

Ultracompact photonic integrated content addressable memory using phase change materials

Md. Ajwaad Zaman Quashef¹ · Md. Kawsar Alam¹

Received: 15 August 2021 / Accepted: 5 February 2022 / Published online: 26 February 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The inherently parallel and ultrafast nature of photonic circuits is naturally well suited for implementations of content addressable memory (CAM) circuits which perform highly parallel, in-memory computations leading to extremely high throughput compare operations at low latency in applications such as network routing, associative learning and CPU caching. Towards improving the latency by orders of magnitude over conventional electronic implementations, we propose a scalable photonic architecture for a CAM based on silicon microring resonators embedded with patches of phase change material $Ge_2Sb_2Te_5$ (GST-225). The CAM cells behave as frequency selective 1×2 switches and emulate the digital XOR operation in a wavelength division multiplexing (WDM) read scheme at speeds limited only by the bit modulation system and photodetector response. Through device and circuit level simulations we show that the proposed reconfigurable nonvolatile architecture can reliably perform tens of Gbps speed read operations in the presence of channel noise, crosstalk and source-receiver nonidealities. The individual CAM cells have a compact footprint of less than 10 μ m² and require no continuous power input to maintain the switch states, which is naturally advantageous given the scalability requirement and long intervals between entry updates in modern CAM systems.

Keywords Content addressable memory · Nonvolatile memory · Silicon photonics · Optical lookup · Microring resonators · Phase change materials

1 Introduction

The seamless integration of chalcogenide phase change materials (PCM) with silicon photonic components is opening new avenues for light-speed computation and novel datacenter applications with all the advantages of nonvolatile memory on a CMOS technology compatible optical platform. Inventive ways of leveraging the exciting properties of PCM are being explored on the optical platform which is already superior over the electrical domain owing to the multiplexing capability, low residual crosstalk and dispersion,

Md. Kawsar Alam kawsaralam@eee.buet.ac.bd; kawsar.alam@alumni.ubc.ca

¹ Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh

not to mention the ultrawide bandwidth. Due to the dramatic refractive index contrast of $Ge_2Sb_2Te_5$ (GST-225) over the vis and near-IR spectral range, long retention time (for at least 10 years) (Wong et al. 2010; Wuttig and Yamada 2007), and its nanosecond scale switching rate with high reproducibility and durability obtained over 10^{12} switching cycles (Burr et al. 2010), this alloy has been used as a suitable candidate for an integrated PCM photonic non-volatile memory. Recently, GST has been integrated directly on silicon and silicon nitride waveguides to implement an all-photonic 5-bit memory cell (Li et al. 2019), a 512-bit photonic random-access memory (Feldmann et al. 2020), a photonic tensor core for convolution processing (Feldmann et al. 2021), an in-memory computing platform (Ríos et al. 2019), mimic synapses and neurons in spiking neural networks (Chakraborty et al. 2018; Cheng et al. 2017; Feldmann et al. 2019), and emulate an abacus calculator (Feldmann et al. 2017).

The fully CMOS-process compatible silicon photonics platform can enable a potential paradigm shift in design of in-memory computing systems due to the inherently parallel nature of light-based calculations. Content addressable memories (CAMs) are a special form of in-memory computing circuit widely used in high-speed searching applications as they can compare an input query against a list of stored data in parallel, returning the address of matching data in the case of a hit. CAMs can also play useful roles in specialized computations, such as hardware compression/ decompression and implementing reverse tagged directories for large-scale multiprocessor coherence (Beamer and Akgul 2009). Though they have also found new utility in the emerging in-memory computing paradigm and deep learning applications (Karam et al. 2015; Ni et al. 2019), commercial success of CAM implementations has mostly been seen in datacenter technologies for network routing and real-time network traffic monitoring (Meiners et al. 2010; Pagiamtzis and Sheikholeslami 2006; Xu et al. 2016). One such technology is the implementation of Address Look-Up (AL) which continues to be traditionally bottlenecked by the speed limitation of SRAM technology operating at a few GHz (Agarwal et al. 2011; Pagiamtzis and Sheikholeslami 2006) as well as by impediments to scaling due to increasing leakage currents. Designs of emerging alternative electronic memory technologies based on phase change material (GST) (Junsangsri et al. 2017), ferroelectric field-effect-transistors (FeFET) (George et al. 2020; Yin et al. 2020; Yin et al. 2019a, b), nanomagnetic logic (Matsunaga et al. 2012), memristors (Grossi et al. 2018; Li et al. 2020; Yang et al. 2019), hold the promise for a better trade-off between low-power consumption and chip area occupancy but without raising any expectations for increased data rates. Crossing into the optical domain in search of higher rates, a photonic integrated CAM capable of 10 Gbps operation was recently introduced (Maniotis and Pleros 2017; Mourgias-Alexandris et al. 2018) whose design produces an optical equivalent of an SRAM-based CAM cell with flip-flops (FF). At each cell, the FF, responsible for storing the bit, requires two coupled semiconductor optical amplifier-Mach Zehnder interferometer (SOA-MZI) switches in a master–slave configuration. In addition, the output is connected to the lower branch of an MZI which constitutes an XOR gate, comprising two SOA. However, this approach requires significant hardware overhead (cell area ~ $6 \times 2 \text{ mm}^2$) (Pitris et al. 2015), therefore achieving low bit density for memory applications. A more recent proof-of-concept for an optoelectronic CAM (Alkabani et al. 2020) achieved much higher compactness and non-requirement of internal light sources as compared to the SRAM-emulating design, by utilizing foundry available single bus microring modulators (MRM) based on pn junction and carrier depletion as the basic elements in a WDM scheme. They proposed the use of electrical writes to store values in an optical search engine by voltage biasing each MRM j to resonance wavelength λ_{0i} (store bit 0) or λ_{1i} (store bit 1) in a row of n MRMs. The micro-ring resonators are coupled in series to a waveguide constituting the CAM's entry (row), where each entry is terminated by a photodetector with output connected to a priority encoder. However, they found that it is required to compromise the switch modulation range (~3 dB) for storing bits of the CAM words to achieve at least 25 GHz modulation speed. Although the bias power for MRM modulation is notionally negligible in high-rate bit-by-bit switching, it adds to significant energy consumption overhead in routing schemes where switching occurs at much lower frequencies than the actual data rate. Furthermore towards denser photonic integration, scaling down the radii of Si resonators from tens of micrometers to wavelength of operation-comparable sizes has been experimentally demonstrated (Liu et al. 2019; Xu et al. 2008), but at the critical expense of ring quality factors (O-factor) due to higher bending losses. Operating these MRRs as modulators demands great increases in the required switching power or else deteriorates the modulation depth (Bogaerts et al. 2012). Therefore, a compact photonic switch with a (1) self-holding capability, (2) moderate extinction ratio, and (3) operating on the basis of optical attenuation and not in spite of it- would be naturally suited for lowfrequency switching applications like content addressing in routing AL; in addition to possessing the vital features of (4) compactness and CMOS-compatibility for easy on-chip integration and scaling alongside fully electronic systems.

Here, we combine the concept of reconfigurable optical phase change memory with silicon microring resonators (MRRs) to propose a photonic CAM architecture benefitting from the above-mentioned features of nonvolatile switches. By setting up the devices to behave as self-holding 1×2 switches, we have emulated the digital XOR (mismatch) operation in a novel all-optical manner and combined the cells in a wavelength division multiplexing (WDM) read scheme to achieve parallelism. The system's read-response is limited only by on-chip modulator and photodetector bandwidths since the time-of-flight of a photon through the low-Q MRRs is negligible (~ few ps), signifying orders of improvement over conventional electronic CAM counterparts. Keeping in mind the low energy and area footprints that will be required to enable denser scaling of photonic integrated circuits (PIC) into the coming decade, we evaluate the performance of the system using ultracompact MRRs (radius $\approx 1.5 \ \mu m$). We also briefly explore reconfigurability of the designed switches in light of recent works on PCM based photonic systems and adopt a low-footprint electro-optic write mechanism that consumes < 200 pJ per element for updating the stored words in the CAM. Finally, through system level simulations, we evaluate the uniformity and reliability of the read operations with different bit permutations in an 8-bit search in the presence of interchannel crosstalk, channel noise, and source-receiver nonidealities in the telecommunication S and C bands (1525–1575 nm).

2 Optical XOR operation

Our simple architecture for the proposed CAM array is based on an SOI add-drop ring resonator/filter with a thin patch of PCM-GST integrated onto the bent waveguide (Fig. 1a). Similar resonator devices have been very recently fabricated on silicon and silicon nitride platforms for both device (Rudé et al. 2013; Stegmaier et al. 2017; Wu et al. 2019; Zhang et al. 2019, 2018; Zheng et al. 2018) and circuit-level evaluation as neuromorphic/in-memory computing systems (Feldmann et al. 2019) and non-volatile memory circuits (Feldmann et al. 2020). However, here we leverage the differential transmission characteristics of the THROUGH port with respect to the IN and ADD ports of the filter at resonance to emulate an XOR-like operation which facilitates the bit search scheme.



Fig. 1 a A perspective view of an add-drop microring resonator with a small patch of GST on top showing its ports, coupling coefficients and materials and b Field profile of fundamental TE mode at 1550 nm for (i) amorphous and (ii) crystalline states of GST in a waveguide bend of $1.55 \ \mu m$ radius. Inset shows cross-section of GST integrated on top of a 0.45×0.25 Si strip waveguide on SiO₂ substrate

Taking into account the refractive indices of GST, Si, and SiO₂, we simulated mode profiles via the Finite-Difference Eigenmode solver in Lumerical MODE (which calculates the spatial profile and frequency dependence of modes by solving Maxwell's equations on a cross-sectional mesh) for the bent waveguide with a 10-nm-thick GST layer topped by a protective indium tin oxide (ITO) layer. The bent waveguide aspect ratio (t/w) of 250/450 nm was chosen to ensure single fundamental TE mode transmission in the 1500–1600 nm (telecom S and C band) range of wavelengths. Figure 1b shows the field profile (|E|) at the cross-section for this mode at 1550 nm, visually demonstrating the difference in optical absorption (represented by $Imag(n_{eff})$) in the GST's (i) amorphous and (ii) crystalline state. As the light couples evanescently to the PCM, the interaction with the material on top of the waveguide decreases exponentially with the distance so that the influence of the protective layer on the device performance can be neglected compared to the PCM.

For our design, we set the minimum radius of the MRRs to be $R = 1.515 \mu m$ to achieve denser integration with the added advantage of miniaturization being that the free spectral range (FSR) of the devices becomes larger, supporting denser WDM channel counts in spite of the increased full width half maximum (FWHM) in absorption-modulated devices. Rings of similar size have been demonstrated previously (Prabhu et al. 2010; Xu et al. 2008) with a more recent experiment (Liu et al. 2019) achieving FSR as large as 93 nm using $R = 0.8 \mu m$, with a moderate FWHM of ~0.8 nm. The modifications involved in the fabrication of the ultracompact MRRs in the mentioned works included reducing the width of the bus waveguide, which increases the spatial period of

the propagating mode due to the lower effective refractive index. This results in a better phase match with the mode in the tightly curved ring waveguide and therefore better coupling. The use of strip waveguides and air cladding instead of ridge waveguides and oxide cladding respectively minimizes the bending loss due to a larger refractive index contrast better confining the mode (Bogaerts et al. 2012).

The microring is designed to be near-critically coupled (parameters in Table 1) when the GST is in the amorphous phase with low absorption such that the resonance of the microring assumes an extinction ratio as high as possible. We used a finite-difference time-domain method based solver (FDTD Solutions from Lumerical Inc.) to solve the Maxwell equations on a discrete spatiotemporal grid and obtain the optical response (Fig. 2) upon switching into the crystalline state: the round-trip loss inside the ring resonator is strongly increased due to the enhanced material absorption in the GST. Because the round-trip losses have increased, the existing coupling gap leads to an undercoupled ring resonator, thus increasing the transmission from IN to THROUGH ports and decreasing the transmission from the ADD to THROUGH ports at resonance.

The following expression for the THROUGH port transmission of an MRR with respect to the IN port at resonance

$$T_{through} = \frac{I_{THROUGH}}{I_{IN}} = \frac{r_2^2 a^2 - 2r_1 r_2 a + r_1^2}{1 - 2r_1 r_2 a + (r_1 r_2 a)^2}$$
(1)

exhibits this strong dependence on the single pass (round trip) amplitude transmission $a \approx exp\left(-\frac{2\pi}{\lambda_{res}}\kappa_{eff,GST}L_{GST}\right)$ which decreases strongly with increasing $\kappa_{eff,GST}$, which is the imaginary part of the effective refractive index of the waveguide material with GST. r_1 , r_2 are the self-coupling coefficients, λ_{res} is the resonance wavelength and L_{GST} is the length of the GST element. The same can be seen for the THROUGH port transmission with respect to the ADD port transmission which is found by interchanging the terms r_1 and r_2 in the regular drop port expression for a double bus MRR:

$$T_{drop} = \frac{I_{THROUGH}}{I_{ADD}} = \frac{k_1^2 k_2^2 a}{1 - 2r_1 r_2 a + (r_1 r_2 a)^2}$$
(2)

where k₁, k₂ are the cross-coupling coefficients. The device also experiences a small resonance shift upon switching (Stegmaier et al. 2017) quantified by $\frac{\Delta \lambda_r}{\lambda_{r,0}} \approx \frac{\Delta n_{eff,GST}}{n_{geff}} \cdot \frac{L_{GST}}{2\pi R}$ which amounts to approximately 0.6 nm for our case. Here, $\lambda_{r,0}$ denotes the resonance wavelength

Table 1 MRR and PCM parameters for FDTD simulation	Dimension	Value
	Radius	1.515-1.59 μm
	Bus WG Cross Section	$0.35 \times 0.25 \ \mu m$
	Ring WG Cross Section	$0.45 \times 0.25 \ \mu m$
	Input Side Coupling Gap	100 nm
	Drop Side Coupling Gap	125 nm
	GST Length	270–400 nm
	GST Width	0.44 µm
	GST & ITO Thickness	10 nm





Fig.2 a In to Through port transmission increases (from 0.027 to 0.602) upon crystallization of GST. **b** Add to Through port transmission increases (0.038 to 0.578) upon amorphization of GST. Readouts are performed at a fixed wavelength, ideally at the resonance wavelength of the MRR

before the switching, $\Delta n_{eff,GST}$ the respective change upon switching, $n_{g,eff}$ the effective group index of the bare waveguide, and *R* the radius of the ring. Since the read is performed at a fixed wavelength, this red shift upon crystallization actually adds to the switching contrast for the rings at the expense of a slightly higher interference with the response of the adjacent larger ring.

Now it is straightforward to demonstrate how the described structure performs the all-optical XOR operation for detecting single bit mismatches between a search query and a stored word/entry in the CAM as shown in Fig. 3. The stored bit B is encoded as the crystal state of the integrated PCM which can be either fully crystalline (B=0) or amorphous (B=1), while the search bit A is represented as the presence of the corresponding resonance frequency λ_i at the IN port (A=1) or ADD port (A=0). Similar to CAMs in the electronic domain



Fig. 3 Single bit mismatches are detected at the Through port as an XOR-like operation between a stored nonvolatile bit encoded in the crystal state of a PCM, and a search bit represented by the presence of an optical pulse at the resonant frequency λ_i of an MRR

(Pagiamtzis and Sheikholeslami 2006; Yin et al. 2019b), the operation requires complementary read signals which can be generated by simply using the opposite modulation scheme of the one used to generate the uncomplemented input pulses. If the stored bit B is 1 (amorphous GST), a HIGH search bit (A=1) means that the pulse from the input port is dropped and does not make it through to the output since the MRR is at a critically coupled state. However, a LOW search bit (A=1) would result in the ADD signal being dropped to the THROUGH port after being attenuated to around half its amplitude (T=0.578 according to Fig. 2b) and evaluated at the detector as a mismatch with that particular stored bit B. Vice-versa for a stored bit B=0 (crystalline GST), a HIGH search bit would result in partial transmission from the IN to the THROUGH port (T=0.602) and evaluated at the detector as a mismatch while a LOW search bit would result in the ADD port signal being sent off-chip via the DROP port, owing to the undercoupled condition of the MRR. We can compensate for the lower transmission from the ADD port compared to that from the IN port by adjusting the corresponding laser power or by simply placing a broadband attenuator at the mouth of the IN bus, in order to ensure uniform response among mismatch permutations. The minimal transmission in the matched bit cases (0.027 and 0.038 from Fig. 2) may be further reduced by precisely engineering the coupling gaps to achieve better critical coupling (Feldmann et al. 2020).

In contrast to a fully optical read scheme, the writing scheme of our photonic CAM is chosen to work via electrical heaters which instantiate the phase change of the GST to change the stored bit B, the details of which have been presented in a later section.

3 CAM array design

We propose the architecture for the CAM array in which the described unit cells are parallelly cascaded after appropriate optimization in the frequency domain. During the content addressing operation, a fast parallel XOR-based comparison operation is performed across the stored data of the table which for example in the case of AL, is to look for a fast match with the packet destination address of the incoming packet of the router. Here we leverage the WDM technique to compute the degree of mismatch (if any) between the bits of an input query and the stored words (entries) in the CAM array. To represent multiple wavelengths, we use multiple ring resonators of increasing ring radii (1.515 to 1.59 μ m) to represent different bits in a row as shown in Fig. 4. A search query and its complement encoded as WDM signals enters the straight waveguide through the IN and ADD ports respectively, and the coupling condition determined by the PCM state on each ring resonator results in a bitwise XOR operation being performed at each read wavelength within the FSR, as described in the previous section. Thus, at the output port we obtain a multiwavelength signal which in the case of a full match will ideally consist of no light, but due to non-zero transmission values (< 5%) in the matched bit states, possesses weak (few μ W) components from all the channels. Wavelengths corresponding to mismatched bits will be present at the Through port and the power detected at the receiver (corresponding to each row/word) will result in evaluation of the mismatch as per the decision levels.

Although we have designed a 'binary' CAM which does not offer the facility of storing don't care (X) states in the table words, the system can be used to eliminate the effect of a particular bit on the word match by sending light with the corresponding frequency in neither of the input buses. In this way a don't care (X) may be searched at multiple specified positions in the query.



Fig. 4 Full match case output, with a don't care bit (x) included in the search query. In the shown case, the first two bits (10) of the 3-bit search word matches with the corresponding bits in the stored word (10) and the 3rd bit is ignored by eliminating light of the corresponding frequency in both the IN and ADD buses. Ideally no light at any wavelength makes it to the THROUGH port/bus (XOR output=000)

Full FDTD simulations were performed for only the smallest and largest rings in a row of n (=CAM entry length) MRRs varying the state of an embedded GST layer between fully amorphous and crystalline. Owing to the large computational expense of resonant devices in FDTD methods, the results for the port transmissions were used to calibrate an adaptive analytical model which takes as input the material and bending losses, coupling coefficients (r, k) and GST lengths as parameters. The material loss due to PCM absorption, quantified by effective indices, was obtained from mode simulations (shown earlier in Fig. 2b) for our chosen geometry and PCM material parameters (Pernice and Bhaskaran 2012). Frequency sweeps were conducted in Lumerical MODE tracking the fundamental TE mode for each GST state. Bending loss and coupling coefficients vs. frequency were calculated in separate FDTD simulations for the designed bend radii and coupling gap sizes. After obtaining all the required parameters, we calculated the through/drop frequency response (from Eqs. 1 and 2) for a row of PCM-MRRs in parallel using the model Eqs. (1) and (2). In this way the full FDTD results were emulated by breaking up the simulation into segments thus reducing the lengthy computation time required for a row of resonators each with two possible PCM states. Comparing the adapted model with full FDTD results for the smallest and largest MRR (Fig. 5a), we find reasonable match for the positions and heights of the transmission peaks/troughs in light of our application of the photonic devices as two-level switches. We noted that our fitted model slightly overestimates the FWHM and underestimates the drop port transmission peaks for the smaller rings, translating to higher crosstalk and thus a miniscule performance drop in circuit operation compared to what would be obtained if we used the results from full FDTD simulations. Figure 5b illustrates the superimposed normalized transmissions for each resonator in a row of 8 MRRs with increasing radii $(1.515-1.59 \ \mu\text{m})$. The PCM patches were designed to be longer for the larger ring (400 nm vs. 270 nm for the smaller ring) in order to compensate for the variation of coupling efficiency with wavelength.

The maximum word size supported by our CAM array design will depend on the number of channels that can be fit into a single FSR while maintaining an acceptable level of crosstalk and is quantified by the *finesse* of the resonator, defined as the ratio of the FSR and linewidth (aka FWHM or 3 dB bandwidth). The overall effect of interchannel crosstalk in our system is to limit the transmission to the output in the mismatched state hence reducing the detected power contrast between match and (single bit) mismatched cases. However, the crosstalk penalty is dependent primarily on the states of the individual PCMs and can be easily analyzed and accounted for in the design process in light of the proposed



Fig.5 a Calibration of an analytical model by cross-matching the calculated frequency responses with those obtained from full FDTD simulations for (i) 1.515 μ m MRR and (ii) 1.59 μ m MRR integrated with PCM-GST (amorphous and crystalline states). **b** Superimposed frequency responses for two degrees of crystallization: GSTa (0%) and GSTc (100%) of 8 PCM integrated MRRs with increasing radii designed to resonate at equally spaced wavelengths (1532 – 1572 nm) within a single FSR (~56 nm)

application. Since the FWHM is larger for lossy rings, MRRs storing the 0-bit value (corresponding to fully crystalline state of the GST) will have a larger overlap on adjacent MRR's frequency responses. We simulate the frequency response of a single ring array (16-bit word) in the industry-grade PIC solver 'Lumerical INTERCONNECT' to visually illustrate the overall through port transmission when a particular word row stores all 1 s (all amorphous) and all 0 s (all crystalline) in Fig. 6. For this purpose, the device parameters (loss, coupling, effective index) extracted from the calibrated analytical model (including the frequency dependency) have been applied to INTERCONNECT's built-in double-bus





MRR element model. The simulated resonance peaks are spaced 2.779 nm apart within an FSR of ~56 nm for the largest ring (having a radius of 1.59 μ m). Selecting a GST length of 350 nm, the FWHM of the largest ring in the (more lossy) crystalline state is 3.43 nmgiving a finesse value of ~16 for our design. Lengths of PCM patches are designed to be longer for larger rings in order to compensate for the larger coupling coefficients at longer wavelengths and keep the extinction ratios across a row as uniform as possible. The resonances of the largest rings in an array are for this reason wider than those of the smallest rings resulting in a slightly lower extinction ratio between the GSTa and GSTc states. We have in a later section shown the photodetector output for all single bit mismatch permutations in a CAM with 8-bit entries to quantitatively evaluate the effect of this deterministic crosstalk on the uniformity and integrity of the search operation.

4 Optical search engine layout and cascadability

To query the match or presence of a search query in a generic CAM table, it is converted to a binary form and bit-encoded as light pulses using a suitable modulation scheme before being input to the search mechanism or engine. The search engine outputs a match line (ML) per table entry that signifies a match (ML=0) or mismatch (ML=1). The match lines are input to a priority encoder that generates the address (or first address in case of multiple matches) at which the match occurred, in which case a Hit output is also asserted. Our designed CAM cells are at the heart of this address-returning system that is capable of simultaneously storing data in a nonvolatile manner and performing the mismatch operation via all-optical XORs. Each n-bit entry in the CAM table is terminated by a photodetector whose photocurrent can be read out to evaluate the match/mismatch of the particular entry with the query. The block diagram of the overall system for returning the address of a matching entry is shown in Fig. 7, including a scheme for cascading word-partitions. By partitioning a longer m*n bit word into m n-bit words and feeding each word-part into an identical CAM array, the search operation can support m-times longer words without requiring increased resolution capability at the detector (otherwise deteriorated by increased crosstalk in the alternative case of using a denser WDM scheme). Electronically joining the match lines corresponding to m different partitions in an m*n bit word comes at the expense of read latency increased by an amount equal to the capacitive delay of an m-input OR gate.



Fig.7 Block diagram of the cascaded system to return the address of the searched word in the case of a match in the CAM table. By cascading smaller arrays with digital OR gates, we can increase the word length supported by the search

5 Reconfiguration of CAM entries

Write operations in CAMs for routing applications are performed during network-topology changes, when the stored routing paths need to be dynamically reconfigured and new data have to be written in the memory. In most CAM applications, including those for memory augmented neural networks, writes are less frequent than searches (Alkabani et al. 2020; Ni et al. 2019) and so the read speed rather than write speed is the main bottleneck for performance. In our non-volatile architecture, writing or reconfiguration translates to amorphization or recrystallization of the thin GST segments integrated on each MRR in a row. Various reconfigurability schemes have been implemented on phase change-based photonic platforms (Kato et al. 2017; Rios et al. 2018; Wu et al. 2019; Xiong et al. 2011; Zhang et al. 2020; Zheng et al. 2020b, 2020a, 2018) to explore the tradeoff among power consumption, speed, insertion loss, footprint and complexity of the routing arrangement. Since our proposed architecture is based on small MRRs to exploit their large free spectral range, we suitably choose an ultracompact integrated heater scheme for electrical writes proposed recently for GST-based switches on an SOI platform (Sb et al. 2018).

Finite-element based simulations were done by Sb et. al. (2018) on instantiation of phase change via voltage pulses applied to a 10 nm copper layer deposited on top of a GST patch with similar dimensions as ours (500 nm) in a 400X250 nm SOI platform (Fig. 8). Pulses with an amplitude of 5 V were applied for 4 ns to the Cu electrode to heat the 30 nm thick cGST layer above its melting temperature and allowed to quench for <1 ns to achieve full amorphization (SET). For full recrystallization (RESET), a 1 V pulse was applied to





hold the GST's temperature above its glass temperature for 100 ns. Energy consumption for set and reset per bit in our case can be roughly estimated to be half of the reported 194 pJ and 163 pJ values respectively due to shorter (270-400 nm) and thinner (10 nm) GST segments. Our WDM-exploiting layout also supports sub-nanosecond optical writes for the amorphization operation (Feldmann et al. 2020; Li et al. 2019; Stegmaier et al. 2017) by utilizing the input buses to send multiplexed write pulses at the resonant frequencies. However, for this scheme, all the devices in a row have to be reset first before setting the individual bits, and since the ring is undercoupled in the RESET state, write power would be partly wasted. We note that replacing the S-band transparent ITO material with copper would incur some insertion losses leading to decreased contrast between the PCM states, which can be recovered by increasing the length or thickness of the GST patch (Rios et al. 2018), or by sub-wavelength patterning of the GST patches rather than simply deposition along the entire width of the ring waveguide (Wu et al. 2019). Considering the overall device footprint and energy available with current technologies, we propose the Cu-heater approach for facilitating the CAM write operations. Electrical reconfiguration of optical PCM switches is currently an area of concentrated research in the search for enabling large-scale optical FPGAs, and silicon PIN heaters as well as ITO and graphene heaters (Kato et al. 2017; Zheng et al. 2020b, 2020a) are emerging as efficient solutions.

6 System level evaluation

We evaluate the read performance of our CAM architecture as part of an optoelectronic system consisting of state-of-the-art laser sources, modulators and photodetectors to study the reliability of the word search operations for various permutations of match and mismatch under channel noise, crosstalk and source/receiver nonidealities. For this purpose, we utilize Lumerical INTERCONNECT as our simulation platform into which the parameters of the MRR-based devices and optoelectronic system are imported. The implementation of the CAM table is shown in Fig. 9 and system parameters are listed in Table 2. We illustrate our design of a four-entry table with 8-bit words to avoid lengthy simulation times. The 8 MRR resonances are equally spaced by $\Delta\lambda$ in the range 1532–1572 nm so that $8\Delta\lambda < FSR$ (= 56 nm). The extracted port transmissions from FDTD simulations for a single ring are used to calibrate an analytical model (discussed earlier in Sec. 3) to estimate the frequency response for all 8 rings (as shown in Fig. 5).

Since our main purpose is to characterize the bit integrity and uniformity between various permutation states for the read operations under noise and crosstalk, choice of modulator



Fig. 9 Partial schematic for the integration of the CAM table implementation in a photonic circuit simulation framework

Table 2Optical source/receiverand noise parameters	Parameter	Value	
	Laser Frequency	1532–1572 nm	
	Linewidth	60 kHz	
	Power	1 mW	
	Laser Noise PSD	10 ⁻¹⁸ W/Hz	
	Photodetector Responsivity	0.8 A/W	
	PD Dark Current	100 nA	
	PD Thermal NEP	1 pA/Hz ^{1/2}	

and photodetector may be arbitrarily selected. However, in the interest of benchmarking the system under familiar conditions, parameters including the photodetector noise (thermal and shot) characteristics are taken from literature on high-speed (> 50 GHz) integrated Ge p-i-n devices (Chen et al. 2016; DeRose et al. 2011). Static digital queries and their bitwise complements are generated via modulation of CWL lasers by NRZ coded signals. Laser output power is arbitrarily chosen as 1 mW for each probe wavelength (aligned to resonance wavelengths in the device's amorphous state). Additionally, a noise spectral density (10^{-18} W/Hz) is added to the laser sources. In a practical system the NEP of the detector and level of two photon absorption in the SOI platform would determine the source powers.

6.1 Photonic match/mismatch evaluation

Read outputs for full match of search query with the four stored words set to simulate full match, single- and double-bit mismatches (Fig. 10) demonstrate the reliability of the



proposed photonic match scheme. Owing to non-zero transmission to the Through port in the case of a bit match when individual MRRs are not fully critically coupled, full match signals possess a non-zero amplitude (~0.1 mA in the shown case). The two different bit permutations for the single-bit mismatch are split due to 1) the dependence of inter-ring crosstalk on the bit states in a single table entry and 2) small differences in transmission seen by input signals from IN and ADD ports, as qualitatively discussed earlier. We further conduct a full sweep over the 2^n by n input space to simulate the n (=8) possible permutations for a single bit mismatch with each of the 2^n (=256) possible queries to estimate the appropriate decision levels at the receiver (Fig. 11). The vellow surface represents the minimum level of the noisy photodetector output current for a mismatch and the blue surface represents the maximum level of the same for a full match. A minimum extinction ratio of 2.054 dB is exhibited between the two cases. Properly engineering the device parameters to achieve better critical coupling, besides material level optimizations of the phase change element are key to further increasing this contrast (Rios et al. 2018; Wu et al. 2019) towards enabling reliable evaluation of the "degree of mismatch" (important for attentional memories (Ni et al. 2019) in a manner similar to a PAM scheme, rather than just simply detecting the occurrence of a mismatch. In this aspect, future work may also explore utilization of intermediate crystallization states of the GST (Chakraborty et al. 2019; Cheng et al. 2018; Li et al. 2019) to assign different weights to mismatch conditions in different bit positions.

6.2 Read latency

The latency for read operations is not limited by the circuit-level implementation of the CAM table itself but rather by the photonic and electro-optic components used to implement the search system. Read latency through the parallel MRR rows is estimated. The time of flight of the photon in an MRR defined as $T_{photon} = \frac{Q}{\pi * f_{res}}$ amounts to ~2.74 ps using the maximum simulated Q-factor (~1700) for the 1.515 µm MRRs in amorphous (low-loss) state. Optical splitter length of approximately 500 µm (Alkabani et al. 2020) contributes another 1.6 ps to the total time of flight. Thus, the upper bound for the parallel XOR operation speed can theoretically exceed 200 GHz though it is currently limited by current photodetector technology (Chen et al. 2016) response times to around 50 GHz (t_{read} = 0.02 ns).

6.3 On-chip area estimate

Compact on-chip footprint is also an advantage in our photonic CAM owing to the simplistic design of the unit cells. Area requirement for storing the table entries scales with the number and bit-length of words as the product of $n*M*Area_{MRR}$ where n is the word length in bits, M is the number of entries and each MRR occupies an average area of ~7.55 μ m², providing a memory density that would be quite difficult to achieve using bulkier components like Mach–Zehnder Modulators or architectures requiring internal light sources and amplifiers. The MRR 'pitch' requirement (~100 μ m in switching applications) due to thermal restrictions (Cheng et al. 2020) is also a constraining factor for the footprint when thermo-optic modulation schemes are used. Since our design is based on self-holding states, the pitch requirement is greatly relaxed.

Table 3 summarizes the characteristics of our CAM cell and compares them with previous works utilizing optoelectronics (Alkabani et al. 2020; Mourgias-Alexandris et al. 2018) and PCM in electronic domain (Junsangsri et al. 2017). As the title of this work suggests, the ultracompact nature of our designed photonic architecture allows us to achieve a bit density of more than 2 orders of magnitude larger than previous photonics-based designs. The nonvolatile capability of the cells to store bits removes the requirement of voltage

Parameter	This Work	PCM (electronic)	SOA-MZI based	MRM-based
Cell Elements	PCM embedded Si MRR	1MOS-1PCM (32 nm node)	InP SOA-MZI	Si MRM
Read time (ns)	0.02	199.34	0.2	0.02
Write time (ns)	5-100	1.092	0.2	0.03
Write Energy (fJ/bit)	~ 10 ⁵	4×10^{5}	-	100-200
Density (Mbit/cm ²)	13.245	700	< 10 ⁻⁶	0.160
Biasing Power	None	None	~mW	$\sim \mu W$
Read Voltage (V)	None(optical)	0.9	6~8	2

 Table 3
 Characteristics of photonic PCM-CAM and Comparison with Literature

biasing of each individual cell (which is an advantage especially compared to mW-level biasing power of MZI-based approaches) and provides a platform for in-memory computing rather than storing the CAM entries in a separate memory. With further optimization of switching techniques or by use of alternative PCMs with lower switching power, further energy benefits can also be aimed for to achieve energy consumption comparable with that of other technologies in the electrical domain.

7 Conclusion

We have proposed a photonic content addressable memory scheme based on a compact nonvolatile switch capable of high-bandwidth search operations. By leveraging the selfholding capability of the phase change material GST integrated with WDM-compatible Si microring add-drop filters, we designed a photonic XOR-operation structure for parallel read-out. We evaluated the reliability of mismatch detection under inter-ring crosstalk and channel/detector noise and engineer the uniformity of the system's read response to various mismatch permutations by appropriate design analyzing the effect of crosstalk. We have also proposed a suitable write scheme for CAM entry reconfiguration through considering a balanced tradeoff between energy consumption, speed, insertion loss and chip footprint. The proposed architecture can be readily cascaded to support longer word searches and integrated with large scale electronics to implement ultrafast searches in hardware. Our work highlights yet another novel application of nonvolatile integrated photonics capable of not only filling a void in high-end datacenter applications but also providing photonic acceleration in search-based operations for specialized heterogeneous computing platforms.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Agarwal, A., Hsu, S., Mathew, S., Anders, M., Kaul, H., Sheikh, F., Krishnamurthy, R.: A 128x128b highspeed wide-and match-line content addressable memory in 32nm CMOS. Eur. Solid-State Circuits Conf. 83–86 (2011)
- Alkabani, Y., Miscuglio, M., Sorger, V.J., El-Ghazawi, T.: OE-CAM: A hybrid opto-electronic content addressable memory. IEEE Photonics J. 12(2), 1–14 (2020)
- Beamer, S., Akgul, M.: Design of a low power content addressable memory (CAM). Project Report: University of California, Berkeley 1–6 (2009).
- Bogaerts, W., de Heyn, P., van Vaerenbergh, T., de Vos, K., Selvaraja, S.K., Claes, T., Dumon, P., Bienstman, P., van Thourhout, D., Baets, R.: Silicon microring resonators. Laser Photonics Rev. 6(1), 47–73 (2012)
- Burr, G.W., Breitwisch, M.J., Franceschini, M., Garetto, D., Gopalakrishnan, K., Jackson, B., Kurdi, B., Lam, C., Lastras, L.A., Padilla, A., Rajendran, B., Raoux, S., Shenoy, R.S.: Phase change memory technology, J. Vac. Sci. Technol. B. Nanotechnol. Microelectron. Mater. Process. Meas. Phenom. 28, 223–262 (2010)

- Chakraborty, I., Saha, G., Roy, K.: Photonic in-memory computing primitive for spiking neural networks using phase-change materials. Phys. Rev. Appl. 11(1), 01406-3-10140631–5 (2019)
- Chakraborty, I., Saha, G., Sengupta, A., Roy, K.: Toward fast neural computing using all-photonic phase change spiking neurons. Sci. Rep. 8(1), 1–9 (2018)
- Chen, H., Verheyen, P., De Heyn, P., Lepage, G., De Coster, J., Absil, P., Roelkens, G., Van Campenhout, J.: -1 V bias 56 Gbps germanium waveguide p-i-n photodetector with silicon contacts. 2016 Opt. Fiber Commun. Conf. Exhib. OFC 2016, 1–3 (2016).
- Cheng, Q., Kwon, J., Glick, M., Bahadori, M., Carloni, L.P., Bergman, K.: Silicon photonics codesign for deep learning. Proc. IEEE 108(8), 1261–1282 (2020)
- Cheng, Z., Ríos, C., Pernice, W.H.P., Wright, C.D., Bhaskaran, H.: On-chip photonic synapse. Sci. Adv. 3, e1700160-1-e1700160-6 (2017)
- Cheng, Z., Ríos, C., Youngblood, N., Wright, C.D., Pernice, W.H.P., Bhaskaran, H.: Device-level photonic memories and logic applications using phase-change materials. Adv. Mater. 30, 1802435-1-1802435–8 (2018)
- DeRose, C.T., Trotter, D.C., Zortman, W.A., Starbuck, A.L., Fisher, M., Watts, M.R., Davids, P.S.: Ultra compact 45 GHz CMOS compatible Germanium waveguide photodiode with low dark current. Opt. Express 19(25), 24897–24904 (2011)
- Feldmann, J., Stegmaier, M., Gruhler, N., Riós, C., Bhaskaran, H., Wright, C.D., Pernice, W.H.P.: Calculating with light using a chip-scale all-optical abacus. Nat. Commun. 8, 1256-1-1256–8 (2017)
- Feldmann, J., Youngblood, N., Karpov, M., Gehring, H., Li, X., Stappers, M., Le Gallo, M., Fu, X., Lukashchuk, A., Raja, A.S., Liu, J., Wright, C.D., Sebastian, A., Kippenberg, T.J., Pernice, W.H.P., Bhaskaran, H.: Parallel convolutional processing using an integrated photonic tensor core. Nature 589, 52–58 (2021)
- Feldmann, J., Youngblood, N., Li, X., Wright, C.D., Bhaskaran, H., Pernice, W.H.P.: Integrated 256 cell photonic phase-change memory with 512-bit capacity. IEEE J. Sel. Top. Quantum Electron. 26(2), 8301807-1-8301807–7 (2020)
- Feldmann, J., Youngblood, N., Wright, C.D., Bhaskaran, H., Pernice, W.H.P.: All-optical spiking neurosynaptic networks with self-learning capabilities. Nature 569, 208–214 (2019)
- George, S., Jao, N., Ramanathan, A. K., Li, X., Gupta, S. K., Sampson, J., Narayanan, V.: Integrated cam-ram functionality using ferroelectric FETs. Proc. - Int. Symp. Qual. Electron. Des. ISQED 2020-March, 81–86 (2020).
- Grossi, A., Vianello, E., Zambelli, C., Royer, P., Noel, J.P., Giraud, B., Perniola, L., Olivo, P., Nowak, E.: Experimental investigation of 4-kb rram arrays programming conditions suitable for tcam. IEEE Trans. Very Large Scale Integr. Syst. 26, 2599–2607 (2018)
- Junsangsri, P., Han, J., Lombardi, F.: Design and comparative evaluation of a pcm-based cam (content addressable memory) cell. IEEE Trans. Nanotechnol. **16**, 359–363 (2017)
- Karam, R., Puri, R., Ghosh, S., Bhunia, S.: Emerging trends in design and applications of memory-based computing and content-addressable memories. Proc. IEEE 103, 1311–1330 (2015)
- Kato, K., Kuwahara, M., Kawashima, H., Tsuruoka, T., Tsuda, H.: Current-driven phase-change optical gate switch using indium-tin-oxide heater. Appl. Phys. Express 10(7), 072201-1-072201-4 (2017)
- Li, C., Graves, C.E., Sheng, X., Miller, D., Foltin, M., Pedretti, G., Strachan, J.P.: Analog contentaddressable memories with memristors. Nat. Commun. 11, 1638-1-1638–8 (2020)
- Li, X., Youngblood, N., Ríos, C., Cheng, Z., Wright, C.D., Pernice, W.H., Bhaskaran, H.: Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell. Optica 6(1), 1–6 (2019)
- Liu, D., Zhang, C., Liang, D., Dai, D.: Submicron-resonator-based add-drop optical filter with an ultralarge free spectral range. Opt. Express 27(2), 416–422 (2019)
- Maniotis, P., Pleros, N.: All-optical ternary-content addressable memory (T-CAM) cell and row architectures for address lookup at 20 Gb/s. Opt. Quantum Electron. 49, 348-1-348–7 (2017)
- Matsunaga, S., Katsumata, A., Natsui, M., Endoh, T., Ohno, H., Hanyu, T.: Design of a nine-transistor/ two-magnetic-tunnel-junction-cell-based low-energy nonvolatile ternary content-addressable memory. Jpn. J. Appl. Phys. 51(2), 02BM06-1-02BM06-5 (2012)
- Meiners, C. R., Patel, J., Norige, E., Torng, E., Liu, A. X.: Fast regular expression matching using small tcams for network intrusion detection and prevention systems. In: Proc. 19th USENIX Secur. Symp. 111–126 (2010)
- Mourgias-Alexandris, G., Vagionas, C., Tsakyridis, A., Maniotis, P., Pleros, N.: All-optical 10Gb/s ternary-CAM cell for routing look-up table applications. Opt. Express 26(6), 7555–7562 (2018)
- Ni, K., Yin, X., Laguna, A.F., Joshi, S., Dünkel, S., Trentzsch, M., Müeller, J., Beyer, S., Niemier, M., Hu, X.S., Datta, S.: Ferroelectric ternary content-addressable memory for one-shot learning. Nat. Electron. 2(11), 521–529 (2019)

- Pagiamtzis, K., Sheikholeslami, A.: Content-addressable memory (CAM) circuits and architectures: a tutorial and survey. IEEE J. Solid-State Circuits 41, 712–727 (2006)
- Pernice, W.H.P., Bhaskaran, H.: Photonic non-volatile memories using phase change materials. Appl. Phys. Lett. 101, 171101-1-171101-4 (2012)
- Pitris, S., Vagionas, C., Kanellos, G. T, Pleros, N., Kisacik, R., Tekin, T., Broeke, R.: Monolithically integrated all-optical SOA-based SR Flip-Flop on InP platform. In: 2015 Int. Conf. Photonics Switch. PS 2015 vol 1, pp 208–210 (2015)
- Prabhu, A.M., Tsay, A., Han, Z., Van, V.: Extreme miniaturization of silicon adddrop microring filters for VLSI photonics applications. IEEE Photonics J. 2(3), 436–444 (2010)
- Rios, C., Stegmaier, M., Cheng, Z., Youngblood, N., Wright, C.D., Pernice, W.H.P., Bhaskaran, H.: Controlled switching of phase-change materials by evanescent-field coupling in integrated photonics. Opt. Mater. Express 8(9), 2455–2470 (2018)
- Ríos, C., Youngblood, N., Cheng, Z., Le Gallo, M., Pernice, W.H.P., Wright, C.D., Sebastian, A., Bhaskaran, H.: In-memory computing on a photonic platform. Sci. Adv. 5(2), eaau57591– eaau57599 (2019)
- Rudé, M., Pello, J., Simpson, R.E., Osmond, J., Roelkens, G., Van Der Tol, J.J.G.M., Pruneri, V.: Optical switching at 1.55 μm in silicon racetrack resonators using phase change materials. Appl. Phys. Lett. 103, 141119-1-141119–4 (2013)
- Sb, G., Yu, Z., Zheng, J., Xu, P., Zhang, W., Wu, Y.: Ultracompact electro-optical modulator-based ge₂sb₂te₅ on silicon. IEEE Photonics Technol. Lett. **30**(3), 250–253 (2018)
- Stegmaier, M., Ríos, C., Bhaskaran, H., Wright, C.D., Pernice, W.H.P.: Nonvolatile all-optical 1 × 2 switch for chipscale photonic networks. Adv. Opt. Mater. 5(1), 2–7 (2017)
- Wong, H.S.P., Raoux, S., Kim, S., Liang, J., Reifenberg, J.P., Rajendran, B., Asheghi, M., Goodson, K.E.: Phase change memory. Proc. IEEE 98, 2201–2227 (2010)
- Wu, C., Yu, H., Li, H., Zhang, X., Takeuchi, I., Li, M.: Low-loss integrated photonic switch using subwavelength patterned phase change material. ACS Photonics 6(1), 87–92 (2019)
- Wuttig, M., Yamada, N.: Phase-change materials for rewriteable data storage. Nat. Mater. 6(11), 824– 832 (2007)
- Xiong, F., Liao, A.D., Estrada, D., Pop, E.: Low-power switching of phase-change materials with carbon nanotube electrodes. Science 332(6029), 568–570 (2011)
- Xu, C., Chen, S., Su, J., Yiu, S.M., Hui, L.C.K.: A survey on regular expression matching for deep packet inspection: applications, algorithms, and hardware platforms. IEEE Commun. Surv. Tutorials 18(4), 2991–3029 (2016)
- Xu, Q., Fattal, D., Beausoleil, R.G.: Silicon microring resonators with 1.5-µm radius. Opt. Express 16(6), 4309–4315 (2008)
- Yang, R., Li, H., Smithe, K.K.H., Kim, T.R., Okabe, K., Pop, E., Fan, J.A., Wong, H.S.P.: Ternary content-addressable memory with MoS2 transistors for massively parallel data search. Nat. Electron. 2, 108–114 (2019)
- Yin, X., Li, C., Huang, Q., Zhang, L., Niemier, M., Hu, X.S., Zhuo, C., Ni, K.: FeCAM: A universal compact digital and analog content addressable memory using ferroelectric. IEEE Trans. on Electron Devices 67(7), 2785–2792 (2020)
- Yin, X., Ni, K., Reis, D., Datta, S., Niemier, M., Hu, X.S.: An ultra-dense 2FeFET TCAM design based on a multi-domain FeFET model. IEEE Tran Circuits Syst. II: Express Briefs 66, 1577–1581 (2019a)
- Yin, X., Reis, D., Niemier, M., Hu, X. S.: Ferroelectric FET based tcam designs for energy efficient computing. In: Proceedings of IEEE Computer Society Annual Symposium on VLSI, ISVLSI 2019b-July, 437–442 (2019b).
- Zhang, H., Yang, X., Lu, L., Chen, J., Rahman, B.M.A., Zhou, L.: Comparison of the phase change process in a GST-loaded silicon waveguide and MMI. Opt. Express 29(3), 3503–3514 (2020)
- Zhang, H., Zhou, L., Xu, J., Lu, L., Chen, J., Rahman, B. M. A.: Silicon microring resonators tuned with GST phase change material. In: Asia Communications and Photonics Conference ACP 2018-October, 1–3 (2018).
- Zhang, H., Zhou, L., Xu, J., Lu, L., Chen, J., Rahman, B. M. A.: All-optical synapses based on silicon microring resonators actuated by the phase change material Ge₂Sb₂Te₅. In: 2019 Conference on Lasers and Electro Optics, CLEO 2019, 3–4 (2019)
- Zheng, J., Fang, Z., Wu, C., Zhu, S., Xu, P., Doylend, J.K., Deshmukh, S., Pop, E., Dunham, S., Li, M., Majumdar, A.: Nonvolatile electrically reconfigurable integrated photonic switch enabled by a silicon PIN diode heater. Adv. Mater. 32, 2001218-1-2001218–8 (2020a)
- Zheng, J., Khanolkar, A., Xu, P., Colburn, S., Deshmukh, S., Myers, J., Frantz, J., Pop, E., Hendrickson, J., Doylend, J., Boechler, N., Majumdar, A.: GST-on-silicon hybrid nanophotonic integrated

circuits: a non-volatile quasi-continuously reprogrammable platform. Opt. Mater. Express 8(6), 1551–1561 (2018)

Zheng, J., Zhu, S., Xu, P., Dunham, S., Majumdar, A.: Modeling electrical switching of nonvolatile phasechange integrated nanophotonic structures with graphene heaters. ACS Appl. Mater. Interfaces 12(19), 21827–21836 (2020b)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.