microprocessors will be starved for memory data [2], [3].

Although it is feasible for electronic interconnection networks to reach per-channel data rates up to 25 Gb/s [4], the power dissipation at such high bandwidths becomes overwhelming and contributes greatly to increased the overall system cost and complexity. Scaling interconnect performance using traditional approaches will continue to exacerbate this imbalance. For example, a typical main memory system consists of multiple chips of synchronous dynamic random access memory (SDRAM) packaged together onto a circuit board called a dual in-line memory module (DIMM), which is capable of providing over 120 Gb/s of peak bandwidth [5]. Multiple DIMMs must be accessed in parallel, requiring an extremely complex electronic bus, to provide the many terabits-per-second of memory bandwidths required by data-intensive applications. Consequently, modern servers are estimated to dissipate 20% of their energy in the memory system alone [6]. The scaling challenges facing electrical interconnects limit the number of DIMMs that can be accessed, and consequently, the total memory bandwidth. Increasing the per-channel SDRAM data rate has been attempted [7], but the resulting system remains limited in use due to its significantly higher energy consumptions.

Further, in the latest generation of DIMMs, the memory module's energy consumption is reduced by using "sleep" states, during which the data buffer and transceivers at each node are powered down when not in use [5]. However, the high latency associated with each DIMM entering or exiting its sleep state can add significant overhead to each memory access. If not managed efficiently, the additional latency can considerably reduce the overall system performance. Therefore, it is necessary to capitalize on innovative interconnect technologies and architectures to redesign processor-to-memory communication.

The use of photonic technology can enable high-bandwidth links, with novel functionalities to reduce off-chip data access latency and power dissipation [8]. The integration of on-chip silicon photonic transceivers [9] will further enable processor-memory communication with the off-chip bandwidth and energy efficiency performance equal to that of on-chip communications [10], [11]; this would be impossible using conventional electronic interconnects. In addition to achieving high per-channel data rates, optical interconnects can significantly improve communication bandwidths through wavelength-division

Manuscript received June 1, 2012; revised September 12, 2012; accepted October 7, 2012. Date of publication October 11, 2012; date of current version April 3, 2013. This work was supported in part by the U.S. Department of Energy under grant DE-SC0005114 and in part by the Samsung Advanced Institute of Technology GRO program.

The authors are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: daniel@ee.columbia.edu; dl2570@columbia.edu; bergman@ee.columbia.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTQE.2012.2224096

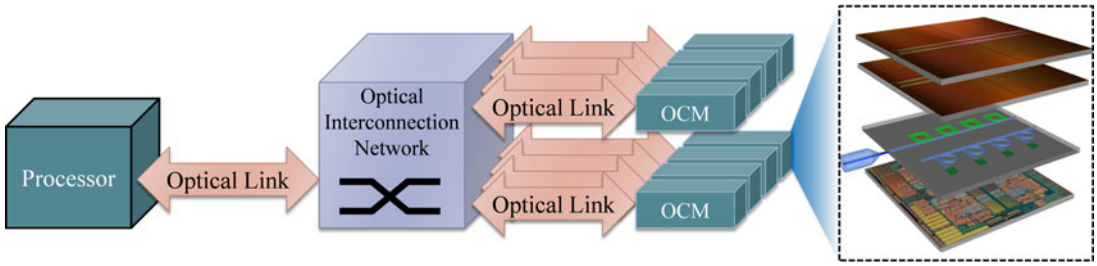

Fig. 1. Block schematic depicting how a next-generation processor can be connected to many OCMM across an optical interconnection network. Inset illustrates a memory module that uses 3-D stacking to integrate SDRAM with photonic transceivers and associated driver circuitry.

multiplexing (WDM), and can therefore support many terabits-per-second of optical bandwidth using a single waveguide or optical fiber [12]. Leveraging the bandwidth density and distance immunity of optics will alleviate pin-count constraints in microprocessors and enable each optically connected memory module (OCMM) to be more physically distant from the processor, thus yielding the potential for a greater number of OCMMs to be connected and accessed in parallel (see Fig. 1). The latency of these links are dictated purely by the fiber's time-of-flight latency [13], allowing efficient, transparent optical networks to provide low memory access latency.

Thus, optical interconnection networks comprise an attractive solution to the communication bottleneck within future large-scale computing systems [14]–[22]. Memory interconnect architectures are especially well suited for the deployment of optical networks, owing to the performance and energy requirements of main memory systems, as well as the necessary flexibility within a network to support potentially diverse and unpredictable traffic patterns. For example, a packet-switched optical network can provide low memory access latency for short messages, while a circuit-switched optical network delivers greater performance for longer messages (e.g., in a streaming application) [17]. The ideal optical interconnect for microprocessor-to-main-memory communication is therefore a hybrid packet- and circuit-switched network. In this novel hybrid approach, memory accesses exceeding a predetermined size threshold use circuit-switched lightpaths, while all other accesses are packet switched. In this way, an optically connected memory (OCM) system can be constructed with greater performance and capacity, while achieving lower memory access latencies and reduced power consumptions.

In this study, we experimentally demonstrate the first hybrid packet and circuit-switched OCM system. We implement a field-programmable gate array (FPGA)-based microprocessor that communicates with three OCMMs across a wavelength-striped 4×4 hybrid packet and circuit-switched optical network. The processor and OCMMs create wavelength-striped memory messages using eight 10-Gb/s electronic transceivers; eight separate wavelength channels are modulated and then combined using WDM to generate 80-Gb/s messages. The resulting OCM system achieves 240-Gb/s aggregate memory bandwidths through the optical network (80 Gb/s per network port). Here, we expand on previous work [18] by developing a custom high-performance OCMM, as well as a custom memory controller that is capable of issuing both packet and circuit memory accesses. The system optimizes communication for each desired transaction size to support a diverse range of applications. This allows small memory accesses and control data to be routed efficiently as packets, and large memory accesses to utilize circuit switching. The energy-efficient OCMMs can also enter a low-power "sleep" state in which the SDRAM and on-board transceivers consume minimal power, and rapidly reenter normal operation when required. The OCMM "sleep state" leverages the low-power mode of DDR3 SDRAM [5] and uses the memory controller's knowledge of memory communication to efficiently place the memory and optical link into low-power states, which impacts the operation of the optical network and reduces the overall system power consumption.

II. Hybrid Packet and Circuit Optical Network

In this implementation, the optical network testbed is comprised of a 4×4 nonblocking switching node that is capable of simultaneously routing optical packets and circuit lightpaths between any of the four input and output ports. This 4×4 optical interconnection network [see Fig. 2(a)] uses 16 semiconductor optical amplifiers (SOAs) as photonic switching elements to transparently route high-bandwidth, wavelength-striped optical messages with nanosecond switching times [23]. The SOAs are controlled by a Xilinx Virtex-V FPGA, which manages output contentions and processes the appropriate packet- or circuit-switching routing information. The packet- and circuit-switching protocols differ primarily in that circuit switching utilize an electronic control plane and packet switching is performed all optically through the use of dedicated header wavelengths.

The optically switched packets use a wavelength-striped message format [see Fig. 2(b)] in which multiple data payload wavelengths are combined with low-speed network control wavelengths using WDM. The network control wavelengths, or *headers*, are modulated at the packet rate, such that each header's value remains constant for the duration of the packet. With one header bit per wavelength and one value per packet duration, this implementation simplifies the routing logic for a small number of purely combinational logic gates and guarantees minimal switching times. Here, we require three header wavelengths: a *frame* signal to indicate the presence of a packet and two *address* bits (in order to map four distinct output ports). The number of required header wavelengths scales efficiently as log (N) for an N $\times$ N network. Each header wavelength is

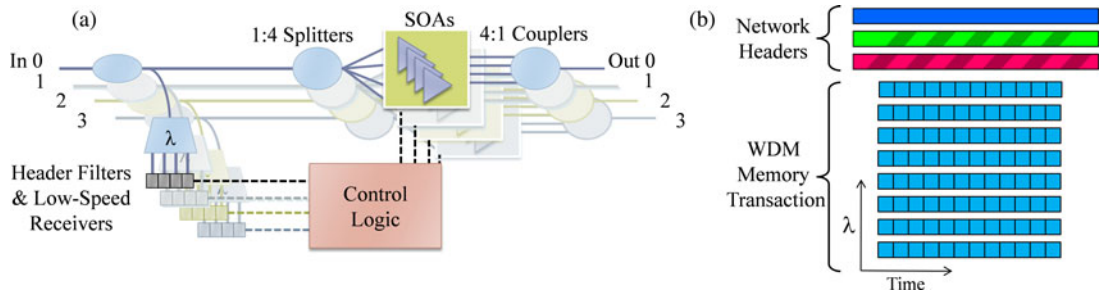

Fig. 2. (a) Architectural block diagram of 4 × 4 optical network testbed. (b) Wavelength-striped memory packet format.

received by a low-speed 155-Mb/s p-i-n receiver with a transimpedance amplifier (TIA) and limiting amplifier (LA), which passes the received header data to the FPGA and enables the rapid switching ON/OFF of the SOAs. Simultaneously, the high-speed data payloads are delayed with a fiber delay line (FDL) that matches the header processing time, which results in transparent just-in-time routing of the optical payload. The total transit time through the 4 × 4 optical network is approximately 40 ns, including the passive optical filters to receive the header wavelengths and the FDLs. The primary source of transit latency through the switch is the time-of-flight through the off-the-shelf components that comprise the testbed, and therefore a future, more integrated switch could achieve significantly lower latencies. Additionally, integration with high-speed logic instead of discrete FPGAs could drastically reduce header processing time for a total switch latency of less than 1 ns.

The 40-ns latency in this study is primarily due to the time-of-flight latency within the experimentally demonstrated testbed, which contains several meters of optical fiber, and is not a physical limitation of the system design. The time required to process the headers is negligible (a few logic gates) and SOAs can be switched in a nanosecond. Future, more integrated systems could significantly reduce the total latency by reducing time-of-flight through the switch, using faster logic gates

Circuit switching is performed by means of an electronic control plane and central arbitration. Network nodes signal to the central arbiter that a circuit path is required through the 4 × 4 optical network. To establish a circuit-switched link, the FPGA maintains a lightpath through one of the 16 SOAs for the duration dictated by the central arbiter. The central arbiter is implemented as dedicated arbitration logic at the FPGA within the switching node. This allows the central arbiter to monitor existing packets within the network and allocate circuit paths without interrupting packets already in transit. Here, only the processor must communicate with the central arbiter so that both processor-to-memory and memory-to-processor circuit paths are controlled by the processor. The OCMMs will only transmit data when directed by the processor, and therefore the processor is aware of all necessary memory transmissions. This is analogous to existing electrically connected memory, in which memory devices are directly managed by a memory controller and will only act in response to explicit instructions from a processor and memory controller. Leveraging this communication protocol guarantees compatibility with commercial memory devices and simplifies both the OCMMs and the arbiter, which in turn reduces overall network complexity and memory access latency.

Future, large-scale interconnection networks, such as optical Omega networks [24], can be created using multiple of the demonstrated 4 × 4 optical switching nodes. The data payloads implemented here consist of eight separate wavelength channels, each modulated at 10 Gb/s, combined into a single 80-Gb/s WDM message [see Fig. 2(b)]. However, the SOA-based switching nodes have been demonstrated to support 40-Gb/s per-channel data rates [25] while supporting hundreds of thousands of network ports [26], and can therefore be deployed in the next generation of HPCs and data centers. For comparison to equivalent electrical links, the optical network has been experimentally measured to consume 1.58 W per port while a 32-port 10-Gb/s electrical switch may require 10 W per port [27]. Individual 10-Gb/s transceivers consume approximately 1 W each [27]. Future OCM implementations can leverage recent advances in silicon photonics [9], [28] to create integrated switches capable of equivalent performance but with much lower transit latencies and power consumption.

III. Experimental Setup and Results

A. Experimental Setup

The experimental demonstration characterizes the performance and efficiency of the proposed hybrid packet- and circuit-switched memory access protocol. We eliminate the power-hungry processor-memory electronic bus and leverage the unique functionalities of our optical interconnection network to offer energy-efficient OCMMs. Each OCMM is accessed all optically and transparently across an implemented optical network using either packet or circuit switching, depending on the memory transaction message sizes. Using circuit switching for smaller memory transactions results in inefficient use of network resources and does not adequately amortize the circuit path setup latency [17]. Smaller messages utilize the previously described wavelength-striped packet switching.

The OCMMs can enter a low-power "sleep" state in which the SDRAM input and output buffers are disabled and the high-speed optical transceivers are idle (not transmitting any data). In this state, the SDRAM consumes only 20% of its normal operating power [5], [29] and the transceiver logic consumes only minimal static power. Here, the memory controller transmits short optical packets to command an OCMM to enter or exit its sleep state; the transition requires less than 10 ns [5].

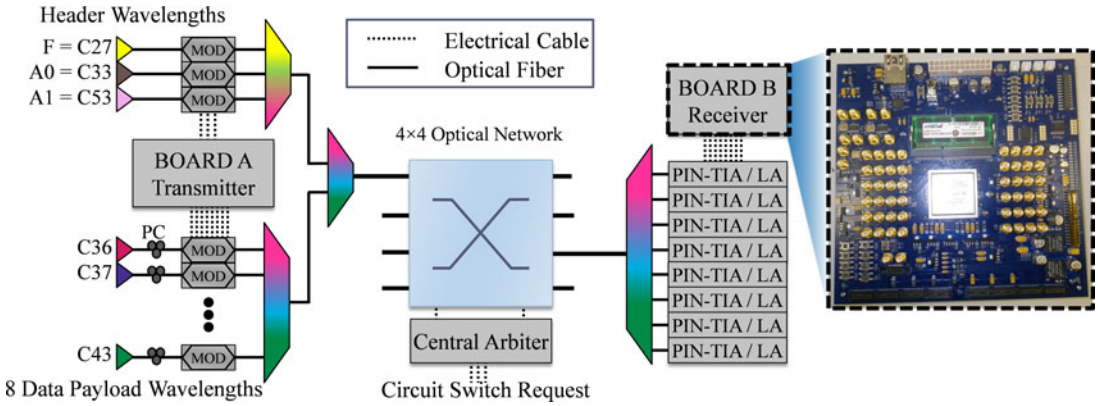

Fig. 3. Experimental setup showing one direction of processor-memory communication, with a photograph of one FPGA-based circuit board (inset). Packet-switched communication use the header wavelengths: frame (F), address 0 (A0), and address 1 (A1); circuit-switched communication utilizes the central arbiter.

The implemented OCM system uses multiple Altera Stratix IV FPGA-based circuit boards to create a processor and three OCMMs (see Fig. 3). All memory accesses are performed over the 4×4 optical network in the testbed. Each OCMM consists of six SDRAM connected to an FPGA, which contains serialization and deserialization functionality and a bank of 8×10-Gb/s bidirectional electrical transceivers. The processor and custom memory controller are implemented using an identical FPGA with 8×10-Gb/s transceivers. Each transceiver bank interfaces with discrete optical components to generate and receive 8×10-Gb/s WDM memory transactions. Each transceiver bank on the FPGA circuit board drives eight $LiNbO_3$ Mach–Zehnder modulators to modulate eight separate wavelengths (ITU C-band channels C36-C43), which are combined using WDM to create 8×10-Gb/s wavelength-striped memory transactions. The transceiver banks also connect to eight p-i-n receivers with TIAs and LAs, which receive the demultiplexed WDM transactions.

For the packet-switched transactions, the FPGAs use three low-speed general purpose input/output pins to drive three SOAs and modulate the frame and address header wavelengths. The three low-speed header wavelengths and 8×10-Gb/s payload wavelengths are combined before being injected into the 4×4 optical network. For circuit-switched transactions, the three header wavelengths are not used, and only the 8×10-Gb/s WDM payloads are injected into the network. The resulting configuration is such that a processor accesses its main memory across a transparent, hybrid packet- and circuit-switched 80-Gb/s WDM optical memory channel. Memory communication and network utilization are optimized by selecting optical packet or circuit communication based on required memory transaction sizes as specified by the processor.

B. Packet and Circuit-Switched Memory Protocol

Our custom memory controller analyzes incoming memory access requests from the processor to efficiently schedule all network communication requests and optimize network utilization. The processor can access the memory space at any of the connected OCMMs. Due to the principle of locality, the access is often to the same OCMM as one or many previous memory accesses. The processor may issue memory access requests faster than the memory controller can analyze them, and therefore the memory controller can buffer up to 16 memory access requests. The memory controller will attempt to reorder buffered memory accesses to issue multiple requests to the same OCMM in a single, longer transaction. When a sufficient number of accesses to an OCMM are combined and exceed a predetermined threshold, the memory access will instead be performed using circuit switching. Any transactions below that threshold are issued as optical packets.

In current state-of-the-art memory modules, memory accesses in modern processors are of a standard size, called bursts, which are typically eight 64-bit memory words. Each burst incurs a memory access overhead latency on the order of tens of nanoseconds, which is due to the standardized SDRAM access protocol [5]. Before the memory controller can transmit data from the processor to memory, for example, it must first issue low-bandwidth SDRAM-specific commands that are required to operate the SDRAM's internal buffers and addressing hardware. This has resulted in a trend of burst sizes doubling with each new generation of SDRAM [5], [29], reaching the current burst size of eight words, which amortizes the SDRAM access latency and maximizes memory bandwidth. Hence, to amortize the additional latency associated with packet- and circuit-switching circuitry utilized here, thus further improving bandwidth within our OCM system, our memory controller assumes a minimum burst size of 32 words for a minimum optical packet size of 2048 bits. In future, integrated OCM systems can reduce packet- and circuit-switching overhead latency while leveraging advanced memory configurations [21], [22] to achieve high memory bandwidths with smaller, more fine-grained memory accesses.

In order to minimize SDRAM command overhead and optimize communication within the optical network, SDRAM command signals are transmitted as optical packets prior to the transmission of memory data. In the case of packet-switched data, the data packets will be transmitted immediately following the command packets and incur an average of 20-ns SDRAM access overhead per packet. However, for circuit switching, the SDRAM command packets are speculatively transmitted while a circuit path setup request is issued to the central arbiter. This hides a portion of the circuit path setup time (in our

implementation, this is approximately 90 ns) within the SDRAM command overhead to reduce the total latency for circuit-switched transactions. Based on the latency overhead of setting up the packets and circuits, we designate a threshold of five aggregated bursts to a single OCMM as the cutoff between packet and circuit accesses. This threshold regulates the flow of data through the optical interconnection network and consequently affects latency, which is directly linked to the system's power dissipation.

The memory controller at the processor node also analyzes the queued memory requests to manage the OCMM sleep states. When no memory requests are pending for a given OCMM, the memory controller will issue an optical sleep command packet to the OCMM. When memory accesses are again desired at a sleeping OCMM, a "wake up" command packet is issued prior to the SDRAM access command packet. The "wake up" process requires up to 10 ns, and therefore the worst case penalty for attempting to access a sleeping OCMM is an additional 10 ns of latency, in addition to the standard SDRAM access overhead. Our memory controller avoids the worst case penalty by leveraging the 16-deep memory access request queue, which enables the memory controller to predict upcoming memory accesses and "wake up" appropriate OCMMs just-in-time for each memory access.

Additionally, for circuit-switched accesses, the extra incurred latency (i.e., 10 ns) can also be hidden within the circuit path setup time by the speculative SDRAM commands. As discussed previously, the use of sleep states in contemporary systems results in additional latency. Here, a novel aspect of our scheme is that the latency overhead associated with entering/exiting a sleep state can be hidden with the optical circuit setup time. Thus, by overlapping optical network latencies with SDRAM access latencies, the system can attain the corresponding power savings and performance improvements without significant latency penalties.

C. Experimental Characterization

To characterize the hybrid optical packet- and circuit-switched OCM system, we program the FPGA-based microprocessor to fill the entire memory address space with predictable bit patterns: a $2^{31}-1$ pseudorandom bit sequence, all 0s, all 1s, and the bit pattern corresponding to the destination memory address. These bit patterns are chosen from both established memory tests and the optical system tests. After the memory address space is full, the processor issues "read from memory" requests to stream all previously stored data back from memory. As the data stream in from memory, a counter within the processor verifies the data test patterns, records the number of correctly verified bits, and calculates the number of bit errors. These counters are used to generate an effective memory-bit-error rate (EMBER) to quantify the functionality and reliability of the hybrid packet- and circuit-switched OCM system. In this way, an error-free operation is achieved when the processor correctly verifies over 1 terabit of memory data from each OCMM, attaining EMBERs less than 10^{-12}.

The order in which the processor accesses each OCMM is random, on an access-by-access basis, such that the memory controller may receive any number of memory access requests for a given OCMM at one time. This creates random traffic that serves to illustrate the flexibility and novel capabilities of the OCM system, and enables the memory controller to reorder memory access requests when possible, and thus generates both packet- and circuit-switched memory transactions. Previous OCM demonstrations have relied on purely circuit-switched optical networks [18], which require applications with predictable memory access patterns such as streaming applications. Here, the use of a hybrid packet- and circuit-switched optical network enables any application to leverage the increased memory capacity and bandwidth offered by OCM. Specific applications would benefit from adjusting the threshold for packet/circuit routing, and a future memory controller may analyze memory and network utilization during runtime to dynamically adjust the threshold.

Randomness is obtained through the use of two linear feedback shift registers (LFSR), one 7 bits and one 8 bits, that are sampled once for each memory access. The serial outputs of the LFSRs are appended together into a 2-bit value that represents which of the three OCMMs will be addressed (00, 01, or 10). The value 11, being an invalid address, causes the processor to address the same OCMM as the previous memory access. This increases the probability that subsequent memory accesses are to the same OCMM, as would be the case for data locality within an application. Using LFSRs is accepted in computing as an acceptable approximation of randomness.

D. Experimental Results

We confirm the error-free operation of the OCM system with EMBERs less than 10^{-12} for all three OCMMs. Fig. 4 shows the optical eye diagrams for the eight 10-Gb/s memory payload channels, depicting clear open eyes for all data payloads. Fig. 5 shows an example of how the processor can communicate with the OCMMs using optical packets and circuits.

The circuit-switching data threshold of five bursts (i.e., 10 240 bits) within our 16-deep buffer results in an average of 24% of memory accesses using circuits rather than packets (evaluated for each terabit of memory traffic). Previous OCM studies [17], [18], [30] have been limited to purely circuit-switched optical networks, which require predictable memory access patterns such as in steaming applications. Thus far, no packet-switched or hybrid packet- and circuit-switched OCM systems have been demonstrated, and therefore a circuit-switched OCM system is used as a baseline comparison. Given identical traffic, such circuit-switched implementations would penalize the 76% of memory traffic that is comprised of smaller sized messages. Each circuit-switched memory access within that 76% would incur up to 70 ns additional latency compared to our hybrid packet- and circuit-switched implementation. Meanwhile, each memory access within the remaining 24% that use circuit switching reduces latency by >10 ns by amortizing setup latency over several aggregated bursts. Overall, the combined use of packet and circuit switching supports more diverse applications and memory access patterns with both short and long data streams as compared to previously demonstrated OCM systems.

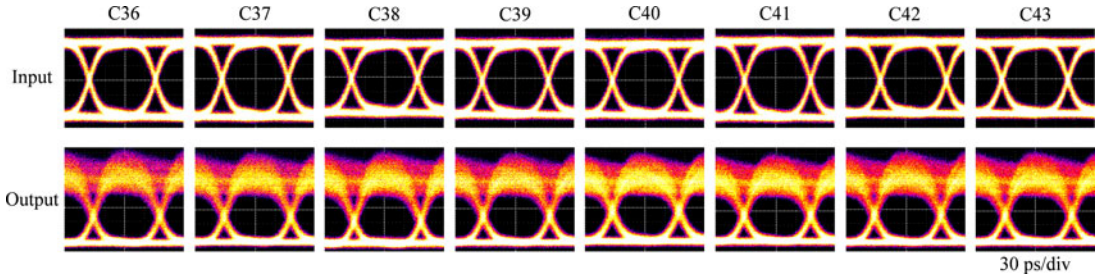

Fig. 4. Optical eye diagrams for the 8 × 10-Gb/s memory payload wavelength channels at one network input port (top) and at one network output (bottom).

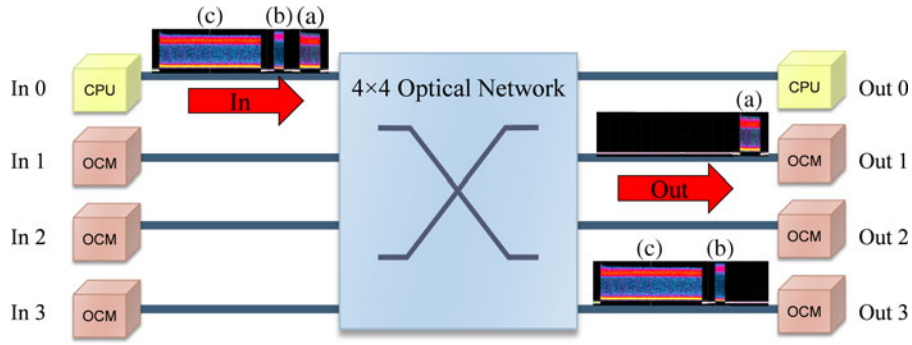

Fig. 5. Processor injects an optical packet into the network (a) addressed to the OCM node at network port 1. Subsequently, the processor issues a short control packet (b), followed by a circuit-switched memory access (c) to the OCM node at network port 3.

The use of sleep states allows for a significant reduction in power consumption of the OCMMs. Each OCMM was determined to be set in sleep state for 35% of the experimental characterization, resulting in an overall memory power savings of 28%. This is due to the fact that although each OCMM is addressed 33% of the time, and therefore unused and potentially able to sleep 66% of the time, the random order of memory accesses prevents the memory controller from putting OCMMs to sleep for their theoretical maximum allowable time. It is possible to increase the percentage of time each OCMM spends in its sleep state, e.g., by allowing the state to persist not only until a request enters the access queue but also until just before the access is actually issued. However, this change would tradeoff critical memory access latency for improved energy savings within the OCMMs, and would be a design choice specific to each individual system. Deploying our OCMMs within large-scale systems, utilizing terabytes of memory per server and many thousands of servers, could thus achieve the extreme levels of energy efficiency required for next-generation data centers and HPCs.

IV. Conclusion

Future HPCs and data centers require novel interconnection networks that can deliver energy-efficient, high-bandwidth memory communication with ultralow latencies across a broad range of applications and communication patterns. Electrically connected memory systems cannot scale to meet the performance requirements of tomorrow's large-scale computers. Further, previous OCM analyses have relied on inflexible circuit-switched networks.

In this study, we experimentally demonstrate the first hybrid packet- and circuit-switched OCM system, with error-free (EMBERs $< 10^{-12}$) transparent routing of 8 × 10-Gb/s wavelength-striped memory transactions between processors and OCM nodes. This implementation efficiently optimizes memory communication based on memory data burst lengths and reduces memory access latency by up to 70 ns per memory transaction compared to purely circuit-switched OCM architectures. Additionally, we show a 28% reduction in the memory node's energy consumption through an optical-packet-controlled OCMM sleep state technique. This study demonstrates the need for energy-efficient, high-performance optical interconnects within future large-scale memory systems, as well as the importance of flexible, hybrid packet- and circuit-switched network architectures.

Acknowledgment

The authors acknowledge valuable discussions with C. P. Lai.

References

[1] P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carlson, W. Dally, M. Denneau, P. Franzon, W. Harrod, J. Hiller, S. Karp, S. Keckler, D. Klein, R. Lucas, M. Richards, A. Scarpelli, S. Scott, A. Snavely, T. Sterling, R. S. Williams, and K. Yelick, "Exascale computing study: Technology challenges in achieving exascale systems," 2008.

[2] R. Ho, W. Mai, and M. A. Horowitz, "The future of wires," *Proc. IEEE*, vol. 89, no. 4, pp. 490–504, Apr. 2001.

[3] The ITRS Technology Working Groups. (2009). International Technology Roadmap for Semiconductors (ITRS) 2009 Edition [Online]. Available: http://www.itrs.net

[4] IEEE P802.3ba 40Gb/s and 100Gb/s Ethernet Task Force. IEEE P802.3ba Specification, (2010). [Online]. Available: http://grouper.ieee.org/groups/802/3/ba/index.html

[5] JEDEC Solid State Technology Association. DDR3 SDRAM Standard, [Online]. Available: http://www.jedec.org/standards-documents/docs/jesd-79-3d

[6] Samsung Electronics. Memory Specification, [Online]. Available: http://www.samsung.com/global/business/semiconductor/Greenmemory/Applications/ServerStorage/ServerStorage DDR3.html
[7] Intel Corp. Specification Addendum. Fully Buffered DIMM, (2006). [Online]. Available: http://www.intel.com/technology/memory/FBDIMM/spec/Intel_FBD_Spec_Addendum_rev_p9.pdf
[8] D. Miller, "Device requirements for optical interconnects to silicon chips," *Proc. IEEE*, vol. 97, no. 7, pp. 1166–1185, Jul. 2009.
[9] L. Chen, K. Preston, S. Manipatruni, and M. Lipson, "Integrated GHz silicon photonic interconnect with micrometer-scale modulators and detectors," *Opt. Exp.*, vol. 17, no. 17, pp. 15248–15256, Aug. 2009.
[10] W. A. Zortman, M. R. Watts, D. C. Trotter, R. W. Young, and A. L. Lentine, "Low-power high-speed silicon microdisk modulators," presented at the Conf. Lasers Electro-Optics, OSA Technical Digest, San Jose, CA, May 2010, Paper CThJ4.
[11] H. D. Thacker, Y. Luo, J. Shi, I. Shubin, J. Lexau, X. Zheng, G. Li, J. Yao, J. Costa, T. Pinguet, A. Mekis, P. Dong, S. Liao, D. Feng, M. Asghari, R. Ho, K. Raj, J. G. Mitchell, A. V. Krishnamoorthy, and J. E. Cunningham, "Flip-chip integrated silicon photonic bridge chips for sub-picojoule per bit optical links," in *Proc. Electron. Compon. Technol. Conf.*, Jun. 2010, pp. 240–246.
[12] A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-capacity optical transmission systems," *J. Lightw. Technol.*, vol. 26, pp. 1032–1045, 2008.
[13] Corning, Inc., Datasheet: Corning SMF-28e optical fiber product information, (2007). [Online]. Available: http://www.princetel.com/datasheets/SMF28e.pdf
[14] A. F. Benner, D. M. Kuchta, P. K. Pepeljugoski, R. A. Budd, G. Hougham, B. V. Fasano, K. Marston, H. Bagheri, E. J. Seminaro, X. Hui, D. Meadowcroft, M. H. Fields, L. McColloch, M. Robinson, F. W. Miller, R. Kaneshiro, R. Granger, D. Childers, and E. Childers, "Optics for high-performance servers and supercomputers," presented at the Optical Fiber Commun. Conf., San Diego, CA, Mar. 2010, Paper OTuH1.
[15] B. J. Offrein and P. Pepeljugoski, "Optics in supercomputers," presented at the *Eur.* Conf. Opt. Commun., Vienna, Austria, Sep. 2009, Paper 3.1.3.
[16] R. Luijten, W. E. Denzel, R. R. Grzybowski, and R. Hemenway, "Optical interconnection networks: The osmosis project," in *Proc. IEEE 17th Annu. Lasers Electro-Optics Soc.*, Nov. 2004, pp. 536–554.
[17] G. Hendry, E. Robinson, V. Gleyzer, J. Chan, L. Carloni, N. Bliss, and K. Bergman, "Circuit-switched memory access in photonic interconnection networks for high-performance embedded computing," in *Proc. Int. Conf. High Perform. Comput. Netw., Storage and Anal.*, Nov. 2010, pp. 1–12.
[18] D. Brunina, C. P. Lai, A. S. Garg, and K. Bergman, "Building data centers with optically connected memory," *J. Opt. Commun. Netw.*, vol. 3, no. 8, pp. A40–A48, Aug. 2011.
[19] Y. Katayama and A. Okazaki, "Optical interconnect opportunities for future server memory systems," in *Proc. IEEE 3rd Int. Symp. High Perform. Comput. Archit.*, 2007, pp. 46–50.
[20] A. Hadke, T. Benavides, S. J. B. Yoo, R. Amirtharajah, and V. Akella, "OCDIMM: Scaling the DRAM memory wall using WDM based optical interconnects," in *Proc. IEEE 16th Int. Symp. High-Perform. Interconnects*, Aug. 2008, pp. 57–63.
[21] S. Beamer, C. Sun, Y. J. Kwon, A. Joshi, C. Batten, V. Stojanovic, and K. Asanovic, "Re-architecting DRAM memory systems with monolithically integrated silicon photonics," in *Proc. 37th Int. Symp. Comput. Archit.*, Jun. 2010, pp. 129–140.
[22] A. N. Udipi, N. Muralimanohar, R. Balasubramonian, A. Davis, and N. P. Jouppi, "Combining memory and a controller with photonics through 3D-stacking to enable scalable and energy-efficient systems," in *Proc. 38th Annu. Int. Symp. Comput. Archit.*, New York, 2011, pp. 425–436.
[23] H. Wang, A. S. Garg, K. Bergman, and M. Glick, "Design and demonstration of an all-optical hybrid packet and circuit switched network platform for next generation data centers," in *Proc. Collocated Nat. Fiber Opt. Eng. Conf. Opt. Fiber Commun.*, Mar. 2010, pp. 1–3.
[24] A. Shacham and K. Bergman, "Building ultralow-latency interconnection networks using photonic integration," *IEEE Micro*, vol. 27, no. 4, pp. 6–20, Jul. 2007.
[25] D. Brunina, C. P. Lai, and K. Bergman, "A data rate- and modulation format-independent packet-switched optical network test-bed," *Photon. Technol. Lett.*, vol. 24, no. 5, pp. 377–379, Mar. 2012.
[26] O. Liboiron-Ladouceur, B. A. Small, and K. Bergman, "Physical layer scalability of WDM optical packet interconnection networks.," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 262–270, Jan. 2006.
[27] Cisco. Cisco Product Data Sheets, (2012). [Online]. Available: http://www.cisco.com/en/US/products/index.html
[28] A. Biberman, N. Sherwood-Droz, X. Zhu, M. Lipson, and K. Bergman, "High-speed data transmission in multi-layer deposited silicon photonics for advanced photonic networks-on-chip," in *Proc. Conf. Lasers and Electro-Opt.*, May 2011.
[29] Micron Technology, Inc., (2008), DDR2 SDRAM Datasheet, [Online]. Available: http://www.micron.com/products/dram/ddr2/
[30] G. Hendry, E. Robinson, V. Gleyzer, J. Chan, L. P. Carloni, N. Bliss, and K. Bergman, "Time-division-multiplexed arbitration in silicon nanophotonic networks-on-chip for high-performance chip multiprocessors," *J. Parallel Distrib. Comput.*, vol. 71, no. 5, pp. 641–650, May 2011.

Daniel Brunina (S'08) received the B.S. and M.S. degrees in computer systems engineering from Boston University, Boston, MA, in 2004 and 2005, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, Columbia University, New York.

His current research interests include designing optically connected memory architectures and optical interfaces for large-scale computing.

Dawei Liu (S'11) received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2006 and the M.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, both in electrical engineering. He received the Professional degree from Columbia University, New York, in 2012.

Keren Bergman (S'87–M'93–SM'07–F'09) received the B.S. degree from Bucknell University, Lewisburg, PA, in 1988, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively, all in electrical engineering.

She is currently a Professor in the Department of Electrical Engineering, Columbia University, New York, where she also directs the Lightwave Research Laboratory. Her research programs involve optical interconnection networks for advanced computing systems, photonic packet switching, and nanophotonic networks on chip.

Prof. Bergman is a Fellow of the Optical Society of America. She is the co-Editor- in-Chief of the OSA/IEEE JOURNAL OF OPTICAL COMMUNICATIONS AND NETWORKING.