

Integrated lithium niobate photonic computing circuit based on efficient and high-speed electro-optic conversion

Received: 13 January 2025

Accepted: 28 July 2025

Published online: 01 September 2025

Check for updates

Yaowen Hu ^{1,2,13} , Yunxiang Song ^{1,3,13} , Xinrui Zhu¹, Xiangwen Guo⁴, Shengyuan Lu¹, Qihang Zhang⁵, Lingyan He⁶, Cornelis A. A. Franken^{1,7}, Keith Powell¹, Hana Warner ¹, Daniel Assumpcao ¹, Dylan Renaud ¹, Ying Wang⁶, Leticia Magalhães¹, Victoria Rosborough⁸, Amirhassan Shams-Ansari^{1,9}, Xudong Li¹, Rebecca Cheng ¹, Kevin Luke⁸, Kiyoul Yang ¹, George Barbastathis ¹⁰, Mian Zhang⁶, Di Zhu ^{11,12}, Leif Johansson⁸, Andreas Beling⁴, Neil Sinclair¹ & Marko Lončar ¹

The surge in artificial intelligence applications calls for scalable, high-speed, and low-energy computation methods. Computing with photons is promising due to the intrinsic parallelism, high bandwidth, and low latency of photons. However, current photonic computing architectures are limited by the speed and energy consumption associated with electronic-to-optical data transfer, i.e., electro-optic conversion. Here, we demonstrate a thin-film lithium niobate (TFLN) computing circuit that addresses this challenge, leveraging both highly efficient electro-optic modulation and the spatial scalability of TFLN photonics. Our circuit is capable of computing at 43.8 GOPS/channel while consuming 0.0576 pJ/OP, and we demonstrate various inference tasks with high accuracy, including the classification of binary data and complex images. Heightening the integration level, we show another TFLN computing circuit that is combined with a hybrid-integrated distributed-feedback laser and heterogeneous-integrated modified uni-traveling carrier photodiode. Our results show that the TFLN photonic platform holds promise to complement silicon photonics and diffractive optics for photonic computing, with extensions to ultrafast signal processing and ranging.

The desire for intelligent systems capable of autonomous learning, reasoning, and adaptation has fueled significant advancements in artificial intelligence (AI), transforming various application landscapes. As the demand for computational resources rapidly grows, traditional electronic computing approaches for AI are reaching their inherent limits in speed and energy efficiency for parallel processing. This limitation has stimulated the exploration of novel computing architectures employing different computational paradigms^{1–3}. One example is photonic computing^{4–9}, which uses photons to perform computational tasks traditionally executed by electronic systems.

Since photons possess unique properties such as high bandwidth enabled by high optical carrier frequencies and inherent parallelism leveraging both the frequency and polarization degrees of freedom, photonic systems have been shown to be capable of computing at unprecedented speeds and energy efficiencies^{10–15}.

Driven by these advantages, as well as the rapid development of integrated photonics as a field, photonic computing has emerged as a promising solution for realizing next-generation computing accelerators. Demonstrations using Mach-Zehnder interferometer arrays^{4,16–18}, free space optics^{5,9,19–21}, silicon photonics with on-chip

A full list of affiliations appears at the end of the paper. e-mail: yaowenhu@pku.edu.cn; ysong1@g.harvard.edu; loncar@seas.harvard.edu

attenuators, photodetectors, and ring banks^{8,22–24}, VCSELs with free space optics²⁵, as well as parallel processors utilizing frequency multiplexing^{6,7,26,27} and phase change materials^{7,28,29}, showcase the versatility and potential of the photonics approach. Despite these advances, a critical challenge persists. Low speed and high energy consumption are typically associated with the electro-optic (EO) conversion of data from electronic memory into the optical domain^{15,30}. However, such EO conversion is unavoidable, as the majority of data is stored and processed electronically. Current approaches utilizing variable attenuators or fast EO modulators^{6,7,16–18,22–25,27–29,31}, typically implemented in silicon photonics or bulk systems, are capable of preparing data at rates ranging from the kHz to GHz but are plagued by high electronic energy consumption and optical losses. Alternatively, approaches using static amplitude masks or spatial light modulators can be more energy efficient yet suffer from low refresh rates^{5,8,9,19–21}. As a result, EO conversion remains the most crucial hurdle for practical implementations of photonic computing in real-world applications. Ultimately, high-performance EO modulation is necessary to achieve a high optical data rate that can match the low optical latency while simultaneously maintaining a low energy budget. However, a photonic accelerator that combines high-speed processing with low energy consumption, addressing the challenges faced by traditional EO conversion, is still missing.

Here, we demonstrate high-speed and energy-efficient photonic computation, leveraging the EO effect in a thin-film lithium niobate (TFLN) photonic circuit^{32,33}. Recent progress in the TFLN platform has enabled development of groundbreaking EO devices including modulators^{34–36}, frequency shifters³⁷, and comb generators^{38,39}, among others. The capabilities of these devices stem from the strong EO effect and are facilitated by the tight confinement of both optical and electronic modes as well as low propagation loss⁴⁰. By seamlessly integrating these high-performance EO devices into an optimized large-scale photonic circuit, we build a TFLN computing accelerator that performs matrix-vector-multiplication (MVM) operations and addresses the speed and power limitations of current computing architectures. Complementary to recent demonstrations of TFLN-based optical neural networks^{41,42}, our approach features a high level of component integration as well as a demonstration of a functional integrated EO circuit with a hybrid-integrated laser source and heterogeneous-integrated fast III-V photodiode. This advance represents an important step towards compact, large-scale TFLN-based photonic circuits enhanced by integrated EO conversion, including, but not limited to, photonic computing.

Results

Accelerator architecture

The accelerator imprints electronic data onto the optical domain by modulating optical amplitudes, performs computations of the data through successive EO modulation, and returns the result back to the electrical domain through optical-to-electronic conversion (Fig. 1a). This workflow is implemented using two high-performance TFLN amplitude modulators connected in series, and photonic advantage is achieved through massively parallel multiplexing in the time domain. In the context of fully-connected deep neural networks (DNNs), where MVM operations for data inference is crucial, an input data vector may be encoded onto the amplitude of light by the first TFLN amplitude modulator. The data-modulated light is then fanned-out into distinct spatial channels, each channel subsequently mapping a machine-learned weight vector onto the data vector using a second TFLN amplitude modulator. The intensity of light passing through both modulators is thus proportional to the product of signals applied to each modulator, which amounts to multiplication of the weight vector with the data vector. This intensity is detected using a photodetector and summed electronically, converting the result back into the electrical domain.

Photonic computing circuit implementation on TFLN

To demonstrate the proposed computing core, we designed individual photonic-integrated components and combined them into a single TFLN circuit. Our accelerator contains $M=2$ computing cores, each with $N=16$ channels realized by a spatial fan-out. Multiplication is established owing to the large set of synchronously-functional EO modulators. In practice, one modulator failed during the fabrication process. Further, numerous passive optical elements (e.g. 17×2 grating couplers, and 15×2 Y-splitters for fan-out splitter trees) are fabricated through high-quality, direct dry etching of TFLN, and smooth waveguide sidewalls over the entire circuit lead to a low optical propagation loss of 0.28 dB/cm, achieved by optimizing the reactive ion etching process (Fig. 2d). Low optical propagation loss critically enables a large-scale TFLN circuit such as ours, as well as energy-efficiency through a one-to-sixteen fan-out splitter tree with added insertion losses of 0.135 dB per channel. The microwave circuit for modulation is accomplished using two layers of gold (heights labeled by h_1 and h_2 in Fig. 2e) with nickel-chromium (NiCr) resistors (Fig. 2e). The bottom layer of gold is required for efficient modulation and direct-current (DC) tunability, while the top layer interfaces with resistors. The use of NiCr ensures high-resistance terminators with high-power handling ability and minimal feature size (Fig. 2f). Each 1-cm long modulator features >40 GHz bandwidth (limited by the 40-GHz detector used), >20 dB microwave reflection suppression, switching voltage to modulator length product ($V_{\pi} \cdot L$) ~ 2.2 V \cdot cm, and >20 dB extinction ratio (Fig. 2g, h). We note that the variations in modulator electrical properties are minimal and do not compromise their EO conversion ability, as evaluated in the following section.

High-speed and energy-efficient photonic computing on TFLN

We first demonstrate high-speed vector operations with low energy consumption (Fig. 3a). For generality, we generate as data and weight vectors two random lists of length 1,000. Both vectors are encoded into the time domain at a rate $r=43.8$ GOPS per channel and multiplied after light passes through the data and weight modulators. We note that the vectors are first converted from digital to analog using a high-speed arbitrary waveform generator, and these electronic signals are delivered to the modulators via contact probes. The twice-modulated optical signal, which represents the multiplication result, is detected by a photodetector and digitized using a high-speed oscilloscope. Direct electronic integration of the sampled time traces gives the photonic dot products, and comparison with expected values allows for assessment of computational accuracy per channel (Fig. 3b and see Supplementary information for details). Slight variations in accuracy are likely due to the passive components in each channel, such as input-output coupling and optical path length differences between channels. Furthermore, the time traces are symbol-wise in good agreement with expectation, confirming the high computational speed per channel of our circuit (Fig. 3c). We evaluate the channel by channel performance of the computing core and find that this computational speed (43.8 GOPS/channel) is well maintained. Operating at this speed, we then evaluate the energy consumption while simultaneously assessing computational accuracy under varying optical power conditions. Our experiments reveal a minimum energy consumption of 0.0576 pJ/OP, demonstrating robust computational accuracy with high energy efficiency for the computing circuit (Fig. 3d and see “Computational accuracy” in Supplementary information for details). This assessment includes energy consumption of the pump laser, microwave energy dissipated by the modulators and detectors, as well as other related components (see Methods and “System characterization: speed and energy consumption calculations” in Supplementary information for details). Altogether, our results confirm the high-speed and energy-efficient multiplication of generalized vectors using the TFLN computing system.

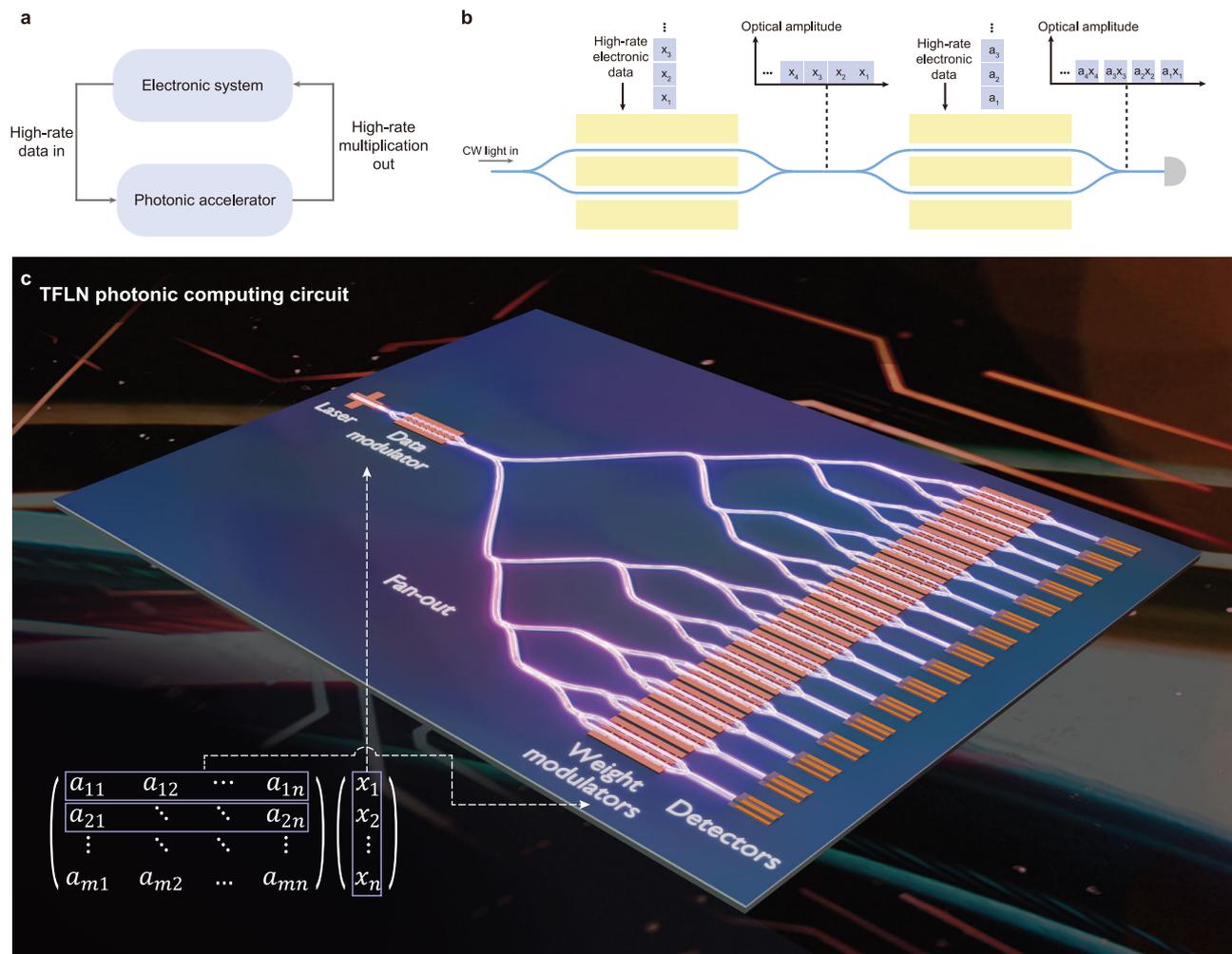


Fig. 1 | Photonic computing accelerator on thin-film lithium niobate. **a** Concept of photonic computing accelerators. Data stored in the electronic system (e.g. a computer) are sent to the photonic accelerator at high rates and are converted into the optical domain. Parallel computations are then performed by the accelerator and results are returned to the electronic system. **b** Illustration of the photonic computing working principle. Continuous-wave light passes through two cascaded amplitude modulators (AMs) which sequentially encode elements of \vec{x} and \vec{a} onto the amplitude of light, effectively performing element-wise multiplication of the two vectors. The components contributing to $\vec{x} \cdot \vec{a}$ are read out by optical-to-

electronic conversion using a low-noise and high-speed detector, and electronic summation of these components finally yields $\vec{x} \cdot \vec{a}$. **c** The vision for a fully integrated computing core based on TFLN photonics, consisting of laser, detectors, and TFLN modulators for high-speed and energy-efficient EO conversion and computation. An input vector is first encoded in the time domain of the optical field through an amplitude modulator and then fanned-out into N spatial channels ($N = 16$ in this figure) to leverage massive spatial parallelism. In each channel, another amplitude modulator is used to multiply weights with the input vector. Finally, detectors convert the multiplication results back into electronic signals.

Inference tasks

We utilize the TFLN circuit to perform computation required by real algorithms. First, we tackle a simple binary classification problem (Fig. 4) over points lying in a two-dimensional plane labeled by “exclusive-or” rule. One such data point $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is input into the circuit for a small number of computing operations to be carried out, and nonlinear activation via an electronic computer is used to obtain the label. The sign of the label indicates the category of the data point inferred by photonics (Fig. 4a). A total of 400 data points is tested, with the classification results shown in the 2D plane (Fig. 4b), achieving an overall accuracy of 93.8% (93.5% on the electronic computer). Further analysis histograms reveal an accuracy of 98.6% (99.1% on the electronic computer) for data points with positive ground truth and 88.4% (87.3% on the electronic computer) for data points with negative ground truth (Fig. 4c). The photonic computing system is thus effective for performing binary classification tasks, supported by the agreement in its classification accuracies with those achieved by traditional electronic computers.

Next, we apply the photonic accelerator to a handwritten digit classification problem over the Modified National Institute of Standards and Technology (MNIST) dataset⁴³. A two-layer feedforward model is chosen to perform the classification (Fig. 5a), with input vectors of size 784 temporally encoded in layer 1, and 10 neurons in layer 2. These input vectors undergo photonic MVM processing followed by electronic softmax activations. Photonic and electronic confusion matrices (Fig. 5b) generated over a 500-image test set indicate that our circuit can efficiently perform computing operations for machine learning inference, as the circuit achieves a classification accuracy of 88%, compared to 92% obtained from electronic computing. In addition to binary classification, our results demonstrate the versatility of our photonic computing system in handling various models, including neural network tasks with vectors of arbitrarily long length. Importantly, we also evaluated the system stability by running identical computation tasks on the photonic circuit for over 20 h, without applying explicit temperature control or adjusting the input/output light coupling to the chip. We observed good stability with a

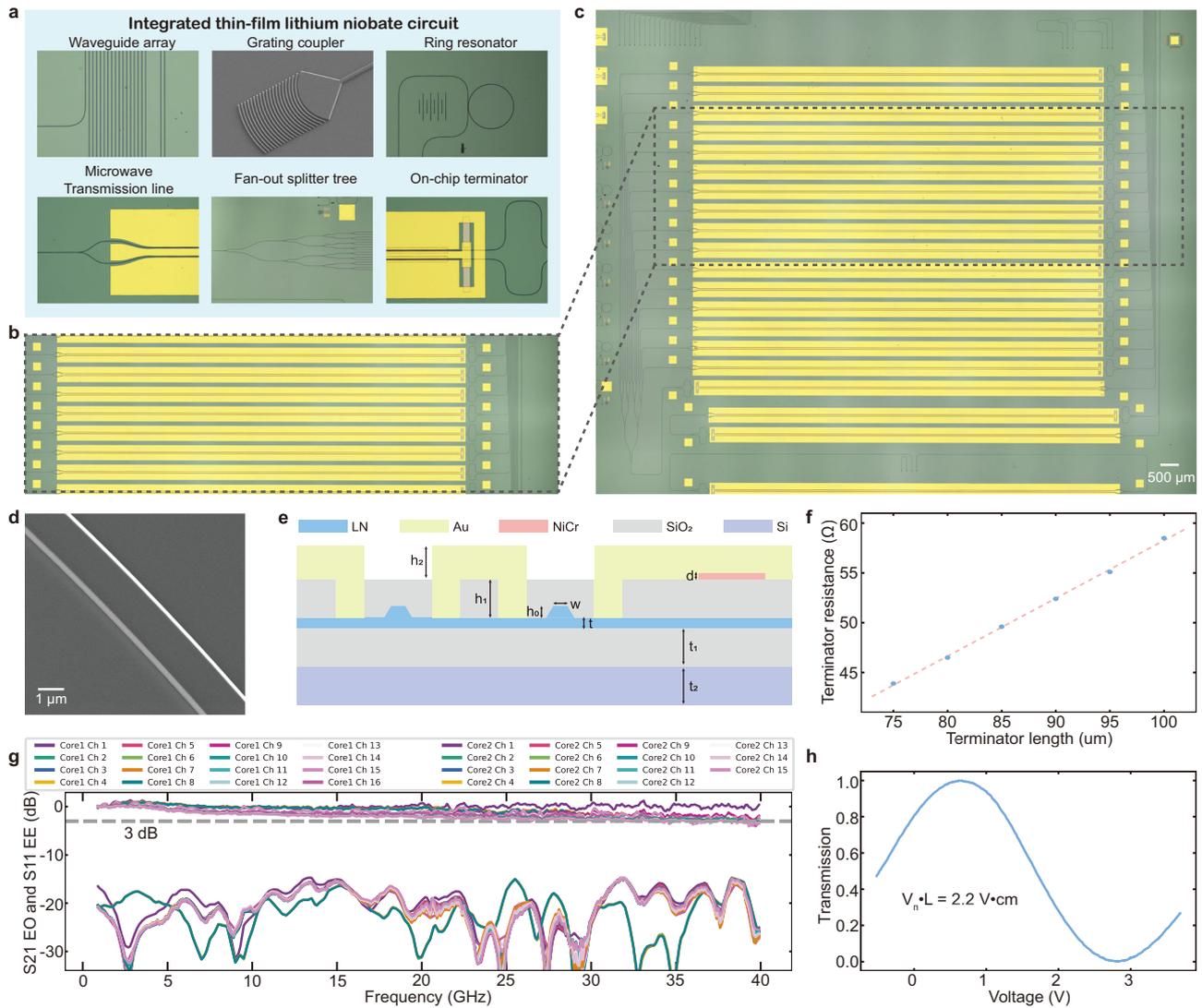


Fig. 2 | Integrated thin-film lithium niobate photonic circuit implementation of computing core. **a** Optical microscope and scanning electron microscope images of the building blocks used in the integrated TFLN circuit: waveguide array (top left) for signal routing; grating coupler (top center) for efficient in- and out-coupling of light; ring resonator (top right) for evaluating the propagation loss and etch quality; fan-out splitter tree (bottom center) to distribute light into distinct spatial channels; microwave transmission line (bottom left) for delivering efficient EO modulation; and on-chip terminator (bottom right) for high-quality microwave impedance matching. **b** Optical microscope image of seven weight modulators in the TFLN accelerator. **c** Full image of one computing core (our circuit contains two such computing cores on the same chip). **d** Scanning electron microscope image of

a high-quality waveguide featuring low propagation loss, enabling low optical energy consumption for the entire circuit. **e** Cross-section illustration of the TFLN circuit, including gold, nickel-chromium (NiCr), lithium niobate, silicon dioxide, and silicon. $d = 100$ nm, $w = 1.5$ μ m, $h_0 = 300$ nm, $h_1 = 1.0$ μ m, $h_2 = 800$ nm, $t = 300$ nm, $t_1 = 4.7$ μ m, $t_2 = 525$ μ m. **f** Measured terminator resistance vs. length. **g** EO forward transmission (S_{21} EO) and electric-electric input reflection (S_{11} EE) response of the modulators in our circuit. Our circuit features a combined total of 31 spatial channels (32 designed, but one failed during the fabrication process). All modulators have a bandwidth beyond 40 GHz (measurement limited by our detector bandwidth). **h** Representative $V_{\pi} \cdot L$ of the modulators (length 1 cm) in our circuit.

standard deviation in the error fluctuation of 0.04% (Fig. 5c and see Supplementary information for details).

Finally, to assess the practical utility of our computing system in solving real-world problems, we extend our demonstrations to image classification over the Canadian Institute for Advanced Research 10-class (CIFAR-10) dataset⁴⁴. Images undergo preprocessing with a convolution layer, a step that could be directly mapped to MVMs and performed with the processor of this work, or accelerated by creating a specialized photonic circuit for convolution operations. In our setup, the circuit is used to propagate preprocessed images through the fully connected layers of the convolutional neural network (CNN). The first layer consists of an input vector of size 1,024 encoded in time, which is subsequently transformed into an intermediate layer with 128 neurons, and then a final output layer with 10 neurons representative of the ten

classes. We tested images in all 10 categories. Examples including trucks, cats, birds, automobiles, frogs, deer, ships, airplanes, and horses are presented, with accurate classification results (Fig. 5d). This confirms our system’s potential for image recognition and analysis, typically requiring large, multi-layer neural networks, since the correct classification of one image requires a large amount of successive computing operations.

Hybrid- and heterogenous-integrated TFLN photonic computing circuit

Next generation high-performance and low power consumption photonic computing circuits would feature lasers and detectors integrated on TFLN chips, alongside active-EO and passive TFLN components. Importantly, wafer-scale fabrication of TFLN (Fig. 6a) is ideally suited

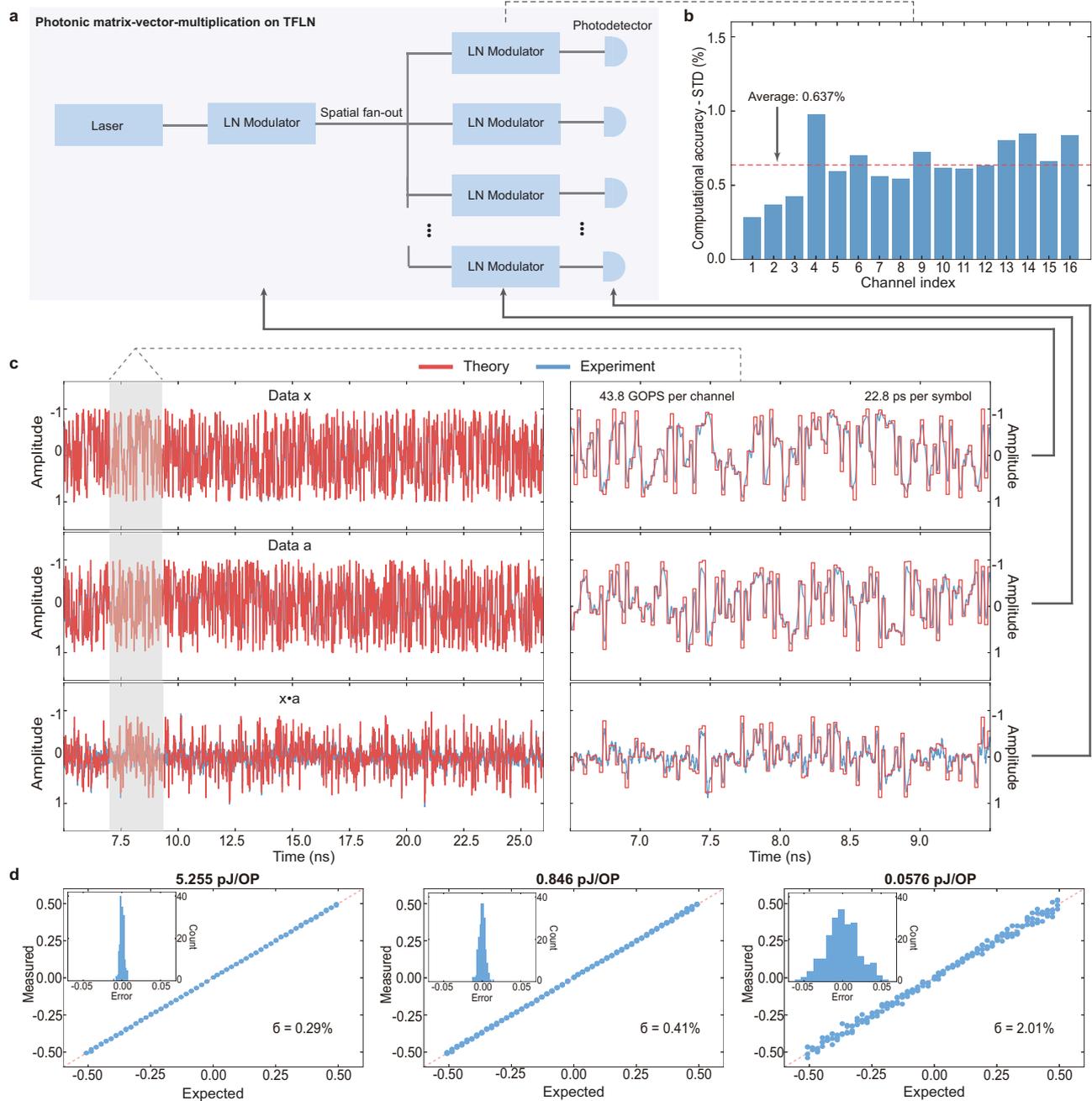


Fig. 3 | High-speed and energy-efficient photonic computing on thin-film lithium niobate. **a** Two-dimensional illustration of the photonic computing core structure. **b** Computational accuracy for different channels in one computing core. The differences in accuracy between channels is minimal though can be further reduced through fine tuning of the system operational parameters. **c** Example waveforms of a temporally-multiplexed computing operation between two random vectors \vec{x} and \vec{a} with 22.8 ps/symbol. Deviations between theory and

experiment can be attributed to limited system bandwidth and residual inaccuracy in time delay between \vec{x} and \vec{a} . Since the final MVM result is a summation of the data trace, it averages out these deviations and further reduces the error.

d Computational accuracy vs. energy consumption by varying the optical power, showing a lowest energy consumption of 0.0576 pJ/OP at 22.8 ps/symbol, while still maintaining computational accuracy. The inset gives error of the computation (Error = Measured-Expected). The σ is the standard deviation of the error.

for the batch production of stand-alone optical computing cores (Fig. 6b, c), and it would unlock significant potential for greater spatial parallelism, in conjunction with the already demonstrated temporal-multiplexing and efficient EO conversion characteristics unique to the TFLN photonic computing platform.

As an important first step in this direction, we fabricated a TFLN photonic computing accelerator chip with heightened degree of integration but featuring a single spatial channel. Our chip consists of active EO modulators and passive components (waveguides, Y-splitters, on-chip terminators, etc.), with the addition of bilayer-taper edge

couplers that are used for hybrid-integration of a distributed feedback (DFB) laser source with the TFLN chip, and heterogeneous-integration of a modified uni-traveling-carrier photodiode (MUTC-PD) evanescently coupled to TFLN signal waveguides (Fig. 6d–h). We calibrate the DFB laser power against injection current and find a lasing threshold of about 50 mA (Fig. 6i) with 0.25 W/A slope efficiency. We also measure the MUTC-PD dark current as function of bias voltage (Fig. 6j) and find a low dark current of 2.7 nA when reverse biased at -2 V for high-speed operation (see Methods and Supplementary information for further details on DFB characterization and PD fabrication). Given these

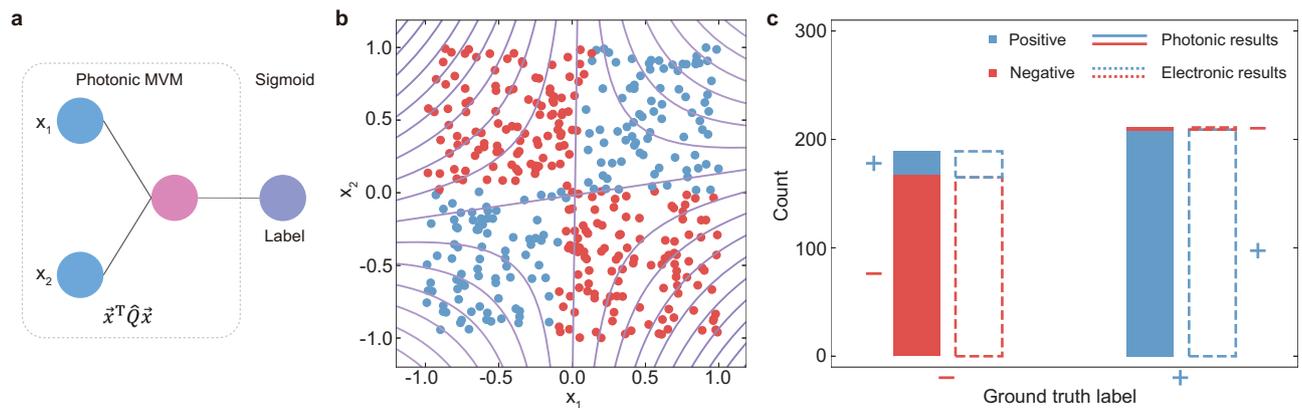


Fig. 4 | Photonic binary classification. **a** Illustration of binary classification conducted over a data set consisting of two-dimensional vectors $\vec{x} = [x_1, x_2]^T$, where each vector is labeled positive or negative based on the exclusive-or condition. To infer the label of some \vec{x} , a number of computing operations are carried out by our photonic accelerator to compute $\vec{x}^T Q \vec{x}$ followed by electronic nonlinear activation. Here, Q is a pre-trained kernel matrix and nonlinearity is electronically applied.

Accuracy of inference thus depends on the accuracy of photonic computations. **b** Classification of 400 randomly selected \vec{x} in the problem space, colored by their photonics-inferred positive (blue) and negative (red) labels. **c** Comparison between photonic and electronic classification results over the 400-vector test set, which shows good agreement between photonic and electronic computing. The photonic circuit (electronic computer) achieves a classification accuracy of 93.8% (93.5%).

performances, we evaluate key performance metrics for computing, such as high-speed and low energy consumption computations, by operating the single spatial channel at a rate $r = 10.33$ GOPS (Fig. 6k) and assessing the computational accuracy vs. various full-system energy consumptions (Fig. 6l). We show high-fidelity encoding and multiplication of pairs of random vectors, as well as computational accuracies characterized by $\sigma < 0.5\%$ across all energy consumptions attempted. Here, the reduction in demonstrated data rate (compared to previously demonstrated 43.8 GOPS per channel) is attributed to imperfect on-chip termination resistance. The larger energy consumptions overall (compared to previously demonstrated 0.0576 pJ/OP) are mostly due to overdriving the laser source in our proof-of-concept demonstration. We note that the DFB source operated just above lasing threshold already provides more optical power than needed for accurate computation utilizing a single spatial channel.

Discussion

In conclusion, our work presents a photonic computing acceleration architecture on TFLN, capable of performing increasingly complex algorithmic tasks, from binary classification, handwritten digit classification, to actual image classification. Testing larger datasets, computing with models containing more parameters, and introducing environmental disturbances during computation could further assess the system's performance. Unlike architectures designed for specific applications such as convolution or vision-based tasks, ours is well suited for general computing tasks. Our experimental demonstration is enabled by a large scale, system-level TFLN circuit. Leveraging the TFLN platform, we include the process of EO conversion in our demonstration to achieve high speed and low energy consumption photonic computing. This is made possible by increasing the scale of TFLN integration while maintaining the high performance of individual components. We then advance our system by fabricating on-chip detectors and replacing the bench-top laser with DFB source butt-coupled to our TFLN chip. Future improvements to enhance performance metrics to match or exceed state-of-the-art photonic accelerators are straightforward: reduce the V_π to $1V$ ³⁶, increase the bandwidth to 100 GHz^{36,45,46}, and minimize the optical propagation loss of waveguides to 0.03 dB/cm^{40,47}, each of which has been previously demonstrated in TFLN devices. State-of-the-art modulator designs leveraging multi-layer capacitance loading and folded electrodes may further improve the modulation efficiency per chip area, resulting in compact EO modulator networks also featuring a high production

yield⁴⁸⁻⁵⁰. While these modulator advances are mainly driven by the needs of optical communication technology, they can readily be adopted in, and will have positive impact on, photonic computing architectures such as ours. Transitioning the accelerator to visible wavelengths may further reduce energy consumption as well as improve scalability, leveraging sub-volt TFLN modulators^{51,52} that may also be shorter in length. Utilizing the frequency degree-of-freedom with TFLN soliton microcombs⁵³ may extend applicability to photonic convolution acceleration. Various optical packaging methods, including photonic wire bonding⁵⁴, can improve the stability of light coupling to the chip, which may enable accurate computations even in the presence of environmental disturbances. For a comprehensive contextualization against state-of-the-art electronic and photonic systems, as well as future projections in performance, see "System characterization: speed and energy consumption calculations" in Supplementary information.

It is worth noting that our approach to address the EO conversion challenge in photonic computing is compatible with other photonic accelerators. Thus, our work may motivate novel hybrid approaches (e.g. combining TFLN EO conversion with free space optics) that require high bandwidth and low power modulation for data encoding. TFLN stands out as the most powerful platform for this task, owing to its exceptional EO conversion capability, thus offering the potential to overcome existing speed and energy bottlenecks in various optical computation schemes. While we did not operate all channels of our computing circuit simultaneously and only examined them separately, due to limited testing equipment, the associated challenge of potential crosstalk between adjacent TFLN modulators has recently been experimentally shown to not play a role⁴⁶. Moreover, the full potential of the platform can be unlocked by leveraging photonic-electronic integration through the nascent lateral and 3D integration strategies⁵⁵⁻⁵⁹, especially integration with high-speed electronic circuits, including multi-channel DACs, ADCs, and FPGAs to utilize the spatial multiplexing capability. In this context, a comprehensive understanding of not just the energy consumptions of novel electronics, but also the uncertainty in their estimations, is critically required for accurately mapping the path forward of TFLN photonics for computing. More broadly, we believe that large-scale TFLN circuits similar to those demonstrated here may hold great promise for applications in vision⁹, sensing²¹, ranging⁶⁰⁻⁶², and even quantum computing⁶³⁻⁶⁵, and we hope that our work will stimulate further exploration of such applications.

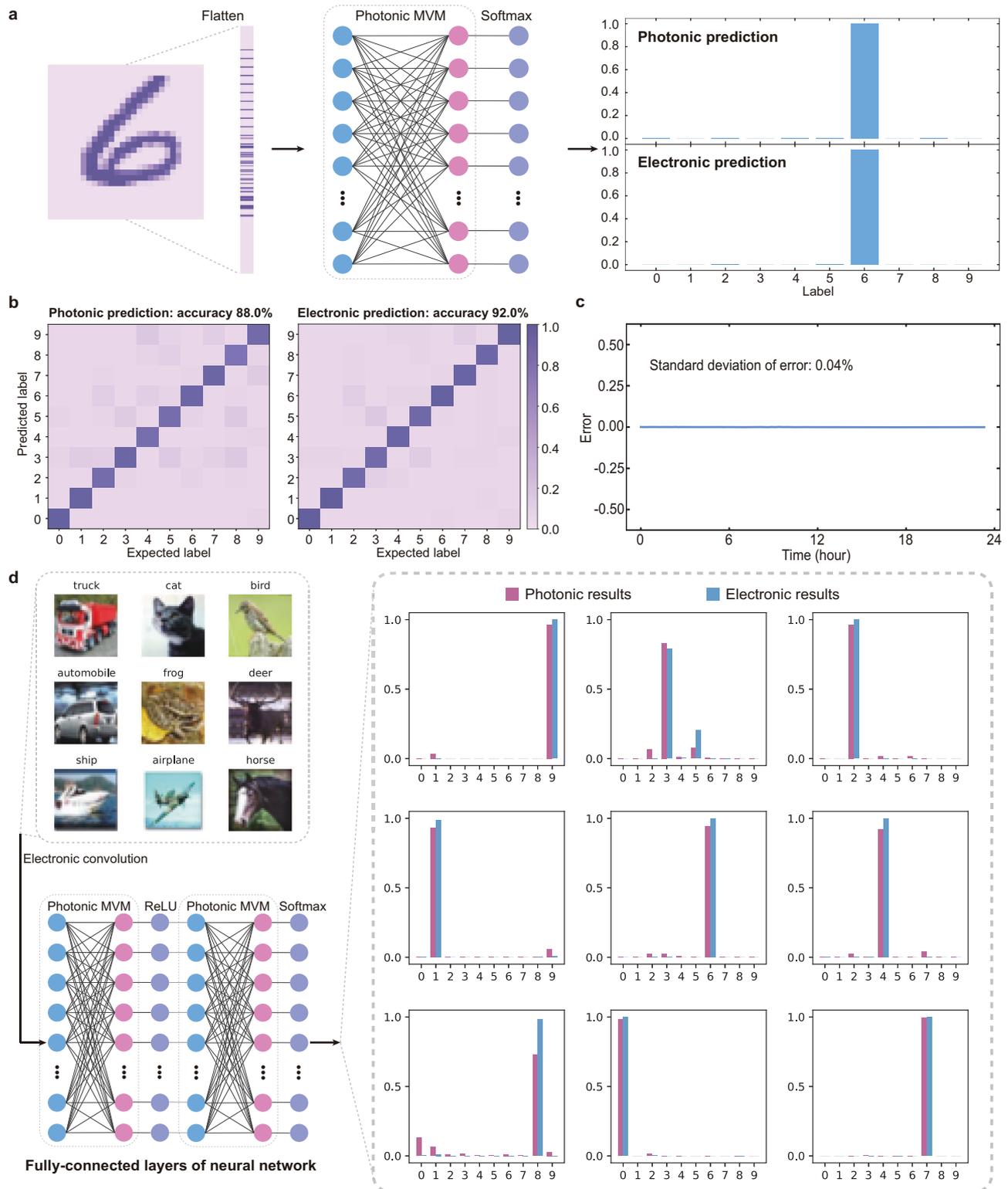


Fig. 5 | Photonic computing for fully-connected layers of photonic neural networks. **a** Classification of an MNIST handwritten digit. The image is flattened into a single vector encoded in the time domain. An example image (number six) is shown on the left. A two-layer photonic neural network is used to perform the classification task (center). The photonic computing result at the end of the network is then sent back to the computer to perform a nonlinear activation (right). The final classification results agree well with the electronically computed result. **b** Statistics of MNIST handwritten digit recognition. 500 MNIST images are selected as the test set and processed. The confusion matrices show a classification accuracy of 88% using our circuit (92% using electronic computer). **c** Stability. The

computing core is programmed to continuously run identical computing tasks over 20 h, experiencing minimal fluctuations of 0.04%. **d** Real image classification. Images resembling real-life objects are selected from the CIFAR-10 database and classified using a convolutional neural network. The images are preprocessed through convolution layers, flattened into vectors, and then sent into our circuit to be classified by the remaining fully-connected layers (bottom left). Nine example figures including a truck, cat, bird, automobile, frog, deer, ship, airplane, and horse (top left), together with the classification results (right) are shown, indicating our photonic computing circuit can accurately process large, multi-layer networks.

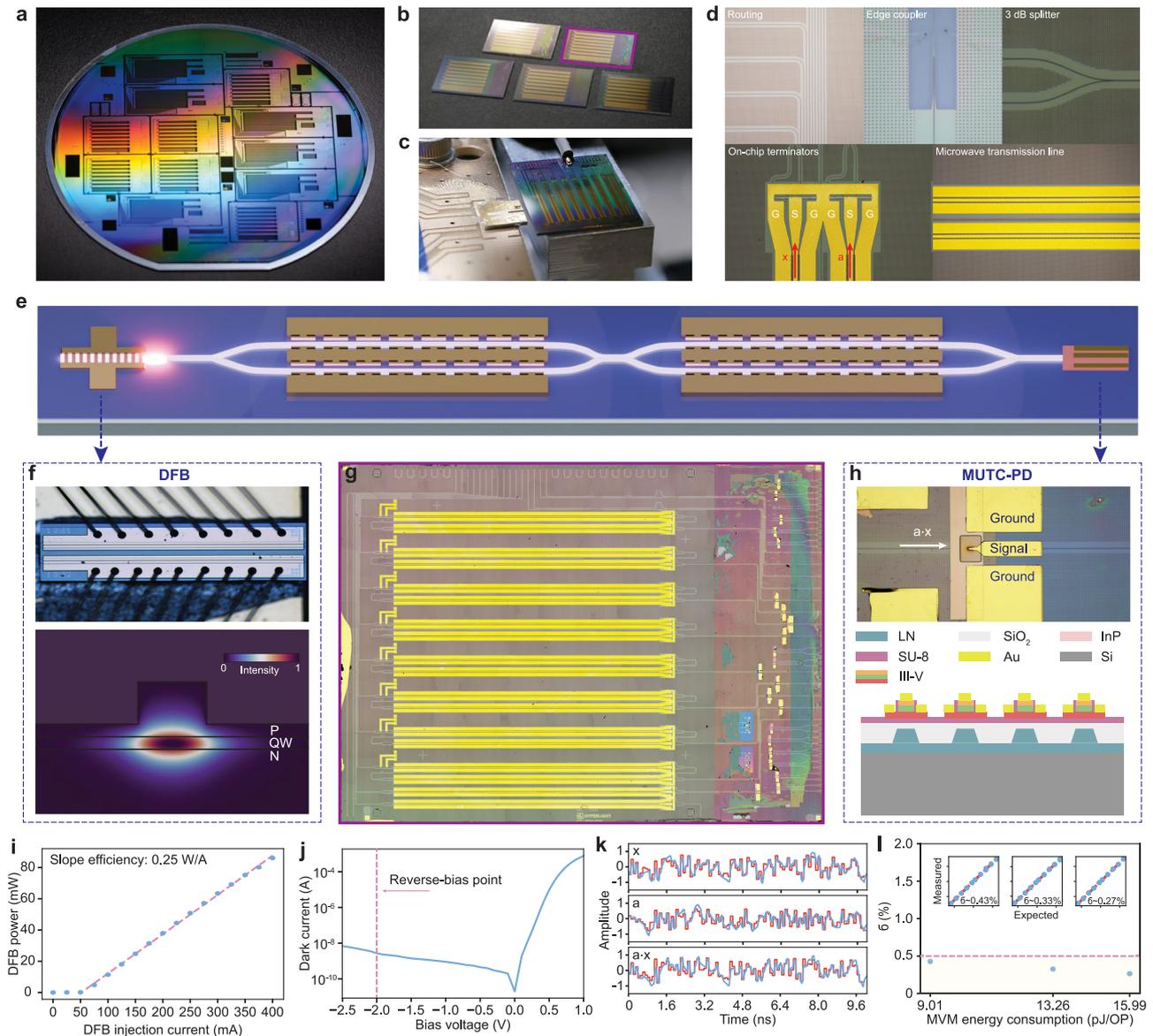


Fig. 6 | Hybrid- and heterogenous- integrated TFLN photonic computing circuit. **a** Wafer-scale fabrication of computing cores comprising a TFLN photonic computing circuit. **b** Chiplets of TFLN computing cores from the wafer-scale process. **c** Measurement setup for characterizing the hybrid- and heterogeneous-integrated system: light from a hybrid-integrated DFB laser source is butt-coupled to the TFLN computing core, while a heterogeneous-integrated MUTC-PD is used to perform optical-to-electronic conversion of the computing signal (the electrical signal is extracted by contact probes). **d** Optical microscope images of essential building blocks used for the integrated TFLN circuit similar to Fig. 2a, except bilayer-taper edge couplers (top middle) are employed as a low-loss interface between the TFLN waveguide mode and the DFB waveguide mode. **e** Schematic of a single channel integrated system. **f** Optical microscope image of DFB laser and 2-D simulation of the DFB waveguide mode. P (N): positively- (negatively-) doped region; QW: quantum well. **g** Optical microscope image of TFLN photonic computing circuit, with an array of cascaded amplitude modulators (left) and an array of MUTC-PDs (right). **h** Optical microscope image of MUTC-PD and schematic of its cross section. **i** DFB output power measured by an integrating sphere vs. injection current, with about 50 mA current threshold and 0.25 W/A slope efficiency. **j** MUTC-PD dark current vs. bias voltage. For high-speed operation, a reverse-bias of -2 V is held for all measurements. **k** Example waveforms of a temporally multiplexed computing operation between two random vectors \vec{x} and \vec{a} with 96.8 ps/symbol (10.33 GOPS per channel). **l** Computational accuracy σ , as previously defined, vs. computing energy consumption as the DFB output power is varied. Three energy consumptions are evaluated based on experimental conditions provided DFB injection currents of 85, 125, and 150 mA, respectively.

Methods

Device fabrication

The devices in our study are fabricated using a commercial X-cut thin-film lithium niobate (TFLN) on insulator wafer (NANOLN). The wafer comprises a 600 nm-thick LN layer, a 4.7 μm buried oxide layer (thermally grown), all mounted on a 525 μm -thick silicon (Si) handle. The fabrication process involves the following steps. (1) Patterning of the optical layer: electron-beam lithography is used to define patterns for the optical layer of the device, including rib waveguides and

microring resonators. Ar⁺-based reactive ion etching is then used to etch the optical layer down by 300 nm. (2) Defining the bottom-layer electrode: microwave electrodes are defined using a combination of photolithography, electron-beam evaporation, and a bilayer lift-off process. Using this approach, an 800 nm-thick layer of gold (Au) is deposited to form the bottom layer of the microwave electrodes. (3) Cladding with SiO₂: the devices are cladded with a 1.0 μm -thick layer of silicon dioxide (SiO₂) using plasma-enhanced chemical vapor deposition (PECVD). (3) Depositing the nickel-chromium resistor: the nickel-

microring resonators. Ar⁺-based reactive ion etching is then used to etch the optical layer down by 300 nm. (2) Defining the bottom-layer electrode: microwave electrodes are defined using a combination of photolithography, electron-beam evaporation, and a bilayer lift-off process. Using this approach, an 800 nm-thick layer of gold (Au) is deposited to form the bottom layer of the microwave electrodes. (3) Cladding with SiO₂: the devices are cladded with a 1.0 μm -thick layer of silicon dioxide (SiO₂) using plasma-enhanced chemical vapor deposition (PECVD). (3) Depositing the nickel-chromium resistor: the nickel-

chromium (NiCr) layer is defined using photolithography with deposition through electron-beam evaporation, followed by a bilayer lift-off process. (4) Defining the top-layer electrode: a second layer of Au is deposited using a combination of photolithography, electron beam evaporation, and bilayer lift-off. The thickness of the top layer gold is 800 nm. The fabrication process results in a TFLN circuit with high-quality optical components and a well-defined microwave circuit, which together enable high-performance computing operations on TFLN.

The on-chip PD is based on the modified uni-traveling carrier (MUTC) structure with optimized transit time limited bandwidth of 165 GHz⁶⁶. They are heterogeneously integrated on TFLN using adhesive bonding with SU-8, involving the following steps. (1) The bare epitaxial structure is bonded on the TFLN photonic chip with ~100 nm-thick SU-8. (2) The InP substrate is removed using HCl wet etch. (3) PD mesas are formed by a combination of standard III-V material dry etch and wet etch recipes. (4) PDs are passivated using SU-8 first before forming RF probe pads to improve yield and dark current. (5) The SU-8 is fully cured and was compatible with other post processing steps needed. We note that the RF pads were designed to be impedance matched to 50 Ω in a ground-signal-ground configuration with 150 μm pitch, matching the available contact probes.

Experimental architecture and setup

Photonic computing for matrix-vector-multiplication using our TFLN circuit is based on the cascaded amplitude modulation of light, which takes advantage of the efficient and high-speed EO modulation characteristics of TFLN. Continuous wave light in the telecommunications C-band is coupled into the chips. This light then passes through the first amplitude modulator (AM) which encodes the data vector \vec{x} onto the light's amplitude. The modulated optical signal is split into multiple spatial channels, using the on-chip fan-out splitting tree, each containing another AM. When these AMs are driven in time-synchrony with the data encoded using the first AM, they apply the weight vector \vec{a} resulting in overall amplitude encoding corresponding to the elements comprising the sum of $\vec{x} \cdot \vec{a}$, which can be converted back to the electrical domain by photodetection. The generated photocurrent tracks amplitude fluctuations of the impinging light up to the detection bandwidth, and the components of $\vec{x} \cdot \vec{a}$ can be directly digitized by an oscilloscope (Keysight UXR1102A). An electronic summation of the components yields the final result, $\vec{x} \cdot \vec{a}$. To enable dot products between signed vectors, we adopt an approach proposed in earlier work²⁰ that converts dot products between signed vectors to that between unsigned vectors, requiring two overhead calculations by the photonic accelerator. This overhead can be removed in future work albeit with higher system-level complexity, for example by integrating balanced photodetection⁶⁷. In our experiment, the electronic signals \vec{x} and \vec{a} are generated by a high-speed arbitrary waveform generator (AWG, Keysight M8196A) and delivered to on-chip microwave transmission lines via high-speed cables and contact probes (GGB 40A-GSG). Due to limitations in the electronic equipment, all interference tasks utilize a single neuron (channel) on the TFLN chip. We note that the synchronization of \vec{x} and \vec{a} electrical drives is accomplished through adjusting the delay between them, up to an integer multiple of the sampling rate of the AWG. Both drives are generated by the same AWG. For the full spatial-multiplexing to be utilized, synchronization between each \vec{x}_i and \vec{a} is possible by using multiple synchronized DACs and delay adjustment.

Data availability

Data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Information. The image datasets MNIST and CIFAR-10 (<https://www.cs.toronto.edu/~kriz/cifar.html>) are publicly available.

References

1. Yao, P. et al. Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).
2. Wan, W. et al. A compute-in-memory chip based on resistive random-access memory. *Nature* **608**, 504–512 (2022).
3. Wright, L. G. et al. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).
4. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
5. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
6. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
7. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
8. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
9. Chen, Y. et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* **623**, 48–57 (2023).
10. Nahmias, M. A. et al. Photonic multiply-accumulate operations for neural networks. *IEEE J. Sel. Top. Quant. Electron.* **26**, 1–18 (2020).
11. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
12. Sui, X., Wu, Q., Liu, J., Chen, Q. & Gu, G. A review of optical neural networks. *IEEE Access* **8**, 70773–70783 (2020).
13. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).
14. Berggren, K. et al. Roadmap on emerging hardware and technology for machine learning. *Nanotechnology* **32**, 012002 (2021).
15. McMahon, P. L. The physics of optical computing. *Nat. Rev. Phys.* **5**, 717–734 (2023).
16. Xu, S. et al. Optical coherent dot-product chip for sophisticated deep learning regression. *Light Sci. Appl.* **10**, 221 (2021).
17. Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).
18. Zhu, H. H. et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044 (2022).
19. Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
20. Wang, T. et al. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **13**, 123 (2022).
21. Wang, T. et al. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photonics* **17**, 408–415 (2023).
22. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
23. Tait, A. N. et al. Silicon photonic modulator neuron. *Phys. Rev. Appl.* **11**, 064043 (2019).
24. Huang, C. et al. A silicon photonic–electronic neural network for fibre nonlinearity compensation. *Nat. Electron.* **4**, 837–844 (2021).
25. Chen, Z. et al. Deep learning with coherent VCSEL neural networks. *Nat. Photonics* **17**, 723–730 (2023).
26. Xu, X. et al. Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks. *Laser Photonics Rev.* **14**, 2000070 (2020).
27. Bai, B. et al. Microcomb-based integrated photonic processing unit. *Nat. Commun.* **14**, 66 (2023).
28. 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).
29. Wu, C. et al. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **12**, 96 (2021).

30. Zhou, H. et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci. Appl.* **11**, 30 (2022).
31. Shi, B., Calabretta, N. & Stabile, R. Deep neural network through an InP SOA-based photonic integrated cross-connect. *IEEE J. Sel. Top. Quantum Electron.* **26**, 1–11 (2020).
32. Zhu, D. et al. Integrated photonics on thin-film lithium niobate. *Adv. Opt. Photonics* **13**, 242 (2021).
33. Hu, Y. et al. Integrated electro-optics on thin-film lithium niobate. *Nat. Rev. Phys.* **7**, 237–254 (2025).
34. Wang, C. et al. Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* **562**, 101–104 (2018).
35. He, M. et al. High-performance hybrid silicon and lithium niobate Mach–Zehnder modulators for 100 Gbit s⁻¹ and beyond. *Nat. Photonics* **13**, 359–364 (2019).
36. Xu, M. et al. Dual-polarization thin-film lithium niobate in-phase quadrature modulators for terabit-per-second transmission. *Optica* **9**, 61 (2022).
37. Hu, Y. et al. On-chip electro-optic frequency shifters and beam splitters. *Nature* **599**, 587–593 (2021).
38. Zhang, M. et al. Broadband electro-optic frequency comb generation in a lithium niobate microring resonator. *Nature* **568**, 373–377 (2019).
39. Hu, Y. et al. High-efficiency and broadband on-chip electro-optic frequency comb generators. *Nat. Photonics* **16**, 679–685 (2022).
40. Zhu, X. et al. Twenty-nine million intrinsic Q-factor monolithic microresonators on thin-film lithium niobate. *Photonics Res.* **12**, A63 (2024).
41. Lin, Z. et al. 120 GOPS photonic tensor core in thin-film lithium niobate for inference and in situ training. *Nat. Commun.* **15**, 9081 (2024).
42. Ou, S. et al. Hypermultiplexed integrated photonics-based optical tensor processor. *Sci. Adv.* **11**, eadu0228 (2025).
43. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
44. Krizhevsky, A. & Hinton, G. *Learning Multiple Layers Of Features From Tiny Images*. (University of Toronto, 2009).
45. Kharel, P., Reimer, C., Luke, K., He, L. & Zhang, M. Breaking voltage–bandwidth limits in integrated lithium niobate modulators using micro-structured electrodes. *Optica* **8**, 357 (2021).
46. St-Arnault, C. et al. Net 3.2 Tbps 225 Gbaud PAM4 O-band IM/DD 2 km transmission using FR8 and DR8 with a CMOS 3 nm SerDes and TFLN modulators. *arXiv* <https://doi.org/10.48550/arXiv.2503.24147> (2025).
47. Zhang, M., Wang, C., Cheng, R., Shams-Ansari, A. & Lončar, M. Monolithic ultra-high-Q lithium niobate microring resonator. *Optica* **4**, 1536 (2017).
48. Song, Y. et al. Integrated electro-optic digital-to-analog link for efficient computing and arbitrary waveform generation. *arXiv* <https://doi.org/10.48550/arXiv.2411.04395> (2024).
49. Della Torre, A. et al. Folded electro-optical modulators operating at CMOS voltage level in a thin-film lithium niobate foundry process. *Opt. Express* **33**, 6747 (2025).
50. Wang, J. et al. Highly tunable flat-top thin-film lithium niobate electro-optic frequency comb generator with 148 comb lines. *Opt. Express* **33**, 23431 (2025).
51. Renaud, D. et al. Sub-1 volt and high-bandwidth visible to near-infrared electro-optic modulators. *Nat. Commun.* **14**, 1496 (2023).
52. Xue, S. et al. Full-spectrum visible electro-optic modulator. *Optica* **10**, 125 (2023).
53. Song, Y., Hu, Y., Zhu, X., Yang, K. & Lončar, M. Octave-spanning Kerr soliton frequency combs in dispersion- and dissipation-engineered lithium niobate microresonators. *Light Sci. Appl.* **13**, 225 (2024).
54. Franken, C. A. A. et al. High-power and narrow-linewidth laser on thin-film lithium niobate enabled by photonic wire bonding. *APL Photonics* **10**, 026107 (2025).
55. Rizzo, A. et al. Massively scalable Kerr comb-driven silicon photonic link. *Nat. Photonics* **17**, 781–790 (2023).
56. Chang, P.-H. et al. A 3D Integrated energy-efficient transceiver realized by direct bond interconnect of co-designed 12 nm FinFET and silicon photonic integrated circuits. *J. Light. Technol.* **41**, 6741–6755 (2023).
57. Daudlin, S. et al. Three-dimensional photonic integration for ultra-low-energy, high-bandwidth interchip data links. *Nat. Photonics* **19**, 502–509 (2025).
58. Ahmed, S. R. et al. Universal photonic artificial intelligence acceleration. *Nature* **640**, 368–374 (2025).
59. Hua, S. et al. An integrated large-scale photonic accelerator with ultralow latency. *Nature* **640**, 361–367 (2025).
60. Riemensberger, J. et al. Massively parallel coherent laser ranging using a soliton microcomb. *Nature* **581**, 164–170 (2020).
61. Zhang, X., Kwon, K., Henriksson, J., Luo, J. & Wu, M. C. A large-scale microelectromechanical-systems-based silicon photonics LiDAR. *Nature* **603**, 253–258 (2022).
62. Zhu, S. et al. Integrated lithium niobate photonic millimetre-wave radar. *Nat. Photonics* **19**, 204–211 (2025).
63. O’Brien, J. L., Furusawa, A. & Vučković, J. Photonic quantum technologies. *Nat. Photonics* **3**, 687–695 (2009).
64. Kues, M. et al. Quantum optical microcombs. *Nat. Photonics* **13**, 170–179 (2019).
65. Wang, J., Sciarino, F., Laing, A. & Thompson, M. G. Integrated photonic quantum technologies. *Nat. Photonics* **14**, 273–284 (2020).
66. Guo, X. et al. High-performance modified uni-traveling carrier photodiode integrated on a thin-film lithium niobate platform. *Photonics Res.* **10**, 1338 (2022).
67. Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M. & Englund, D. Large-Scale Optical Neural Networks Based on Photoelectric Multiplication. *Phys. Rev. X* **9**, 021032 (2019).

Acknowledgements

The authors thank Kyle Richard, Neil Hoffman, Mark Roberts, and John Dorighi (Keysight Technologies, Inc.) for technical support. This work is supported in part by DARPA LUMOS (HR001-20-C-0137), NSF (EEC-1941583, OMA-2137723, OMA-2138068), ONR (N00014-22-C-1041), NASA (231180A/ 80NSSC22K0262). D.Z. acknowledges support from the Harvard Quantum Initiative, NRF and A*STAR Quantum Engineering Program (NRF2022-QEP2-01-PO7), and NRF Fellowship (NRF-NRFF15-2023-0005). S.L. acknowledges support from the A*STAR NSS (PhD) scholarship. L.M. acknowledges support from the Capes-Fulbright and Behring Foundation fellowships. These views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Distribution Statement “A” (Approved for Public Release, Distribution Unlimited).

Author contributions

Y.H. conceived the project. Y.H. and Y.S. performed the experiment with S.L. and X.Z. assisting. X.Z., S.L., and Y.S. fabricated the TFLN circuit with Y.H. assisting. X.G. fabricated the photodetector. Q.Z. and G.B. developed the binary classification problem. Y.H., L.H., X.G. designed the circuit for III-V TFLN heterogeneous integration. C.F., K.P., H.W., D.A., D.R., Y.W., L.M., V.R., A.S., X.L., R.C., K.L., K.Y., M.Z., D.Z., L.J., N.S. helped with the project. Y.H., Y.S., and M.L. wrote the manuscript with contributions from all authors. M.L. and A.B. supervised the project.

Competing interests

L.H., Y.W., K.L., M.Z., and M.L. are involved in developing lithium niobate technologies at HyperLight Corporation. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62635-8>.

Correspondence and requests for materials should be addressed to Yaowen Hu, Yunxiang Song or Marko Lončar.

Peer review information *Nature Communications* thanks Lin Chang, Stefano Paesani, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. ²State Key Laboratory for Mesoscopic Physics and Frontiers Science Center for Nano-optoelectronics, School of Physics, Peking University, Beijing 100871, China. ³Quantum Science and Engineering, Harvard University, Cambridge, MA 02138, USA. ⁴Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22903, USA. ⁵Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ⁶HyperLight Corporation, 675 Massachusetts Ave, Cambridge, MA 02139, USA. ⁷Laser Physics and Nonlinear Optics Group, Department of Science and Technology, MESA+ Institute of Nanotechnology, University of Twente, Enschede, The Netherlands. ⁸Freedom Photonics, 41 Aero Camino, Goleta, CA, USA. ⁹DRS Daylight Solutions, 16465 Via Esprillo, San Diego, CA, USA. ¹⁰Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ¹¹Department of Materials Science and Engineering, National University of Singapore, Singapore 117575, Singapore. ¹²A*STAR Quantum Innovation Centre (Q.InC), Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore. ¹³These authors contributed equally: Yaowen Hu, Yunxiang Song.

✉ e-mail: yaowenhu@pku.edu.cn; ysong1@g.harvard.edu; loncar@seas.harvard.edu