Integrated nonreciprocal magneto-optics with ultra-high endurance for photonic in-memory computing

Paolo Pintus[1,2,*], Mario Dumont[1], Vivswan Shah[3], Toshiya Murai[4], Yuya Shoji[5], Duanni Huang[1], Galan Moody[1], John E. Bowers[1], Nathan Youngblood[3,#]

[1]Department of Electrical and Computer Engineering, Univ. of California, Santa Barbara, CA, USA
[2]Department of Physics, University of Cagliari, Italy,
[3]Department of Electrical and Computer Engineering, Univ. of Pittsburgh, Pittsburgh, PA, USA
[4] Platform Photonics Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Japan
[5]Laboratory for Future Interdisciplinary Research of Science and Technology, Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, Japan
[*]e-mail: ppintus@ece.ucsb.edu , paolo.pintus@unica.it
[#]e-mail: nathan.youngblood@pitt.edu

Abstract

Processing information in the optical domain promises advantages in both speed and energy efficiency over existing digital hardware for a variety of emerging applications in artificial intelligence and machine learning. A typical approach to photonic processing is to multiply a rapidly changing optical input vector with a matrix of fixed optical weights. However, encoding these weights on-chip using an array of photonic memory cells is currently limited by a wide range of material-level and device-level issues, such as the programming speed, extinction ratio, and endurance among others. Here, we propose a new approach to encoding optical weights for in-memory photonic computing using magneto-optic memory cells comprised of heterogeneously integrated cerium-substituted yttrium iron garnet (Ce:YIG) on silicon micro-ring resonators. We show that leveraging the non-reciprocal phase shift in such magneto-optic materials offers several key advantages over existing architectures, providing a fast (1 ns), efficient (143 fJ per bit), and robust (2.4 billion programming cycles) platform for on-chip optical processing.

Introduction

The growing divide between the demand for computing resources and the performance of digital hardware necessitates the development of post-CMOS architectures that can achieve ultra-high computational throughput at ultra-low energies. An extreme example of this comes from the field of deep learning, where the computation required to train state-of-the-art deep neural networks grew by over 300,000× between 2015 and 2020, doubling every 3.4 months [1,2]. Over the same period, the computing performance of CMOS microprocessors has started to reach saturation, encountering the limitation of the maximum power dissipated per unit area [3,4], while the efficiency of graphics processing units (GPUs) has grown by only 300× [5]. Additionally, for emerging applications, such as high-speed qubit classification, plasma control in fusion reactors, real-time signal processing, and autonomous navigation, processing data with ultra-low latencies is highly challenging using conventional digital approaches that rely on sequential arithmetic operations [6]. Although analog in-memory computing offers a solution to the problem of sequential operations

for matrix multiplication [7], its latency increases with the memory array size due to the capacitance of metal interconnects [8].

Processing information with light offers a compelling alternative to conventional computing because optical interconnects, such as fibers and waveguides, lack capacitance. This absence of capacitance eliminates the energy-bandwidth tradeoff that restricts the data transfer rate of electrical interconnects and enables exceptionally high bandwidth density when combined with wavelength division multiplexing. Additionally, by breaking free from the confines of binary logic and encoding, linear operations (e.g., convolutions, matrix multiplications, Fourier transforms, random projections, etc.) can be reduced to a single optical transmission measurement with extremely low energy consumption—even less than one photon per multiply-accumulate operation [9,10]. These combined advantages of analog computing in the optical domain enable the dramatic scaling necessary for continued innovations in artificial intelligence (AI) and machine learning—both in terms of compute density (operations per chip area) and energy efficiency (operations per watt).

Many distinct photonic architectures have been proposed and demonstrated that attempt to address the major bottleneck facing the field of AI and computing, namely matrix-vector multiplication (MVM). The vast majority of these approaches can be classified as "weight-stationary" photonic processors [11] where a matrix of programmable optical weights—typically encoded in a 2D array of nonvolatile memory elements or optical modulators [12–19]—is used to perform a linear transformation on a vector of optical inputs. This design approach has the benefit of performing computation in the memory array itself, not unlike analog computing in the electrical domain using crossbar arrays of resistive random access memory (RRAM) [20], which can significantly reduce data movement and latency, while improving energy efficiency.

However, a crucial (yet seldom mentioned) limitation of these weight-stationary approaches is the time and energy required to update the fixed weights of the matrix to implement useful computing algorithms. Unlike electronic crossbar arrays which aim to fully map the entire neural network into analog weight banks to minimize reprogramming of the array, optical weight banks have much lower storage density (e.g., ~0.01 bits/μm^2 [21]) and can only store a tiny fraction of the necessary parameters on-chip (e.g., a 16×16 phase-change memory array [22] or 64×64 Mach-Zehnder interferometer array [12]). Thus, to accommodate the large number of parameters required by real-world applications, the photonic weight bank must be reprogrammed many times for each matrix operation that exceeds the dimensions of the weight bank. This requires photonic memory cells which can be deterministically programmed quickly, efficiently, and with high endurance or they will dramatically reduce throughput and energy efficiency of the entire system [11,23,24].

To address these challenges, we propose a resonance-based photonic architecture (inspired by the "broadcast and weight" design [13]) which leverages the non-reciprocal phase shift in magneto-optical materials to implement photonic in-memory computing. In this architecture (**Figure 1a**), a y-splitter excites both the clockwise (CW) and counterclockwise (CCW) propagating modes of a microring resonator (MRR) with a magneto-optic cladding layer (Cerium-substituted Yttrium Iron Garnet or "Ce:YIG"). The interaction of the optical mode with Ce:YIG causes a non-reciprocal

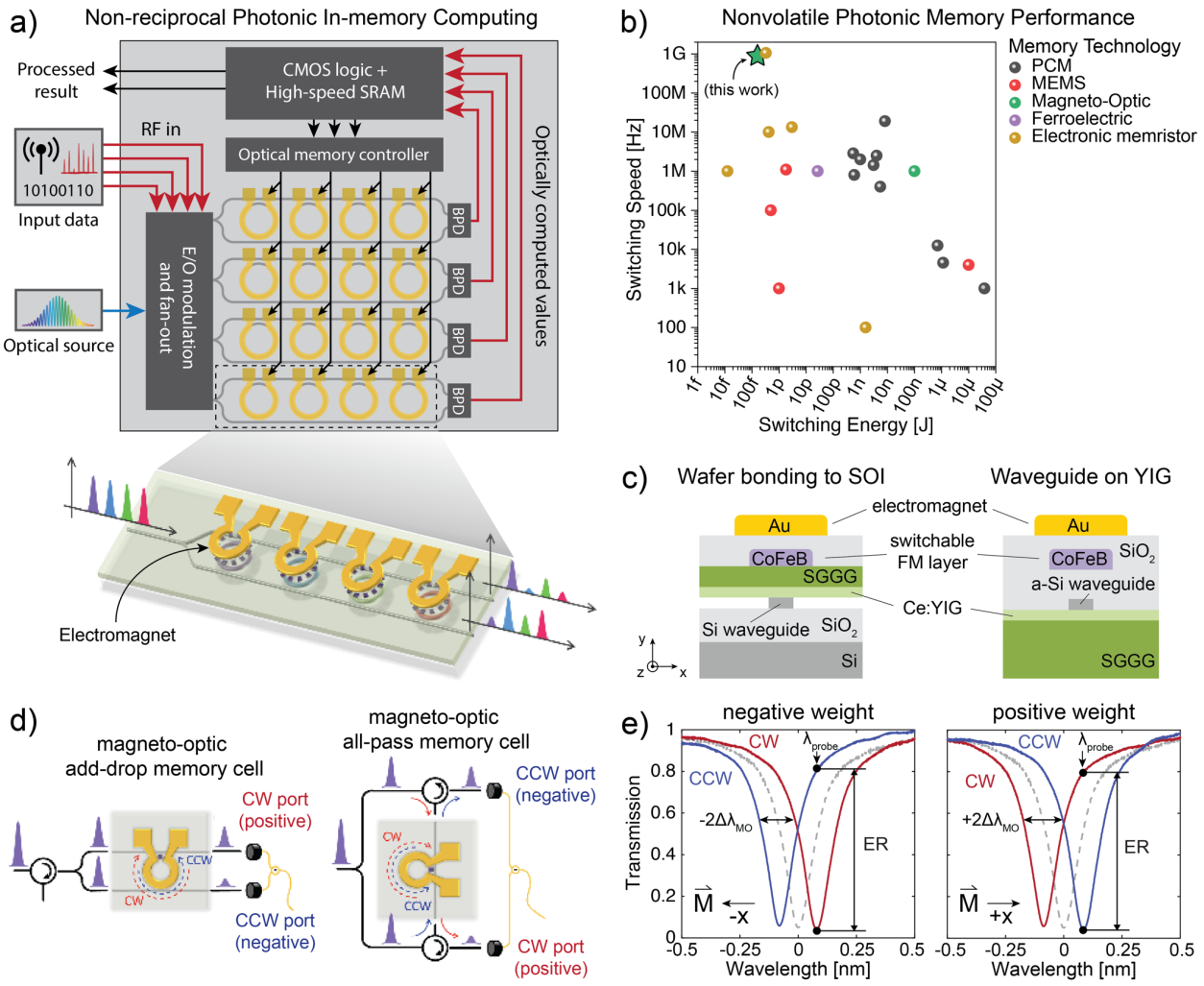

Figure 1: Non-reciprocal photonic in-memory computing. a) Our vision for a non-reciprocal photonic computing platform which leverages high-speed magneto-optic memory arrays to enable fast weight updates. A single row from our proposed architecture is illustrated below which leverages non-reciprocal memory cells to compute a dot-product optically. **b)** Speed versus energy comparison of state-of-the-art nonvolatile photonic memory technologies which have been demonstrated on chip. **c)** Cross section of non-reciprocal magneto-optic memory. Either wafer bonding and chemical mechanical polishing (left) or growth and patterning of amorphous silicon on Ce:YIG substrates (right) can be used to heterogeneously integrate Ce:YIG with silicon waveguides. In both cases, high-quality Ce:YIG is guaranteed by growing the garnet on a native substrate of substituted gadolinium gallium garnet (SGGG). **d)** Illustration of two implementations used to demonstrate magneto-optic memory cells where both counter-propagating modes are excited in the MRR. **e)** Positive and negative weights are encoded by switching the magnetization direction and amplitude, which results in opposite resonance shifts for the clockwise (CW) and counterclockwise (CCW) counter-propagating modes.

phase shift for the two counterpropagating modes which is visible as a split resonance shift with opposite signs. The direction of the non-reciprocal phase shift is dependent on the sign of an applied external magnetic field which allows both positive and negative weights to be encoded (**Figure 1e**). A circulator is used to prevent feedback to the optical source for the case of an add-drop MRR or, in the case of an all-pass MRR, a pair of circulators are used (see **Figure 1d**). When

scaled to an entire array of add-drop MRRs, only one broadband isolator at the input is necessary. While in the experimental characterization of the memory cells we used fiber optic circulators, the proposed magneto-optic platform offers the distinct advantage of seamlessly integrating broadband optical isolators on chip [25–28] in parallel with the non-reciprocal memory array.

This approach has several benefits. First, by using the non-reciprocal effect in magneto-optic materials, we can efficiently achieve ~1 GHz programming speeds, nonvolatility, and multi-level encoding as we demonstrate in this work (see **Figure 1b** and **Table 1** for a comparision with state-of-the-art). Secondly, the cycling endurance of magnetic-based memory is known to be orders of magnitude greater than competing technologies [29], an outstanding challenge for many nonvolatile optical memories [30]. Finally, unlike the case of add-drop MRR weights based on reciprocal optical effects [13] (e.g., thermo-optic or plasma-dispersion effects), the differential signal is measured from the through port transmission for both the CW and CCW modes, improving the symmetry and extinction ratio for both positive and negative weighting.

Device Design

Integration of the Ce:YIG layer with the photonic circuit can be achieved either through wafer bonding and polishing [31] or by deposition and patterning of an amorphous silicon layer [32] as illustrated in **Figure 1c**. Programming the state of the memory cell requires a radial in-plane magnetic field which is supplied by an integrated Au electromagnet. To maintain a nonvolatile state without dissipating power, a ferromagnetic thin-film (CoFeB) patterned into an array of bar magnets can be integrated with the electromagnet on-chip [32,33]. **Figure 1e** shows example spectra from a non-reciprocal memory cell under a negative (left) and positive (right) radial applied magnetic field. In the case of positive magnetization ($M_x > 0$), the forward propagating CCW mode (i.e., same propagation direction as current flow in the electromagnet) red shifts while the CW mode blue shifts. If the optical probe is red detuned from resonance when $M_x = 0$, the resulting differential transmission encodes a negative weight. The opposite is true for negative magnetization ($M_x < 0$). For a critically coupled MRR, this approach can achieve high transmission contrast, limited by the extinction ratio of the CW and CCW modes.

The functionality of our non-reciprocal memory cell can be extended beyond the single dot-product shown in **Figure 1a** to MVM operations. In **Figure 2** we compare two broadcast and weight architecture designs featuring non-reciprocal (**Figure 2a**) and reciprocal (**Figure 2b**) MRR-based weights. Here, the matrix operation $\boldsymbol{W}\vec{x} = \vec{b}$ is achieved through fan-out of the optical input vector to each row of $\boldsymbol{W}$. This input vector is then multiplied by the wavelength-dependent transmission of each resonator in the row before being summed through incoherently combining the transmitted power at each row's pair of output waveguides as originally proposed by Tait et al. [13]. An important distinction between the two approaches can be visualized in the transmission spectra of the bus waveguides. While the differential photocurrent compares the CW and CCW through ports in **Figure 2a**, the differential transmission of the through and drop ports is used to compute $\vec{b}$ in **Figure 2b**. This is an important distinction since the drop port of the reciprocal MRR weight reaches its maximum extinction ratio at a phase shift of π when the optical probe is centered at resonance. Thus, to achieve high optical contrast between the through and drop ports (i.e., to

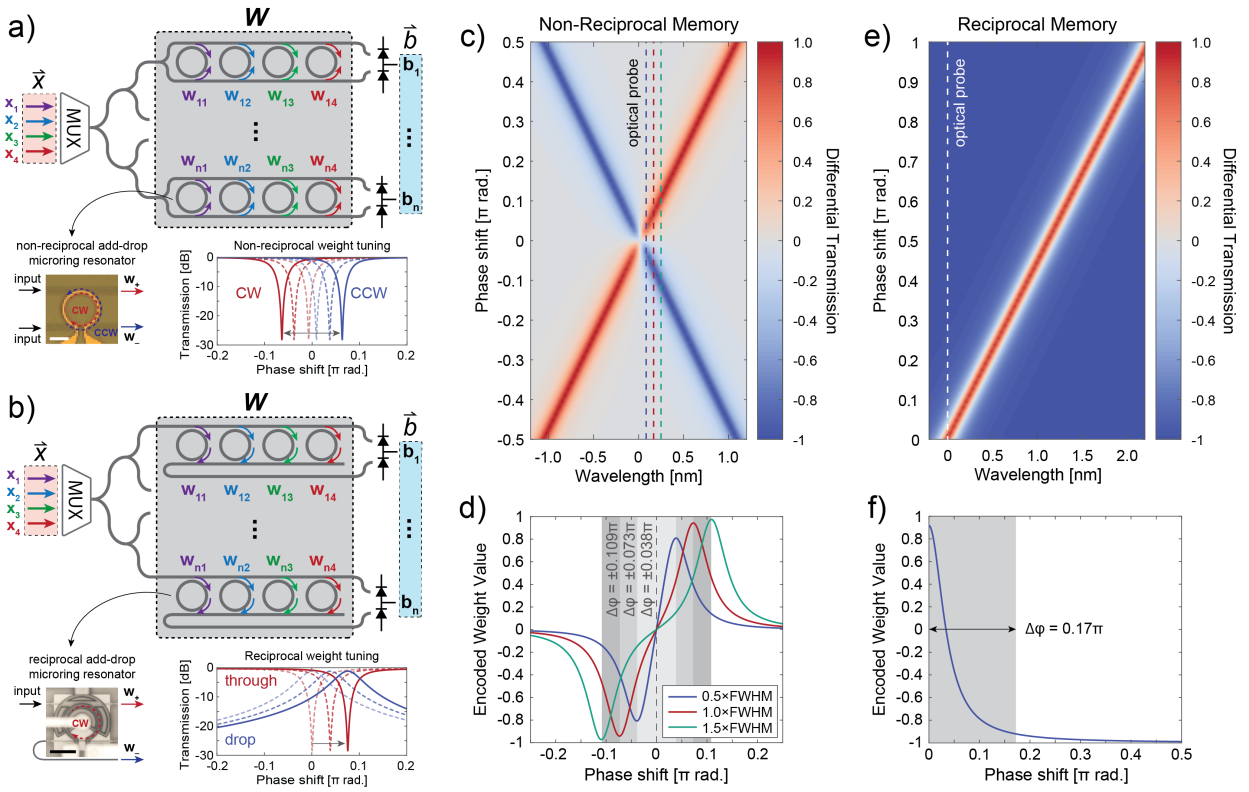

Figure 2: Reciprocity in photonic computing. a) Non-reciprocal photonic computing platform leveraging an integrated magneto-optic memory array. Matrix-vector multiplication is achieved by taking the differential transmission between the clockwise (CW) and counter-clockwise (CCW) propagating modes. **b)** Reciprocal "Broadcast and weight" architecture which uses the difference between the through and drop ports of an add-drop microring resonator to encode the values of matrix ***W***. **c)** Simulated map of differential through-port transmission for a non-reciprocal memory cell with Q = 10,000. **d)** Encoded weight value of non-reciprocal memory cell for an optical probe spaced 0.5, 1.0, and 1.5 FWHM away from resonance (dashed lines in **d**). Notably, the encoded value is an anti-symmetric function centered at zero phase-shift. **e)** Simulated map of the difference between the through and drop ports of a reciprocal memory cell with Q = 10,000. **f)** Encoded weight value of reciprocal memory cell when the optical probe is held at resonance (white dashed line in **e**). A larger phase shift is needed to achieve symmetric weighting compared to the non-reciprocal case.

improve bit-precision of the weight), a much larger phase shift is required in the case of a reciprocal MRR compared to a non-reciprocal MRR. While this may not be a limiting factor for a single memory cell, reducing the phase shift required to achieve the full range of positive and negative weight values in the array reduces the optical crosstalk between neighboring resonances [34] and alleviates the challenge of achieving strong optical modulation on-chip.

This distinction is highlighted in **Figure 2c-f** where we simulate the differential transmission of both reciprocal and non-reciprocal optical memory with the same quality factor (Q = 10,000). In **Figure 2c**, we see that the differential transmission for both positive and negative values is an anti-symmetric function centered at $\varphi = 0$. This function is shown in **Figure 2d** for three different optical probe wavelengths: 0.5×, 1.0×, and 1.5× full-width half maximum (FWHM) linewidths red

detuned from resonance when $\varphi = 0$. As the detuning of the probe increases, the maximum and minimum weight values increase in magnitude, while the linearity of the weighting function near $\varphi = 0$ decreases. The differential transmission for the case of a reciprocal memory cell is shown in **Figure 2e**. The reciprocal weighting function is symmetric and centered at $\varphi = 0$, requiring a resonance shift of ~0.5×FWHM to reach negative values when the probe is centered at resonance (**Figure 2f**). Thus, to achieve an equal range of positive and negative weights (i.e., minimal power penalty [34]), a phase shift of $\Delta\varphi = 0.17\pi$ is required for an ideal reciprocal MRR with the same quality factor of Q = 10,000.

Modelling and Experimental Results

The transmission spectra of magneto-optic memory cell is modeled using the transfer matrix method [35] where the effective index of the CW and CCW modes is computed using a finite element method [36]. The resonance of the MRR is controlled by the current in the integrated electromagnet, giving rise to a magnetic field and a Joule heating effect. The magneto-optic and thermo-optic effect alter the effective index of the modes, and their impact is described using a perturbative approach. A comprehensive description of the model employed is provided in Section 1 of the Supplementary Material.

The modeling results of **Figure 2** were experimentally verified for a MRR with Ce:YIG (Q ≈ 10,000) and a gold integrated electromagnet. For this demonstration, the ring radius is 35 μm and the waveguide cross-section is a 600 nm × 220 nm silicon ridge with a 400 nm thick top-cladding of Ce:YIG. A 10 nm thin silicon oxide layer separate the Silicon from the Ce:YIG layer (see Supplementary Materials for more details). **Figure 3a** shows the optical transmission spectra for the fundamental transverse magnetic (TM) CW and CCW propagating modes as a function of applied current. The resonance position of the spectra shows both a linear and quadratic dependence on the applied current corresponding to the magneto-optic and thermo-optic effects, respectively. In **Figure 3b-c**, we separate the resonance shifts of the CW and CCW modes into their non-reciprocal (magneto-optic) and reciprocal (thermo-optic) components. While the thermo-optic effect red-shifts both CW and CCW spectra, the magneto-optic effect induces shifts in opposite directions for the two modes. For a set current, the thermo-optic shift is estimated as the average shift of the two spectra compared to the case of no current. The magnitude of the magneto-optic shift is half of the measured resonance split between the CW and CCW modes. In **Figure 3b-c**, we overlay the mathematical model with the measurement results, showing an excellent agreement between theory and experiments. **Figure 3d** plots the FWHM linewidth of the resonator for the CW and CCW modes, showing a similar quality factor for both modes. By varying the applied current in the electromagnet, we observe slight changes in the extinction ratio in **Figure 3a** and linewidth in **Figure 3d.** These variations are caused by the non-reciprocal loss in the Ce:YIG, where the optical loss changes depending on the direction of light propagation and the transverse magnetic field [36–38]. For more details, please refer to Section 1 of the Supplementary Materials.

In **Figure 3e**, we plot the differential optical transmission of the CW and CCW modes for three different optical probe wavelength positions: 0.5×, 1.0×, and 1.5× FWHM linewidths red detuned from resonance when $\varphi = 0$ (i.e., no magnetic field). As expected, there is good overlap between

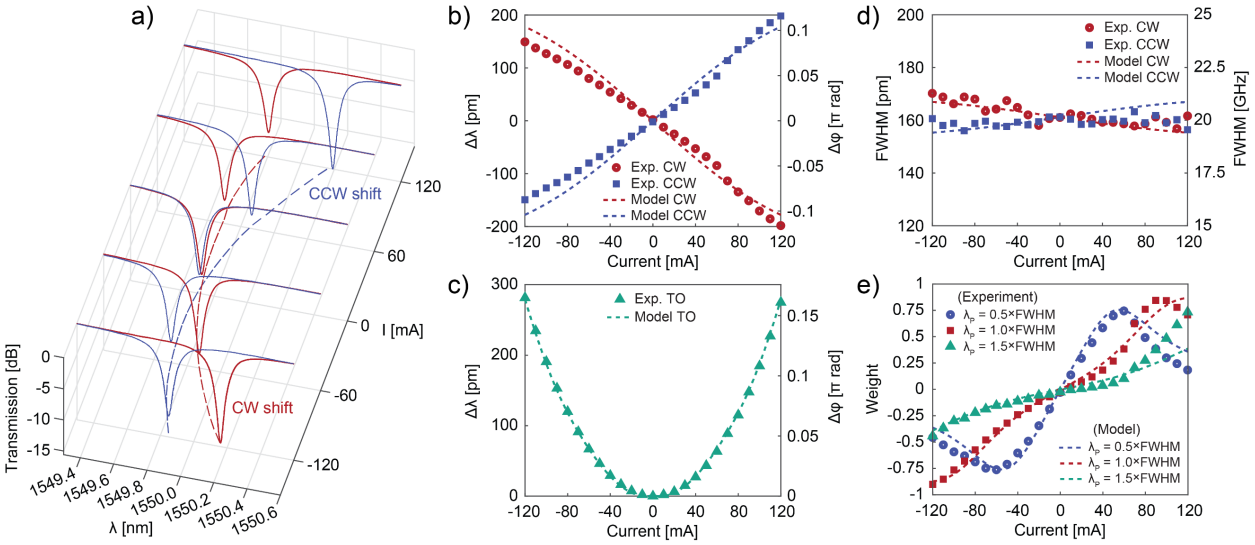

Figure 3: Experimental characterization of non-reciprocal optical memory. a) Spectral response of a non-reciprocal magneto-optic memory cell for different fixed currents (Q ≈ 10,000). Both linear and quadratic resonance shifts are observed due to the magneto-optic and thermo-optic effects, respectively. **b)** Magneto-optic spectral shift for the CW and CCW modes. Dashed lines correspond to our analytical model of the expected magneto-optic phase shift excluding thermal effects. **c)** Thermo-optic phase shift extracted from **a)** due to heating of the electromagnet while a constant current is applied. **d)** Extracted resonance linewidth as a function of applied current for the CW and CCW modes. Slight changes in the non-reciprocal loss for different propagation directions change the quality factor of the MRR. **e)** Encoded weight for a probe wavelength red detuned 0.5×, 1.0×, and 1.5× FWHM away from the central resonance when no magnetic field is applied. Experimental weight values are in good agreement with the modeled results in **Figure 2d**.

the experimental results and the simulated weighting functions for the case of an ideal resonator with similar quality factor (**Figure 2d**). Again, we see that increasing the red detuning of the optical probe increases the dynamic range of the weight function due to reduced insertion loss while also decreasing the linearity between the maximum and minimum weight values. More details on the modelling of the weight function is reported in Section 1 of the Supplementary Material. At larger current values, we observe a deviation from the expected weight value. We attribute this to the fact that for large detuning, the two modes have different FWHM, resulting in a distinct weighting for the two directions. This is due to the minor wavelength dependence for the coupling between the resonator and bus waveguide, which can be mitigated by designing adiabatic couplers or reducing the detuning of the probe wavelength from resonance.

Having characterized the steady-state response of our non-reciprocal optical memory, we next demonstrate high-speed weight updates through characterization of the memory cell's dynamic response. For high-speed characterization beyond ~1 MHz, the magneto-optic response dominates while the dynamic thermo-optic response becomes negligible. This can be seen in **Figure 4a** where the dynamic optical transmission of a CW probe for two current pulses can be observed at different time scales (1 ms versus 10 ns pulse width). In the case of a slow 1 ms current pulse, the optical transmission includes both a blue shift (increase in transmission) from the fast magneto-optic response and a red shift (decrease in transmission) from the slow thermo-optic response. For the case of a fast 10 ns current pulse and red detuned probe, the slow decay from the thermal response

of the ring disappears (rise time ~50 μs and fall time 92 μs), and we only observe the fast magneto-optic response with a rise and fall time of less than 1 ns, and a ferromagnetic resonance of 0.55 GHz. Since the estimated time response of the integrated electromagnet circuit is 6 ps, the ringing observed in the optical response is attributed to magneto-optic response of the Ce:YIG, resulting in a rise/fall time of 0.95 ns. More details are reported in Section 3 of the Supplementary Material.

Due to the high-speed magneto-optic response and the soft in-plane magnetic axis of the Ce:YIG, the memory cell can be programmed with low energy. **Figure 4b-c** show an open eye diagram measured for both the CW and CCW optical probe for 500 Mbps and 1 Gbps modulation speeds (see Section 2 and Figure S5 of the Supplementary Material for details of the measurement setup). From **Figure 4b**, we see that for a weight update rate of 500 Mbps (i.e., 2 ns pulse width), we can

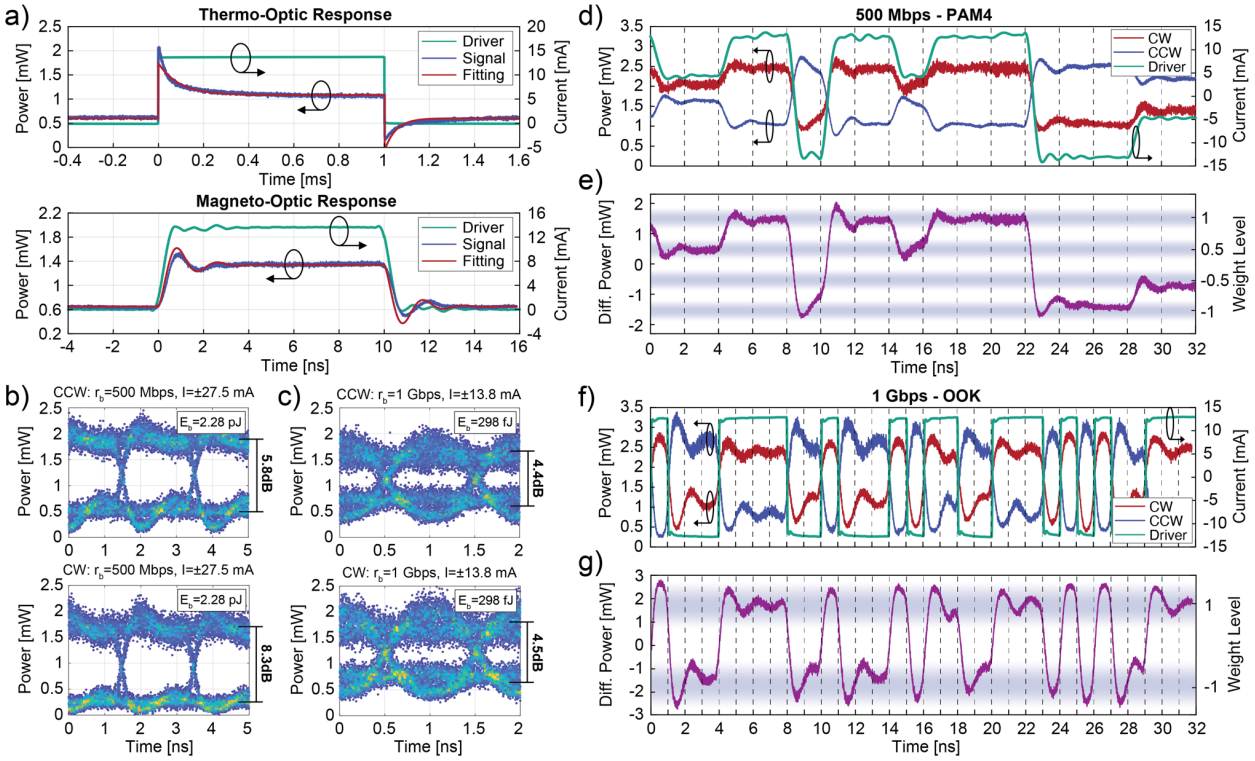

Figure 4: Dynamic response of non-reciprocal optical memory. a) Comparison of dynamic response of thermo-optic and magneto-optic effects demonstrating five orders of magnitude difference in response time (red detuned CW probe used for both measurements). From the thermo-optic response, the fall time is 92 μs while the rise time is 50 μs. For high-speed modulation above ~1 MHz, the thermo-optic effect is negligible provided the average power dissipated remains constant. The magneto-optic response can be fit with a second order response with a natural angular frequency of 3.6 Grad/s and a dimensionless damping factor equal to 0.29. From these results, we estimate a rise/fall time of 0.95 ns and a ferromagnetic resonance of 0.55 GHz (see Supplementary Material for more details). **b-c)** Eye diagrams for clockwise and counterclockwise propagating modes for pseudorandom binary sequence (PRBS) modulation at **b)** 500 Mbps and **c)** 1 Gbps speeds. **d)** Simultaneous measurement of CW and CCW transmission for PAM4 modulation at 500 Mbps. **e)** Plot of differential optical power between CW and CCW signals demonstrating the ability to rapidly update non-reciprocal multi-bit optical weights. **f)** Simultaneous measurement of CW and CCW transmission for binary modulation (On-Off Keying) at 1 Gbps. **g)** Differential optical power between CW and CCW signals demonstrating programming speeds up to 1 ns.

achieve an open eye diagram with an extinction ratio as high as 8.3 dB for a programming energy as low as 2.28 pJ. Reducing the pulse width to 1 ns and amplitude by 0.5×, reduces the programming energy to a mere 298 fJ as shown in **Figure 4c**. This corresponds to a ~8× improvement in energy efficiency since $E_b = I^2R \cdot \Delta t + LI^2/2$, where I is the programming current, R is the resistance of the electromagnet (~1.43 Ω), L is the inductance (0.3 nH), and Δt is the duration of the programming pulse. It is also worth noting that for our non-reciprocal memory cell, we achieve significant modulation of the optical signal with a peak voltage as low as ±21.5 mV for ±13.8 mA current pulses, representing an extremely low programming voltage which is compatible with the most advanced CMOS nodes.

In **Figure 4d-e**, we demonstrate multi-level optical weighting using a 4-level (2-bit) pulse amplitude modulation (PAM4) programming signal with maximum current amplitude of ±13.8 mA, corresponding to a record low programming energy of 143 fJ per bit. In these experiments, we capture the transmission of the CW and CCW modes simultaneously using the experimental setup described in Supplementary Figure S6. The differential transmitted power between the CW and CCW modes is shown in **Figure 4e** where we clearly observe four distinct transmission levels, allowing us to achieve to two positive and two negative optical weights given our 2-bit electrical input. Using the high-speed magneto-optic effect, we can achieve optical weight updates as fast as 1 ns as shown in **Figure 4f-g**. In our case, the maximum programming speed is limited by the ferromagnetic resonance of the Ce:YIG. Although in this device the maximum time response is 1 GHz, a much faster programming speed can be reached using other magneto-optic material system that can support modulation rate of tens of gigahertz [39,40].

In a final demonstration, we show the nonvolatile response of our non-reciprocal memory cell when integrated with a switchable ferromagnetic layer. In this experiment, patterned CoFeB magnetic stripes are integrated in the cladding above the MRR which provides a programmable, nonvolatile magnetic field. The shape anisotropy and orientation of the micron sized CoFeB bar magnets provides the static radial magnetic field needed to induce a non-reciprocal optical phase shift in the memory cell [32]. In this demonstration, the silicon MRR has a cross-section of a 1000 nm × 220 nm with a radius of 50 μm. Our measurement setup and device cross section are shown in **Figure 5a**. To program the state of the memory cell, current is applied to the integrated electromagnet which encodes the magnetic field strength in the CoFeB magnetic domains. After the current is removed, the laser is swept to obtain the nonvolatile MRR spectral shift of the CW mode (spectra shown in **Figure 5c-e**). For an optical probe positioned at the resonance dip of the MRR (dashed line in **Figure 5c-d**), we observe the hysteretic behavior in the optical transmission shown in **Figure 5b**. When the programming current is increased from negative to positive values, we observe an increase in optical transmission for currents greater than 0 mA (red points in **Figure 5b**). The transmission eventually saturates above 200 mA and remains constant due to saturation of the CoFeB magnetic layer (see illustration in **Figure 5a**). As we change the direction of the applied current from positive to negative values, the transmission remains constant until the negative current values, highlighting the nonvolatile response of the memory cell. The results in **Figure 5b** show at least 11 distinct optical transmission levels corresponding to a nonvolatile memory cell capable of storing ~3.5 bits. However, this value is limited by our experimental setup rather than the device itself. In Section 4 of the Supplementary Materials, we provide a theoretical

analysis of the magnetic, thermal, and optical noise and expect the maximum bit precision of our device to exceed 13 bits.

The intermediate spectra obtained for a sweep from negative to positive current values is shown in **Figure 5c** while the spectra for the reverse sweep from positive to negative values is shown in **Figure 5d**. The dashed line in **Figure 5c-d** shows the position of the optical probe used to plot the hysteresis of the through port transmission (**Figure 5b**). At this probe wavelength, we observe a maximum extinction ratio of 16.2 dB between the minimum and maximum nonvolatile states. To better visualize the spectral dependence on programming current, we sequentially plot the optical

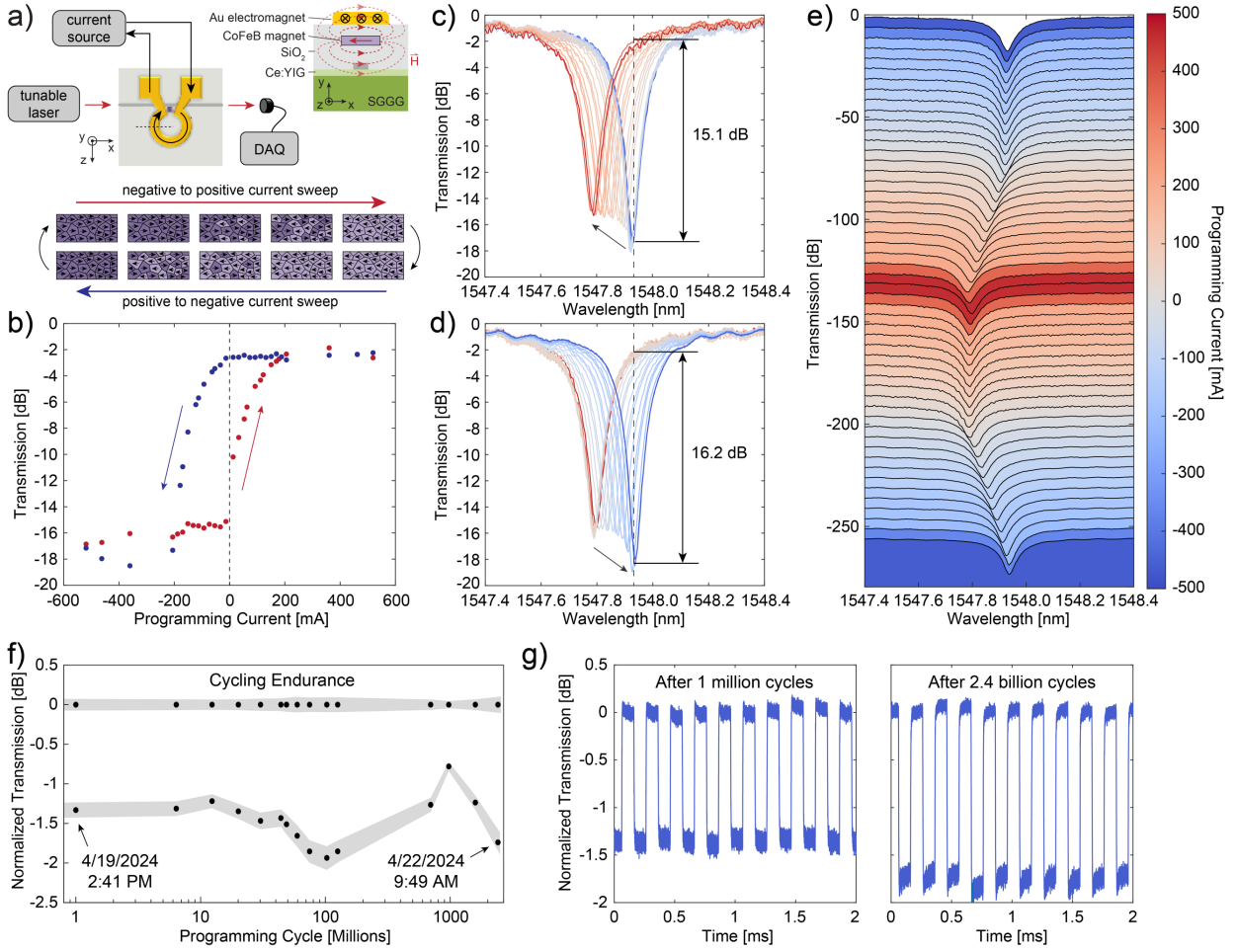

Figure 5: Nonvolatile magneto-optic storage. a) Experimental setup for nonvolatile weight encoding. The magnetic field from the Au electromagnet aligns the magnetic domains of the CoFeB ferromagnetic layer according to the amplitude and polarity of the applied current. **b)** Hysteresis of optical transmission for a non-reciprocal memory cell with CoFeB bar magnets integrated on-chip. For these transmission measurements, the probe wavelength was fixed (black dashed lines in **c** and **d**). **c)** Negative to positive programming current sweep showing a blue shift for positive currents. **d)** Positive to negative current sweep showing a return back to the original resonance position. **e)** Waterfall plot of transmission spectra for sequential programming currents sweeping first in the positive and then in the negative directions. **f)** Demonstrated cycling endurance for over 2.4 billion write and erase cycles. Time domain measurements in **g)** show the operation of the memory cell without observing any degradation. Variation in the observed extinction ratio is attributed to thermal drift of the unpackaged device over the 3-day measurement.

transmission spectra for the full current sweep as a waterfall graph in **Figure 5e**. As the current increases from negative to positive values, the MRR resonance blue shifts until saturation. Reversing the current from positive to negative values causes a red shift back to the original resonance position.

To demonstrate the ultra-high cycling endurance of our non-reciprocal memory cell, we programmed an arbitrary function generator to cycle between write and erase pulses at a rate of 10 kHz, with an amplitude of ±5V, and a pulse width of 500 ns. The optical transmission was captured manually during the 3-day experiment and is shown in **Figure 5f** (see Supplementary Materials for more experimental details). After 2.4 billion write and erase cycles, the device continued to function without any sign of degradation (**Figure 5g**). This is more than three orders of magnitude improvement over prior photonic memory technologies, highlighting the benefit of using optically-coupled magnetic media for non-volatile data storage. While we do see a variation in the extinction ratio of the device in **Figure 5f-g**, we attribute this to slight thermal and mechanical drift of the unpackaged device during the 3-day experiment. To confirm the long-term nonvolatile stability of our device, we also compared the nonreciprocal spectral shift of the CW and CCW modes after programming and see data retention over a 4-day measurement (see Section 2.2 of the Supplementary Materials). Opto-electronic packaging with active thermal control is expected to counter drift of the MRR resonance peak which would address this observed variation in extinction ratio.

Conclusion

We have demonstrated the first instance of a nonvolatile magneto-optical memory cell that features non-reciprocity for in-memory computing in the optical domain. This unique combination of fast, fatigue-free programmability with nonvolatile weights addresses current limitations of existing integrated approaches to optical memory which have yet to combine (1) nonvolatility, (2) multi-bit storage, (3) high switching speed, (4) low switching energy, and (5) high endurance in a single platform. Additionally, leveraging the inherent non-reciprocity of the magneto-optic effect in Ce:YIG enables symmetric and high contrast encoding of both positive and negative optical weights, alleviating the constraints of amplitude-only weight encoding.

In **Table 1** we compare state-of-the-art demonstrations of various nonvolatile photonic memory technologies capable of being integrated on-chip. It is worth noting that while some recent demonstrations of waveguide-integrated electronic memristors (termed “memresonators”) have shown great promise for fast and efficient nonvolatile switching between two or three states [41,42], the maximum cycling endurance of such devices has not been demonstrated beyond 1,000 write-erase cycles. These devices are also not well suited for photonic computing applications where multi-level, nonvolatile storage is necessary. While phase-change memory has shown great promise for optical computing platforms due to its compact footprint, multi-level storage, and ease of integration [19,43–45], the limited cycling endurance, high switching energies and limited switching speeds remain outstanding challenges. Other nonvolatile photonic memory cells leveraging charge trapping, MEMS, or ferroelectric materials have sub-MHz programming speeds which are not yet practical for computing applications due to significant write latencies.

Memory Technology	Switching Speed	Switching Energy (J/bit)	Bit Precision	Footprint (μm^2)	ER[b] (dB)	IL (dB)	Switching Cycles	Ref.
Trapped charge	0.6 s	Set: 11.4 pJ Reset: 17.2 pJ	4-bits	314	13	2	>30	[46]
MEMS	10 µs	500 fJ	1-bit	9,650	37	0.13	100	[47]
Ferroelectric (HZO)	>1 s[(a)]	--[(a)]	1-bit	~2,000,000	40	2.75	--	[48]
Ferroelectric (BTO)	1 µs[(b)]	30 pJ	3.5-bits	>20,000	12	>0.07	300	[49]
PCM	408 ns 220 µs	Am: 5.55 nJ Cry: 1.15 µJ	3.5-bits	4.73	~3	~1	1,500	[50]
PCM	50 ns 200 ns	Am: 8.8 nJ Cry: 6.9 nJ	4-bits	25	4.13	1.78	100	[44]
Memristor	1 µs	13 fJ	1-bit	2	9.2	25	1000	[51]
Memristor	100 ns	Set: 1.3 pJ Reset: 400 fJ	1-bit	314	~12	~5	800	[42]
Memristor	0.3 ns 0.9 ns	Set: 150 fJ Reset: 360 fJ	1.5-bit	314	~14	~2	1,000	[41]
Magneto-optic	**1 ns**	**143 fJ**	**3.5-bits**	**4900**	**16.2**	**1.8**	**2.4 billion**	**This work**

Table 1: Comparison of electrically programmable, nonvolatile memory technologies which have been experimentally demonstrated on an integrated photonic platform. [(a)]Switching speed and energy not reported. [(b)]Ferroelectric BTO requires a 0.7 s initialization proceedure prior to writing the final state.

Compared to these competing photonic memory technologies, our non-reciprocal magneto-optic memory cell offers an efficient nonvolatile storage solution which could provide unlimited read/write endurance at sub-ns programming speeds. Our initial results used integrated electromagnets to switch a nonvolatile ferromagnetic layer, however, we expect future implementations employing spin-orbit-torque (SOT) or spin-torque-transfer (STT) effects could further improve the switching efficiency of our magneto-optic memory cells and provide a direct optical interface to emerging magnetic and spintronic memory technologies. While bonding Ce:YIG on silicon-on-insulator wafer and depositing amorphous silicon on Ce:YIG are currently the best fabrication techniques for integrating high-quality Ce:YIG in silicon photonics, recent strides in the monolithic integration of magneto-optic garnet on silicon and silicon nitride substrates offer a pathway to further enhancements in the near future [52]. By precisely depositing magneto-optic materials on specific areas, we can further reduce the insertion loss and achieve higher integration density for nonreciprocal photonic computing.

Methods

A 500-nm-thick single-crystalline Ce:YIG ($CeY_2Fe_5O_{12}$) was epitaxially grown on a wafer of (111)-oriented (Ca, Mg, Zr)-substituted gadolinium gallium garnet (SGGG) using an RF sputtering method at 750 °C. This magneto-optic garnet has a large Faraday rotation of 4800 deg/cm at 1550 nm and was used for all the devices presented in this work.

Devices characterized in Figure 3 and Figure 4 were fabricated by bonding a Ce:YIG/SGGG on a 220 nm-thick silicon-on-insulator (SOI) wafer with 2 µm buried oxide. The SOI wafer is patterned using a 248 nm ASML 5500 deep-ultraviolet stepper, and dry etched using a Bosch process

(Plasma-Therm 770) to form the waveguides and resonators. Patterned SOI and Ce:YIG/SGGG samples are rigorously cleaned, and activated with O_2 plasma (EVG 810). Ce:YIG is directly bonded onto the SOI wafer using a flip-chip bonder (Finetech) and then annealed at 200 °C for 6 h under 3 MPa pressure to strengthen the bond. The required alignment accuracy is fairly tolerant (~200 μm). After bonding, a 1 μm layer of SiO_2 is sputtered everywhere on the chip as an upper cladding. Next, the SGGG substrate is thinned by mounting the sample against a flat chuck and polishing (Allied Technologies) using a series of increasingly fine lapping films. The thickness of SGGG is monitored using a micrometer and confirmed to be ~5 μm with a separate Dektak profilometry measurement. Variation of thickness across the sample is roughly ±1.5 μm due to imperfect levelling of the chuck. The patterns for gold coils and contacts are defined on the backside of the SGGG with a 365 nm GCA i-line wafer stepper. Then, 22 nm Ti is deposited as an underlayer, followed by 1.5 μm Au using electron-beam evaporation, and the metal coils and contacts are released with a lift-off procedure. Finally, the sample is diced and the facets are polished.

Nonvolatile magneto-optic memory cell characterized in Figure 5 is fabricated growing amorphous silicon (a-Si) on a Ce:YIG/SGGG wafer. A 10-nm-thick SiO_2 buffer layer is deposited on a Ce:YIG/SGGG wafer via plasma-enhanced chemical vapor deposition (PE-CVD). Next, a 220-nm-thick a-Si:H guiding layer is deposited via PE-CVD with a gas mixture of SiH_4 at 300 °C. Subsequently, a 200-nm-thick SiO_2 layer is deposited as a hard mask to protect the a-Si:H layer, and a 300-nm-thick positive resist (ZEP-520A) as well as a charge-dissipating agent (ESPACER) is coated onto the substrate. Waveguide patterns are exposed to the resist using an electron beam lithography (EBL) system. The waveguide patterns are transferred to a SiO_2 hard mask via reactive ion etching using CF_4 and a-Si:H waveguides are formed using SF_6. A 750-nm-thick SiO_2 layer is deposited on the top of the waveguide core to isolate the guided mode from the integrated thin-film magnet to avoid optical absorption. EBL is performed to transfer the magnet patterns of an array of 20 μm × 5 μm stripes. A 10-nm-thick Ru buffer layer followed by a 300-nm-thick CoFeB thin-film magnet are deposited using an RF facing target sputtering method at room temperature with Ar. Next, the stripe array of thin-film magnets is formed using the lift-off process. The longer side of each stripe, which is the easy magnetization axis of CoFeB, is aligned perpendicular to the waveguide. Finally, after the deposition of an 80-nm-thick SiO2 layer using PE-CVD, a 25-μm-wide and 700-nm-thick Cr/Au electromagnet for magnetizing integrated magnets was formed via electron beam vapor deposition.

Data Availability

Data sets generated during the current study are available from the corresponding author on reasonable request.

Code Availability

The complete simulation code and all simulation files required to reproduce the results presented in Figure 2 is available at https://nonreciprocalringresonators.github.io

Conflicts of Interest

The authors declare the following competing interests: N.Y. and P.P. have filed a patent application related to this work.

Author Contributions

N.Y. and P.P. conceived the idea and the experiments. N.Y., P.P., G.M., and J.E.B. oversaw the collection of data. T.M., Y.S., and D.H. fabricated devices. Y.S. grew Ce:YIG thin films used in device fabrication. V.S. and P.P. performed mathematical modelling and numerical simulations of the devices. P.P., M.D., Y.S., and T.M. performed optical characterization of fabricated devices. N.Y. and P.P. wrote the manuscript with input and feedback from all the authors.

Acknowledgements

This work was supported in part by the U.S. National Science Foundation under Grants ECCS-2210168/2210169 and CISE-2105972. N.Y. acknowledges support from the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681. P.P., G.M., and J.E.B. acknowledge support from the Air Force Office of Scientific Research under award no. FA9550-21-1-0042 and FA9550-20-1-0150. P.P. also acknowledges for support from the Autonomous Region of Sardinia though the program "Mobilità Giovani Ricercatori (MGR)" of the University of Cagliari, the Italian Ministry of University and Research through the PRIN PNRR 2022 project "MAGneto-Optic Integrated Computing (MAGIC)" (CUP F53D23008340001), and the Fondazione di Sardegna through the project "Investigation of Novel Magneto-Optic Materials and Devices for Silicon Photonic Integrated Circuits" (CUP F73C23001820007). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force.

References

1. Amodei, D. & Hernandez, D. AI and Compute. https://openai.com/blog/ai-and-compute/ (2018).

2. Thompson, N. C., Greenewald, K., Lee, K. & Manso, G. F. The Computational Limits of Deep Learning. *MIT INITIATIVE ON THE DIGITAL ECONOMY RESEARCH BRIEF* **4**, (2020).

3. Shalf, J. The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* vol. 378 1–15 Preprint at https://doi.org/10.1098/rsta.2019.0061 (2020).

4. Rupp, K. 50 Years of Microprocessor Trend Data. https://github.com/karlrupp/microprocessor-trend-data.

5. Mehonic, A. & Kenyon, A. J. Brain-inspired computing needs a master plan. *Nature* vol. 604 255–260 Preprint at https://doi.org/10.1038/s41586-021-04362-w (2022).

6. Shastri, B. J. *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics* **15**, 102–114 (2021).

7. Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat Nanotechnol* (2020) doi:10.1038/s41565-020-0655-z.

8. Agarwal, S. *et al.* Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and Its Application to Sparse Coding. *Front Neurosci* **9**, (2016).

9. Sludds, A. *et al.* Delocalized photonic deep learning on the internet's edge. *Science (1979)* **378**, 270–276 (2022).

10. Wang, T. *et al.* An optical neural network using less than 1 photon per multiplication. *Nat Commun* **13**, 123 (2022).

11. Demirkiran, C. *et al.* An Electro-Photonic System for Accelerating Deep Neural Networks. *ACM J Emerg Technol Comput Syst* **19**, 1–31 (2023).

12. Peng, B., Hua, S., Su, Z., Xu, Y. & Shen, Y. A 64×64 integrated photonic accelerator. in *2022 IEEE Photonics Conference (IPC)* MB4.4 (IEEE, Vancouver, 2022).

13. Tait, A. N., Nahmias, M. A., Shastri, B. J. & Prucnal, P. R. Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing. *Journal of Lightwave Technology* **32**, 4029–4041 (2014).

14. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).

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

16. Huang, C. *et al.* A silicon photonic–electronic neural network for fibre nonlinearity compensation. *Nat Electron* **4**, 837–844 (2021).

17. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).

18. Xu, X. *et al.* 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).

19. Wu, C. *et al.* Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat Commun* **12**, 96 (2021).

20. Ielmini, D. & Wong, H.-S. P. In-memory computing with resistive switching devices. *Nat Electron* **1**, 333–343 (2018).

21. Zhang, W. *et al.* Silicon microring synapses enable photonic deep learning beyond 9-bit precision. *Optica* **9**, (2022).

22. Feldmann, J. *et al.* Integrated 256 Cell Photonic Phase-Change Memory With 512-Bit Capacity. *IEEE Journal of Selected Topics in Quantum Electronics* **26**, 1–7 (2020).

23. Youngblood, N. Coherent Photonic Crossbar Arrays for Large-Scale Matrix-Matrix Multiplication. *IEEE Journal of Selected Topics in Quantum Electronics* **29**, 1–11 (2022).

24. Yang, G. *et al.* Processing-in-Memory Using Optically-Addressed Phase Change Memory. in *ACM/IEEE International Symposium on Low Power Electronics and Design* (2023).

25. Huang, D. *et al.* Integrated broadband Ce:YIG/Si Mach–Zehnder optical isolators with over 100 nm tuning range. *Opt Lett* **42**, 4901 (2017).

26. Huang, D. *et al.* Dynamically reconfigurable integrated optical circulators. *Optica* **4**, 23 (2017).

27. Shoji, Y. & Mizumoto, T. Magneto-optical non-reciprocal devices in silicon photonics. *Sci Technol Adv Mater* **15**, 014602 (2014).

28. Mitsuya, K., Shoji, Y. & Mizumoto, T. Demonstration of a Silicon Waveguide Optical Circulator. *IEEE Photonics Technology Letters* **25**, 721–723 (2013).

29. Resch, S. *et al.* On Endurance of Processing in (Nonvolatile) Memory. in *Proceedings of the 50th Annual International Symposium on Computer Architecture* 1–13 (ACM, New York, NY, USA, 2023). doi:10.1145/3579371.3589114.

30. Youngblood, N., Ríos Ocampo, C. A., Pernice, W. H. P. & Bhaskaran, H. Integrated optical memristors. *Nat Photonics* **17**, 561–572 (2023).

31. Pintus, P. *et al.* An integrated magneto-optic modulator for cryogenic applications. *Nat Electron* **5**, 604–610 (2022).

32. Murai, T., Shoji, Y., Nishiyama, N. & Mizumoto, T. Nonvolatile magneto-optical switches integrated with a magnet stripe array. *Opt Express* **28**, 31675 (2020).

33. Shoji, Y., Taniguchi, S. & Yajima, S. Nonvolatile photonic switch with magnetic materials on a silicon photonic platform [Invited]. *Opt Mater Express* **13**, 2489 (2023).

34. Tait, A. N. *et al.* Microring Weight Banks. *IEEE Journal of Selected Topics in Quantum Electronics* **22**, 312–325 (2016).

35. Heebner, J., Grover, R., Ibrahim, T. & Ibrahim, T. A. *Optical Microresonators: Theory, Fabrication, and Applications*. (Springer Science & Business Media, 2008).

36. Pintus, P. Accurate vectorial finite element mode solver for magneto-optic and anisotropic waveguides. *Optics Express* vol. 22 15737 Preprint at https://doi.org/10.1364/OE.22.015737 (2014).

37. Onbasli, M. C. *et al.* Optical and magneto-optical behavior of Cerium Yttrium Iron Garnet thin films at wavelengths of 200–1770 nm. *Sci Rep* **6**, 23640 (2016).

38. Zaets, W. & Ando, K. Optical waveguide isolator based on nonreciprocal loss/gain of amplifier covered by ferromagnetic layer. *IEEE Photonics Technology Letters* **11**, 1012–1014 (1999).

39. Freeman, M. R. Picosecond pulsed-field probes of magnetic systems (invited). *J Appl Phys* **75**, 6194–6198 (1994).

40. Rey-De-Castro, R. *et al.* Subpicosecond Faraday effect in Cd 1-xMn xTe and its application in magneto-optical sampling. *Appl Phys Lett* **85**, 3806–3808 (2004).

41. Tossoun, B. *et al.* High-speed and energy-efficient non-volatile silicon photonic memory based on heterogeneously integrated memresonator. *Nat Commun* **15**, 551 (2024).

42. Fang, Z. *et al.* Fast and Energy-Efficient Non-Volatile III-V-on-Silicon Photonic Phase Shifter Based on Memristors. *Adv Opt Mater* (2023) doi:10.1002/adom.202301178.

43. Ríos, C. *et al.* In-memory computing on a photonic platform. *Sci Adv* **5**, eaau5759 (2019).

44. Zhou, W. *et al.* In-memory photonic dot-product engine with electrically programmable weight banks. *Nat Commun* **14**, 2887 (2023).

45. 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).

46. Song, J.-F. *et al.* Integrated photonics with programmable non-volatile memory. *Sci Rep* **6**, 22616 (2016).

47. Edinger, P. *et al.* A Bistable Silicon Photonic Mems Phase Switch For Nonvolatile Photonic Circuits. in *2022 IEEE 35th International Conference on Micro Electro Mechanical Systems Conference (MEMS)* 995–997 (IEEE, 2022). doi:10.1109/MEMS51670.2022.9699739.

48. Taki, K. *et al.* Non-volatile optical phase shift in ferroelectric hafnium zirconium oxide. (2023).

49. Geler-Kremer, J. *et al.* A ferroelectric multilevel non-volatile photonic phase shifter. *Nat Photonics* **16**, 491–497 (2022).

50. Fang, Z. *et al.* Ultra-low-energy programmable non-volatile silicon photonics based on phase-change materials with graphene heaters. *Nat Nanotechnol* **17**, 842–848 (2022).

51. Emboras, A. *et al.* Atomic Scale Plasmonic Switch. *Nano Lett* **16**, (2016).

52. Zhang, Y. *et al.* Monolithic integration of broadband optical isolators for polarization-diverse silicon photonics. *Optica* **6**, 473 (2019).