Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Electrical Programmable Low-loss high cyclable Nonvolatile Photonic Random-Access Memory

Jiawei Meng
The George Washington University
Nicola Peserico
The George Washington University
Xiaoxuan Ma
The George Washington University
Yifei Zhang
MIT https://orcid.org/0000-0002-4928-2921
Cosmin-Constantin Popescu
Massachusetts Institute of Technology https://orcid.org/0000-0001-8575-4781
Myungkoo Kang
University of Central Florida
Mario Miscuglio
The George Washington University
Kathleen Richardson
University of Central Florida
Juejun Hu
Massachusetts Institute of Technology https://orcid.org/0000-0002-7233-3918
Volker Sorger (sorger@gwu.edu)
George Washington University https://orcid.org/0000-0002-5152-4766

Article

Keywords: Random Access Memory, Integrated Photonics, Phase Change Material, Cyclability, Neural Networks, Tensor Operations

Posted Date: April 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1527814/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

The retention-of-state functionality provided by memories is fundamental to any Turing machine and neural network, hence is critical for any information system today. While emerging optical machine learning accelerators and photonic neuromorphic computing paradigms provide promising signal processing and computing performance, the lack of a photon-photon force in the universe makes storing optical information challenging. Fortunately, phase change materials provide such a missing memristive nonvolatile function via their reconfigurable crystalline structure and allow for rapid optical READ paradigms. However, demonstrations of photonic memory are limited by high optical loss, low state-cyclability, and rely on cumbersome non-CMOS like optical programmability. To overcome all three shortcomings and unlock the full potential of optical information storage and access, here we introduce a photonic random-access memory featuring vanishing low optical loss, demonstrate more than half a million switching cycles, a 100x improvement over state-of-art, and realize electrical programmability on-chip. The exceedingly low optical absorption (0.0015 dB/μm) is achieved via a novel broadband transparent phase change material, $Ge_2Sb_2Se_5$ integrated atop a nanophotonic waveguide of a silicon chip. We show a highly efficient signal modulation (0.2 dB/μm) achieved by realizing a newly designed paired micro-heaters along both sides of waveguide, which allows for electronic-standard programmability of these photonic memories. When interrogated by an optical beam, they offer picosecond-short memory READ latency. Furthermore, we demonstrate a partial amorphization scheme realizing multi-state memory levels on a single heater enhancing footprint efficiency. Lastly, we verify the energy and switching speed and show how each trades-off with heater-to-waveguide proximity and signal strength, respectively. Such as CMOS-near electronically programmed and optical read photonic random-access memory with low-optical loss yet efficient programmability can become a crucial building block for network edge AI system of the looming industry 4.0 era.

Introduction

Since the inception of logic machines thousands of years ago, humanity strived to invent systems capable of performing mathematical and logic operations. In fact, some argue that the ability to execute ‘logic’ is mandatory for any entity displaying a form of ‘intelligence’ and ‘consciousness’. Indeed, even the most fundamental mathematical algebraic operations (summation, subtraction, division, multiplication) that run quadrillion of times quietly in the background daily as part of our modern life performing higher-order complex operations such as multiply-accumulate (MAC), vector-matrix multiplications (VMM), tensor operations, convolutions, for cryptographic algorithms, machine learning and AI, or simply running a CPU or GPU on any modern-day processor, all of these do require retention-of-state or, in short, memory functionality. The amount and performance (e.g., energy and latency, speed, for programmability, read, and reset operations) of the memory naturally depend on the underlying algorithm to be executed; for instance, a push-down automaton offers access only to the latest-written state but not previous ones, while any Turing machine is characterized by the ability to access any formerly saved state, leading to

random-access memory (RAM). Regardless of the compute or processing system, all do require memory functionality.

Now, the recent trend in computing is favoring heterogeneous systems where certain operations are accelerated by a dedicated machine rather than a generalized system – with the gains being 'dedicated' performance (e.g., higher TOPS/W). In addition, there is also new interest in data processing at the network edge foreshadowing the industry 4.0 era where co-design of hard- and software will intertwine in an internet-of-everything approach. Certainly, demand for machine intelligence and AI has skyrocketed and is continuing to do so for the foreseeable future. All this information processing demands more memory performance, especially in emerging non-van Neumann systems such as neural networks or tensor-core processors[1–3].

Given this golden era of hardware development, mixed-signal and analog signal processing systems and accelerators are being considered, including electronic and photonic approaches. Indeed, optics and, more relevantly, those implementations utilizing photonic integrated circuits (PIC) that leverage the manufacturing know-how from the electronic semiconductor industry, have shown to offer a number of compelling benefits when it comes to signal processing and compute-acceleration to include i) ps-short latency leading to real-time-compute enabling rapid decision-making[4], ii) parallelization strategies such as offered by wavelength multiplexing or using millions of optical channels in digital-mirror-display technology[5], iii) free Fourier transformation performed by a lens[6], leading to iv) reduced algorithmic complexity scaling (e.g. $O(N^2)$ vs. $O(N^4)$ for a convolution operation, where N is the input size of the data and of the filtering kernel, K, i.e. assuming $N = K$, for illustrative purposes) enabled by a naturally-occurring execution of the convolution theorem, v) seamless MAC operations[7] enabled by time-of-flight light-matter multiplications[8] and photodetector summation, all awhile vi) leveraging mature semiconductor processes. Nonetheless, any optical accelerator or signal processor must be designed from a system perspective and will always be an electronic-photonic hybrid. Especially compute-in-memory schemes and neuromorphic computing paradigms requires seamless and efficient local memory storage and rapid access.

However, realizing memory functionality in the optical domain for these emerging hybrid accelerators is challenging, given that no photon-photon force exists in the universe. The brute-force solution to use delay lines as memory are futile given the lightspeed velocity of photons and can only lead to information buffering but not to nonvolatility [9]. While slow-light paradigms may extend the buffer-time, they usually introduce optical absorption, and random-access is challenging thus rendering them not useful[10]. Fortunately, information can be stored in close proximity to an optical bus and simply accessed (i.e., READ operation) by passing a signal through such a waveguide bus. Here the required memory WRITE/RESET operations can be realized using electro-optic[11,12], or electro-thermal effects[13], while the programming trigger stimulus can be either electrically or optically signals a stimulus. However, simply using an electro-optic modulator as photonic memory does not provide nonvolatility since the memory state is lost as soon as the voltage is removed.

Opportunely, both random-access and nonvolatility can be introduced using Phase-Change Materials (PCMs). While demonstrations utilizing GST prove the validity of an optical PCM-based memory[14–18], there are a number of unsolved shortcomings that severely limit the vision of a photonic RAM (P-RAM) including high optical loss, low state-cyclability, and an inability of electrical programmability that, once solved, can lead to high efficiency and an approach that is well-aligned with standards of the semiconductor industry. Nonvolatility is especially of interest for efficient machine learning accelerators that perform inference or classification tasks since the trained 'weights', which could be performed by PCM-based P-RAMs, oftentimes are only updated rarely or seldom at the most.

For the operation of a P-RAM, the PCM is switched between two structural states, either in amorphous state or crystalline state, with distinct optical and electrical properties at each. Those states can be reversibly switched and cycled through appropriate thermal or optical stimulation with long-term stability[19]. One of the commonly used PCM materials for photonic memory is GST (Ge-Sb-Te) which exhibits a relatively large contrast of both refractive index (Δn) and optical absorption (Δk) when switched between amorphous and crystalline states.[20,21, 22,23] However, GST is characterized by a high absorption coefficient even in the amorphous state. For large photonic networks such as the one implementing deep NNs, the multi-layer design requires the photonic memory containing the kernel memory to be very low loss[24], an aspect that cannot be met by GST-based photonic memory due to its high absorption coefficient which results in high static power consumption for the operation of networks.

To solve this pain point, here, we introduce and demonstrate an ultra-low loss and highly cyclable nonvolatile electrically programmed on-chip P-RAM based on the phase-change material $Ge_2Sb_2Se_5$ (GSSe). This material offers a vanishingly low optical loss enabling bit-scalability and efficient photonic circuits. We demonstrate a high WRITE/RESET cyclability of half a million switching events without any display of degradation. Position-optimized electrical on-chip micro-heaters allow for a compact and CMOS-near memory programmability. The design allows to scale-up the number of memory bits by adding more micro-heaters, and we also show single micro-heater multi-state programmability by a partial amorphization for applications with less demanding signal-to-noise ratio yet require a footprint-conscious chip design. Unlike previous work on photonic PCM memories, here we verify the available signal modulation strength of the optical READ signal passing through the P-RAM as a function of programming speed and map-out the optical signal modulation-to-loss ratio as a function of micro-heater proximity to the waveguide, hence optimizing the memory design. For future successful chip implementations, P-RAMs must be electrically (rather than all-optically) programmable due to packaging synergies aligned with industry standards for future mass production. Adding photonic compute-in-memory capability via local P-RAM building block to the PIC toolbox enables a plethora of possibilities in signal processing and computing machine learning accelerators, AI-edge modules, but also for analog-RF signal processing, cryptographic accelerators or bio-medical devices, and wearable devices required to process information or make decisions.

Results

The PCM chosen as the key role in our proposed P-RAM is GSSe since it presents a broadband transparent region for telecommunication wavelengths while in its amorphous state. This GSSe amorphous state is characterized by a remarkably low extinction coefficient (2.0×10^{-5} at 1550 nm wavelength, as shown in Fig. 1e), enabling near-lossless devices monolithically co-integrated with PICs. This low absorption coefficient is over two orders of magnitude lower than regularly employed GST at 1550 nm. Meanwhile, when in its crystalline state, the absorption coefficient increases to 0.14 (see Supplementary Note 7), which results in an absorption contrast between the two states of 0.13.

To demonstrate our nonvolatile P-RAM, a thin layer of GSSe film is directly deposited on the top of a planarized silicon waveguide, as shown in Fig. 1a. The obtained memory states are programmed by selectively 'writing/erasing' portions of the GSSe film via local electro-thermal heating. This allows electrically driven change of GSSe's structural state (crystalline/amorphous) and consequently results in the strong imaginary-part variation of the effective refractive index, leading to a significant optical absorption change.

In our scheme, heat is applied to the material externally via joule heating of a tungsten-titanium (W/Ti) metal layer in contact with a 20 nm aluminum oxide dielectric layer over the GSSe film, to protest GSSe from oxidization, whose mode and thermal profile are simulated, as shown in Fig. 1f. According to the type of transition wanted, different pulse train profiles are applied to the metal wire via electrical connections to the device. With the 3D mode simulation through COMSOL (see Supplementary Note 8), we optimized the position of the heaters regarding the waveguide to minimize the ohmic losses due to the presence of metal and concurrently lower the threshold voltage for delivering the necessary amount of heat for inducing the phase transition. The optimized heaters configuration consists of two non-plasmonic tungsten resistive heaters placed in contact with a thin spacer of aluminum oxide deposited on top of the GSSe film, as shown in Fig. 1g,h. The heaters are placed 500 nm away on the side of the waveguide, thus providing heat to the film locally, which not only lowers the switching threshold but also temporally stores the heat for successive pulses.

To reach the multi-state power output response, a sequence of paired heaters is placed along the waveguide in series. Each pair of heaters are individually tuned to Joule heat locally the GSSe material for solid-state phase transition. Whereas, in the crystalline state, the GSSe becomes much lossier with a linear absorption coefficient of ~ 0.2 dB/μm obtained by experimental data, compared to the close to zero insertion loss in the amorphous state. This P-RAM configuration takes advantage of this near-lossless characteristic of the GSSe material in its amorphous state, as the optical signal loss in the waveguide is minimal even for long strips. We precisely control the state of each portion of material by tuning each pair of heaters to obtain a stepwise extinction ratio. When N pairs of heaters are placed, a total of N + 1 memory states of power intensity response are realized.

As shown in Fig. 1a,b), our photonic memories comprise one single 30-nm thin GSSe pad with paralleled pairs of W-Ti micro-heaters arranged along the waveguide, as each pair of heaters correspond to a quantized state. The double-sided heater design leads to the highest thermal energy efficiency for the

phase transition of GSSe. Furthermore, this design prevents the extra optical insertion loss introduced by the metal heaters, since the metal strips are not directly deposited over the waveguide, but instead have a few hundred nm horizontal distance from the side of the waveguide.

We also present an alternative layout, shown in Fig. 1d), that comprises 30-nm-thin and 5000-nm-wide programable PCM wires arranged in a grating fashion (duty cycle 50%), with a series of single-sided heaters to Joule-heat each PCM wire, exploiting the same electrical local Joule heating concept. The single-side heater concept shown here is mainly used for P-RAM with a high order of bit number resolution which requires a larger amount of GSSe material cells along with the required number of micro-heaters, metal pads, and routing.

By using this layout, an all-electrical controlled 4-bit P-RAM element is implemented. Considering the highest state as the condition in which all GSSe wires are in the amorphous state, 15 reprogrammable wires are sufficient for implementing the 4-bit memory with an overall length of just 80 μm, excluding electrical circuitry. The insertion loss, defined as the optical power loss when all the wires are in an amorphous state, is only 0.12 dB for the 4-bit multilevel memory. The optical power transmitted decreases when the GSSe wires are written/SET (switching to crystalline), leading to discrete power levels for each quantized state. The insertion losses for the multistate memory device with different quantization states are measured and shown in Fig. 2a) which realized 16 quantization states for 4-bit. The photonic memory implemented in this configuration provides a uniform quantization. For a 4-bit P-RAM, the quantization step is 0.75 dB/state with the total maximum extinction ratio of 12 dB achieved along with 16 output power states.

To further enhance the speed and energy efficiency of electro-thermally switched PCM P-RAM, we optimized the micro-heaters position. We tested different distances between waveguide and heaters, from 0.125 μm to 5 μm with the previously shown design of double-sided heaters. With the same amount of electrical energy applied to each heater pair, the total extinction ratio (ER) achieved decreased with the increasing distance, while concurrently the unit insertion loss introduced by the GSSe cell decreased, as shown in Fig. 2b,e). To balance the phase transition energy efficiency and insertion loss (IL), we calculated the Figure-Of-Merit (FOM) as ER/IL for each distance, and the optimized position is shown in Fig. 2c), indicating 500 nm distance as the best value. At this distance, we compared the FOM of our proposed device along with three other PCM photonic memories as shown in Fig. 2e,f). With the same theoretical unit insertion loss, our devices achieved the highest extinction ratio.

To evaluate the endurance of our device, a cyclability measurement was conducted and a total of half-million Writing-Resetting cycles was successfully achieved, as shown in Fig. 2d), with stable power responses in either state. The main limitation which prevents memory from achieving higher Writing-Resetting cycles was the failure of micro-heaters on the two sides of the waveguide. With the large number of heating-cool down cycles for GSSe Writing-Resetting, the initial tungsten micro-heaters were easily broken due to oxidization under the fast temperature change, as shown in the lower right subfigure of Fig. 2d). To overcome the issue, we replaced the heater material from tungsten to tungsten-titanium

with a 200nm thick dual layer of aluminum deposited over the W/Ti on the routing part. The Al layer reduces the electrical resistance, enhances heat uniformity, and protects the W/Ti heaters. Meanwhile, a 600 nm thick layer of Al_2O_3 is deposited over the device to prevent further oxidization and physical bend of metal[25]. Such structure allows the heater to survive after half-million cycles, as shown in the upper right sub-figure of Fig. 2b). Up to our knowledge, this is the longest cycle test of a PCM P-RAM integrated into a photonic circuit with stable Writing-Resetting photonic responses.

For GSSe material, the amorphization temperature is the melting point (> 900K), while for crystallization a certain temperature (~ 600K) must be applied and kept constant for approximately $20\mu s$[26]. Crystallization is achieved by applying the pulse setting shown in Fig. 3c) to keep the material temperature consistent in the desired range for over $20\mu s$, while the amorphization is achieved by adding a threshold voltage 10-12V (~ 2 μs) to the local heater up to 900K. The voltage range takes into consideration the fabrication variability of micro-heaters. The real-time continuous Writing-Resetting measurement for two states of memory is shown in Fig. 3a).

Since the total extinction ratio that we want to achieve is proportional to the area of the GSSe cell covering the waveguide, the phase transition time required is also proportional to the desired extinction ratio, for the different thermal volumes of PCM material. We then experimentally mapped the amount of ER that can be achieved as a function of transition time at the falling edge of the real-time transmission trace in Fig. 3a), and the results are shown in Fig. 3b). To achieve 0.2 dB ER, 0.5 ms is needed compared to 500 ms required for 6 dB total ER response.

Besides the concept of multi-heater pairs that we proposed for the multi-states P-RAM, here we propose another design concept for the optical memory, as shown in Fig. 4a). As we described, the multi-state optical power response is realized by tuning the ratio of GSSe film over the waveguide through local Joule-heating to introduce a different level of insertion loss[27]. Based on this theory, we developed the non-equal heater pair memory cell. The shorter heater on the right side works for the amorphization of the GSSe cell. The COMSOL electrical-thermal simulation results, as shown in Fig. 4b), indicate that with different energies applied to the heater, the above-melting-point hot area changes, which introduces the different amorphous areas and by so the different absorption levels. The fitting equation as shown in Fig. 4b indicates the different extinction ratio achieved is propagated to the thermal decay time $\frac{1}{e}\tau$ as long as the device heat structure gives us a hint about how the real terminal structure of P-RAM influences the rime response of extinction ratio achieved. Figure 4d) displays the numeric results which more clearly show the distribution of hot area (> 900K) with the increasing of electrical energy applied to the micro-heater. On the other side, the longer heater on the left side works as the resetting button to change the full GSSe cell into its crystalline state and erases all the previously stored information set by the right heater. The experimental result for 6 different states (2.58-bit) has been achieved as shown in Fig. 4c) for different energy levels from 80 to 400 nJ, and further measurement and optimization will keep going on for the higher-order, to reach 5-bit 32 states P-RAM, as the theoretical highest bit resolution is mainly limited by the resolution of photodetector for which can distinguish small neighbor extinction ratio

over the noise level. The clear advantage of this design is the reduction of electrical tracks and pads, allowing a better integration into photonic deep NN. On the other hand, the main drawback is the difficulty of programming compared to the multi-heater pair design. In the first design, the programming pulse setting for each heater pair is fixed, and the ultimate power response is the result of an accumulation for all parallel paired heaters. For this design, there is not an *a priori* known relationship between ER and applied energy. To get the multi-state power response with fixed step size, a pulse setting for each state needs to be individually set, introducing an extra program difficulty.

Discussion

Here, we presented a novel ultralow insertion loss phase change material GSSe, implementing nonvolatile electrical-controlled photonic memory as various reconfigurable devices following a similar concept. We also proved the device's endurance with over half a million switching cycles. As the initial cyclability test was mainly limited by the durability of metal heaters, as they were physically broken after a large number of heating and cooling cycles, we improved the lifetime of heaters with a thick oxide layer covered on the top to reach 500,000 cycles result. Novel structures and devices could be optimized to enhance further the photonic memory cyclability by improving the design of the materials stack for the heaters.

A few key P-RAM performance characteristics have been compared with two other demonstrated P-RAM approaches, as shown in Table 1. Though this work has larger setting energy and smaller unit extinction ratio compared to the all-optical setting GST-based photonic memory, we hold the best figure-of-merit (ER/IL) due to our ultra-low insertion loss benefitting from the transparent GSSe material and a novel double-sided metal heater design. Moreover, we have successfully demonstrated half-million Writing-Resetting cycles with very stable performance which is far more than other P-RAM's cyclability results, as shown in Table 1.

From simulations, we expected an extinction ratio of 0.4 $dB/\mu m$, while from the experimental demonstration we obtained about 0.2 dB/μm. The main reason for this difference is due to the heat distribution applied to the PCM cell through the micro-heater, as not the whole PCM reached the transition temperature. As shown in Fig. 4b), the heat spread follows an ellipse shape which results in a non-uniform temperature map, that caused a lower extinction ratio compared to the simulations. The different crystalline-amorphous ratios caused by the non-uniform heating led to our duty-cycle proposed P-RAM design, which compromises a smaller volume of PCM, and by so a more uniform heat distribution.

As we described before, the total extinction ratio of our P-RAM could be achieved is based on the length of GSSe cell which could be transitioned through a micro-heater. Then the highest bit resolution that could be achieved by each device is limited by the minimum detected dynamic extinction ratio for every single state through an optical power meter. Based on our current measurement setup, the minimum detectable power range is 35 pW which means that we could achieve 1 binary state as small as over 35 pW difference in theory. Then for traditional 4- or 5-bit memory, the length of the active region for each memory could be sub-micrometer long and could be cascaded for different bit resolutions required.

The nonvolatility of our P-RAM results in zero static power consumption for state maintenance and exceptionally low insertion loss introduced by active PCM material GSSe. Meanwhile, the setting energy of our P-RAM is also relatively low, computed around 0.3 $nJ/(dB*\mu m)$. As we discussed previously, the bit resolution is limited by the minimum dynamic extinction ratio in dB that could be detected over the system noise level, which means that the required energy for each bit Writing-Resetting and the required footprint could be as small as nW level as shown in Table 1 for our device.

In large-scale photonic computing architectures, such as high order matrix MAC operation required for deep neural networks, the stringent energy requirements motivate the implementation of multiple photonic memories for weight bank[7,29]. For these challenges, our devices can perform even orders of magnitude better than volatile memories in terms of energy consumption and footprint.

Besides the low operating energy consumption for high dimension photonic tensor operations, our proposed P-RAM takes advantage of all-electrical micro-heaters, and by so reducing the packaging complexity compared to all-optical laser heating PRAMs compared in Table 1. When tens of thousands of P-RAMs need to be implemented, electrical control is the only feasible way for memory programming and large-scale photonic circuit packaging[30].

Conclusion

In summary, we designed, fabricated, and experimentally tested a new class of electrically programmed photonic nonvolatile random-access memory integrated into a photonic chip. These memories feature an underexplored phase-change material, GSSe, offering a vanishingly small optical loss at the amorphous crystallinity state when READ in the telecommunication-relevant spectral band. We further showed how these memories allow for multi-bit programmability via optimizing the on-chip micro-heaters, which enables a CMOS-near memory packaging solution offering electrical memory programmability, unlike earlier demonstrations relying on all-optical programmability. Furthermore, we demonstrated a record-high half-million switching cycles showing stable operations. We also investigated and verified the programming speed of these memories as a function of optical signal modulation strength and obtain millisecond-like speeds when a strong multi-dB signal level is required, such as relevant for high baud-rate photonic AI accelerators. Moreover, we find an efficient 0.3 nJ/μm programmability and a signal strength-to-optical loss ratio of 70 or about 5 times higher than other work. Such nonvolatile zero-static power-consuming photonic RAMs on a Silicon Photonic chip along with CMOS-near packaging option and electrical programmability showing high cyclability and picosecond short READ latency are a promising technology for applications in optical signal processing and computing such as weights for photonic neural networks, optical architectures switching for telecommunication and data-centers, bio-medical photonic applications, or for executing cryptographic algorithms.

Methods

Fabrication:

20 nm GSSe thin film layer was deposited by using single-source thermal evaporation and a 20 nm layer of Al_2O_3 was deposited by using atomic layer deposition (ALD) as a protective coating to prevent GSSe from oxidation. The tungsten-titanium microheater was fabricated in the nanofabrication and imaging center at George Washington University. A 200 nm think tungsten titanium layer is sputtered. Then another 200 nm think Al is deposited over the W/Ti route to decrease the overall resistance, increase the heat spread over the micro-heaters, and to protect the W/Ti layer from oxidation. Then a thick 400 nm Al_2O_3 layer is deposited over the full circuit using the ALD for oxidation prevention. Contact pad windows are opened using oxide layer plasma dry etch for electrical probes to connect with circuits for micro-heaters driving.

Electro-thermal simulation/optical mode simulation and microheater modeling:

The Joule heating process and heat dissipation model were performed using a three-dimensional finite-element method in COMSOL Multiphysics. We used the AC/DC Joule-heating module coupled with the heat transfer module, which accounts for surface radiation as well as thermal boundary resistance.

Declarations

Acknowledgments

This work was performed in part at the George Washington University Nanofabrication and Imaging Center (GWNIC). Thin film material analysis is supported from NIST Center for Nanoscale Science and Nanotechnology (CNST), and James N. Hilfiker from J.A. Woollam Co. V.J.S. is supported by AFOSR (FA9550-20-1-0193) under the Presidential Early Career Award in Science and Engineering (PECASE). We acknowledge utilizing planarized photonic integrated waveguide chips provided by D. Van Thourhout and appreciated ellipsometry support from Y. Gui.

Contributions

V.J.S. conceived the project. M.M. and J.M. conceptualized the design. J.M. and C.P. fabricated the devices. J.M. performed the measurements and data analysis. J.M. and M.M. modeled the devices and provided theoretical analysis. X.M. and N.P. supported photonic chip measurements and test setups. Y.Z., C.P. and J.J. provided material processing support. M.K. and K.R. supplied the phase-change material. All authors discussed the results and commented on the manuscript.

References

1. Varnava, C. Photonic devices compute in memory. *Nature Electronics 2019 2:3* **2**, 91–91 (2019).

2. Feldmann, J. *et al.* Calculating with light using a chip-scale all-optical abacus. *Nature Communications* **8**, 1–8 (2017).

3. Chakraborty, I., Saha, G., Sengupta, A. & Roy, K. Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons. *Scientific Reports* **8**, 1–9 (2018).

4. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics 2017 11:7* **11**, 441–446 (2017).

5. Shastri, B. J. *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics 2021 15:2* **15**, 102–114 (2021).

6. Dalir, H. *et al.* Massively parallel amplitude-only Fourier neural network. *Optica, Vol. 7, Issue 12, pp. 1812-1819* **7**, 1812–1819 (2020).

7. Miscuglio, M. & Sorger, V. J. Photonic tensor cores for machine learning. *Applied Physics Reviews* **7**, 031404 (2020).

8. George, J. K. *et al.* Neuromorphic photonics with electro-absorption modulators. *Optics Express, Vol. 27, Issue 4, pp. 5181-5191* **27**, 5181–5191 (2019).

9. Photonic Radio Frequency Memory Using Frequency Shifting Recirculating Delay Line Structure. https://opg.optica.org/jlt/abstract.cfm?uri=jlt-32-1-99.

10. Wang, X. *et al.* Continuously tunable ultra-thin silicon waveguide optical delay line. *Optica, Vol. 4, Issue 5, pp. 507-515* **4**, 507–515 (2017).

11. Suer, C. *et al.* Sub-wavelength GHz-fast broadband ITO Mach–Zehnder modulator on silicon photonics. *Optica, Vol. 7, Issue 4, pp. 333-335* **7**, 333–335 (2020).

12. Hong, J., Qiu, F., Cheng, X., Spring, A. M. & Yokoyama, S. A high-speed electro-optic triple-microring resonator modulator. *Scientific Reports 2017 7:1* **7**, 1–6 (2017).

13. Hashemi, H., Nakai, M. & Chung, S. Low-power thermo-optic silicon modulator for large-scale photonic integrated systems. *Optics Express, Vol. 27, Issue 9, pp. 13430-13459* **27**, 13430–13459 (2019).

14. Cheng, Z. *et al.* Device-Level Photonic Memories and Logic Applications Using Phase-Change Materials. *Advanced Materials* **30**, (2018).

15. Rios, C. *et al.* Integrated all-photonic non-volatile multi-level memory. *Nature Photonics* **9**, 725–732 (2015).

16. Li, X. *et al.* Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell. *Optica* **6**, 1 (2019).

17. Xu, P., Zheng, J., Doylend, J. K. & Majumdar, A. Low-Loss and Broadband Nonvolatile Phase-Change Directional Coupler Switches. *ACS Photonics* **6**, 553–557 (2019).

18. Zhang, H. *et al.* Miniature Multilevel Optical Memristive Switch Using Phase Change Material. *ACS Photonics* vol. 6 2205–2212 (2019).

19. Zhang, Y. *et al.* Myths and truths about optical phase change materials: A perspective. *Applied Physics Letters* vol. 118 (2021).

20. Alexoudi, T., Kanellos, G. T. & Pleros, N. Optical RAM and integrated optical memories: a survey. *Light: Science & Applications 2020 9:1* **9**, 1–16 (2020).

21. Zhang, H. *et al.* Comparison of the phase change process in a GST-loaded silicon waveguide and MMI. *Optics Express, Vol. 29, Issue 3, pp. 3503-3514* **29**, 3503–3514 (2021).

22. Wu, C. *et al.* Low-Loss Integrated Photonic Switch Using Subwavelength Patterned Phase Change Material. *ACS Photonics* **6**, 87–92 (2019).

23. Li, P. *et al.* Reversible optical switching of highly confined phonon–polaritons with an ultrathin phase-change material. *Nature Materials 2016 15:8* **15**, 870–875 (2016).

24. Sui, X., Wu, Q., Liu, J., Chen, Q. & Gu, G. A review of optical neural networks. *IEEE Access* **8**, 70773–70783 (2020).

25. Xu, M., Slovin, G., Paramesh, J., Schlesinger, T. E. & Bain, J. A. Thermometry of a high temperature high speed micro heater. *Review of Scientific Instruments* **87**, (2016).

26. Zhang, Y. *et al.* Broadband transparent optical phase change materials for high-performance nonvolatile photonics. *Nature Communications* **10**, (2019).

27. Ríos, C. *et al.* Ultra-compact nonvolatile photonics based on electrically reprogrammable transparent phase change materials. (2021).

28. [2201.05439] Broadband nonvolatile electrically programmable silicon photonic switches. https://arxiv.org/abs/2201.05439.

29. Feldmann, J. *et al.* Parallel convolutional processing using an integrated photonic tensor core. *Nature 2020 589:7840* **589**, 52–58 (2021).

30. Peserico, N., de Lima, T. F., Prucnal, P., & Sorger, V. J. (2022). Emerging devices and packaging strategies for electronic-photonic AI accelerators: opinion. Optical Materials Express, 12(4), 1347-1351.

Table 1

Table 1 is available in the supplementary files section.

Figures

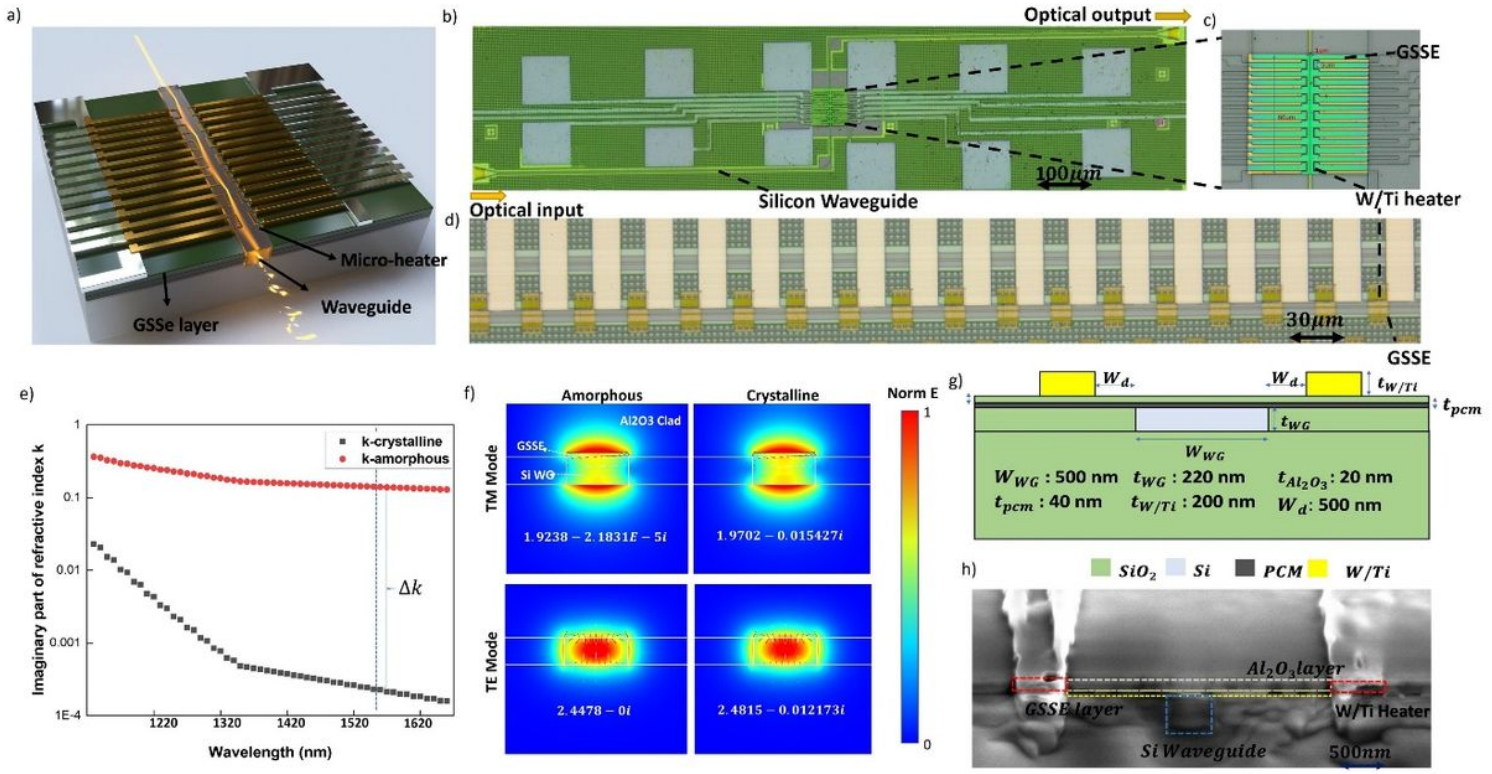


Figure 1

Structure explanation of P-RAM with device optical characterization. (a) 3D schematic of a planarized waveguide with a 30 nm GSSe layer on top of the waveguide and multiple parallel double-sided tungsten-titanium heaters (b). Detailed optical image of GSSe on waveguide with discrete double-sided heaters (c) Zoom-in detailed image of Fig b. Discrete double-sided heaters are arranged along the waveguide over the GSSe film. (d) Detailed optical image of GSSe strip array with single-sided heaters for measurement of high order bit memory. (e) Experimentally obtained (ellipsometry) optical properties of GSSe film. Absorption coefficient contrast (imaginary part of the refractive indices) of GSSe for crystalline and amorphous states. The GSSe shows a strong unity Δk, while simultaneously showing a small, induced loss at the amorphous state. (f) Normalized electric field mode profile of hybrid Si-GSSe waveguide for TE and TM mode with 0.54dB/mm absorption coefficient between amorphous and crystalline state. The imaginary part of the effective refractive index in the amorphous state is which leads to an exceedingly small unit of passive absorption of the memory. (g) 2D cross-section schematic of the lateral thermoelectric switching configuration. One pair of tungsten-titanium heaters is deposited on the side of the Si waveguide on top of the GSSe film. The electrical current running through the W-Ti circuit dissipates energy in the form of local heat, inducing local material phase transition. (h) Cross-section SEM image of the device.

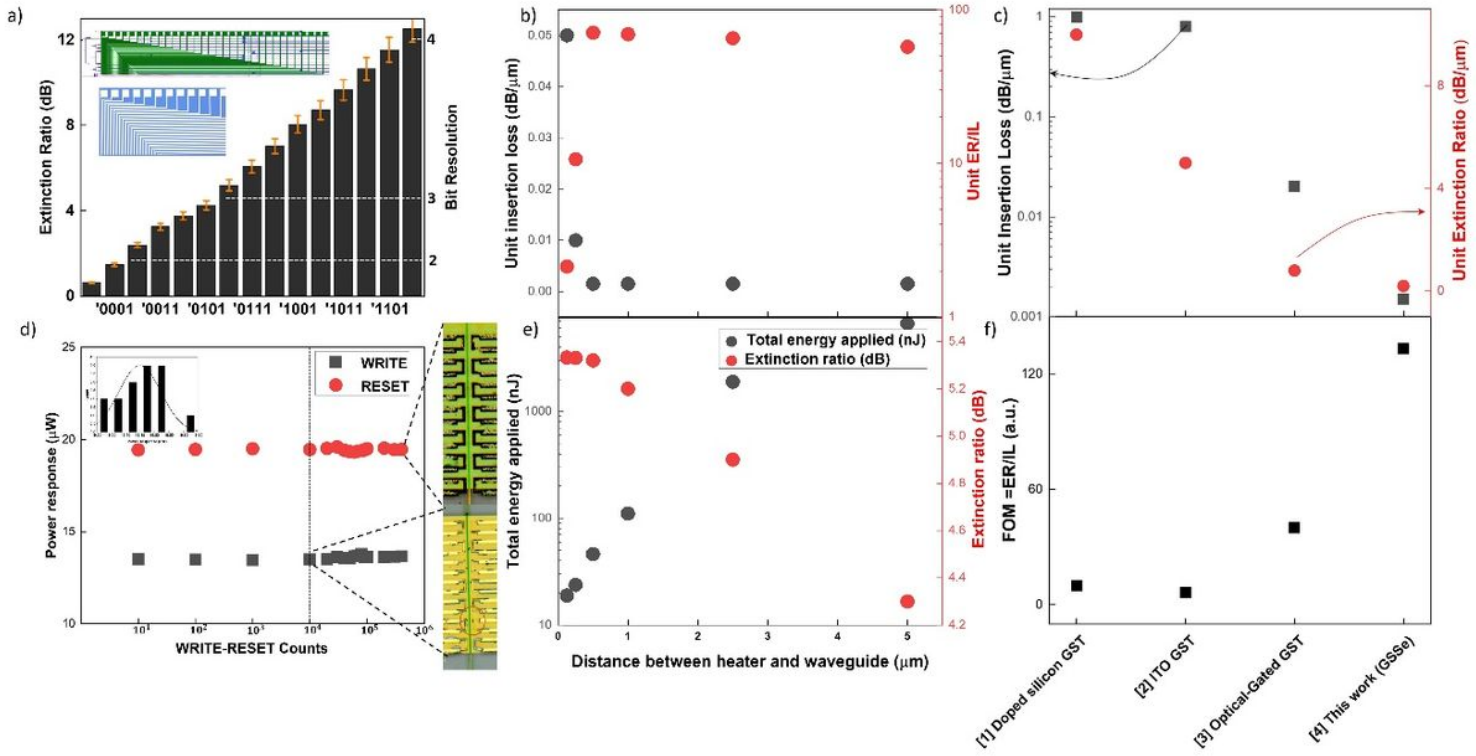


Figure 2

P-RAM performance for bit resolution, energy, cyclability, and FOM. (a) Optical power response for a 4-bit photonic memory as a function of digital states, for an increasing number of crystalline-wire the Extinction Ratio (ER) increases linearly and uniformly. (b) Unit insertion loss (IL) and extinction ratio (ER) per unit insertion loss vs. heater position. (c) Unit insertion loss and unit extinction ratio comparison between PCM-based photonic memories. (d) Bi-State optical responses change exceeding 500,000 switching cycles. For heaters exposed to air with no Al_2O_3 layer protection, the maximum Writing-Resetting cycles achieved is 10,000 and then heaters were broken due to heavy oxidization or physical deformation as shown. With a thick Al_2O_3 layer on top of the heaters, the maximum cycle reached is 500,000 and heaters are still alive as shown in the lower right subfigure. (e) Heater performance vs position of the heater. The distance between the edge of the waveguide and the double heater is swept from 125 nm to 5000 nm. Left axis: Total energy applied vs. heater-waveguide distance for reaching 6-dB extinction ratio. Right axis: With the same applied energy, the ER change corresponds to the heater position. f) Figure of merit comparison for different PIC-based nonvolatile photonic memories.

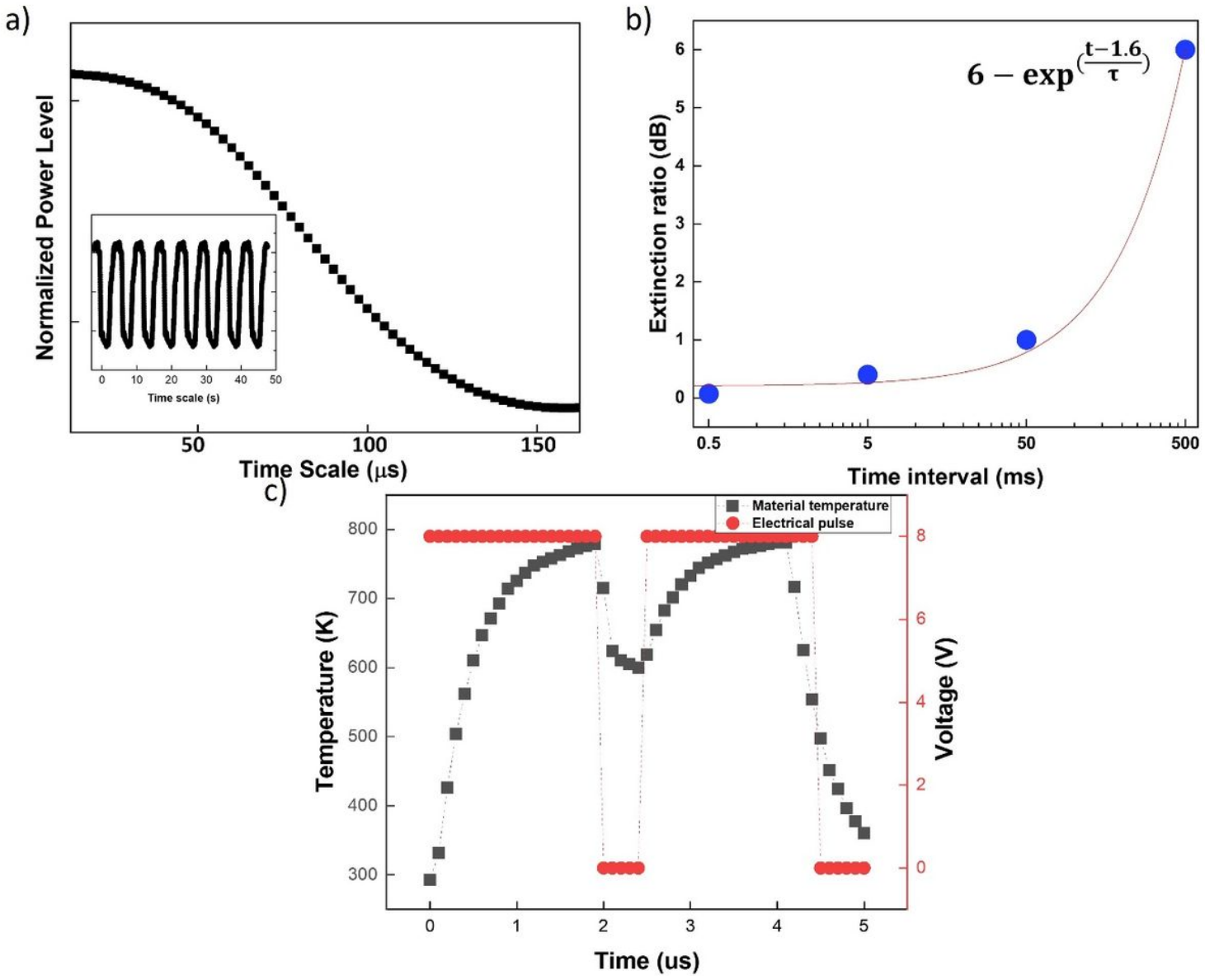


Figure 3

P-RAM speed response and writing pulse set up. (a) Time-dependent trace of transmission with the power response down edge in μs level. (b) The time taken for reaching different levels of extinction ratio varies as shown in the figure from 0.5 ms to 500 ms. (c) Simulated pre-programmed voltage pulses are applied for each heater and a two-sided neighbor works simultaneously for GSSe to transient from amorphous to crystalline state.

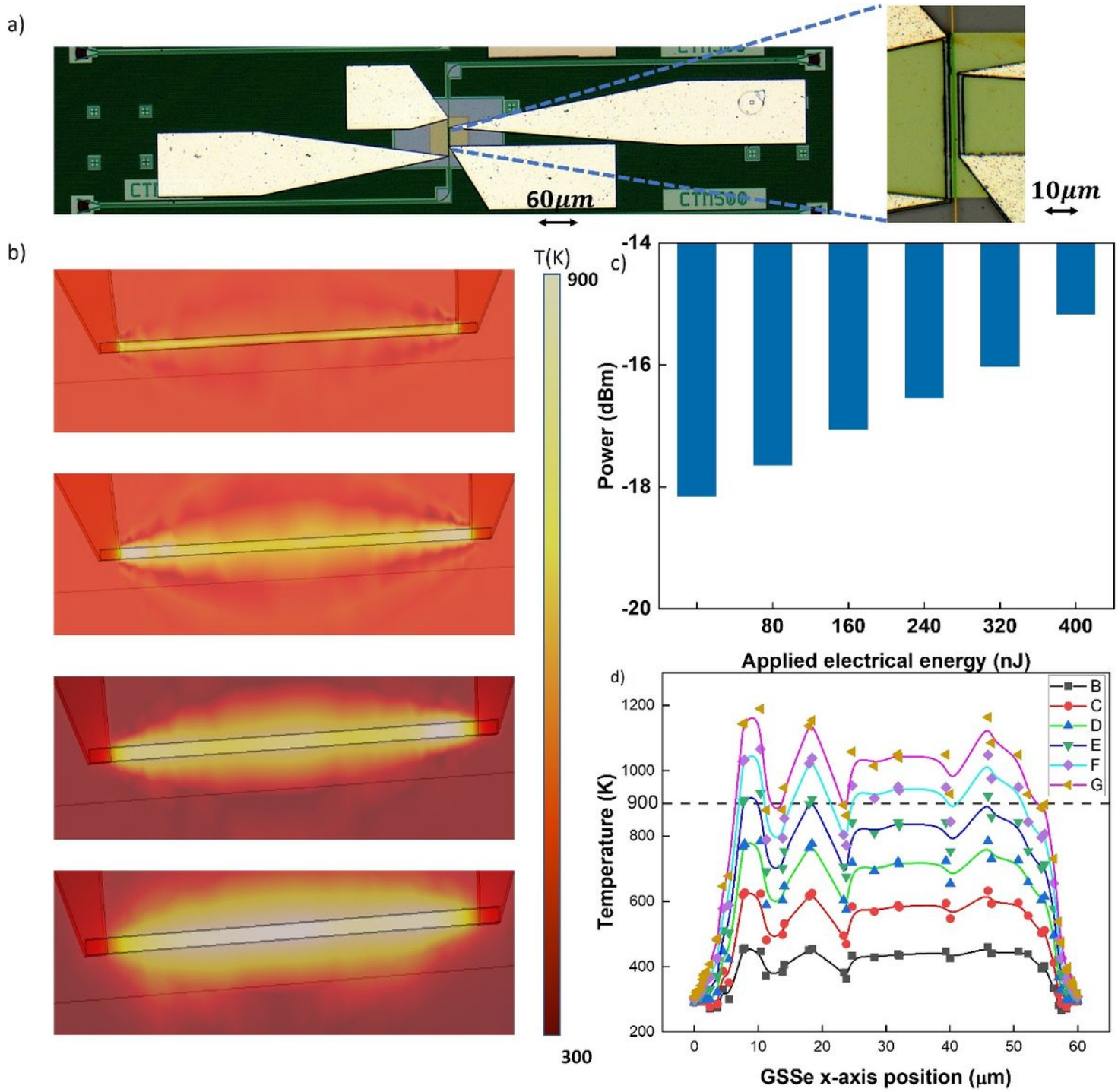


Figure 4

Single heater pair, multi-states power response P-RAM. (a) Schematic of single pair heater multi-level power response device with a different crystalline-amorphous ratio on a single GSSe pad. Detailed optical image of asymmetric paired heaters stands along the waveguide over the GSSe pad. The longer heater on the left side is used as the resetting heater to change the phase of all GSSe film from amorphous to crystalline. The shorter heater on the right side is used as the setting heater to control the area of amorphous state film by applying different levels of Joule-heating energy to the W-Ti microheater. As more energy is applied, a larger area of GSSe film will be transient to the amorphous state and the total

absorption will decrease. Based on different levels of absorption, a multi-level power response function is achieved. (b) COMSOL thermal simulation of temperature distribution along with the heater also indicates the area ratio of GSSe material in the amorphous and crystalline state. (c) Measured 6-states power responses achieved. (d) GSSe material temperature distribution along with the heater from COMSOL thermal simulation with different energy applied to micro-heater.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table1.jpg
- PRAMSOM2022Jiawei.docx