Architecting Optically Controlled Phase Change Memory

ADITYA NARAYAN, Boston University
YVAIN THONNART and PASCAL VIVET, Univ. Grenoble Alpes, CEA, List, France
AYSE COSKUN and AJAY JOSHI, Boston University

Phase Change Memory (PCM) is an attractive candidate for main memory, as it offers non-volatility and zero leakage power while providing higher cell densities, longer data retention time, and higher capacity scaling compared to DRAM. In PCM, data is stored in the crystalline or amorphous state of the phase change material. The typical electrically controlled PCM (EPCM), however, suffers from longer write latency and higher write energy compared to DRAM and limited multi-level cell (MLC) capacities. These challenges limit the performance of data-intensive applications running on computing systems with EPCMs.

Recently, researchers demonstrated optically controlled PCM (OPCM) cells with support for 5 *bits/cell* in contrast to 2 *bits/cell* in EPCM. These OPCM cells can be accessed directly with optical signals that are multiplexed in high-bandwidth-density silicon-photonic links. The higher MLC capacity in OPCM and the direct cell access using optical signals enable an increased read/write throughput and lower energy per access than EPCM. However, due to the direct cell access using optical signals, OPCM systems cannot be designed using conventional memory architecture. We need a complete redesign of the memory architecture that is tailored to the properties of OPCM technology.

This article presents the design of a unified network and main memory system called COSMOS that combines OPCM and silicon-photonic links to achieve high memory throughput. COSMOS is composed of a hierarchical multi-banked OPCM array with novel read and write access protocols. COSMOS uses an Electrical-Optical-Electrical (E-O-E) control unit to map standard DRAM read/write commands (sent in electrical domain) from the memory controller on to optical signals that access the OPCM cells. Our evaluation of a 2.5D-integrated system containing a processor and COSMOS demonstrates 2.14× average speedup across graph and HPC workloads compared to an EPCM system. COSMOS consumes 3.8× lower read energy-per-bit and 5.97× lower write energy-per-bit compared to EPCM. COSMOS is the first non-volatile memory that provides comparable performance and energy consumption as DDR5 in addition to increased bit density, higher area efficiency, and improved scalability.

CCS Concepts: • **Hardware → Emerging optical and photonic technologies**; *Emerging architectures;*

Additional Key Words and Phrases: Phase Change Memory, silicon-photonics, 2.5D computing system, non-volatile memory

This is a new article, not an extension of a conference paper. This work was funded by NSF CCF 2131127 and NSF CCF 1716352 grants.

Authors' addresses: A. Narayan, A. Coskun, and A. Joshi, Boston University, 8 Saint Mary's Street, Boston, MA, 02215; emails: {adityan, acoskun, joshi}@bu.edu; Y. Thonnart and P. Vivet, Univ. Grenoble Alpes, CEA, List, Grenoble, France; emails: {yvain.thonnart, pascal.vivet}@cea.fr.

1544-3566/2022/12-ART48 $15.00
https://doi.org/10.1145/3533252

ACM Reference format:
Aditya Narayan, Yvain Thonnart, Pascal Vivet, Ayse Coskun, and Ajay Joshi. 2022. Architecting Optically Controlled Phase Change Memory. *ACM Trans. Arch. Code Optim.* 19, 4, Article 48 (December 2022), 26 pages. https://doi.org/10.1145/3533252

1 INTRODUCTION

Today's data-driven applications that use graph processing [30, 53, 56, 79], machine learning [15, 29], or privacy-preserving paradigms [3, 19, 82] demand memory sizes on the order of hundreds of *GBs* and bandwidths on the order of TB/s. The widely used main memory technology, DRAM, is facing critical technology scaling challenges and fails to meet the increasing bandwidth and capacity demands of these data-driven applications [37, 40, 41, 48, 58, 95]. **Phase Change Memory (PCM)** is emerging as a class of **non-volatile memory (NVM)** that is a promising alternative to DRAM [33, 39, 46, 47, 71, 72]. PCMs outperform other NVM candidates owing to their higher reliability, increased bit density, and better write endurance [13, 16, 61, 92].

In PCMs, data is stored in the state of the phase change material, i.e., crystalline (logic 1) or amorphous (logic 0) [64, 93]. A SET operation triggers a transition to crystalline state, and a RESET operation triggers a transition to amorphous state. PCMs also enable **multi-level cells (MLC)** using the partially crystalline states. Higher MLC capacity enables increased bit density ($bits/mm^2$). PCM cells are typically controlled electrically (we refer to them as EPCM cells), where different PCM states have distinct resistance values. EPCM cells are SET or RESET by passing the corresponding current through the phase change material (via the bitline) to trigger the desired state transition. The state of the EPCM cells is read out by passing a read current and measuring the voltage on the bitline. Main memory systems using EPCM cells are designed using the same microarchitecture and read/write access protocol as DRAM systems [44, 85]. EPCM systems, however, experience resistance drift over time and so are limited to 2 $bits/cell$ [13, 17], have $3-4\times$ higher write latency than DRAM leading to lower performance [5, 44], consume high power due to the need for large on-chip charge pumps [35, 66, 90], and have lower lifetime than DRAM due to faster cell wearout [70].

Recent advances in device research have demonstrated optically controlled PCM cells (we refer to them as OPCM cells) [18, 26, 27, 78]. OPCM cells exhibit higher MLC capacity than EPCM cells (up to 5 $bits/cell$ [52]). Moreover, high-bandwidth-density silicon-photonic links [84, 87], which are being developed for processor-to-memory communication, can directly access these OPCM cells, thereby yielding higher throughput and lower energy-per-access than EPCM. These two factors make OPCM a more attractive candidate for main memory than EPCM.

Given that in OPCM the optical signals in silicon-photonic links directly access the OPCM cells, the traditional row-buffer-based memory microarchitecture and the read/write access protocol encounter critical design challenges when adapted for OPCM. We need a complete redesign of the memory microarchitecture and a novel access protocol that is tailored to the OPCM cell technology. **In this article, we propose a COmbined System of Optical Phase Change Memory and Optical LinkS, COSMOS, which integrates the OPCM technology and the silicon-photonic link technology, thereby providing seamless high-bandwidth access from the processor to a high-density memory.** Figure 1 shows a computing system with COSMOS. COSMOS includes a hierarchical multi-banked OPCM array, E-O-E control unit, silicon-photonic links, and laser sources. The multi-banked OPCM array uses 3D optical integration to stack multiple banks vertically, with 1 bank/layer. The cells in the OPCM array are directly accessed using silicon-photonic links that carry optical signals, thereby eliminating the need for **electrical-optical (E-O)** and **optical-electrical (O-E)** conversion in the OPCM array. These optical signals are generated

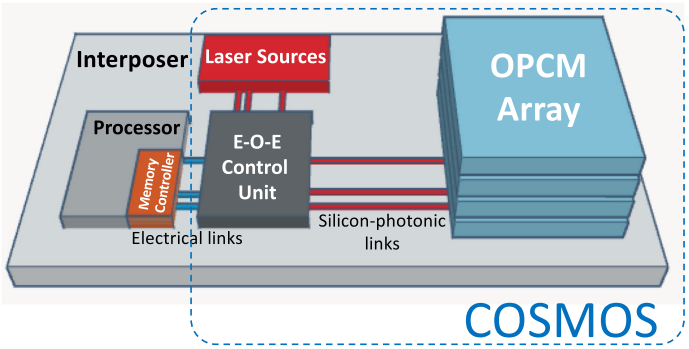

Fig. 1. Overview of a 2.5D-integrated computing system with OPCM array stack as the main memory, E-O-E control unit chiplet, processor chiplet, and laser sources chiplet.[1]

by an E-O-E control unit that serves as an intermediary between the **memory controller (MC)** in the processor and the OPCM array. This E-O-E control unit is responsible for mapping the standard DRAM protocol commands sent by the MC onto optical signals and then sending these optical signals to the OPCM array.

The major contributions of our work are as follows:

(1) We architect the COSMOS, which consists of a hierarchical multi-banked OPCM array, where the cells are accessed directly using optical signals in silicon-photonic links. The OPCM array design combines **wavelength-division-multiplexing (WDM)** and **mode-division-multiplexing (MDM)** properties of optical signals to deliver high memory bandwidth. Moreover, the OPCM array contains only passive optical elements and does not consume power, thus providing cost and efficiency advantages.
(2) We propose a novel mechanism for read and write operation of cache lines in COSMOS. A cache line is interleaved across multiple banks in the OPCM array to enable high-throughput access. The write data is encoded in the intensity of optical signals that uniquely address the OPCM cell. The readout of an OPCM cell uses a three-step operation that measures the attenuation of the optical signal transmitted through the cell, where the attenuation corresponds to a predetermined bit pattern. Since the read operation is destructive, we design an opportunistic writeback operation of the read data to restore the OPCM cell state.
(3) We design an E-O-E control unit to interface COSMOS with the processor. This E-O-E control unit receives standard DRAM commands from the processor and converts them into the OPCM-specific address, data, and control signals that are mapped onto optical signals. These optical signals are then used to read/write data from/to the OPCM array. The responses from the OPCM array are converted by the E-O-E control unit back into standard DRAM protocol commands that are sent to the processor.

Evaluation of a 2.5D system with a multi-core processor and COSMOS demonstrates 2.15× higher write throughput and 2.09× higher read throughput compared to an equivalent system with EPCM. This increased memory throughput in COSMOS reduces the memory latency by 33%. For graph and **high performance computing (HPC)** workloads, when compared to EPCM, COSMOS has 2.14× better performance, 3.8× lower read energy-per-bit, and 5.97× lower write energy-per-bit. Moreover, COSMOS provides a scalable and non-volatile alternative to DDR5 DRAM systems, with

[1]COSMOS-based computing system is agnostic of the integration technology. However, 3D-integrated systems raise thermal concerns and 2D systems result in large system footprint and communication overheads.

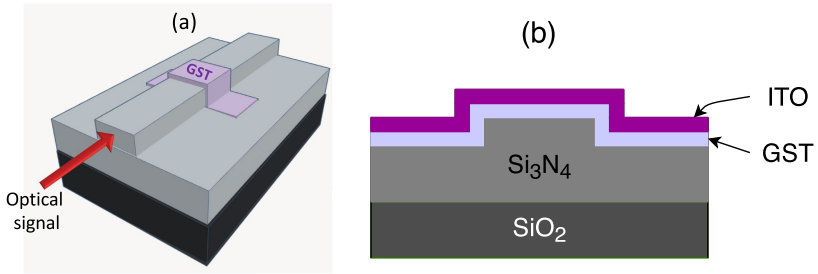

Fig. 2. (a) 3D view of GST-based PCM cell. (b) Cross-sectional view of GST deposited on a Si_3N_4 waveguide.

similar performance and energy consumption for read and write accesses. With DRAM technology undergoing critical scaling challenges, COSMOS presents the first non-volatile main memory system with improved scalability, increased bit density, high area efficiency, and comparable performance and energy consumption as DDR5 DRAM.

2 BACKGROUND

In this section, we discuss the basic operation of an OPCM cell along with its properties and the silicon-photonic links that enable optical signals to directly access the OPCM cells.

2.1 OPCM Cell

$Ge_2Sb_2Te_5$ (GST) is a well-known phase change material that exhibits high contrast in the electrical property (resistance) and the optical property (refractive index) between its two states, in addition to long data retention time and nanoscale size [55, 75, 93]. Thus, GST has been widely used as a storage element in a PCM cell (EPCM and OPCM cells). An OPCM cell consists of only a GST element and does not use a separate access transistor as an EPCM cell. Figure 2 shows the structure of an OPCM cell, where the GST is integrated on a waveguide [52, 78]. The waveguides are fabricated using a Si_3N_4 layer deposited over a SiO_2 layer [51]. The GST layer is covered with a layer of **Indium-Tin-Oxide (ITO)** to prevent oxidation. The optical signals to read and write the OPCM cell lie in the C band ($1530nm-1565nm$) and L band ($1565nm-1625nm$) of the telecommunication spectrum.

2.2 Write Operation in OPCM Cells

For write operation, the optical signal traversing through the waveguide is coupled to the GST element. The energy of this optical signal heats the GST element and triggers a state transition. For RESET operation, i.e., switching the GST element to an amorphous state (a-GST), an optical pulse of $180pJ$ energy is applied to the GST element for $25ns$ [52]. For SET operation, i.e., switching the GST element to a fully crystalline state (c-GST), an optical pulse with an energy of $130pJ$ is applied to the GST element for $250ns$ [52]. The transition of the GST state to a partially crystalline state requires different values of pulse energies ($60pJ-130pJ$) applied for varying durations ($50ns-250ns$) [52].

2.3 Read Operation in OPCM Cells

The readout mechanism for an OPCM cell uses the high contrast in the refractive indices of a-GST (3.56) and c-GST (6.33) [57]. When an optical signal is passed through the GST element, the higher refractive index of c-GST results in an increased optical absorption by the GST element. Rios et al. [78] demonstrate that c-GST absorbs 79% of the input optical signal and allows transmission of only 21% of the optical signal. In contrast, a-GST transmits 100% of the optical signal.

The transmission of partially crystalline states lies between 100% and 21% [78]. An OPCM cell is, therefore, read out by sending a sub-*ns* optical pulse through the GST element and measuring the transmitted optical intensity of the output pulse. This transmitted intensity corresponds to a pre-determined bit pattern, thus allowing the readout of the stored data in the GST element.

2.4 High MLC Capacity of OPCM Cells

In OPCM cells, the read operation uses the refractive index of the GST state to determine the stored value. Unlike the resistance value used in EPCM cells, the refractive index experiences minimal to no drift over time [52, 78]. This enables designing OPCM cells with multiple stable partially crystalline states with each having a unique refractive index. Prior works have demonstrated that it is possible to reliably program an OPCM cell to contain more than 34 unique partially crystalline states [52, 99], which enables an OPCM cell to have an MLC capacity of up to 5 $bits/cell$. Using a higher capacity MLC enables the read and write operation of a higher number of bits per access than EPCM, thereby increasing the memory throughput.

2.5 Silicon-photonic Links

In a computing system that uses a main memory composed of OPCM cells, optical signals in silicon-photonic links can directly read/write the cells. The silicon-photonic links provide higher bandwidth density at negligible data-dependent power compared to electrical links [8, 10, 42]. In addition, these silicon-photonic links have single-cycle latency, in contrast to electrical links that often take 3–4 cycles each for a memory request and a memory response. Moreover, we can multiplex multiple optical signals (up to 32 signals) in a single waveguide, resulting in dense WDM [45]. **MicroRing Resonators (MRRs)** can modulate these optical signals at data rates up to $12Gbps$ [4, 67, 86] giving a peak memory throughput of $384Gbps$ per link. Therefore, it is possible to design densely multiplexed silicon-photonic links that can directly access the OPCM cells, further increasing the memory throughput.

3 MOTIVATION

In this section, we motivate the need for a novel memory microarchitecture and access protocol for OPCM by first describing the typical EPCM architecture and then explaining why such an architectural design is impractical for OPCM arrays. Figure 3 shows the architecture of EPCM [39, 44]. The EPCM array is a hierarchical organization of banks, blocks, and sub-blocks [44]. During read or write operations, the EPCM first receives a row address. The row address decoder reads the appropriate row from the EPCM array into a row buffer. The EPCM next receives the column address, and the column address multiplexer selects the appropriate data block from the row buffer. The bitlines of the selected data block are connected to the write drivers for write operation or to the sense amplifiers for read operation. For write operation, the charge pumps supply the required drive voltage to the write drivers, which corresponds to SET or RESET operation. For read operation, a read current is first passed through the GST element in the EPCM cell through an access transistor [44]. Then, sense amplifiers determine the voltage on the bitline to read out logic 0 or logic 1.

Naively adapting the EPCM architecture for OPCM, by just replacing the EPCM cells with OPCM cells raises latency, energy, and thermal concerns, thereby rendering such a design impractical. To understand these concerns, let us consider an OPCM array that uses the EPCM architecture from Figure 3 with either an optical row buffer or an electrical row buffer. Such an OPCM array architecture has following limitations:

Limitations with optical row buffer: An optical row buffer can be designed using a row of GST elements whose states are controlled using optical signals. When a row is read from the

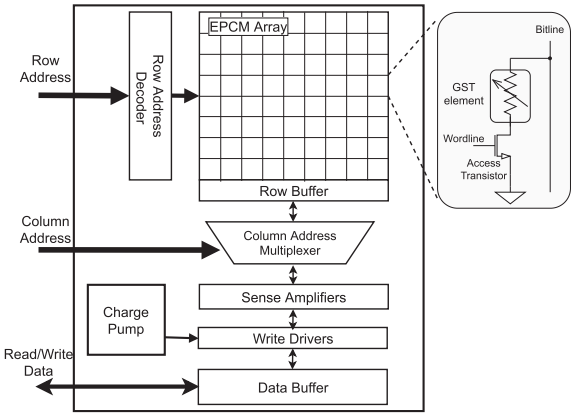

Fig. 3. A typical EPCM architecture [44].

OPCM array using an optical signal, the data is encoded in the signal's intensity. This intensity is not large enough to update the state of the GST elements in the optical row buffer. So, the read value first needs to be converted into an electrical signal. Based on this value, an optical signal with the appropriate intensity is generated to write the value into the optical row buffer. Essentially, we perform an extra O-E and E-O conversion. This necessitates the use of photodetectors, receivers, transmitters, and optical pulse generators, which adds to the energy and latency of a memory access. Hence, an optical row buffer is not a viable option.

Limitations with electrical row buffer: An electrical row buffer can be designed either using capacitor cells as in DRAM or using phase change materials controlled using electrical current as in EPCM. In both cases, the row buffer is accessed using electrical signals (assuming electrical links between the processor and memory). This increases the access latency and energy and creates thermal issues as follows:

(1) **Impact on read latency:** Upon receiving a row address from the MC on electrical links, the address first needs to be converted to an optical pulse, which is then used to read data from OPCM cells. After optical readout of an entire row from OPCM array, the data has to be converted back into electrical domain to store it in the row buffer. These two operations require an E-O and an O-E conversion, respectively, inside the OPCM array. These E-O/O-E conversions adds a latency of 25−30 cycles for each read access [6].
(2) **Impact on write latency:** When writing data from the row buffer to the OPCM array, a set of sense amplifiers reads the data from the electrical row buffer. This row buffer data is then mapped onto optical signals with appropriate intensities using a pulse generation circuitry within memory. The optical signals are then used to write the data to the OPCM cells. Therefore, the write operation requires three E-O/O-E conversions, which adds a latency of 40−45 cycles for each write access [6].
(3) **Impact on read/write energy:** The energy spent in the peripheral circuitry for optical signal generation and readout, as well as in the circuitry for E-O-E conversion increases the active power dissipation within memory [6, 60, 63]. Since each read/write operation encounters multiple E-O-E conversions, the energy per read and write access rises considerably high ($>200pJ/bit$) [24].
(4) **Thermal issues:** The MRRs used in the OPCM array are highly sensitive to thermal variations [65]. The thermal variations due to active electrical circuits within memory lowers the

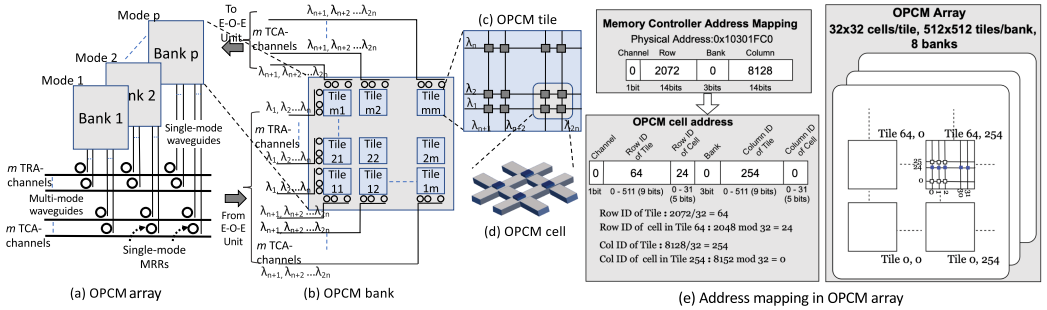

Fig. 4. (a) A multibanked-OPCM uses p optical modes to access p banks. (b) An OPCM bank is an array of $m \times m$ tiles. Every tile is accessed by a TRA-channel and a TCA-channel, each channel containing n optical signals. (c) An OPCM tile is an array of $n \times n$ cells. Every cell is accessed by a unique pair of optical signals. (d) OPCM cells are placed at every waveguide crossing. (e) Address mapping of the physical address to cells in the OPCM array. The physical address corresponds to OPCM cells in the shaded blue row of OPCM array.

reliability of the MRR operation. Such a design calls for active thermal and power management in OPCM, which contributes to a power overhead of $10-30W$ [2].

Furthermore, using silicon-photonic links in combination with OPCM requires additional E-O and O-E conversions on the MC and the OPCM array with this EPCM architecture that exacerbate the above discussed problems. Hence, we argue for the need to redesign the microarchitecture and the read/write access mechanisms that are tailored to the properties of the OPCM cell technology and the associated silicon-photonic link technology.

4 COSMOS ARCHITECTURE

In this section, we describe the microarchitecture of the high-throughput OPCM array in COSMOS. **The key innovation of our proposed microarchitecture is enabling direct access of OPCM cells by the optical signals in the silicon-photonic links.** This direct access avoids the extra E-O and O-E conversions that are required if we were to adapt the EPCM architecture for COSMOS. Our OPCM array microarchitecture is a hierarchical multi-banked design that maximizes the degree of parallelism for read and write accesses within the array using a combination of WDM and MDM. A distinguishing feature of our OPCM array design is that it does not contain any active circuits that consume power, i.e., it only contains passive optical devices. Figure 4 illustrates the detailed microarchitecture of our proposed OPCM array in COSMOS that uses GST as the phase change material. We base our architectural design on prior OPCM cell prototype designs [26, 27, 52, 78], which demonstrate the switching of OPCM cells between multiple states with high reproducibility. The confidence of cell read/write is mainly limited by the variations in cell switching and by the SNR of readout circuits. For 4-bit OPCM cells, prior works show minimal variations in cell switching and high SNR, resulting in high confidence of read/write. We describe each component of the proposed architecture, particularly focusing on how to read and write an OPCM cell in the optical domain with minimal E-O and O-E conversions.

4.1 OPCM Tile

An OPCM tile (see Figure 4(c)) consists of an $n \times n$ array of GST elements, i.e., OPCM cells. The GST elements are placed on top of waveguide crossings, as shown in Figure 4(d). This organization enables every OPCM cell to be accessed using a unique pair of optical signals: one on the associated row and one on the associated column. We need a total of n unique optical signals with wavelengths

$\lambda_1, \lambda_2, \ldots, \lambda_n$ that are routed in the rows (one per row waveguide) and n unique optical signals with wavelengths $\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_{2n}$ that are routed in the columns (one per column waveguide). Wavelengths λ_1 to λ_n together form the **Tile Row Access (TRA)**-channel, and wavelengths λ_{n+1} to λ_{2n} together form the **Tile Column Access (TCA)**-channel. A TRA-channel (and similarly each TCA-channel) is mapped to one or more waveguides, depending on the number of wavelengths that can be multiplexed in a waveguide. Owing to MLC, each OPCM cell stores b_{cell} bits. The total capacity of an OPCM tile is $n^2.b_{cell}$. A maximum of n cells can be read/written in parallel from a single tile, which gives us a peak throughput of $n.b_{cell}$ bits per read/write access for a tile.

4.2 OPCM Bank

Figure 4(b) shows the organization of an OPCM bank. The OPCM bank is composed of an array of $m \times m$ OPCM tiles and has a total capacity of $m^2.n^2.b_{cell}$ bits. The OPCM bank uses m TRA-channels, one for each row in the bank, and m TCA-channels, one for each column in the bank to communicate with the E-O-E control unit. Each TRA-channel uses λ_1 to λ_n, and each TCA-channel uses λ_{n+1} to λ_{2n}. We design a hierarchical array of OPCM cells (m^2 tiles with n^2 OPCM cells per tile) instead of a large monolithic array ($m^2.n^2$ OPCM cells), as designed by Feldman et al. [26, 27] to decrease the laser power required by the optical signals. With our proposed design, the laser sources only need to support $2n$ unique optical signals (in the range of λ_1 to λ_{2n}) instead of the $m.2n$ unique optical signals that would be required in a large monolithic array. We utilize MRRs to couple the optical signals of each TRA-channel and TCA-channel to its corresponding tile. We need n MRRs that are tuned to λ_1 to λ_n in each of the m TRA-channels and n MRRs that are tuned to λ_{n+1} to λ_{2n} in each of the m TCA-channels.

4.3 Multi-banked OPCM Array

Figure 4(a) shows the proposed multi-banked organization of the OPCM array using MDM. We interleave a cache-line across multiple banks. There are p banks, each supporting one of the p spatial modes of the $2n$ optical signals. Bank 1 only uses mode 1 of all optical signals $\lambda_1, \ldots \lambda_n$ and $\lambda_{n+1}, \ldots \lambda_{2n}$, Bank 2 only uses mode 2 of all optical signals, and so on. The waveguides connecting the OPCM to the E-O-E control unit are multi-mode waveguides, which carry all the p spatial modes of optical signals. We employ single-mode MRRs [89, 96] that couple a single spatial mode of optical signals from the multi-mode waveguide to a bank. Multiple prior works have exploited MDM property of optical signals coupled with WDM to design high-bandwidth-density silicon-photonic links [54, 91].

4.4 Address Mapping in COSMOS

Figure 4(e) shows an example mapping of the physical address received by the MC to the physical location of cells within the OPCM array in COSMOS. A cache line of $64B$ is stored in a total of 128 OPCM cells with $4bits/cell$. We interleave the cache line across 4 different banks. Within a bank, we map the 128-bit chunk of a cache line to a tile. The tile has 32×32 cells, and so we map that 128-bit chunk to an entire row within a tile. The row (column) field of physical address in the MC is mapped to the row ID of tile (column ID of tile) field and the row ID of cell (column ID of cell) field. In Figure 4(e), we show how the different fields of the physical address `0x10301FC0` are mapped to bank ID, row ID of tile, column ID of tile, row ID of cell, and column ID of cell.

5 ACCESS PROTOCOL IN COSMOS

To enable high-throughput access of OPCM cells within the OPCM array, we propose a novel read and write access protocol for COSMOS. When the MC issues a read or write operation, the row

address and column address are entered into the Row Address Queue and Column Address Queue, respectively, and the write data is entered into the Data Buffer in the E-O-E control unit.

5.1 Writing a Cache Line to OPCM Array

To write a cache line to the OPCM array, the E-O-E control unit identifies the bank ID, the row ID, and column ID of the tile, and the row ID and column ID of the cell within a tile using the address mapping. In our example with 32×32 array of cells in a tile, when writing 128-bit chunk of a cache line, we end up updating all the cells in a row (any misaligned accesses are handled on the processor side). Hence, for writes at cache line granularity, the column ID within a tile is not used. The E-O-E control unit determines the optical intensity that is required at each OPCM cell in the row to write the 128-bit chunk of the cache line. It then breaks down the optical intensity into two signals: one with a constant intensity of I_0 and the other with a data-dependent intensity of I_i, where $i = 1, 2, \ldots, 128$. The E-O-E control unit modulates the constant intensity I_0 onto the optical signal corresponding to the row (selected by the row ID of cell) within a tile. The E-O-E control unit then modulates the data-dependent optical intensities (i.e., $I_1, I_2, \ldots, I_{128}$) onto the optical signals corresponding to the 4 tiles spread across 4 banks with 32 columns per tile. The E-O-E control unit transmits the row signal I_0, and the column optical signals $I_1, I_2, \ldots, I_{128}$ in parallel to write the cache line in the OPCM array. The superposition of the optical signals, i.e., $I_0{+}I_1, I_0{+}I_2, \ldots, I_0{+}I_{128}$ updates the state of the OPCM cells. Note that, since a cache line is spread across 4 banks, the E-O-E control unit modulates data on optical signals to write to an OPCM tile in each of these 4 banks. None of the optical signals individually carries sufficient intensity to trigger a state transition at any cell, so none of the other cells along the row or column are affected.

5.2 Reading a Cache Line from OPCM Array

To read a cache line from OPCM array, the E-O-E control unit transmits sub-ns optical pulses along all the columns in a tile that contain the cache line and measures the pulse attenuation. However, there are multiple OPCM cells along each column and so the output intensity of optical signals will be attenuated by all cells in that column. It is, therefore, not possible to determine the OPCM cell values using a one-pulse readout. Hence, we use a three-step process for read operation of OPCM array in COSMOS. (1) To read a cache line, the E-O-E control unit first determines the bank ID, row ID, and column ID of tile, and row ID and column ID of cell. The E-O-E control unit transmits a read pulse RD_1 through all the columns in a tile containing the cache line. Note that, since a cache line is spread across 4 banks, the E-O-E control unit transmits RD_1 on the 4 different optical modes corresponding to the 4 banks. Each read pulse is attenuated by all the OPCM cells in the column. The attenuated pulses are received by the E-O-E control unit, which records the intensities of these attenuated pulses as $I_{1,1}, I_{2,1}, \ldots, I_{128,1}$. These intensities are converted into electrical voltage and stored as $V_{1,1}, V_{2,1}, \ldots, V_{128,1}$. (2) The E-O-E control unit then transmits a RESET pulse to the OPCM cells of the cache line, i.e., all the cells along a row within a tile. All the cells along the row are now amorphized and have 100% optical transmission. (3) The E-O-E control unit then sends a second read pulse RD_2 through all the columns of a tile containing the cache line. Each read pulse is again attenuated by all OPCM cells in the column. Given that step 2 amorphized all OPCM cells of the cache line, the output pulse intensities are different from those in step 1. The attenuated pulses are received by the E-O-E control unit, which records the intensities of these attenuated pulses as $I_{1,2}, I_{2,2}, \ldots, I_{128,2}$. These intensities are converted into electrical voltage and stored as $V_{1,2}, V_{2,2}, \ldots, V_{128,2}$. The E-O-E control unit computes the difference of the stored voltages of steps 1 and 3, i.e., $V_{1,1}{-}V_{1,2}, V_{2,1}{-}V_{2,2}, \ldots, V_{128,1}{-}V_{128,2}$. This difference is used to determine the cache line data stored in the OPCM cells.

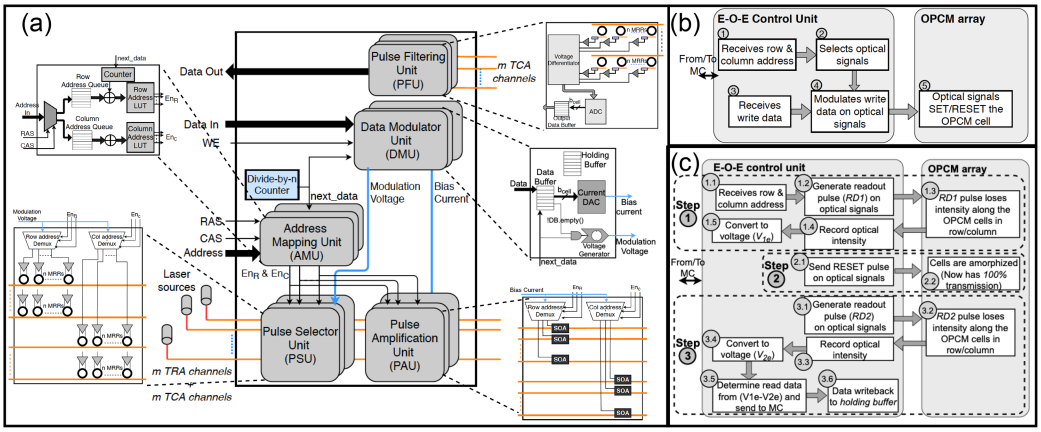

Fig. 5. (a) E-O-E control unit design. DMU: Generates the modulation voltage and the bias current corresponding to read/write data. AMU: Determines optical signals that correspond to read/write address. PSU: Selects the optical signals. PAU: Amplifies the optical signals using the bias current. PFU: Filters the optical signals to read cell data. Different micro-steps performed in E-O-E control unit and OPCM array during (b) write operation and (c) read operation.

5.3 Opportunistic Writeback After Read

The RESET operation in step 2 of the read operation destructs the original data in the OPCM cells. We, therefore, perform an opportunistic writeback of the cache line to the OPCM cells. After completing the three steps of the read operation, the read data and the address are saved into a *holding* buffer in the E-O-E control unit. When there are no pending read or write operations from the MC, the E-O-E control unit reads the data and its address from the *holding* buffer and writes the data back to the OPCM array. This writeback operation does not block any critical pending read and write operations coming from the MC. The dependencies in read and write requests between the *holding buffer* and the data buffer are handled in the E-O-E control unit. For a Read-After-Read case, the second read operation reads the data from the *holding buffer* if present. If the data is not in the *holding buffer,* then the second read operation just uses the three-step process + writeback (described above) to complete the read operation. For a Write-After-Read case, if the write address matches the read address and there is an entry for that read in the *holding buffer*, then the corresponding entry in the *holding buffer* is invalidated. The write data is then entered into the data buffer and then written into the appropriate OPCM array.

6 E-O-E CONTROL UNIT DESIGN

Our proposed E-O-E control unit provides the interface between the processor and the OPCM array. The MC sends standard DRAM access protocol commands to the E-O-E control unit. The E-O-E control unit maps these commands onto optical signals that read/write the data from/to OPCM array. **Though we can design a COSMOS-specific MC and the associated read/write protocol, our goal is to enable the COSMOS operation with a standard MC in any processor.** The E-O-E control unit uses the following five sub-units to read from and write to the OPCM array: **data modulation unit (DMU), address mapping unit (AMU), pulse selector unit (PSU), pulse amplification unit (PAU)**, and **pulse filtering unit (PFU)**. Each OPCM bank has a dedicated set of these five sub-units in the E-O-E control unit. Figure 5(a) shows the design of the E-O-E control unit in COSMOS and the internals of these sub-units.

Figure 5(b) illustrates the sequence of operations in the E-O-E control unit for write operation to a bank containing 512×512 tiles with 32×32 cells per tile (same design as that used in

Figure 4(e)). The AMU in the E-O-E control unit first receives the row address and then the column address from MC (Step 1). Depending on the addresses, the PSU in the E-O-E control unit selects the appropriate optical signals using the address mapping explained in Section 4.4 (Step 2). The PSU selects one optical signal for the row and 32 optical signals for the 32 columns in the row to write to 32 cells in a tile. In parallel with the write address, the DMU in the E-O-E control unit receives the write data from the MC (Step 3). The DMU generates a unique bias current for each of the 32 optical signals depending on write data and applies the currents to the **semiconductor optical amplifiers (SOA)** in the PAU (Step 4). The SOAs amplify the optical signals to the required intensities. These amplified signals and the optical signal (corresponding to the row) traverse through the silicon-photonic links to the appropriate OPCM cells in the bank and SET/RESET the cell (Step 5). The E-O-E control unit incurs a latency of T_{EO} cycles to map the address and data onto optical signals, resulting in a peak throughput of $1/T_{EO}$. It should be noted that the physical location of a cell in the OPCM array in COSMOS determines the level of losses that will be experienced by an optical signal that is writing to the cell. These losses in turn dictate the amplification of that optical signal in the E-O-E control unit. To address this, the E-O-E control unit uses the address mapping (refer to Figure 4(e)) to map the physical address to the corresponding OPCM cell that needs to be written. Based on the physical location of the cell, the DMU in the E-O-E control unit looks up a pre-programmed LUT, which holds the amplification factor required for each cell. The DMU applies a bias current as a function of this amplification factor to the PAU, which amplifies the optical signals to the required level.

Figure 5(c) illustrates the sequence of operations in the E-O-E control unit for the three-step read operation from a bank. In the first step, the AMU receives the row and column addresses from MC and selects the appropriate 32 optical signals in the PSU using the address mapping explained in Section 4.4 (Step 1.1). The DMU generates a low-intensity readout pulse (RD_1) and the PAU modulates this pulse on the 32 optical signals (Step 1.2). The optical signals traverse through the silicon-photonic link and then through the columns in the tile. The optical signals lose intensity as they pass through all the OPCM cells in their associated columns (Step 1.3). The intensities of these attenuated signals are recorded by the PFU (Step 1.4). The PFU then converts the optical intensities into electrical voltages, $V_{1,1}$, $V_{2,1}, \ldots, V_{32,1}$ (Step 1.5). In the second step, the DMU generates the RESET pulse. This RESET pulse is mapped onto the appropriate optical signals, and these signals are sent to the OPCM array (Step 2.1). The signals traverse through the silicon-photonic links and amorphize the OPCM cells corresponding to the read address (Step 2.2). In the third step, the DMU generates another readout pulse (RD_2) and the PAU modulates this pulse on a set of 32 optical signals (Step 3.1). These signals traverse through the silicon-photonic links and then through the appropriate columns in the tile. These signals, too, lose intensity as they pass through all the OPCM cells in their associated columns (Step 3.2). The PFU records these attenuated signals (Step 3.3) and converts these optical signals into electrical voltages $V_{1,2}$, $V_{2,2}, \ldots, V_{32,2}$ (Step 3.4). Finally, the PFU computes $V_{1,1}-V_{1,2}$, $V_{2,1}-V_{2,2}, \ldots, V_{32,1}-V_{32,2}$ to determine the data (Step 3.5) and sends the data to the MC. The PFU also writes this data back to the *holding buffer* in the DMU (Step 3.6).

7 EVALUATION METHODOLOGY

7.1 Multicore System with COSMOS

Our simulations for COSMOS are primarily based on parameters derived from prior multi-bit prototypes [26, 78]. These works demonstrate the scalability and precision of up to 5-bit/cell OPCM arrays under different load conditions using state-of-the-art optical devices for signal modulation and filtering. Moreover, the cell-to-cell static variability on refractive indices of GST elements have been shown to be minimal in these works [52, 78]. Due to the lack of active circuitry within the

Table 1. Architectural Details of the Simulated System

Processor, On-chip caches	
Cores	8-core, 2.5 GHz x86 ISA, Out-of-Order, 192 ROB entries, dispatch/fetch/issue/commit width=8
L1 caches	32 kB split L1 I$ and D$, 2-way, 1-cycle hit, 64 B, LRU, write-through, MSHR: 4 instruction & 32 data
L2 cache	Shared L2$, 2 MB, 8-way, 8-cycle hit, 64 B, LRU, write-back, MSHR: 32 (I & D)
Main memory ($2GB$)	
EPCM [20]	4 banks, 8 devices/rank, 1 rank/channel, bus width = 64, burst length = 4 $t_{SET} = 120ns$, $t_{RESET} = 50ns$, $t_{read} = 60ns$, $t_{BURST} = 4ns$
OPCM array in COSMOS [52, 78]	8 banks, 1 rank/channel, 1 device/rank, bus width = $32 \times b_{cell}$, burst length = 8 $t_{SET} = 160ns$, $t_{RESET} = 25ns$, $t_{read} = 25ns$, $t_{BURST} = 1ns$, $t_{EOE} = 5ns$

OPCM array, the dynamic variations in COSMOS due to thermal gradient is negligible. The minimal impact of these variations on GST cell operation enable high-fidelity optical detection and SET/RESET operation of OPCM arrays. As part of our future work, we plan to further explore the impact of these variations on reliability for larger-scale OPCM arrays at an architectural level. In our simulations, we use OPCM cell parameters (MLC, pulse intensity, and GST size) from real prototypes [26, 27, 78, 101], losses in optical elements based on prior demonstrations [9, 31, 52, 81], silicon-photonic link parameters (signals/waveguide, data rates, MRR sizes) from prior chip prototypes [8–10, 12]. In addition to 4-bit OPCM cells, we evaluate the potential performance benefits of an 8-bit OPCM cell. Though designing optical circuitry for high-precision filtering of 8-bit OPCM cells is a challenge, our goal is to motivate the potential benefits of higher-density OPCM arrays.

We use an 8-core processor for our evaluation. We primarily evaluate COSMOS with 4-bit MLC OPCM cells (given that OPCM cell with 5 $bits/cell$ has been prototyped [52]) against an EPCM with 2 $bits/cell$. We choose 2 $bits/cell$ instead of 4 $bits/cell$ [61] for EPCM, as prior works [13, 17] have shown that a cell density higher than 2 $bits/cell$ leads to unreliable EPCM designs. Table 1 details the processor and memory configurations. For processor-memory networks, we consider electrical as well as silicon-photonic links, with $1GT/s$ transfer rate per link. We obtain a peak bandwidth of $64GB/s$ in EPCM and $256GB/s$ in COSMOS. Peak bandwidth in COSMOS is calculated as the product of data rate, bus width (64 lines between process and memory), OPCM MLC capability as each optical signal can read/write 4bits/cell and the number of parallel banks ($1GT/s \times 64lines \times 4bits/cell \times 8banks = 256GBps$).

The OPCM array in COSMOS is organized as a single rank connected to a memory channel via the E-O-E control unit. Each one of the 8 OPCM banks has its dedicated set of DMU, ATU, PSU, PAU, and PFU in the E-O-E control unit. The average SET latency is $t_{SET} + t_{EOE}$, $165ns$, the RESET latency is $t_{RESET} + t_{EOE}$, $30ns$, and the read latency is t_{read} (time for three-step read operation) + t_{EOE}, i.e., $30ns$. A maximum of $t_{SET}/t_{EOE} = 32$ writes can be issued from the E-O-E control unit to OPCM in parallel. So, we can write $32 \times b_{cell}$ bits in parallel. A maximum of $t_{read}/t_{EOE} = 5$ reads can be issued from the E-O-E control unit to OPCM in parallel. So, we can read $5 \times b_{cell}$ bits in parallel. We use a holding buffer that is large enough (16 cache line slots from our evaluations) to avoid stalling any read/write memory requests from the MC.

7.2 Simulation Framework

We model the architectural specifications of the system in gem5 [14]. We conduct full-system simulations in gem5 with Ubuntu 12.04 OS and Linux kernel v4.8.13. We fast-forward to the end of

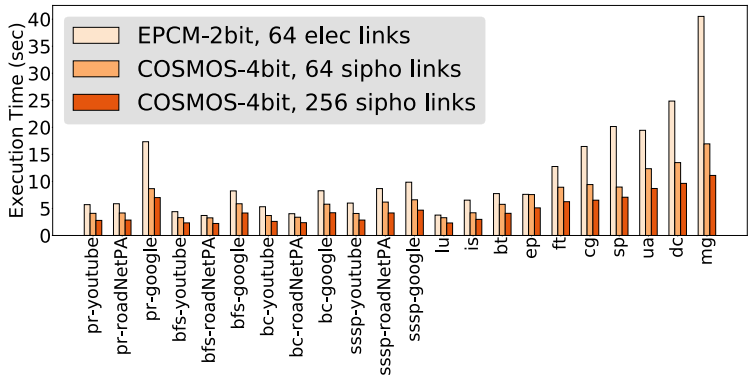

Fig. 6. Performance comparison of COSMOS with EPCM.

Linux boot and execute each workload for 10 billion instructions. The main memory models with the different timing parameters for DDR5 are modeled in DRAMSim2 [77]. For modeling EPCM and OPCM, we integrate NVMain2.0 [68] in gem5.

7.3 Workloads

We simulate graph applications from GAP-BS benchmark [11] and HPC applications from NAS-PB benchmark [7]. We evaluate the graph applications on three different input datasets from SNAP repository [49]: Google web graph (*google*), road network graph of Pennsylvania (*roadNetPA*), and YouTube online social network (*youtube*). For HPC applications from NAS-PB benchmark, we use the large dataset. We execute 8 threads of these applications in a workload.

8 EVALUATION RESULTS

8.1 COSMOS vs. EPCM

8.1.1 Performance. We compare EPCM (2bit MLC or EPCM-2bit) that uses 64 processor-to-memory electrical links with COSMOS (4bit OPCM cells, or COSMOS-4bit) that also uses 64 processor-to-memory silicon-photonic links, and with COSMOS-4bit that uses 256 processor-to-memory silicon-photonic links. Figure 6 shows the overall performance (execution time in seconds) for systems with these three configurations. Compared to the EPCM-2bit with 64 electrical links, COSMOS-4bit with 64 silicon-photonic links has on average 1.52× better performance across all workloads. This performance improvement is due to the higher *bits/access* throughput of COSMOS resulting from higher MLC capacity and the single-cycle latency in silicon-photonic links. Increasing the number of silicon-photonic links from 64 to 256 further improves the system performance. Compared to EPCM-2bit using 64 electrical links, we observe performance improvement of 2.14× on average for graph and HPC workloads with COSMOS-4bit using 256 silicon-photonic links. These performance benefits are due to denser WDM in silicon-photonic links. *The key takeaway from this comparison is that even though the OPCM cells suffer from long write latency similar to EPCM cells, the superior MLC capacity of OPCM cells that are directly accessed by high-bandwidth-density silicon-photonic links improves the system performance in COSMOS.*

8.1.2 Throughput. Figures 7(a) and 7(b) show the read and write throughput, respectively, of COSMOS-4bit with 256 silicon-photonic links and EPCM-2bit with 64 electrical links. Compared to EPCM-2bit with 64 electrical links, COSMOS-4bit with 256 silicon-photonic links theoretically has 8× higher peak throughput, i.e., 2× due to higher MLC capacity and the 4× due to the increased number of processor-to-memory links. Therefore, it is possible to issue increased number

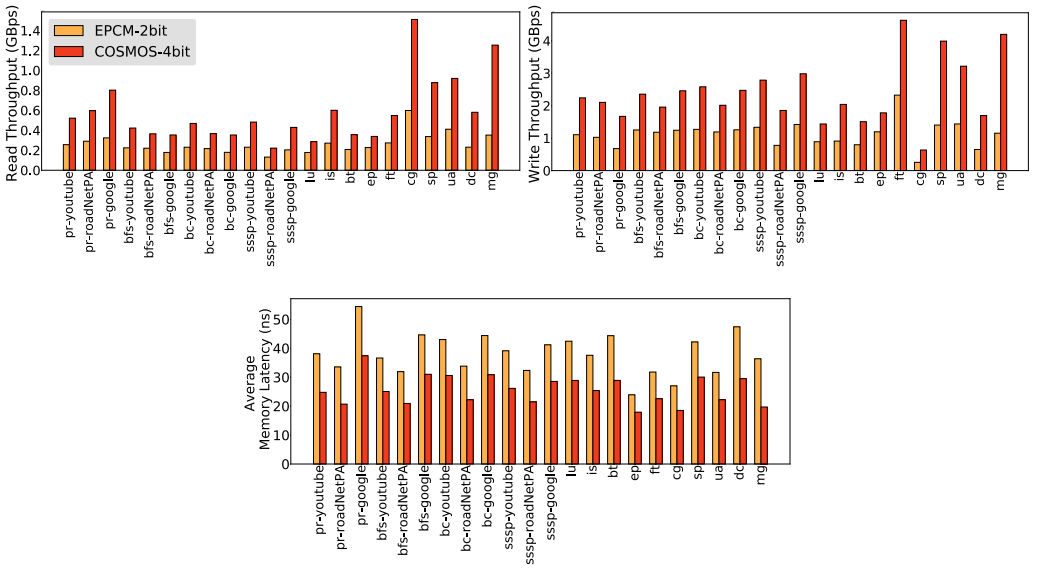

Fig. 7. (a) Read throughput, (b) Write throughput, (c) Average memory latency.

of parallel read and write operations in COSMOS-4bit. As a result, from Figure 7(a) and Figure 7(b), we observe that COSMOS-4bit can achieve 2.09× higher read throughput and 2.15× higher write throughput, respectively, than EPCM-2bit for graph and HPC workloads. This increased read and write throughput of COSMOS-4bit hides the long write latencies. Figure 7(c) shows that the average memory latency (read+write) of COSMOS-4bit is 33% lower than EPCM-2bit across all workloads. *The key insight from this study is the increased read and write throughput provided by the higher MLC capacity and the silicon-photonic links hide the long write latencies of OPCM cells in COSMOS.*

8.1.3 Energy Consumption. The primary contributors to the overall power consumption during the read and write operations are the different active components in the E-O-E control unit and the laser sources that drive the silicon-photonic links. The OPCM array in COSMOS consists of only passive optical devices, so it does not consume any active or idle power. The electrical power consumed in the laser source is proportional to its optical output power, which in turn depends on the optical losses in the path of the optical signal and the minimum power required to switch the farthest GST element. Table 2 lists the optical losses in the various components and the maximum switching power required at the GST element in **decibels (dB)**. The various optical losses and SOA gains are obtained from prior characterization works [9, 31, 52, 81]. By accounting for the wall-plug efficiency, we calculate the minimum required laser power per optical signal as $0.95mW$. Aggregating the laser power for all optical signals required in a $2GB$ COSMOS system, we get a total laser power of $16.38W$.

In the E-O-E control unit, the current-DAC in DMU and the ADC in PFU consume $0.3mW$ each [74]. For OPCM-4bit, 32 write operations can be issued in parallel per bank, i.e., we can write $32 \times b_{cell} \times 8 = 128B$ in parallel with an average write latency of $160ns$. That aggregates to writing 2 cache lines of $64B$ each in parallel. A cache line is interleaved across 4 banks and is row aligned in an OPCM tile. Therefore, we need 4 row optical signals and 4×32 column optical signals to write a cache line. Therefore, the total power of the laser, SOAs, and DACs in the E-O-E control unit for writing 2 cache lines in parallel aggregates to $334.8mW$. This equates to $40.68pJ/bit$ for writing to COSMOS-4bit.

Table 2. Optical Power Budget for 2*GB* COSMOS

Loss/gain component	Single	Total
Coupling loss	$-1dB$	$-1dB$ [9]
MRR drop loss (E-O-E control)	$-0.5dB$ [31]	$-0.5dB$
MRR through loss (E-O-E control)	$-0.05dB$ [31]	$-3.2dB$
Propagation loss (Laser to SOA)	$-0.3dB/cm$ [81]	$-0.09dB$
SOA gain	$+20dB$	$+20dB$
Propagation loss (SOA to OPCM)	$-0.3dB/cm$ [81]	$-0.09dB$
Bending loss	$-0.167dB$ [81]	$-0.167dB$
MRR drop loss (OPCM)	$-0.5dB$ [31]	$-0.5dB$
MRR through loss (OPCM)	$-0.05dB$ [31]	$-3.2dB$
Propagation loss (in OPCM)	$-0.03dB/cm$ [81]	$-4.91dB$
Max. power required to SET the GST	$\frac{135pJ}{250ns}$ [52]	$-2.67dBm$
Power per optical signal		$-7.22dBm = 0.19mW$
Laser wall-plug efficiency		20%
Total laser power		$16.38W$

The table shows optical power losses and SOA gain along the optical path from laser source to OPCM cells.

Table 3. Energy-per-bit for Read and Write Accesses

Energy-per-bit (pJ/bit)	EPCM-2bit	COSMOS-4bit
Write	243	40.68
Read	44.5	11.6
Opportunistic Writeback	NA	40.68

For read operation, up to 5 read operations can be issued in parallel per bank, i.e., $5 \times b_{cell} \times 8 = 20B$ bits in parallel, with a read latency of $25ns$. The total power of the laser, SOA, DAC, and ADC in E-O-E control for 5 parallel read operations is $9.3mW$, resulting in a read energy of $11.6pJ/bit$ for COSMOS-4bit. The energy consumed in the electrical links connecting the processor and the E-O-E control unit is $<1pJ/bit$ [21]. For EPCM, we use parameters from the HSpice models in prior work [39] and model them in NVSim [24] to estimate the energy-per-bit for read and write operations. The opportunistic writeback operation in COSMOS uses the same energy as that required for write operation. Table 3 shows the energy-per-bit for EPCM-2bit and COSMOS-4bit. The read and write energy-per-bit of COSMOS-4bit are 3.8× and 5.97× lower, respectively, than that of EPCM-2bit.

8.2 Sensitivity Analysis of COSMOS

8.2.1 MLC Values. Rios et al. gave the first demonstration of a 2-bit OPCM cell operation [78]. Advances in optical signaling and control have resulted in the demonstration of denser multilevel OPCM cells. Li et al. demonstrated 5–6 bits per OPCM cell [52]. Further prototypes have demonstrated scalable integration of OPCM cell arrays in silicon and silicon nitride platforms [27, 51]. With the maturity in optical integration technologies, we also evaluate the performance of 8-bit OPCM cells to provide a forward-looking comprehensive view of the potential benefits of developing higher bit density OPCM cells compared to DRAM. We compare the performance of COSMOS that uses OPCM cells with different MLC capacities, ranging from 2 *bits/cell* to 8 *bits/cell*, for

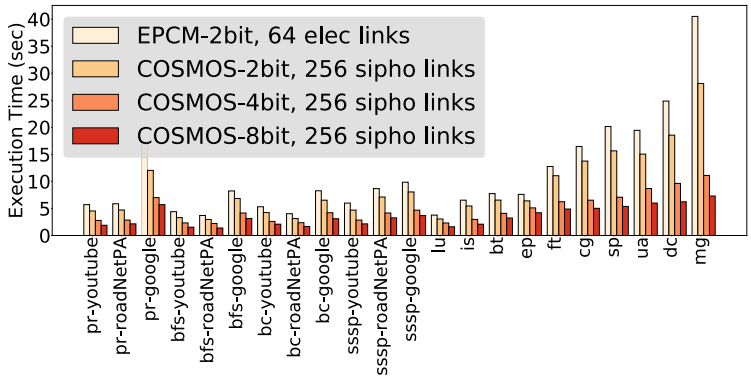

Fig. 8. Performance comparison of COSMOS with different MLC.

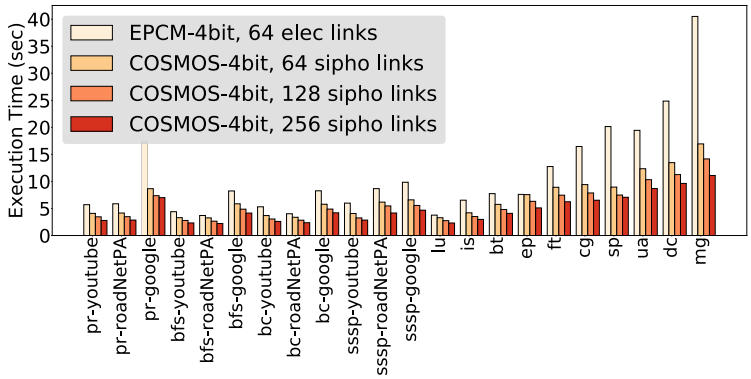

Fig. 9. Performance comparison of COSMOS with different number of silicon-photonic links.

the same number of silicon-photonic links (see Figure 8). The performance across applications increases, on average, by 39.2% and 26.4% as the MLC capacity of OPCM cells increases from 2 *bits/cell* to 4 *bits/cell* and from 4 *bits/cell* to 8 *bits/cell*, respectively. An OPCM cell with higher MLC capacity will provide higher memory throughput.

8.2.2 Number of Silicon-photonic Links. We compare the performance of COSMOS-4bit with different number of silicon-photonic links (see Figure 9). Multiplexing a higher number of optical signals in silicon-photonic links enables parallel read and write accesses of a higher number of OPCM cells. Due to this increased throughput, the overall system performance improves as the number of silicon-photonic links increases. We observe a performance improvement of 29.3% (on average) for COSMOS-4bit with 256 silicon-photonic links over COSMOS-4bit with 64 links.

8.2.3 Holding Buffer. As discussed earlier, in absence of the *holding buffer*, the read data needs to be written back to the OPCM cells immediately after readout due to the destructive read operation. Therefore, the complete read operation incurs a total latency of readout latency (25*ns*) + writeback latency (160*ns*). In contrast, when the E-O-E control unit uses a *holding buffer*, the read data is stored in the *holding buffer* at the end of read operation. The data from the *holding buffer* is written back to the OPCM cells only when the DB in the E-O-E control unit is empty, ensuring that the writeback operation does not stall any critical read and write operations. Using the

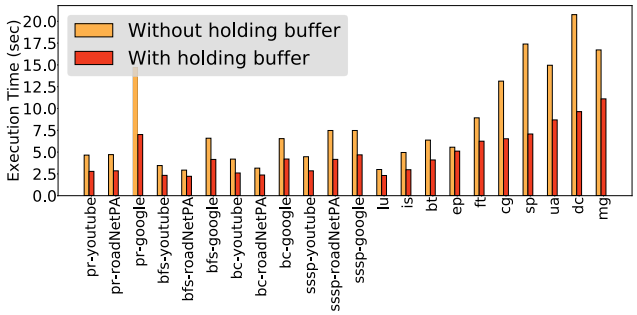

Fig. 10. Performance comparison of COSMOS with and without holding buffer for opportunistic writeback in read operation.

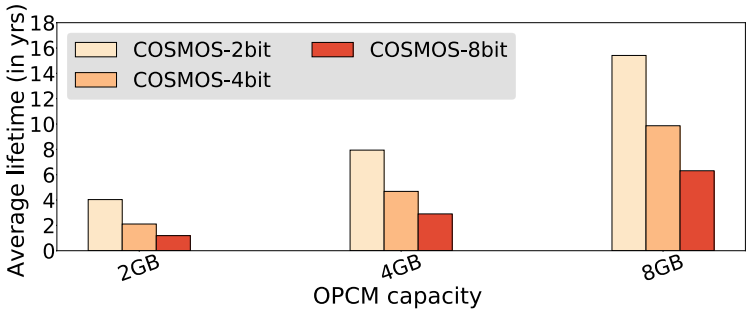

Fig. 11. Average lifetime (in years) of COSMOS with different MLC capacities of OPCM cells and different memory capacities.

highest read and write rate of the workloads that we evaluated, we determine that a *holding buffer* with 16 cache line slots, i.e., $1KB$, is enough to avoid any memory read/write stalls. The *holding buffer* occupies <1,000 μm^2 area and can be integrated into the E-O-E control unit with minimal overhead. Figure 10 shows that using a holding buffer in COSMOS provides 59.2% average performance uplift.

8.3 Endurance Analysis of COSMOS

Similar to EPCM, OPCM cells have lower endurance due to cell wearout. The OPCM cell endurance depends on how often we write to that cell [70]. Given that the read operation in COSMOS also includes a write (RESET) in step 2, the endurance of OPCM cells also depends on the read rate. We estimate the COSMOS lifetime using the equation proposed by Qureshi et al. [71]:

$$Y = \frac{S.W_m}{B.F.2^{25}},$$

where Y is lifetime in years, W_m is maximum allowable writes per cell (10^6 for OPCM cells [52, 78]), B is write rate in bytes/cycle (average read+write rate across graph and HPC workloads), F is core frequency in Hz ($1GHz$), and S is COSMOS size in bytes ($2GB$, $4GB$, and $8GB$).

Figure 11 plots the average lifetime for OPCM with different MLC capacities. Here, we assume that for a given memory size, all MLC options use the same number of silicon-photonic links. Hence, the COSMOS with 8-bit OPCM cells has higher effective throughput than the COSMOS with 4-bit OPCM cells, and so an application running on COSMOS-8bit runs faster than an application running on COSMOS-4bit. As a result, for an application, even if the absolute number of

Table 4. Dimensions of Optical Devices in the OPCM Array

Optical device	Dimension
GST	$500nm \times 500nm$ [75, 78]
Separation between adjacent GSTs	$50nm$ [32]
MRR diameter	$5\mu m$ [50]

Table 5. Bit Density ($bits/mm^2$) of Memory Technologies

Memory technology	Area of $2GB$ memory	Bit density ($bits/mm^2$)
DDR4	$224mm^2$ [1]	$9.14MB/mm^2$
HBM2.0	$91.99mm^2$ [38]	$22.26MB/mm^2$
EPCM-2bit	$336mm^2$ (simulated [24])	$6.095MB/mm^2$
3D OPCM-4bit	$268.43mm^2$ (calculated)	$7.63MB/mm^2$
3D OPCM-8bit	$67.1mm^2$ (calculated)	$30.52MB/mm^2$

memory writes is same for both COSMOS-8bit and COSMOS-4bit, the average number of *writes/second* to COSMOS-8bit is higher than the average number of *writes/second* to COSMOS-4bit. Hence, the lifetime of COSMOS-8bit is lower than that of the COSMOS-4bit, and similarly, the lifetime of COSMOS-4bit is lower than that of COSMOS-2bit.

8.4 Area Analysis of the OPCM Array

To design the OPCM array in COSMOS, we use the prototype of a GST element developed by Rios et al. [75, 78] and the MRR dimensions from prior work, as shown in Table 4. We use 3D stacking for OPCM array, with different banks stacked vertically (one bank per layer). The multi-mode waveguides in the interposer are routed vertically, and at each layer single-mode MRRs filter out the mode of all optical signals that belong to its corresponding bank. For a $2GB$ 4-bit OPCM array with 8 banks, a single bank consists of 1,024 tiles with 32 cells/tile and a row and column of MRRs, as shown in Figure 4(b).[2] A bank, therefore, is composed of $1{,}024 \times 32$ GSTs along a row/column with $(1{,}024 \times 32 - 1) \times 50nm$ of separation between GSTs and a single row/column of MRRs at the beginning. Using the dimensions of these optical devices listed in Table 4, we calculate the area of a $2GB$ OPCM array and its bit density and report it in Table 5.

We compare the area and bit density of the 3D-stacked OPCM array in COSMOS with DDR4, 3D-stacked HBM2.0, and EPCM-2bit memory system (see Table 5).[3] With current OPCM cell footprints, 3D-stacked OPCM-4bit has 1.2× and 2.9× lower bit density than DDR4 and HBM2.0, respectively, and 1.25× higher bit density than EPCM-2bit. 3D-stacked OPCM-8bit has 3.4×, 1.4×, and 5× higher bit density than DDR4, HBM2.0, and EPCM-2bit, respectively. Nevertheless, device-level research efforts have demonstrated that GST elements are highly scalable and can retain the electrical and optical characteristics at amorphous and crystalline states [73, 88]. An aggressive chip prototype with $200nm \times 200nm$ GST element with $50nm$ separation has been recently fabricated [32]. These aggressive optical fabrication technologies promise achieving several orders higher densities for OPCM arrays than current DRAM technologies.

[2]The tile size is limited by the number of unique optical signals in C and L bands with sufficient guardbands (32 in our case). The number of banks depends on the number of unique electromagnetic modes that can be supported (8 in our case).
[3]DDR5 area models were not publicly available at the time of submitting the manuscript. So, we report a comparison with DDR4.

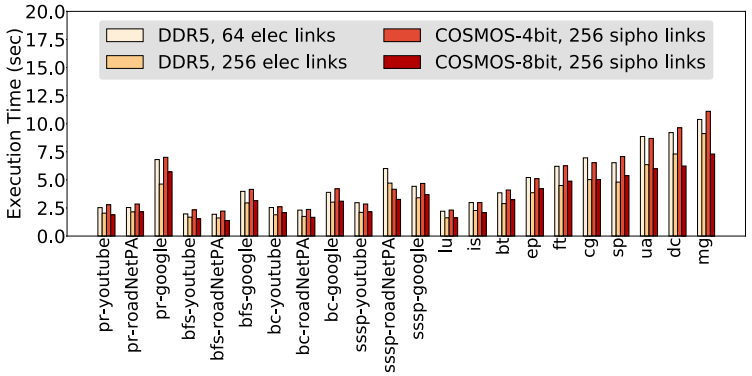

Fig. 12. Performance comparison of OPCM with DDR5.

8.5 COSMOS vs. DRAM

The overarching goal of COSMOS is to replace DRAM systems that are used widely in computing systems. We noted that though all other NVM systems (in their current form) provide non-volatility, data persistence, and high scalability, their poor performance negates their benefits and makes them impractical to replace DRAM systems. We, therefore, compare the performance and energy of current state-of-the-art DRAM systems, DDR5 with 64 electrical links, DDR5 with 256 silicon-photonic links [12], COSMOS-4bit with 256 silicon-photonic links, and COSMOS-8bit with 256 silicon-photonic links. Figure 12 shows the overall system performance across the four configurations. For DDR5, replacing 64 electrical links with 256 silicon-photonic links provides 24% average performance improvement. This improvement results from the higher throughput due to dense WDM and single-cycle latency of silicon-photonic links. With COSMOS-4bit with 256 silicon-photonic links, we obtain 1.2% improvement in performance compared to DDR5 with 64 electrical links. This is in stark contrast to EPCM-2bit, which performs 4–5× worse than DDR5. COSMOS-8bit with 256 silicon-photonic links performs 24.7% better than DDR5 with 64 electrical links and 1.8% better than DDR5 with 256 silicon-photonic links. Here, the increased read and write throughput due to the higher MLC capacity and dense WDM silicon-photonic links reduces the average memory access latency of COSMOS and in turn improves performance. Figure 7(c) shows the average memory latency in COSMOS is 33.64*ns* across all workloads, which is lower than DDR5 DRAM (48*ns*).

Though we evaluate DDR5 memory with silicon-photonic links, such a system encounters several design challenges. To support silicon-photonic links in DDR5, memory requests from MC require an E-O conversion in MC and an O-E conversion in memory, and memory responses from DDR5 require an E-O conversion in memory and an O-E conversion in MC. Effectively, we need two extra conversions on the memory side. The active peripheral circuitry to support E-O-E conversions within memory increases the power density and raises thermal concerns. Due to the high thermal sensitivity of MRRs, there is a need for active thermal management. The power and resulting thermal concerns affect the reliability of optical communication in DRAM systems.

We observe that COSMOS with 4 *bits/cell* OPCM array demonstrates similar performance and energy characteristics as current state-of-the-art DDR5 systems, while COSMOS with 8 *bits/cell* OPCM array improves performance. This is particularly exciting, as COSMOS exhibits zero leakage power, better scaling, and non-volatility, making it a viable replacement for DRAM in the near future.

Table 6. Survey of Research Efforts to Improve Write Performance and Write Energy for Using EPCM as Main Memory

	Fine-grained power budgeting [34]	Write truncation [36]	Logical decoupling & mapping [97]	Proactive SET [69]	Partition-aware scheduling [83]	Double-XOR mapping [25]	Boosting rank parallelism [5]	**COSMOS**
Performance gains	76%	26%	19.2%	34%	28%	12%	16.7%	2.31×
Energy reductions	NR	NR	14.4%	25%	20%	NR	NR	4×

The performance gains and energy reductions are shown in comparison to a naive EPCM system. (NR: Not reported).

9 RELATED WORK

9.1 Phase Change Memories

Several works have proposed architectural and management policies to address the PCM challenges and have designed EPCM systems either as a standalone main memory, as part of hybrid DRAM-PCM systems or as a storage memory between DRAM and flash memory [5, 25, 33, 34, 36, 39, 43, 46, 47, 69, 71, 72, 83, 85, 94, 97]. Most of these efforts have focused on addressing the long write latency and high write energy. A summary of these efforts is shown in Table 6. Hybrid DRAM-PCM systems leverage the higher bit density in PCMs for improved performance, but at the cost of higher write energy [33, 46, 47, 71, 72]. To address PCM cell wearout, the techniques to enhance the write endurance include rotation-based wear leveling [70], process variation-aware leveling [23, 102], and writeback minimization and endurance management [28]. Due to lower write endurance, PCM cells are also susceptible to malicious write attacks. Common strategies employed in EPCMs to thwart these attacks and improve reliability include write-efficient data encryption [98], multi-way wear leveling [100], write-verify-write [62], or randomized address mapping [80]. These techniques can be readily deployed in OPCM. While several approaches discussed above address EPCM limitations, EPCM is not yet a viable alternative for DRAM due to their scalability and reliability challenges, high energy overhead, and constrained bandwidth density.

In Table 6, we see that optical control of PCMs combined with silicon-photonic links significantly improves performance and lowers energy without using any of the complementary methods provided in prior work. Applying these complementary methods to OPCM will further improve its performance and lower energy.

9.2 Silicon-photonic Links and OPCM Cells

Silicon-photonic links have enabled high bandwidth-density and low-energy communication between processor and memory [9, 10, 12, 22, 59, 84, 86, 87]. To provide high DRAM internal bandwidth, Beamer et al. [12] proposed a joint silicon-photonic link and electro-photonic DRAM design. However, the O-E-O conversion in DRAM adds to the latency. Optical control of memory cells can avoid this O-E-O conversion and enable signals in the silicon-photonic links to directly access the cells and deliver higher memory throughput.

Several recent efforts have prototyped GST-basd PCM cells with optical control. Rios et al. demonstrate the optical control of multi-bit GST-based PCMs with fast readout and low switching energies [78]. Zhang et al. [101] present an approach to selectively couple optical signals from MRR to GST. Feldman et al. [26, 27] design a prototype of a monolithic OPCM array based on waveguide crossing but not a comprehensive memory microarchitecture and access protocols. Subsequent efforts demonstrate higher bit density per GST [52], in-memory computing on PCM cells

using optical signals [76], basic arithmetic operations in OPCM [26, 27], and a behavioral model for neuromorphic computing [18]. **We are the first to propose a comprehensive OPCM microarchitecture with custom read/write access protocols and design an E-O-E control unit to interface the OPCM array with the processor.**

10 CONCLUSION

EPCM systems suffer from long write latencies and high write energies, yielding poor performance and high energy consumption for data-intensive applications. In contrast, OPCM technology provides the opportunity to design high-performance and low-energy memory systems due to its higher MLC capacity and the direct cell access via high-bandwidth-density and low-latency silicon-photonic links. Adapting the current EPCM design architecture for OPCM systems, however, raises major issues in terms of latency, energy, and thermal concerns, thereby rendering such a design impractical. We are the first to architect a complete memory system, COSMOS, which consists of an OPCM array microarchitecture, a read/write access protocol tailored for OPCM technology, and an E-O-E control unit that interfaces the OPCM array with the MC. Our evaluations show that, compared to an EPCM system, our proposed COSMOS system provides 2.09× higher read throughput and 2.15× higher write throughput, thereby reducing the execution time by 2.14×, read energy by 1.24×, and write energy by 4.06×.

We show that COSMOS designed with state-of-the-art technology provides similar performance and energy as DDR5. This is a significant finding, as future higher-density OPCM cells are expected to provide better performance. Our promising first version of a COSMOS architecture opens doors for new architecture-level, circuit-level, and system-level methods to enable practical integration of OPCM-based main memory in future computing systems. Moreover, the high-throughput and scalable OPCM technology ushers in interesting research opportunities in persistent memory, in-memory computing, and accelerator-specific memory designs.

REFERENCES

[1] [n.d.]. DDR4 area. Retrieved from https://www.micron.com/products/dram/ddr4-sdram/.

[2] José L. Abellán, Ayse K. Coskun, Anjun Gu, Warren Jin, Ajay Joshi, Andrew B. Kahng, Jonathan Klamkin, Cristian Morales, John Recchio, Vaishnav Srinivas, et al. 2016. Adaptive tuning of photonic devices in a photonic NoC through dynamic workload allocation. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst.* 36, 5 (2016), 801–814.

[3] Abbas Acar, Hidayet Aksu, A. Selcuk Uluagac, and Mauro Conti. 2018. A survey on homomorphic encryption schemes: Theory and implementation. *Comput. Surv.* 51, 4 (2018), 1–35.

[4] Andrew Alduino, Ling Liao, Richard Jones, Michael Morse, Brian Kim, Wei-Zen Lo, Juthika Basak, Brian Koch, Hai-Feng Liu, Haisheng Rong, Matthew Sysak, Christine Krause, Rushdy Saba, Dror Lazar, Lior Horwitz, Roi Bar, Stas Litski, Ansheng Liu, Kevin Sullivan, Olufemi Dosunmu, Neil Na, Tao Yin, Frederick Haubensack, I-Wei Hsieh, John Heck, Robert Beatty, Hyundai Park, Jock Bovington, Simon Lee, Hat Nguyen, Hinmeng Au, K. Nguyen, P. Merani, M. Hakami, and M. Paniccia. 2010. Demonstration of a high speed 4-channel integrated silicon photonics WDM link with hybrid silicon lasers. In *Proceedings of the Integrated Photonics Research, Silicon and Nanophotonics and Photonics in Switching*. Optical Society of America, PDIWI5.

[5] Mohammad Arjomand, Mahmut T. Kandemir, Anand Sivasubramaniam, and Chita R. Das. 2016. Boosting access parallelism to PCM-based main memory. In *Proceedings of the International Symposium on Computer Architecture*. 695–706.

[6] Meisam Bahadori, Robert Polster, Sébastien Rumley, Yvain Thonnart, José-Luis Gonzalez-Jimenez, and Keren Bergman. 2016. Energy-bandwidth design exploration of silicon photonic interconnects in 65nm CMOS. In *Proceedings of the Optical Interconnects Conference*. 2–3.

[7] David H. Bailey, Eric Barszcz, John T. Barton, David S. Browning, Robert L. Carter, Leonardo Dagum, Rod A. Fatoohi, Paul O. Frederickson, Thomas A. Lasinski, Rob S. Schreiber, H. D. Simon, V. Venkatakrishnan, and S. K. Weeratunga. 1991. The NAS parallel benchmarks. *Int. J. Supercomput. Applic.* 5, 3 (1991), 63–73.

[8] T. Barwicz, H. Byun, F. Gan, C. W. Holzwarth, M. A. Popovic, P. T. Rakich, M. R. Watts, E. P. Ippen, F. X. Kärtner, H. I. Smith, J. S. Orcutt, R. J. Ram, V. Stojanovic, O. O. Olubuyide, J. L. Hoyt, S. Spector, M. Geis, M. Grein, T. Lyszczarz,

and J. U. Yoon. 2007. Silicon photonics for compact, energy-efficient interconnects. *J. Optical Network.* 6, 1 (2007), 63–73.

[9] Christopher Batten, Ajay Joshi, Jason Orcutt, Anatol Khilo, Benjamin Moss, Charles W. Holzwarth, Miloš A. Popovic, Hanqing Li, Henry I. Smith, Judy L. Hoyt, F. X. Kartner, Rajeev J. Ram, Vladimir Stojanović, and Krste Asanović. 2008. Building many-core processor-to-DRAM networks with monolithic CMOS silicon photonics. In *Proceedings of the Symposium on High Performance Interconnects*. 21–30.

[10] C. Batten, A. Joshi, V. Stojanovic, and K. Asanovic. 2012. Designing chip-level nanophotonic interconnection networks. *IEEE J. Emerg. Select. Topics Circ. Syst.* 2, 2 (2012), 137–153.

[11] Scott Beamer, Krste Asanović, and David Patterson. 2015. The GAP benchmark suite. *arXiv preprint arXiv:1508.03619* (2015).

[12] Scott Beamer, Chen Sun, Yong-jin Kwon, Ajay Joshi, Christopher Batten, Vladimir Stojanovic, and Krste Asanovi. 2009. Re-architecting DRAM with monolithically integrated silicon photonics. In *Proceedings of the International Symposium on Computer Architecture*. 129–140.

[13] Ferdinando Bedeschi, Rich Fackenthal, Claudio Resta, Enzo Michele Donze, Meenatchi Jagasivamani, Egidio Cassiodoro Buda, Fabio Pellizzer, David W. Chow, Alessandro Cabrini, Giacomo Matteo Angelo Calvi, Roberto Faravelli, Andrea Fantini, Guido Torelli, Duane Mills, Roberto Gastaldi, and Giulio Casagrande. 2008. A multi-level-cell bipolar-selected phase-change memory. In *Proceedings of the International Solid-State Circuits Conference*. 428–625.

[14] Nathan Binkert, Bradford Beckmann, Gabriel Black, Steven K. Reinhardt, Ali Saidi, Arkaprava Basu, Joel Hestness, Derek R. Hower, Tushar Krishna, Somayeh Sardashti, Rathijit Sen, Korey Sewell, Muhammad Shoaib, Nilay Vaish, Mark D. Hill, and David A. Wood. 2011. The gem5 simulator. *ACM SIGARCH Comput. Archit. News* 39, 2 (2011), 1–7.

[15] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).

[16] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, and B. Rajendran. 2010. Phase change memory technology. *J. Vacuum Sc. Technol. B, Nanotechnol. Microelectron.: Mater., Process., Measur., Phenom.* 28, 2 (2010), 223–262.

[17] A. Cabrini, S. Braga, A. Manetto, and G. Torelli. 2009. Voltage-driven multilevel programming in phase change memories. In *Proceedings of the International Workshop on Memory Technology, Design, and Testing*. 3–6.

[18] Santiago G.-C. Carrillo, Emanuele Gemo, Xuan Li, Nathan Youngblood, Andrew Katumba, Peter Bienstman, Wolfram Pernice, Harish Bhaskaran, and C. David Wright. 2019. Behavioral modeling of integrated phase-change photonic devices for neuromorphic computing applications. *APL Mater.* 7, 9 (2019), 091113.

[19] Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. 2017. Homomorphic encryption for arithmetic of approximate numbers. In *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 409–437.

[20] Youngdon Choi, Ickhyun Song, Mu-Hui Park, Sanghoan Chang Hoeju Chung, Beakhyoung Cho, Jinyoung Kim, Younghoon Oh, Dukmin Kwon, Jung Sunwoo, Junho Shin, Yoohwan Rho, Changsoo Lee, Min Gu Kang, Jaeyun Lee, Yongjin Kwon, Soehee Kim, Jaewhan Kim, Yong jun Lee, Qi Wang, Sooho Cha, Sujin Ahn, Hideki Horii, Jaewook Lee, KiSeung Kim, Han-Sung Joo, KwangJin Lee, Yeong-Taek Lee, Jei-Hwan Yoo, and Gitae Jeong. 2012. A 20nm 1.8V 8Gb PRAM with 40MB/s program bandwidth. In *Proceedings of the International Solid-State Circuits Conference*. 46–48.

[21] Ayse Coskun, Furkan Eris, Ajay Joshi, Andrew B. Kahng, Yenai Ma, Aditya Narayan, and Vaishnav Srinivas. 2020. Cross-layer co-optimization of network design and chiplet placement in 2.5 D systems. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst.* 39, 12 (2020), 5183–5196.

[22] Yigit Demir, Yan Pan, Seukwoo Song, Nikos Hardavellas, John Kim, and Gokhan Memik. 2014. Galaxy: A high-performance energy-efficient multi-chip architecture using photonic interconnects. In *Proceedings of the International Conference on Supercomputing*. 303–312.

[23] Jianbo Dong, Lei Zhang, Yinhe Han, Ying Wang, and Xiaowei Li. 2011. Wear rate leveling: Lifetime enhancement of PRAM with endurance variation. In *Proceedings of the Design Automation Conference*. 972–977.

[24] Xiangyu Dong, Cong Xu, Yuan Xie, and Norman P. Jouppi. 2012. NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst.* 31, 7 (2012), 994–1007.

[25] Yu Du, Miao Zhou, Bruce R. Childers, Daniel Mossé, and Rami Melhem. 2013. Bit mapping for balanced PCM cell programming. In *Proceedings of the International Symposium on Computer Architecture*. 428–439.

[26] J. Feldmann, M. Stegmaier, N. Gruhler, C. Ríos, H. Bhaskaran, C. D. Wright, and W. H. P. Pernice. 2017. Calculating with light using a chip-scale all-optical abacus. *Nat. Commun.* 8, 1 (2017), 1–8.

[27] Johannes Feldmann, Nathan Youngblood, Xuan Li, C. David Wright, Harish Bhaskaran, and Wolfram H. P. Pernice. 2019. Integrated 256 cell photonic phase-change memory with 512-bit capacity. *IEEE J. Select. Topics Quant. Electron.* 26, 2 (2019), 1–7.

[28] Alexandre P. Ferreira, Miao Zhou, Santiago Bock, Bruce Childers, Rami Melhem, and Daniel Mossé. 2010. Increasing PCM main memory lifetime. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition*. 914–919.

[29] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu. 2019. An adaptive ensemble machine learning model for intrusion detection. *IEEE Access* 7 (2019), 82512–82521. DOI: https://doi.org/10.1109/ACCESS.2019.2923640

[30] Joseph E. Gonzalez, Yucheng Low, Haijie Gu, Danny Bickson, and Carlos Guestrin. 2012. PowerGraph: Distributed graph-parallel computation on natural graphs. In *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation*. 17–30.

[31] Paolo Grani and Sandro Bartolini. 2014. Design options for optical ring interconnect in future client devices. *ACM J. Emerg. Technol. Comput. Syst.* 10, 4 (2014), 1–25.

[32] Peiman Hosseini, C. David Wright, and Harish Bhaskaran. 2014. An optoelectronic framework enabled by low-dimensional phase-change films. *Nature* 511, 7508 (2014), 206–211.

[33] Gangyong Jia, Guangjie Han, Jinfang Jiang, and Li Liu. 2016. Dynamic adaptive replacement policy in shared last-level cache of DRAM/PCM hybrid memory for big data storage. *IEEE Trans. Industr. Inform.* 13, 4 (2016), 1951–1960.

[34] Lei Jiang, Youtao Zhang, Bruce R. Childers, and Jun Yang. 2012. FPB: Fine-grained power budgeting to improve write throughput of multi-level cell phase change memory. In *Proceedings of the International Symposium on Microarchitecture*. 1–12.

[35] Lei Jiang, Bo Zhao, Jun Yang, and Youtao Zhang. 2014. A low power and reliable charge pump design for phase change memories. *Proc. Int. Symp. Comput. Archit.* 42, 3 (2014), 397–408.

[36] Lei Jiang, Bo Zhao, Youtao Zhang, Jun Yang, and Bruce R. Childers. 2012. Improving write operations in MLC phase change memory. In *Proceedings of the International Symposium on High Performance Computer Architecture*. 1–10.

[37] Uksong Kang, Hak-Soo Yu, Churoo Park, Hongzhong Zheng, John Halbert, Kuljit Bains, S. Jang, and Joo Sun Choi. 2014. Co-architecting controllers and DRAM to enhance DRAM process scaling. The Memory Forum, Vol. 14.

[38] Joonyoung Kim and Younsu Kim. 2014. HBM: Memory solution for bandwidth-hungry processors. In *Proceedings of the Hot Chips Symposium*. 1–24.

[39] Nam Sung Kim, Choungki Song, Woo Young Cho, Jian Huang, and Myoungsoo Jung. 2019. LL-PCM: Low-latency phase change memory architecture. In *Proceedings of the Design Automation Conference*. 1–6.

[40] Seong Keun Kim, Sang Woon Lee, Jeong Hwan Han, Bora Lee, Seungwu Han, and Cheol Seong Hwang. 2010. Capacitors with an equivalent oxide thickness of $<0.5nm$ for nanoscale electronic semiconductor memory. *Adv. Funct. Mater.* 20, 18 (2010), 2989–3003.

[41] Seong Keun Kim and Mihaela Popovici. 2018. Future of dynamic random-access memory as main memory. *MRS Bull.* 43, 5 (2018), 334.

[42] A. V. Krishnamoorthy, H. Schwetman, X. Zheng, and R. Ho. 2015. Energy-efficient photonics in future high-connectivity computing systems. *J. Lightw. Technol.* 33, 4 (2015), 889–900.

[43] Manuel Le Gallo, Abu Sebastian, Roland Mathis, Matteo Manica, Heiner Giefers, Tomas Tuma, Costas Bekas, Alessandro Curioni, and Evangelos Eleftheriou. 2018. Mixed-precision in-memory computing. *Nat. Electron.* 1, 4 (2018), 246–253.

[44] Benjamin C. Lee, Engin Ipek, Onur Mutlu, and Doug Burger. 2009. Architecting phase change memory as a scalable DRAM alternative. In *Proceedings of the International Symposium on Computer Architecture*. 2–13.

[45] B. G. Lee, X. Chen, A. Biberman, X. Liu, I. Hsieh, C. Chou, J. I. Dadap, F. Xia, W. M. J. Green, L. Sekaric, Y. A. Vlasov, R. M. Osgood, and K. Bergman. 2008. Ultrahigh-bandwidth silicon photonic nanowire waveguides for on-chip networks. *IEEE Photon. Technol. Lett.* 20, 6 (2008), 398–400.

[46] Hyung Gyu Lee, Seungcheol Baek, Chrysostomos Nicopoulos, and Jongman Kim. 2011. An energy-and performance-aware DRAM cache architecture for hybrid DRAM/PCM main memory systems. In *Proceedings of the International Conference on Computer Design*. 381–387.

[47] Soyoon Lee, Hyokyung Bahn, and Sam H. Noh. 2011. Characterizing memory write references for efficient management of hybrid PCM and DRAM memory. In *Proceedings of the International Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems*. 168–175.

[48] Charles Lefurgy, Karthick Rajamani, Freeman Rawson, Wes Felter, Michael Kistler, and Tom W. Keller. 2003. Energy management for commercial servers. *Computer* 36, 12 (2003), 39–48.

[49] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. Retrieved from http://snap.stanford.edu/data.

[50] Cheng Li, Rui Bai, Ayman Shafik, Ehsan Zhian Tabasy, Geng Tang, Chao Ma, Chin-Hui Chen, Zhen Peng, Marco Fiorentino, Patrick Chiang, and Samuel Palermo. 2013. A ring-resonator-based silicon photonics transceiver with

bias-based wavelength stabilization and adaptive-power-sensitivity receiver. In *Proceedings of the International Solid-State Circuits Conference*. 124–125. DOI: https://doi.org/10.1109/ISSCC.2013.6487665

[51] Xuan Li, Nathan Youngblood, Zengguang Cheng, Santiago Garcia-Cuevas Carrillo, Emanuele Gemo, Wolfram H. P. Pernice, C. David Wright, and Harish Bhaskaran. 2020. Experimental investigation of silicon and silicon nitride platforms for phase-change photonic in-memory computing. *Optica* 7, 3 (2020), 218–225.

[52] Xuan Li, Nathan Youngblood, Carlos Ríos, Zengguang Cheng, C. David Wright, Wolfram H. P. Pernice, and Harish Bhaskaran. 2019. Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell. *Optica* 6, 1 (2019), 1–6.

[53] Yucheng Low, Danny Bickson, Joseph Gonzalez, Carlos Guestrin, Aapo Kyrola, and Joseph M. Hellerstein. 2012. Distributed GraphLab: A framework for machine learning and data mining in the cloud. *Proc. VLDB Endow.* 5, 8 (Apr. 2012), 716–727. DOI: https://doi.org/10.14778/2212351.2212354

[54] Lian-Wee Luo, Noam Ophir, Christine P. Chen, Lucas H. Gabrielli, Carl B. Poitras, Keren Bergmen, and Michal Lipson. 2014. WDM-compatible mode-division multiplexing on a silicon chip. *Nat. Commun.* 5, 1 (2014), 1–7.

[55] Ho-Ki Lyeo, David G. Cahill, Bong-Sub Lee, John R. Abelson, Min-Ho Kwon, Ki-Bum Kim, Stephen G. Bishop, and Byung-ki Cheong. 2006. Thermal conductivity of phase-change material $Ge_2Sb_2Te_5$. *Appl. Phys. Lett.* 89, 15 (2006), 151904.

[56] Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. 2010. Pregel: A system for large-scale graph processing. In *Proceedings of the International Conference on Management of Data*. 135–146.

[57] Ann-Katrin U. Michel, Peter Zalden, Dmitry N. Chigrin, Matthias Wuttig, Aaron M. Lindenberg, and Thomas Taubner. 2014. Reversible optical switching of infrared antenna resonances with ultrathin phase-change layers using femtosecond laser pulses. *ACS Photon.* 1, 9 (2014), 833–839.

[58] Onur Mutlu. 2013. Memory scaling: A systems architecture perspective. In *Proceedings of the International Memory Workshop*. 21–25.

[59] Aditya Narayan, Yvain Thonnart, Pascal Vivet, Ajay Joshi, and Ayse K. Coskun. 2020. System-level evaluation of chip-scale silicon photonic networks for emerging data-intensive applications. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition*. 1444–1449.

[60] Aditya Narayan, Yvain Thonnart, Pascal Vivet, César Fuguet Tortolero, and Ayse K. Coskun. 2019. WAVES: Wavelength selection for power-efficient 2.5D-integrated photonic NoCs. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition*. 516–521.

[61] T. Nirschl, J. B. Philipp, T. D. Happ, G. W. Burr, B. Rajendran, M.-H. Lee, A. Schrott, M. Yang, M. Breitwisch, C.-F. Chen, E. Joseph, M. Lamorey, R. Cheek, S.-H. Chen, S. Zaidi, S. Raoux, Y. C. Chen, Y. Zhu, R. Bergmann, H.-L. Lung, and C. Lam. 2007. Write strategies for 2 and 4-bit multi-level phase-change memory. In *Proceedings of the International Electron Devices Meeting*. 461–464.

[62] Hiroki Noguchi, Kazutaka Ikegami, Satoshi Takaya, Eishi Arima, Keiichi Kushida, Atsushi Kawasumi, Hiroyuki Hara, Keiko Abe, Naoharu Shimomura, Junichi Ito, Shinobu Fujita, Takashi Nakada, and Hiroshi Nakamura. 2016. 4Mb STT-MRAM-based cache with memory-access-aware power optimization and write-verify-write/read-modify-write scheme. In *Proceedings of the International Solid-State Circuits Conference*. 132–133. DOI: https://doi.org/10.1109/ISSCC.2016.7417942

[63] Masaya Notomi, Kengo Nozaki, Akihiko Shinya, Shinji Matsuo, and Eiichi Kuramochi. 2014. Toward fJ/bit optical communication in a chip. *Optics Commun.* 314 (2014), 3–17.

[64] Stanford R. Ovshinsky. 1968. Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.* 21, 20 (1968), 1450.

[65] Kishore Padmaraju and Keren Bergman. 2014. Resolving the thermal challenges for silicon microring resonator devices. *Nanophotonics* 3, 4–5 (2014), 269–281.

[66] Gaetano Palumbo and Domenico Pappalardo. 2010. Charge pump circuits: An overview on design strategies and topologies. *IEEE Circ. Syst. Mag.* 10, 1 (2010), 31–45.

[67] Awanish Pandey and Shankar Kumar Selvaraja. 2018. Four channel 48Gbps multicasting in a coupled Si ring resonator with tunable channel spacing. In *Proceedings of the Conference on Lasers and Electro-Optics/Pacific Rim*. Optical Society of America, W2D–4.

[68] Matthew Poremba, Tao Zhang, and Yuan Xie. 2015. NVMain 2.0: A user-friendly memory simulator to model (non-) volatile memory systems. *IEEE Comput. Archit. Lett.* 14, 2 (2015), 140–143.

[69] Moinuddin K. Qureshi, Michele M. Franceschini, Ashish Jagmohan, and Luis A. Lastras. 2012. PreSET: Improving performance of phase change memories by exploiting asymmetry in write times. In *Proceedings of the International Symposium on Computer Architecture*.

[70] Moinuddin K. Qureshi, John Karidis, Michele Franceschini, Vijayalakshmi Srinivasan, Luis Lastras, and Bulent Abali. 2009. Enhancing lifetime and security of PCM-based main memory with start-gap wear leveling. In *Proceedings of the International Symposium on Microarchitecture*. 14–23.

[71] Moinuddin K. Qureshi, Vijayalakshmi Srinivasan, and Jude A. Rivers. 2009. Scalable high performance main memory system using phase-change memory technology. In *Proceedings of the International Symposium on Computer Architecture*. 24–33.

[72] Luiz E. Ramos, Eugene Gorbatov, and Ricardo Bianchini. 2011. Page placement in hybrid memory systems. In *Proceedings of the International Conference on Supercomputing*. 85–95.

[73] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y. Chen, R. M. Shelby, M. Salinga, D. Krebs, S. Chen, H. Lung, and C. H. Lam. 2008. Phase-change random access memory: A scalable technology. *IBM J. Res. Devel.* 52, 4.5 (2008), 465–479. DOI: https://doi.org/10.1147/rd.524.0465

[74] Angad S. Rekhi, Brian Zimmer, Nikola Nedovic, Ningxi Liu, Rangharajan Venkatesan, Miaorong Wang, Brucek Khailany, William J. Dally, and C. Thomas Gray. 2019. Analog/mixed-signal hardware error modeling for deep learning inference. In *Proceedings of Design Automation Conference*. 1–6.

[75] Carlos Rios, Peiman Hosseini, C. David Wright, Harish Bhaskaran, and Wolfram H. P. Pernice. 2014. On-chip photonic memory elements employing phase-change materials. *Adv. Mater.* 26, 9 (2014), 1372–1377.

[76] Carlos Ríos, Nathan Youngblood, Zengguang Cheng, Manuel Le Gallo, Wolfram H. P. Pernice, C. David Wright, Abu Sebastian, and Harish Bhaskaran. 2019. In-memory computing on a photonic platform. *Sci. Adv.* 5, 2 (2019), eaau5759.

[77] Paul Rosenfeld, Elliott Cooper-Balis, and Bruce Jacob. 2011. DRAMSim2: A cycle accurate memory system simulator. *IEEE Comput. Archit. Lett.* 10, 1 (2011), 16–19.

[78] Carlos Ríos, Matthias Stegmaier, Peiman Hosseini, Di Wang, Torsten Scherer, C. David Wright, Harish Bhaskaran, and Wolfram H. P. Pernice. 2015. Integrated all-photonic non-volatile multi-level memory. *Nat. Photon.* 9, 11 (2015), 725.

[79] Semih Salihoglu and Jennifer Widom. 2013. GPS: A graph processing system. In *Proceedings of the International Conference on Scientific and Statistical Database Management*. 1–12.

[80] Nak Hee Seong, Dong Hyuk Woo, and Hsien-Hsin S. Lee. 2010. Security refresh: Prevent malicious wear-out and increase durability for phase-change memory with dynamically randomized address mapping. In *Proceedings of the International Symposium on Computer Architecture*. 383–394.

[81] Kuanping Shang, Shibnath Pathak, Binbin Guan, Guangyao Liu, and S. J. B. Yoo. 2015. Low-loss compact multilayer silicon nitride platform for 3D photonic integrated circuits. *Optics Expr.* 23, 16 (2015), 21334–21342.

[82] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. 2016. Edge computing: Vision and challenges. *IEEE Internet Things J.* 3, 5 (2016), 637–646.

[83] Shihao Song, Anup Das, Onur Mutlu, and Nagarajan Kandasamy. 2019. Enabling and exploiting partition-level parallelism (PALP) in phase change memories. *ACM Trans. Embed. Comput. Syst.* 18, 5s (2019), 1–25.

[84] Chen Sun, Mark T. Wade, Yunsup Lee, Jason S. Orcutt, Luca Alloatti, Michael S. Georgas, Andrew S. Waterman, Jeffrey M. Shainline, Rimas R. Avizienis, Sen Lin, Benjamin R. Moss, Rajesh Kumar, Fabio Pavanello, Amir H. Atabaki, Henry M. Cook, Albert J. Ou, Jonathan C. Leu, Yu-Hsin Chen, Krste Asanović, Rajeev J. Ram, Miloš A. Popović, and Vladimir M. Stojanović. 2015. Single-chip microprocessor that communicates directly using light. *Nature* 528, 7583 (2015), 534–538.

[85] Ishan G. Thakkar and Sudeep Pasricha. 2017. DyPhase: A dynamic phase change memory architecture with symmetric write latency and restorable endurance. *IEEE Trans. Comput.-Aid. Des. Integ. Circ. Syst.* 37, 9 (2017), 1760–1773.

[86] Yvain Thonnart, Stéphane Bernabé, Jean Charbonnier, Christian Bernard, David Coriat, César Fuguet, Pierre Tissier, Benoît Charbonnier, Stéphane Malhouitre, Damien Saint-Patrice, Myriam Assous, Aditya Narayan, Ayse Coskun, D. Dutoit, and Pascal Vivet. 2020. POPSTAR: A robust modular optical NoC architecture for chiplet-based 3D integrated systems. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition*. 1456–1461.

[87] M. Wade, E. Anderson, S. Ardalan, P. Bhargava, S. Buchbinder, M. L. Davenport, J. Fini, H. Lu, C. Li, R. Meade, C. Ramamurthy, M. Rust, F. Sedgwick, V. Stojanovic, D. Van Orden, C. Zhang, C. Sun, S. Y. Shumarayev, C. O'Keeffe, T. T. Hoang, D. Kehlet, R. V. Mahajan, M. T. Guzy, A. Chan, and T. Tran. 2020. TeraPHY: A chiplet technology for low-power, high-bandwidth in-package optical I/O. *IEEE Micro* 40, 2 (2020), 63–71. DOI: https://doi.org/10.1109/MM.2020.2976067

[88] J. Wang, L. Wang, and J. Liu. 2020. Overview of phase-change materials based photonic devices. *IEEE Access* 8 (2020), 121211–121245. DOI: https://doi.org/10.1109/ACCESS.2020.3006899

[89] Shipeng Wang, Xianglian Feng, Shiming Gao, Yaocheng Shi, Tingge Dai, Hui Yu, Hon-Ki Tsang, and Daoxin Dai. 2017. On-chip reconfigurable optical add-drop multiplexer for hybrid wavelength/mode-division-multiplexing systems. *Optics Lett.* 42, 14 (2017), 2802–2805.

[90] Oi-Ying Wong, Hei Wong, Wing-Shan Tam, and C. W. Kok. 2012. A comparative study of charge pumping circuits for flash memory applications. *Microelectron. Reliab.* 52, 4 (2012), 670–687.

[91] Xinru Wu, Chaoran Huang, Ke Xu, Chester Shu, and Hon Ki Tsang. 2017. Mode-division multiplexing for silicon photonic network-on-chip. *J. Lightw. Technol.* 35, 15 (2017), 3223–3228.

[92] Matthias Wuttig, Harish Bhaskaran, and Thomas Taubner. 2017. Phase-change materials for non-volatile photonic applications. *Nat. Photon.* 11, 8 (2017), 465–476.
[93] Matthias Wuttig and Noboru Yamada. 2007. Phase-change materials for rewriteable data storage. *Nat. Mater.* 6, 11 (2007), 824–832.
[94] Fei Xia, Dejun Jiang, Jin Xiong, Mingyu Chen, Lixin Zhang, and Ninghui Sun. 2014. DWC: Dynamic write consolidation for phase change memory systems. In *Proceedings of the International Conference on Supercomputing*. 211–220.
[95] Cong Xu, Dimin Niu, Naveen Muralimanohar, Rajeev Balasubramonian, Tao Zhang, Shimeng Yu, and Yuan Xie. 2015. Overcoming the challenges of crossbar resistive memory architectures. In *Proceedings of the International Symposium on High Performance Computer Architecture*. 476–488.
[96] Yue-De Yang, Yu Li, Yong-Zhen Huang, and Andrew W. Poon. 2014. Silicon nitride three-mode division multiplexing and wavelength-division multiplexing using asymmetrical directional couplers and microring resonators. *Optics Exp.* 22, 18 (2014), 22172–22183.
[97] Hanbin Yoon, Justin Meza, Naveen Muralimanohar, Norman P. Jouppi, and Onur Mutlu. 2014. Efficient data mapping and buffering techniques for multilevel cell phase-change memories. *ACM Trans. Archit. Code Optim.* 11, 4 (2014), 1–25.
[98] Vinson Young, Prashant J. Nair, and Moinuddin K. Qureshi. 2015. DEUCE: Write-efficient encryption for non-volatile memories. In *Proceedings of the Architectural Support for Programming Languages and Operating Systems*. 33–44.
[99] Nathan Youngblood, Carlos Ríos, Emanuele Gemo, Johannes Feldmann, Zengguang Cheng, Anna Baldycheva, Wolfram H. P. Pernice, C. David Wright, and Harish Bhaskaran. 2019. Tunable volatility of $Ge_2Sb_2Te_5$ in integrated photonics. *Adv. Funct. Mater.* 29, 11 (2019), 1807571.
[100] Hongliang Yu and Yuyang Du. 2012. Increasing endurance and security of phase-change memory with multi-way wear-leveling. *IEEE Trans. Comput.* 63, 5 (2012), 1157–1168.
[101] Hanyu Zhang, Linjie Zhou, Jian Xu, Liangjun Lu, Jianping Chen, and B. M. A. Rahman. 2018. All-optical non-volatile tuning of an AMZI-coupled ring resonator with GST phase-change material. *Optics Lett.* 43, 22 (2018), 5539–5542.
[102] Mengying Zhao, Lei Jiang, Youtao Zhang, and Chun Jason Xue. 2014. SLC-enabled wear leveling for MLC PCM considering process variation. In *Proceedings of the Design Automation Conference*. 1–6.

Received 11 September 2021; revised 14 April 2022; accepted 21 April 2022