

All-photonic in-memory computing based on phase-change materials

Carlos Ríos^{1,2,†,*}, Nathan Youngblood^{1,†}, Zengguang Cheng¹, Manuel Le Gallo³, Wolfram H.P. Pernice⁴, C. David Wright⁵, Abu Sebastian³, and Harish Bhaskaran^{1,*}

¹Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

²Current Address: Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

³IBM Research - Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

⁴Institute of Physics, University of Muenster, Heisenbergstr. 11, 48149 Muenster, Germany

⁵Department of Engineering, University of Exeter, Exeter EX4 QF, UK

*carios@mit.edu, harish.bhaskaran@materials.ox.ac.uk

† These authors contributed equally

Abstract: We experimentally demonstrate, for the first time, co-located data storage and processing (i.e. in-memory computing) on an integrated photonic platform based on nonvolatile phase-change materials. © 2019 The Author(s)

OCIS codes: (230.3120) Integrated optics devices; (200.0200) Optics in computing; (210.4680) Optical memories.

1. Introduction

Breaking the processor-memory dichotomy in Von-Neumann architectures by computing directly on the memory elements – known as in-memory computing – would tremendously transform the computing landscape [1]. Hardware-based implementations of tasks such as scalar multiplication, bulk-bitwise operations, and correlation detection are some of the emerging applications of such architectures, as it has been demonstrated in electronic in-memory computing [2]. An integrated photonic implementation has the potential to further exploit the potential of this architecture by providing an even faster solution with the extra benefits of bandwidth and wavelength multiplexing on a chip. To achieve this, on-chip nonvolatile multilevel memories are required, a gap that has been successfully bridged by embedding phase-change materials (PCMs) in all-photonic chip-scale devices for information processing [3–5]. In this paper, we demonstrate the first instance of a photonic computational memory for scalar multiplication of two numbers and describe the optimization of PCM-based multilevel memory. This represents a milestone for optical processing in memory given their potential for applications in solution of systems of linear equations, machine learning, and deep learning [1,6].

2. Results

We used the photonic memory device sketched in Fig. 1a as the active device to demonstrate multiplication between two scalars (i.e. “ $a \times b$ ”). We do so by mapping the variable “ a ” to the power of an input pulse P_{in} and “ b ” to the transmittance T_{GST} of the device. Prior to the input pulse P_{in} , we set the value of T_{GST} with a Write pulse P_{Write} which nonvolatility switches between different mixtures of amorphous and crystalline $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) [7]. The corresponding output pulse P_{Out} contains the resulting multiplication of $P_{in} \times T_{GST}$, which is amplitude encoded. We used pulse energies for P_{Write} and P_{in} above and below the device’s switching threshold (i.e. the minimum energy to partially amorphize fully crystalline GST), respectively. In this manner, P_{in} pulses are modulated by the transmittance imposed by GST, but they do not further modify the transmission state of the device. Our device uses evanescent-field coupling between the waveguide mode and the phase-change memory cell placed on top of the waveguide to crystallize or amorphize GST by annealing over 150°C and melt-quenching over $\sim 600^\circ\text{C}$, respectively using P_{Write} [3].

To obtain reliable behavior of GST even after many hundreds of cycles, which is crucial to obtain low errors in the calculations, we optimized the following to boost the device performance [8]:

1. We maximized the SNR of the optical transmission measurement and thus, we were able to codify a larger number of transmission levels in a single cell (equivalent to a 4-bit variable).
2. We introduced a new switching mechanism consisting of a single double-step pulse to *Erase* from any transmittance level directly to the baseline. This double-step pulse represents a lower energy (sub-nJ) and faster (125ns) approach to that of a train of decreasing energy pulses (nJ and microseconds) [3]. However, to transition between two T_{GST} levels it was necessary to always *Erase* to the baseline beforehand.
3. We demonstrated the lack of optical drift up to 10^4 s (i.e. a true nonvolatile behavior), which is unique in this device architecture and one of its biggest advantages, especially if compared with the resistance drift of its

electrical counterpart. This means that for our optical in-memory computing approach, no energy is wasted in periodically correcting the state of the memory [4].

4. We demonstrated that the GST transmittance can be operated in the linear regime—that is, above the switching threshold and below the saturation region using $P_{Write} \in [180\text{pJ}, 354\text{pJ}]$. This unique property allows us to have a linear mapping between the scalar number “b” and the transmittance value T_{GST} . This property avoids fitting functions and demanding postprocessing to retrieve the calculations, which is required in the case of electronic counterparts due to the nonlinear pseudo-ohmic behavior of GST.

The outstanding performance of our optimized device allowed the demonstration of in-memory multiplication that we show in Fig. 1b with reduced postprocessing, smaller calculations errors, and at the speed of light. This proof-of-concept made use of 13 repeatable transmittance levels in the nonvolatile memories, and 33 unique values for the power of input pulses. A postprocessing step was necessary to remove the baseline transmission offset due to the finite transmission of fully crystalline GST. This can be further improved by utilizing photonic devices with improved extinction ratios. The error distribution shown in the inset of Fig. 1b shows good correspondence between measured and exact values, with a standard deviation of just 2.3%. We attribute great part of this error to electrical fluctuations in the generation of the P_{Write} pulses via electro-optical modulation, which introduce uncertainty in both the actual T_{GST} that is reached and the power of each P_{in} .

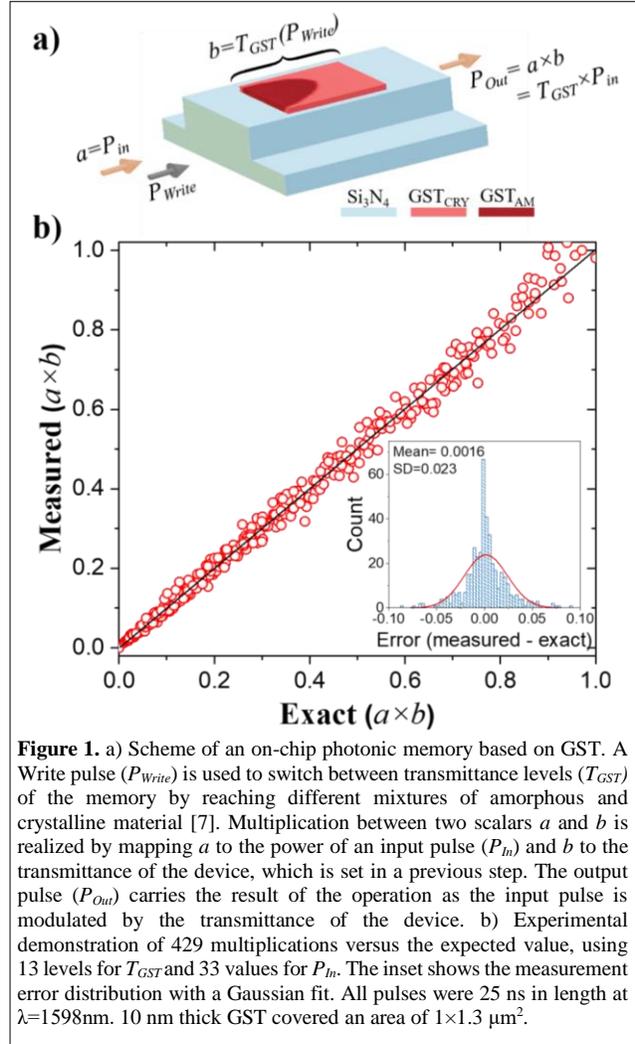


Figure 1. a) Scheme of an on-chip photonic memory based on GST. A Write pulse (P_{Write}) is used to switch between transmittance levels (T_{GST}) of the memory by reaching different mixtures of amorphous and crystalline material [7]. Multiplication between two scalars a and b is realized by mapping a to the power of an input pulse (P_{in}) and b to the transmittance of the device, which is set in a previous step. The output pulse (P_{Out}) carries the result of the operation as the input pulse is modulated by the transmittance of the device. b) Experimental demonstration of 429 multiplications versus the expected value, using 13 levels for T_{GST} and 33 values for P_{in} . The inset shows the measurement error distribution with a Gaussian fit. All pulses were 25 ns in length at $\lambda=1598\text{nm}$. 10 nm thick GST covered an area of $1 \times 1.3 \mu\text{m}^2$.

These results confirm the potential of phase-change materials in photonic computational hardware paradigms—including the ability to perform calculations in the same physical location as memory, using light. While there is room for improvement, the current capabilities of this kind of photonic integrated device hold promise for all-optical nonvolatile data storage [3], on-chip photonic synapse [9], all-optical and electro-optical switches [10,11], and, as we have also demonstrated here, optically performing computational tasks with co-located memory and processing [5,8].

3. References

1. F. L. Traversa and M. Di Ventra, "Universal Memcomputing Machines," *IEEE Trans. Neural Networks Learn. Syst.* **26**, 2702–2715 (2015).
2. M. Le Gallo, *et al.* "Mixed-Precision In-Memory Computing," *Nat. Electron.* **1**, (2017).
3. C. Ríos, M. Stegmaier, *et al.* "Integrated all-photonic non-volatile multi-level memory," *Nat. Photonics* **9**, 725–732 (2015).
4. M. Wuttig, H. Bhaskaran, and T. Taubner, "Phase-change materials for non-volatile photonic applications," *Nat. Photonics* **11**, 465–476 (2017).
5. J. Feldmann, M. Stegmaier, *et al.* "Calculating with light using a chip-scale all-optical abacus," *Nat. Commun.* **8**, 1256 (2017).
6. M. Di Ventra and Y. V. Pershin, "The parallel approach," *Nat. Phys.* **9**, 200–202 (2013).
7. C. Ríos, M. Stegmaier, *et al.* "Controlled switching of phase-change materials by evanescent-field coupling in integrated photonics [Invited]," *Opt. Mater. Express* **8**, 2–7 (2018).
8. C. Ríos, N. Youngblood, *et al.* "In-memory computing on a photonic platform," *ArXiv* **1801.06228**, (2018).
9. Z. Cheng, C. Ríos, W. H. P. Pernice, C. David Wright, and H. Bhaskaran, "On-chip photonic synapse," *Sci. Adv.* **3**, 1–7 (2017).
10. M. Stegmaier, C. Ríos, H. Bhaskaran, C. D. Wright, and W. H. P. Pernice, "Nonvolatile All-Optical 1X2 Switch for Chipscale Photonic Networks," *Adv. Opt. Mater.* **5**, 2–7 (2017).
11. Q. Zhang, *et al.* "Broadband nonvolatile photonic switching based on optical phase change materials : beyond the classical figure-of-merit," *Opt. Lett.* **43**, 94–97 (2018).