Title
Arbitrarily programmable wave propagation on a photonic chip

Authors
Stein, Martin M
Onodera, Tatsuhiro
Ash, Benjamin A
et al.

Publication Date
2024-06-18

DOI
10.1117/12.3017660

Arbitrarily programmable wave propagation on a photonic chip

Martin M. Stein[1,*], Tatsuhiro Onodera[1,2,*], Benjamin A. Ash[1], Mandar M. Sohoni[1], Melissa Bosch[1], Ryotatsu Yanagimoto[1,2], Marc Jankowski[2,3], Timothy P. McKenna[2,3], Tianyu Wang[1], Gennady Shvets[1], Maxim R. Shcherbakov[1], Logan G. Wright[1,2], Peter L. McMahon[1,4]

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA
[2]NTT Physics and Informatics Laboratories, NTT Research, Inc., Sunnyvale, CA, USA
[3]E. L. Ginzton Laboratory, Stanford University, Stanford, CA, USA
[4]Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, NY, USA
[*]These authors contributed equally.

ABSTRACT

On-chip photonic-neural-network processors promise benefits in both speed and energy efficiency but have not yet reached the scale to compete with electronic processors. The dominant paradigm is to build integrated-photonic processors using discrete components connected by single-mode waveguides. A far more compact alternative is to avoid discrete components and instead sculpt a complex and continuous microphotonic medium in which computations are performed by multimode waves controllably propagating in two dimensions. We show our realization of this approach with a device whose refractive index as a function of space can be rapidly reprogrammed. We demonstrate optical computations much larger and more error-resilient than previous photonic chips relying on discrete components. We argue that beyond photonic-neural-network processors, devices with such arbitrarily programmable index distributions enable the realization of a wide range of photonic functionality.

Keywords: Optical neural networks, programmable integrated photonics, lithium niobate

Introduction

The size of deep neural network (DNN) models has been increasing exponentially over the past decade, leading to exponentially increasing energy costs for their execution. Limits to energy costs now impose a practical constraint on how large models can be,[2] strongly motivating the exploration of alternative, energy-efficient computing approaches for executing DNNs. Optical neural networks (ONNs) that specialize in performing the computationally most expensive operation, matrix-vector multiplications (MVMs), with optics instead of electronics are one promising candidate approach.[3–6]

Integrated photonics is a leading platform for optical neural networks due to its compact form factor, excellent phase stability, availability of high-bandwidth modulators and detectors, manufacturability, and ease of integration with electronics.[3,4,7–11] The dominant paradigm for designing integrated photonic neural networks is to construct networks of discrete, programmable photonic components—such as Mach–Zehnder interferometers, microring resonators, or phase-change-memory cells—connected by single-mode waveguides.[6] The scale of such chips has been limited by at least two factors: (1) the large spatial footprint of individual components and the inefficiency of dedicating a substantial portion of the chip's area to non-programmable interconnection regions comprising well-isolated waveguides that connect relatively sparsely arranged programmable elements, and (2) the systems-integration complexity of controlling each discrete component with electronic wires carrying the trainable parameters of the neural network.

We could achieve far greater spatial efficiency[12,13] if, instead of building the integrated photonic neural network from discrete components, we treated the entire chip as a blank slate that we could arbitrarily and reprogrammably sculpt. Here lies the central challenge that our work tackles: for such a chip to perform an MVM with a programmable matrix, we need to be able to continuously program the chip's refractive-index distribution, $n(x, z)$.[12–17] How can we make a photonic chip whose refractive-index distribution is programmable, ideally in a way that avoids the integration complexity of introducing electronic wiring?

This conference proceeding is a summary of our recent work,[1] adapting figures and text thereof.

Machine Learning in Photonics, edited by Francesco Ferranti, Mehdi Keshavarz Hedayati, Andrea Fratalocchi, Proc. of SPIE Vol. 13017, 130170M · © 2024 SPIE
0277-786X · doi: 10.1117/12.3017660

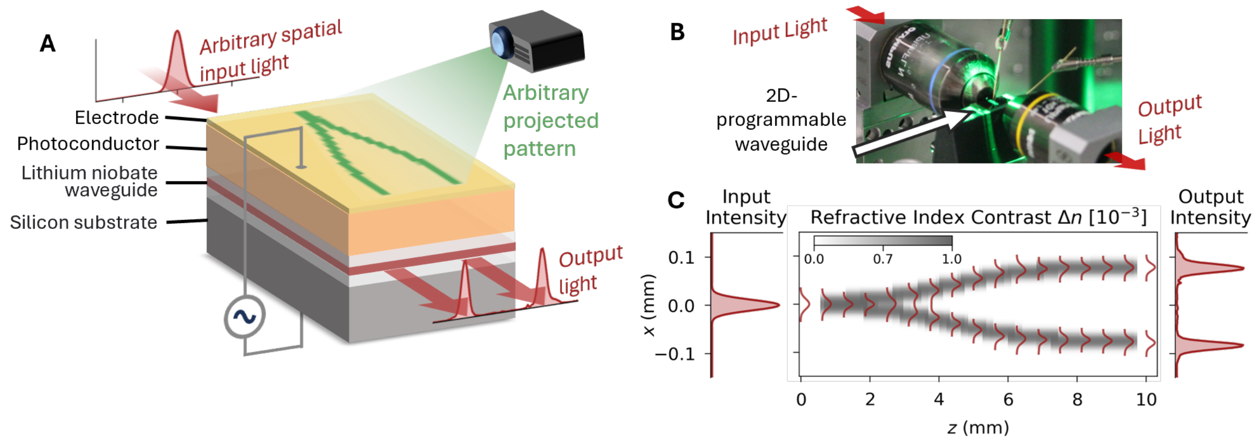

Figure 1. **Operating principle of the 2D-programmable waveguide.** (**A**) The 2D-programmable waveguide linearly transforms arbitrary optical input fields via wave propagation through a lithium niobate slab waveguide whose two-dimensional refractive-index modulation $\Delta n(x, z)$ can be continuously and arbitrarily programmed (up to practical limits on resolution and the maximum modulation) by an illumination pattern that is projected onto the device (shown in green). (**B**) A photograph of our 2D-programmable waveguide in our experimental setup. (**C**) Experimental realization of a Y-branch splitter on the 2D-programmable waveguide, which splits the input light into two equal output beams. The projected pattern shown in (A) directly corresponds to the induced refractive-index modulation (in gray). A simulation of the wave propagating through the device is overlaid with the pattern (in red). Figure adapted from ref.[1]

Operating principle of the device

We present a photonic chip with a waveguide that is fully programmable in two dimensions: a *2D-programmable waveguide*. The chip uses massively parallel electro-optic modulation to program $n(x, z)$ across ~10,000 individual regions of a lithium niobate slab waveguide. A distinguishing feature of our device is its programmability, setting it apart from typical inverse-designed photonic devices, which are fixed after manufacturing. We achieve programmability optically, decoupling the electronic wiring for programming from the photonic chip: a pattern of light shone on top of our device creates a spatially-varying refractive-index modulation $\Delta n(x, z)$ in the slab waveguide. Inspired by optoelectronic tweezers, this is achieved by using the principle of photoconductive gain[18, 19] to induce a refractive-index modulation via the strong electro-optic effect in lithium niobate.

Our device is composed of a lithium niobate slab waveguide and a photoconductive film, which are sandwiched between a pair of electrodes across which a bias voltage is applied (Fig. 1A). For regions of the chip that are illuminated at intensities of tens of $\mathrm{mW/cm^2}$, the photoconductor's impedance drops substantially, increasing the voltage across the slab waveguide (and the electric field within), thereby locally changing the refractive index of lithium niobate by approximately 10^{-3}. Illumination patterns extend over an area of $9\,\mathrm{mm} \times 1\,\mathrm{mm}$ with a resolution of $9\,\mu\mathrm{m}$ and can be updated at a rate of around $3\,\mathrm{Hz}$.

To illustrate the operating principle of our device, we projected a pattern in the shape of a Y-branch splitter onto the 2D-programmable waveguide (Fig. 1A). The projected pattern instantiated a refractive-index distribution of a Y-branch splitter. We coupled a single input Gaussian beam into the device using a beamshaper and measured the intensity of the output light with a camera.

MNIST handwritten-digit classification

We train multimode photonic structures within the chip that perform neural-network inference. The structures realized by our 2D-programmable waveguide are similar to inverse-designed nanophotonic devices:[16, 20] they are computer-optimized, two-dimensional metastructures that control multimode wave propagation. We show that we can train the refractive-index distribution so that the complex wave propagation through the device performs a desired neural-network inference.

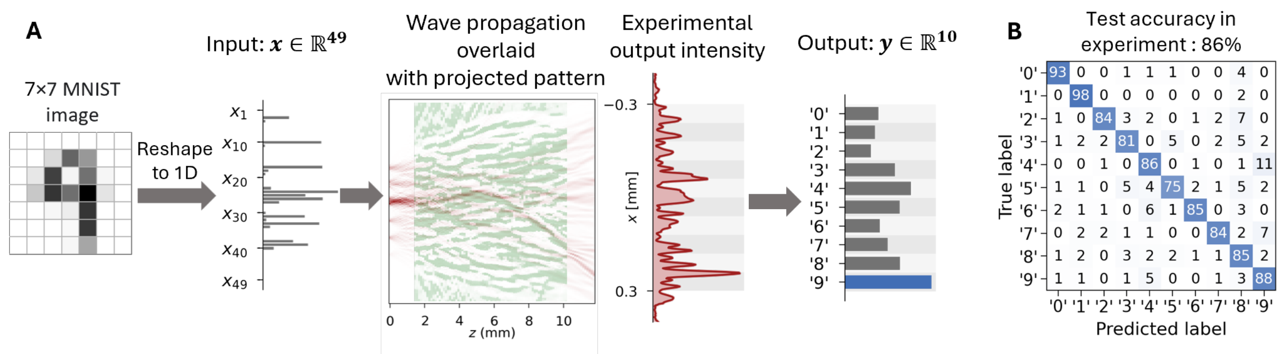

Figure 2. **MNIST classification with multimode wave propagation in the 2D-programmable waveguide.** (**A**) Exemplary classification of an image of a handwritten digit. Left: The image is reshaped to 1D and amplitude-encoded into 49 spatial Gaussian modes to produce the initial optical field distribution. Center: Simulated wave propagation (red) in the chip after training of the projected pattern (overlaid in green). Right: The experimentally measured output intensity. (**B**) Confusion matrix showing the test accuracy in experiment after training. Figure adapted from ref.[1]

In Fig. 2, we present our experimental results on MNIST handwritten-digit classification. The task consists of classifying 14-by-14-pixel images of handwritten digits from 0 to 9. We divided the MNIST dataset in the standard manner into 60,000 training images and 10,000 test images. We down-sampled each MNIST image to 7-by-7 pixels, then reshaped them to 49-dimensional input vectors. We encoded the 49-dimensional input vectors in the amplitudes of 49 equidistant spatial Gaussian modes at the input facet of the device. We measured the intensity at the output facet with a camera and binned the camera pixels into 10 different regions, with each region corresponding to one digit. The predicted digit for an input is given by the region that receives the most optical power. The refractive-index distribution was trained via a backpropagation algorithm[21] to steer the most power toward the region corresponding to the correct digit.

As shown in Fig. 2B, the system achieved 86% accuracy on the test dataset after 10 epochs of training, which took about 10 hours on the experimental setup. This falls 4 percentage points short of the 90% accuracy that a one-layer digital neural network achieves on this downsampled MNIST classification task, likely due to imperfect modeling and experimental drifts. This result suggests that complex wave propagation in our device can be harnessed to perform computations comparable to that of a single-layer neural network with a 49×10 matrix of trainable parameters.

Discussion and outlook

We have introduced and demonstrated a 2D-programmable photonic processor comprising a lithium niobate slab waveguide whose refractive-index distribution, $n(x, z)$, can be continuously programmed. The device design enables programming by massively parallel electro-optic modulation with approximately 10,000 degrees of freedom. We used our chip to perform neural-network inference by training the refractive-index distribution and consequently the multimode wave propagation through the chip.

We believe that our device concept, with its ability to programmably control multimode wave propagation, may create new opportunities in the fields of optical computing and optical information processing.[5,6,22] Any photonic device that can be specified as an inhomogeneous refractive-index distribution can in principle be realized. It may ultimately even be possible to make a device that combines programmable linear wave propagation (this work), programmable nonlinear wave propagation (a natural extension of this work to having programmable $\chi^{(2)}(x, z)$), and programmable gain/loss (demonstrated in ref. [17]), giving rise to a reconfigurable on-chip platform capable of realizing almost every functionality we have in free-space optics.

ACKNOWLEDGMENTS

We acknowledge funding from the Air Force Office of Scientific Research (FA9550-22-1-0378), the National Science Foundation (CCF-1918549) and NTT Research. This work was performed in part at the Cornell NanoScale Facility, an NNCI member supported by NSF Grant NNCI-2025233.

REFERENCES

[1] Onodera, T., Stein, M. M., Ash, B. A., Sohoni, M. M., Bosch, M., Yanagimoto, R., Jankowski, M., McKenna, T. P., Wang, T., Shvets, G., et al., "Scaling on-chip photonic neural processors using arbitrarily programmable wave propagation," *arXiv preprint arXiv:2402.17750* (2024).

[2] Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., and Dean, J., "Carbon emissions and large neural network training," *arXiv:2104.10350* (2021).

[3] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., et al., "Deep learning with coherent nanophotonic circuits," *Nature Photonics* **11**, 441–446 (2017).

[4] Tait, A. N., de Lima, T. F., Zhou, E., Wu, A. X., Nahmias, M. A., Shastri, B. J., and Prucnal, P. R., "Neuromorphic photonic networks using silicon photonic weight banks," *Scientific Reports* **7**, 7430 (Aug. 2017).

[5] Wetzstein, G., Ozcan, A., Gigan, S., Fan, S., Englund, D., Soljačić, M., Denz, C., Miller, D. A., and Psaltis, D., "Inference in artificial intelligence with deep optics and photonics," *Nature* **588**, 39–47 (2020).

[6] Shastri, B. J., Tait, A. N., Ferreira de Lima, T., Pernice, W. H., Bhaskaran, H., Wright, C. D., and Prucnal, P. R., "Photonics for artificial intelligence and neuromorphic computing," *Nature Photonics* **15**, 102–114 (2021).

[7] Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H., and Pernice, W. H., "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature* **569**, 208–214 (2019).

[8] Feldmann, J., Youngblood, N., Karpov, M., Gehring, H., Li, X., Stappers, M., Le Gallo, M., Fu, X., Lukashchuk, A., Raja, A. S., et al., "Parallel convolutional processing using an integrated photonic tensor core," *Nature* **589**, 52–58 (2021).

[9] Huang, C., Fujisawa, S., de Lima, T. F., Tait, A. N., Blow, E. C., Tian, Y., Bilodeau, S., Jha, A., Yaman, F., Peng, H.-T., et al., "A silicon photonic–electronic neural network for fibre nonlinearity compensation," *Nature Electronics* **4**(11), 837–844 (2021).

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

[11] Bandyopadhyay, S., Sludds, A., Krastanov, S., Hamerly, R., Harris, N., Bunandar, D., Streshinsky, M., Hochberg, M., and Englund, D., "Single chip photonic deep neural network with accelerated training," *arXiv:2208.01623* (2022).

[12] Larocque, H. and Englund, D., "Universal linear optics by programmable multimode interference," *Optics Express* **29**, 38257–38267 (2021).

[13] Khoram, E., Chen, A., Liu, D., Ying, L., Wang, Q., Yuan, M., and Yu, Z., "Nanophotonic media for artificial neural inference," *Photonics Research* **7**, 823–827 (2019).

[14] Hughes, T. W., Williamson, I. A. D., Minkov, M., and Fan, S., "Wave physics as an analog recurrent neural network," *Science Advances* **5**, eaay6946 (Dec. 2019).

[15] Nakajima, M., Tanaka, K., and Hashimoto, T., "Neural Schrödinger equation: Physical law as deep neural network," *IEEE Transactions on Neural Networks and Learning Systems* **33**, 2686–2700 (June 2022).

[16] Nikkhah, V., Pirmoradi, A., Ashtiani, F., Edwards, B., Aflatouni, F., and Engheta, N., "Inverse-designed low-index-contrast structures on a silicon photonics platform for vector–matrix multiplication," *Nature Photonics* (2024).

[17] Wu, T., Menarini, M., Gao, Z., and Feng, L., "Lithography-free reconfigurable integrated photonic processor," *Nature Photonics* **17**, 710–716 (Apr. 2023).

[18] Chiou, P. Y., Ohta, A. T., and Wu, M. C., "Massively parallel manipulation of single cells and microparticles using optical images," *Nature* **436**, 370–372 (2005).

[19] Wu, M. C., "Optoelectronic tweezers," *Nature Photonics* **5**, 322–324 (May 2011).

[20] Molesky, S., Lin, Z., Piggott, A. Y., Jin, W., Vucković, J., and Rodriguez, A. W., "Inverse design in nanophotonics," *Nature Photonics* **12**, 659–670 (Oct. 2018).

[21] Wright, L. G., Onodera, T., Stein, M. M., Wang, T., Schachter, D. T., Hu, Z., and McMahon, P. L., "Deep physical neural networks trained with backpropagation," *Nature* **601**, 549–555 (2022).

[22] McMahon, P. L., "The physics of optical computing," *Nature Reviews Physics* **5**, 717–734 (Oct. 2023).