Photonic Integrated Circuit for Matrix Inversions and Multiplications

Minjia Chen(1), Chunhui Yao(1), Adrian Wonfor(1), Shuai Yang(1), Mark Holm(2), Qixiang Cheng(1) *, Richard Penty(1)

(1) Electrical Engineering Division, Department of Engineering, University of Cambridge, 9 JJ Thomson Avenue, Cambridge, CB3 0FA, UK
(2) Huawei Technologies (Sweden) AB, Isafjordsgatan 34, 164 40 Kista, Sweden
* Email: qc223@cam.ac.uk

Abstract *Photonic computing has the potential to surpass the speed of electronic computing. Here we demonstrate the first all-optical iterative photonic integrated processor for matrix inversions and multiplications. A 4×4 matrix is inverted with 93% accuracy and a speed of 8.5×10^5 inversions per second.*

Introduction

Modelling systems using linear equations is one of the most frequently used methods in science and engineering. Numerically solving such problems usually involves computationally expensive matrix calculations including matrix inversions and multiplications [1,2]. Limitations in speed and power consumption of the digital electronic computing platforms have caused researchers to seek new pathways to accelerate such tasks. Photonic computing systems that enable in-propagation computing have recently gained increasing interest owing to the rapid development of photonic integration platforms. In the past decades, efforts have been made in free-space [3], all-fibre-based [4], and integrated waveguide [5] systems. Free-space optical systems allow a relatively high matrix size but are generally bulky. All-fibre optical networks have the problem of phase fluctuation arising from environmental effects, thus prohibiting the use of a coherent source/receiver and the extension to complex-valued computations. In contrast, photonic integration platforms offer compact size, enabling ultra-fast computations, and simple phase and polarization control. This enables full utilization of high parallelism of optical signal processing, and good reconfigurability for implementing different computations using a single chip.

Since the 2010s, numerous demonstrations of on-chip matrix-vector multipliers have been reported [5–8]. However, few explorations of optical matrix inversions have been made, despite the fact that matrix inversion is more computationally taxing than matrix multiplication and is widely used in numerically solving equations [1], communication system applications [9], control systems [10] and cryptography [11]. Since iterative methods are usually used for optical matrix inversion systems [12], the computing speed can be significantly improved if all-optical loop-back is applied to avoid optical/electrical/optical (OEO) conversions between iterations.

In this paper, we present the first photonic integrated circuit that is capable of performing direct computations for both matrix inversions and multiplications. We implement 4×4 matrix inversions on a silicon nitride (SiN) photonic chip, with an accuracy of 93% and a speed of 8.5×10^5 inversions per second. In addition, the chip is also used for 4×4 matrix multiplications with 90% accuracy. Our results pave a promising way towards ultra-fast photonic integrated generic matrix processors.

Principles and chip design

We have proposed an iterative Richardson

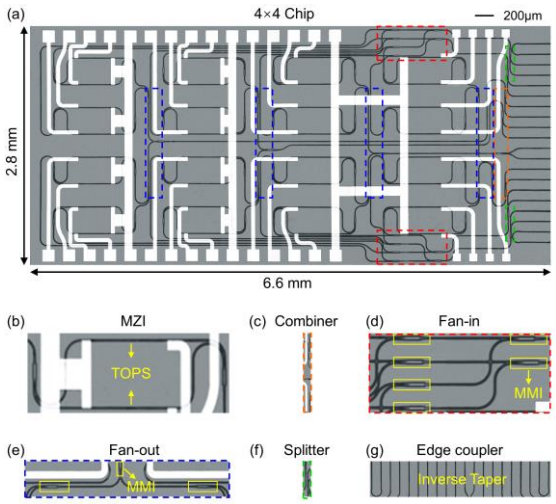

Fig. 1:. Micrographs of (a) the 4×4 SiN integrated chip. (b) A Mach-Zehnder Interferometer (MZI) unit equipped with two thermo-optic phase shifters (TOPSs). (c) a Combiner block. (d) a Fan-in block consisting of cascaded multi-mode interferometers (MMIs). (d) a Fan-out block consisting of MMIs. (f) a Splitter block. (g) Edge couplers with the inverse taper structure.

processor for matrix inversion as described by Eq. (1) [12]:

$$\boldsymbol{X}^{(k+1)} = (\boldsymbol{I}_N - \omega\boldsymbol{A})\boldsymbol{X}^{(k)} + \omega\boldsymbol{I}_N \qquad (1)$$

where A is the N×N matrix to be inverted, $\boldsymbol{I}_N$ is the N×N identity matrix, ω is a parameter used to adjust the convergence of the inversion algorithm, $\boldsymbol{X}^{(k+1)}$ and $\boldsymbol{X}^{(k)}$ (k=0,1,2,…) are output matrices after k+1 and k iterations, and $\omega\boldsymbol{I}_N$ is both the initial input matrix ($\boldsymbol{X}^{(0)}$) that initiates the computation and the additional term which needs to be added in each iteration.

Fig. 1a shows a micrograph of the 4×4 photonic integrated chip for computing matrix inversions and multiplications, which has a footprint of 2.8×6.6 mm^2. Fig. 1b-g shows enlarged views of the components integrated on-chip, including Mach-Zehnder interferometers (MZIs) equipped with two thermo-optic phase shifters (TOPSs), the signal fan-in blocks comprising cascaded multi-mode interferometers (MMIs), the combiner blocks comprising single-stage 2×1 MMI couplers, the signal fan-out blocks consisting of cascaded MMIs, the splitter block consisting of single-stage 1×2 MMI coupler and edge couplers with the inverse taper structure.

To prepare for the inversion, the weight matrix $\boldsymbol{I}_N - \omega\boldsymbol{A}$ is pre-loaded onto the on-chip weight bank comprising 16 MZIs. Specifically, the transmissions of MZIs are changed by applying voltages to the TOPSs according to pre-calibrated lookup tables. Then the input light signal representing one column of $\omega\boldsymbol{I}_N$ is launched into the chip via the edge couplers to initiate the inversion. The signal is split into 4 copies in the fan-out blocks (blue dashed boxes) and sent to the weight bank. After passing the weight bank and being combined at the fan-in blocks (red dashed boxes), a matrix-vector multiplication (MVM) operation is achieved, corresponding to the multiplication of $(\boldsymbol{I}_N - \omega\boldsymbol{A})$ and one column of $\boldsymbol{X}^{(k)}$. The MVM results are split in the 1×2 splitter blocks (green dashed boxes), one of which is sent to off-chip components via edge couplers for amplification, filtering and detection, and the others are used for monitoring purposes. After compensating for propagation, combining, splitting, and coupling losses, the signals are sent back to the chip and combined with the initial one column of $\omega\boldsymbol{I}_N$ in the combiner blocks (orange dashed boxes). This process is repeated several times until the outputs converge.

Experimental setup

Fig. 2a shows the schematic of the experimental setup. A 1550 nm continuous-wave laser (CW) source is used as the input signal. A modulator (Mod) is used to turn the input signal into short pulses so that the computation results in each iteration can be recorded in the 4-channel oscilloscope (4-ch OSC). Upon entering the chip, the signal goes through a combiner block, a fan-out block, the weight bank, a fan-in block, and a splitter block in turn. The 4 signals representing the MVM results are coupled out of the chip. 4 erbium-doped fibre amplifiers (EDFAs) are used to compensate for on-chip losses and fibre-to-chip coupling losses in each path. 4 Bandpass filters (BPFs, 0.1 nm bandwidth) are then used to suppress amplified spontaneous emission (ASE) noise from the amplification process. The filtered signals are split into two parts in the 1×2 splitters. One part is photoconverted by photodetectors (PDs) into currents and amplified by transimpedance amplifiers (TIAs) into voltages, which are captured by the 4-ch OSC. The other part is sent back to the chip's first fan-in block to form loop operations. Polarisation controllers (PCs) are used to align the polarisations of the light signal to the chip.

Fig. 2b displays the packaged chip, with the SiN chip highlighted in the white dashed box. Bonding wires are used to provide electrical control to the TOPSs from the customised printed circuit board (PCB). An ultra-high numerical aperture fibre array (UHNA FA) is used for coupling light signals into and out of the chip.

Fig. 2c presents an example transmission curve of an MZI unit. Red dots show the measured transmission and the blue line shows the fitted curve. The transmission-voltage (T-V)

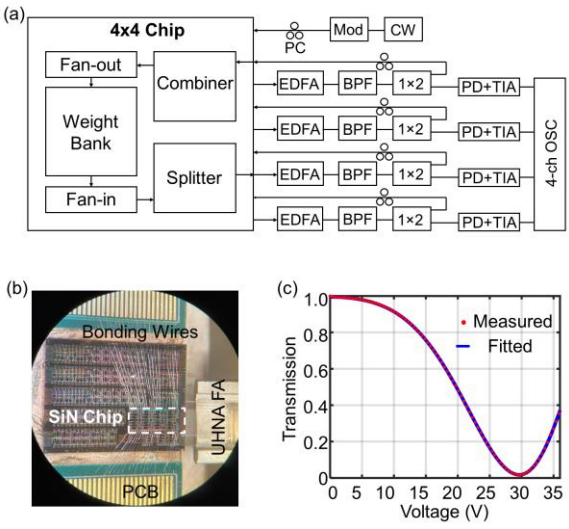

Fig. 2:. (a) Experimental setup of the all-optical iterative matrix inverter. (b) Packaged chip. (c) Calibrated exemplary MZI transmission curve

relationship is depicted in Eq. (2):

$$T = 0.51\cos(0.02V^2 + 0.42) + 0.53 \qquad (2)$$

Results

We implement 4×4 matrix inversion and multiplication using the photonic integrated chip. Fig. 3a-3b exhibit the 4×4 matrix inversion and multiplication results respectively.

In Fig. 3a, the inversion accuracy in terms of norm is: $(1 - \|\boldsymbol{A}_{\boldsymbol{meas}}^{-1} - \boldsymbol{A}_{\boldsymbol{ideal}}^{-1}\| / \|\boldsymbol{A}_{\boldsymbol{ideal}}^{-1}\|) \times 100\% = 93\%$. In this experiment, the time for signals to propagate one circulation is around 130 ns. The computation converges after 9 iterations, corresponding to an inversion rate of 8.5×10^5

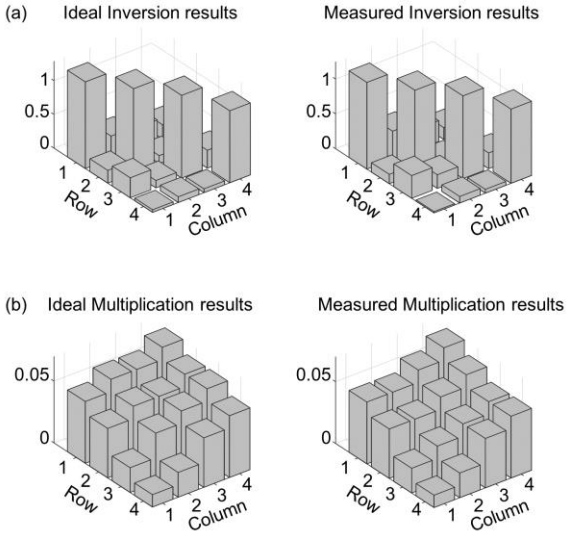

Fig. 3:. (a) Inversion results of a randomly generated 4×4 matrix. (b) Multiplication results of two randomly generated 4×4 matrices.

inversions per second.

In Fig. 3b, the multiplication accuracy in terms of norm is: $(1 - \|\boldsymbol{Y}_{\boldsymbol{meas}} - \boldsymbol{Y}_{\boldsymbol{ideal}}\| / \|\boldsymbol{Y}_{\boldsymbol{ideal}}\|) \times 100\% = 90\%$.

Conclusions

We demonstrate the first all-optical iterative photonic integrated processor for matrix inversions and multiplications. We implement 4×4 matrix inversions on a silicon nitride (SiN) photonic chip, with an accuracy of 93%, and a speed of 8.5×10^5 inversions per second, which is approximately an order of magnitude faster than traditional electronic computers (A desktop computer with a 2.9GHz Intel i7 core calculates 4x4 matrix inversions at a speed of 9.2×10^4 inversions per second). In addition, the chip is also used for 4×4 matrix multiplications with an accuracy of 90%. We show a pathway towards generic photonic matrix processors.

Acknowledgements

This work was supported by Huawei Technologies (Sweden) AB (G107576). The authors thank CORNERSTONE for providing free access to their second SiN MPW run (funded by the CORNERSTONE 2 project under Grant EP/T019697/1).

References

[1] G. H. Golub and C. F. Van Loan, *Matrix computations*, Fourth edition. in Johns Hopkins studies in the mathematical sciences. Baltimore: The Johns Hopkins University Press, 2013.

[2] D. C. Montgomery, *Introduction to linear regression analysis.*, 3rd ed. / Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. New York: New York, 2001.

[3] H. Rajbenbach, Y. Fainman, and S. H. Lee, 'Optical implementation of an iterative algorithm for matrix inversion', *Appl. Opt.*, vol. 26, no. 6, p. 1024, Mar. 1987, doi: 10.1364/AO.26.001024.

[4] K. Wu, C. Soci, P. P. Shum, and N. I. Zheludev, 'Computing matrix inversion with optical networks', *Opt. Express*, vol. 22, no. 1, p. 295, Jan. 2014, doi: 10.1364/OE.22.000295.

[5] F. Ashtiani, A. J. Geers, and F. Aflatouni, 'An on-chip photonic deep neural network for image classification', *Nature*, vol. 606, no. 7914, pp. 501–506, Jun. 2022, doi: 10.1038/s41586-022-04714-0.

[6] A. N. Tait *et al.*, 'Neuromorphic photonic networks using silicon photonic weight banks', *Sci Rep*, vol. 7, no. 1, p. 7430, Aug. 2017, doi: 10.1038/s41598-017-07754-z.

[7] D. A. B. Miller, 'Establishing Optimal Wave Communication Channels Automatically', *Journal of Lightwave Technology*, vol. 31, no. 24, pp. 3987–3994, Dec. 2013, doi: 10.1109/JLT.2013.2278809.

[8] N. C. Harris *et al.*, 'Quantum transport simulations in a programmable nanophotonic processor', *Nature Photon*, vol. 11, no. 7, pp. 447–452, Jul. 2017, doi: 10.1038/nphoton.2017.95.

[9] C. Tang, C. Liu, L. Yuan, and Z. Xing, 'High Precision Low Complexity Matrix Inversion Based on Newton Iteration for Data Detection in the Massive MIMO', *IEEE Communications Letters*, vol. 20, no. 3, pp. 490–493, Mar. 2016, doi: 10.1109/LCOMM.2015.2514281.

[10] X.-W. Zhang, L. Zuo, M. Li, and J.-X. Guo, 'High-Throughput FPGA Implementation of Matrix Inversion for Control Systems', *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6205–6216, Jul. 2021, doi: 10.1109/TIE.2020.2994865.

[11] X. Lei, X. Liao, T. Huang, H. Li, and C. Hu, 'Outsourcing Large Matrix Inversion Computation to A Public Cloud', *IEEE Transactions on Cloud Computing*, vol. 1, no. 1, pp. 1–1, Jan. 2013, doi: 10.1109/TCC.2013.7.

[12] M. Chen, Q. Cheng, M. Ayata, M. Holm, and R. Penty, 'Iterative photonic processor for fast complex-valued matrix inversion', *Photon. Res.*, vol. 10, no. 11, p. 2488, Nov. 2022, doi: 10.1364/PRJ.468097.