

DOI: [10.29026/oea.2021.200060](https://doi.org/10.29026/oea.2021.200060)

All-optical computing based on convolutional neural networks

Kun Liao¹, Ye Chen¹, Zhongcheng Yu¹, Xiaoyong Hu^{1,2*},
Xingyuan Wang^{3*}, Cuicui Lu⁴, Hongtao Lin^{5*}, Qingyang Du⁶, Juejun Hu⁶
and Qihuang Gong^{1,2}

The rapid development of information technology has fueled an ever-increasing demand for ultrafast and ultralow-energy-consumption computing. Existing computing instruments are pre-dominantly electronic processors, which use electrons as information carriers and possess von Neumann architecture featured by physical separation of storage and processing. The scaling of computing speed is limited not only by data transfer between memory and processing units, but also by RC delay associated with integrated circuits. Moreover, excessive heating due to Ohmic losses is becoming a severe bottleneck for both speed and power consumption scaling. Using photons as information carriers is a promising alternative. Owing to the weak third-order optical nonlinearity of conventional materials, building integrated photonic computing chips under traditional von Neumann architecture has been a challenge. Here, we report a new all-optical computing framework to realize ultrafast and ultralow-energy-consumption all-optical computing based on convolutional neural networks. The device is constructed from cascaded silicon Y-shaped waveguides with side-coupled silicon waveguide segments which we termed “weight modulators” to enable complete phase and amplitude control in each waveguide branch. The generic device concept can be used for equation solving, multifunctional logic operations as well as many other mathematical operations. Multiple computing functions including transcendental equation solvers, multifarious logic gate operators, and half-adders were experimentally demonstrated to validate the all-optical computing performances. The time-of-flight of light through the network structure corresponds to an ultrafast computing time of the order of several picoseconds with an ultralow energy consumption of dozens of femtojoules per bit. Our approach can be further expanded to fulfill other complex computing tasks based on non-von Neumann architectures and thus paves a new way for on-chip all-optical computing.

Keywords: convolutional neural networks; all-optical computing; mathematical operations; cascaded silicon waveguides

Liao K, Chen Y, Yu ZC, Hu XY, Wang XY et al. All-optical computing based on convolutional neural networks. *Opto-Electron Adv* 4, 200060 (2021).

¹State Key Laboratory for Mesoscopic Physics & Department of Physics, Collaborative Innovation Center of Quantum Matter, Beijing Academy of Quantum Information Sciences, Nano-optoelectronics Frontier Center of Ministry of Education, Peking University, Beijing 100871, China; ²Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China; ³College of Mathematics and Physics, Beijing University of Chemical Technology, Beijing 100029, China; ⁴Beijing Key Laboratory of Nanophotonics and Ultrafine Optoelectronic Systems, School of Physics, Beijing Institute of Technology, Beijing 100081, China; ⁵College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310027, China; ⁶Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*Correspondence: XY Hu, E-mail: xiaoyonghu@pku.edu.cn; XY Wang, E-mail: wang_xingyuan@mail.buct.edu.cn;

HT Lin, E-mail: hometown@zju.edu.cn

Received: 29 September 2020; Accepted: 14 December 2020; Published online: 25 November 2021



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021. Published by Institute of Optics and Electronics, Chinese Academy of Sciences.

Introduction

The demand for ultrahigh-speed and energy-efficient computing¹ has been increasing exponentially driven by the rapid development of advanced engineering calculations, economic data analysis, and cloud computing. Traditional electronic processors, the pre-dominant computing platform to date, adopt the von Neumann architecture²⁻³ where storage and processing units are physically separated. In von Neumann processors, limited data communication bandwidth between the memory and processing units as well as RC delay of integrated circuits have become major barriers towards continuing scaling of computing speed⁴⁻⁵. Moreover, heat dissipation due to resistive losses in electrical wires severely compromises energy efficiency of traditional electronic processors⁶. These limitations make it difficult to realize high speed and low energy consumption simultaneously⁷⁻⁸. Specialized processors, such as graphic process units designed for mathematical calculation tasks and field programmable gate arrays specialized for arithmetic logic operations, utilize modified Von Neumann architecture^{5,9}, which however still suffer from the speed and energy consumption limitations. All-optical computing adopting photons as information carriers offers a promising alternative approach. To date, optical computing usually relies on third-order optical nonlinearity to implement all-optical control, which requires ultrafast response and giant third-order optical nonlinearity in photonic materials¹⁰⁻¹¹. However, ultrafast response time and giant nonlinearity often presents an inherent trade-off in optical materials such that larger nonlinear susceptibility typically can only be attained at the expense of slower response time. The trade-off imposes a major challenge to constructing integrated photonic processors following the von Neumann architecture, often mandating complicated heterogeneous integration of various photonic devices in a single chip. Therefore, exploring new architectures and unconventional computing schemes for all-optical computing becomes imperative.

Here, we report a new strategy to realize ultrafast and ultralow-energy-consumption all-optical computing including equation solving, multifunctional logic operations based on optical convolutional neural network (CNN). Inspired by biological brains¹²⁻¹³, optical neural networks have been used to carry out image classification¹⁴⁻¹⁵, speech recognition¹⁶ and self-learning tasks¹⁷. Up to now, the scheme of three-dimensional all-optical

diffraction network can process a large amount of data and obtain rich characteristic information, there is still a long way to go to achieve on-chip integration due to the spatial diffraction characteristics of the light field. Besides, in the strategy of programmable network through thermo-optical modulation, the potential problems caused by ohmic loss make the processing speed and energy consumption unable to be equivalent to the all-optical control network. For the first time, we propose an all-optical computing chip based on physically-fixed CNN. Optical CNNs possesses a non-von Neumann architecture, which underlies its ultrafast computing time and ultralow energy consumption. The proposed optical CNNs are able to perform computing tasks through convolution operations between layers without the aid of nonlinearity layers, conducive to multi-tasking processing and significantly reduced energy consumption¹⁸⁻¹⁹. Furthermore, owing to their powerful prediction capability, a single network can solve a specific class of computing problems rather than one single task. This scheme has built a new platform for all-optical computing, on which almost all signal processing functions are allowed to be implemented.

The optical CNN consists of cascaded silicon Y-shaped waveguides with side-coupled silicon waveguide segments designed to control the amplitude and phase of light in the waveguide branches. This conceptually and architecturally simple design uniquely affords both ultrafast computing time and low energy consumption. Importantly, the design is also scalable to handle CNNs with arbitrary network complexity. Our scalable optical CNN architecture presents a universal platform for implementing CNN-related functions leveraging the vast asset base of algorithms that have been matured in the field of computer science research (Supplementary information Section 1). Another important advantage of CNNs is that they can protect signals from distortion compared with fully-connected neural networks, as CNNs only contain local connections. As a proof-of-concept, we experimentally implemented the network design through several computation tasks including transcendental equations solvers, multifunctional logic gate operators, and half-adders.

Results and discussion

Scalable network configuration

To realize CNNs in an on-chip platform, we designed an all-optical network to emulate the convolutional opera-

tions (as shown in Fig. 1(a)). The signals fed into the network are encoded in the form of light amplitude distribution in discrete input waveguides. The network weights optimized to yield the target solutions are implemented through convolution operation between layers, i.e. $\tilde{X}^l = \tilde{W}_{l,l-1}\tilde{X}^{l-1}$, where \tilde{X}^l and \tilde{X}^{l-1} represent the light amplitude distributions in the l^{th} and $(l-1)^{\text{th}}$ layers, respectively, and $\tilde{W}_{l,l-1}$ gives the weight that dictates the way signals are passed from the l^{th} layer to the $(l-1)^{\text{th}}$ layer. Results of the computing tasks are given in the form of light amplitude distribution in a set of discrete output waveguides. To the best of our knowledge, this is the first implementation of a physically-fixed CNN on a full optical implementation chip.

The CNN is constructed from cascaded element structures comprising Y-shaped silicon waveguides side-coupled with silicon weight modulators. As an example, the schematic structure of the all-optical transcendental equation solver based on CNN is shown in Fig. 1(b). There are three layers of the element structure arrays. Each element structure connected to two adjacent element structures in the adjacent layers. Weight modulat-

ors are used to regulate the weights of the network according to the coupled mode theory. The weight modulator waveguide (as shown in Fig. 2(a)) has the same width as the transmission waveguide to ensure efficient coupling and large amplitude modulation. As Fig. 2(b) shows, the magnitude of weight w , which stands for the amplitude transmittance of the signal light in the transmission waveguide within the element structure, can be continuously tuned from 0.025 to 0.955 by varying the length (a) of weight modulator and the gap width (b) between the transmission waveguide and the weight modulator. Similarly, the phase of weight ω representing the phase of the signal light at the waveguide output port can be continuously adjusted from 0 to 2π by changing a and b as well (as shown in Fig. 2(c)). Importantly, the amplitude and phase of weight w can be independently adjusted to achieve arbitrary control of the signal light. For further discussions of the weight modulation method, please see Supplementary information Section 2.

It's worth mentioning that other complex mathematical operations can be systematically designed into the unified optical CNN architecture by cascading the Y-

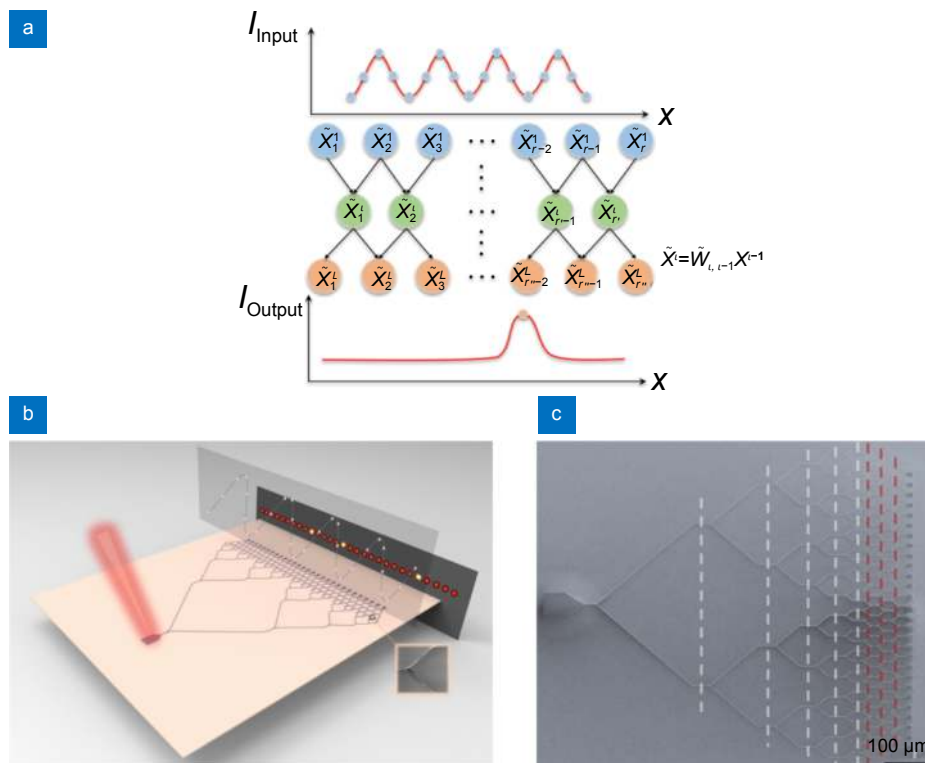


Fig. 1 | General architecture of the all-optical computing framework. (a) The CNN architecture showing the connections between adjacent layers: $\tilde{X}^l = \tilde{W}_{l,l-1}\tilde{X}^{l-1}$, where \tilde{X}^l and \tilde{X}^{l-1} represent the l^{th} and $(l-1)^{\text{th}}$ layer's optical signal respectively, $\tilde{W}_{l,l-1}$ shows the weight that determines how signals are propagated from l^{th} layer to $(l-1)^{\text{th}}$ layer. (b) Schematic diagram of the all-optical transcendental equation solver. (c) Top-view SEM image of the all-optical transcendental equation solver, where the scale bar is 100 μm . Here, the white dotted lines mark the five layers for waveform discretization, and the red dotted lines separate the three layers of the optical CNN structure.

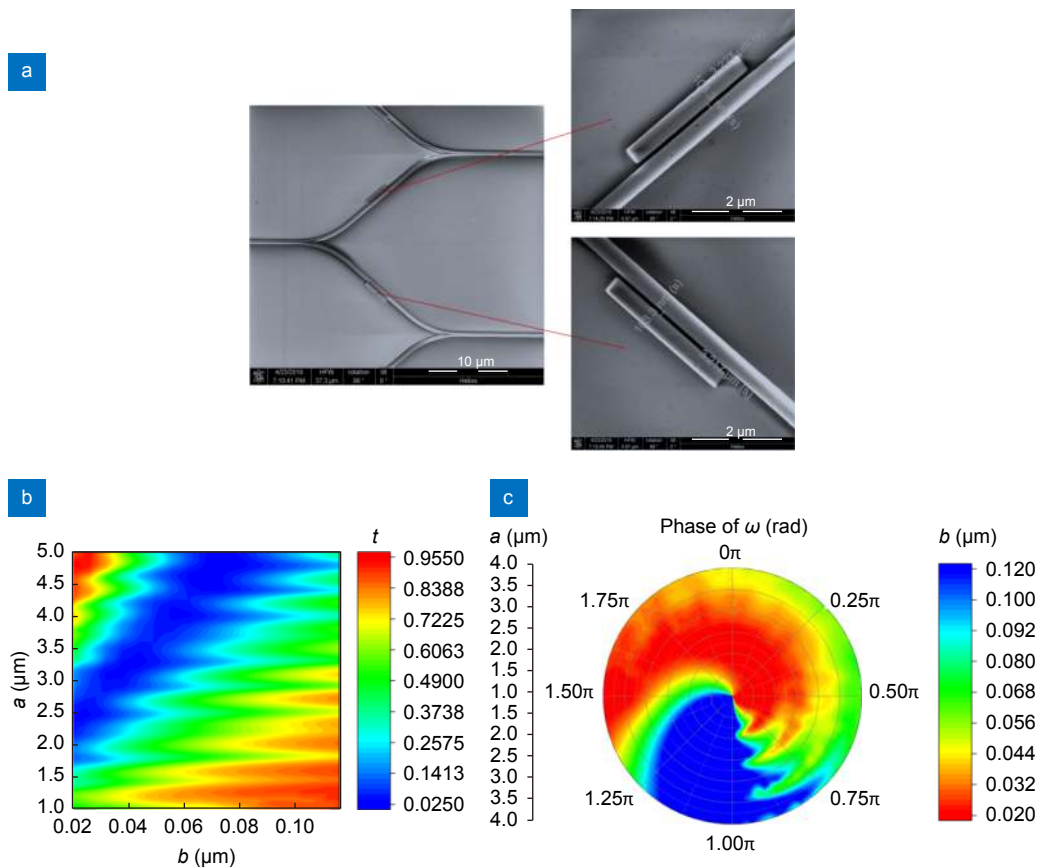


Fig. 2 | Weight regulation. (a) SEM image of Y-shaped waveguides side-coupled silicon weight modulators. Two arms of the “Y” structure waveguide correspond to two kinds of weights. By regulating the length a of weight modulator and the gap width b between two waveguides: (b) The magnitude of weight ω can be continuously adjusted from 0.03 to 0.95. (c) The phase of weight ω can be continuously adjusted from 0 to 2π .

shaped element structures. In the followings, we elaborate several examples of our optical CNN design being implemented as transcendental equation solvers, multifarious logic gate operators, and half-adders. It should be noted that signals operated in the network are the complex amplitudes of the light field, and what are measured in the experiment are light intensities at output ports. Therefore, the nonlinearity is introduced in the process of the measurement to realize the various functionalities of the devices although there are only convolutional layers in our networks.

All-optical transcendental equation solver

Since equations are effective tools for describing system states and processes, solving equations²⁰ can inform the state of the system under investigation and predict the trajectory of system evolution. Since transcendental equations can only be numerically solved except for a few cases, numerical solution of transcendental equations is still an important subject in mathematical calculations. We have developed a solver that can predictably

solve transcendental equations using optical CNN with exceptional computational performance.

A transcendental equation with the form of a trigonometric function is selected because in general, any arbitrarily complex mathematical expressions can be decomposed into trigonometric functions by Fourier decomposition, which means that we can solve any other transcendental equations in principle. The all-optical transcendental equation solver is used to solve the equation with a variable parameter k :

$$\cos(2kx) + 4 = \tan(kx) . \quad (1)$$

We choose to represent the input waveform (in this case $\cos(2kx) + 4$) by encoding its values at a set of evenly spaced x points with light amplitudes in a waveguide array. Firstly, light from a single input waveguide is passed through a 5-layer cascaded Y-branch structures to generate the discretized waveform of $\cos(2kx) + 4$ in the range from 3 to $3 + 8 \times i/N$, where i is an integer in the range from 0 to 29, and N is the total number of waveguides. Then the output signal representing the discretized waveform is input to the CNN with a total

layer number of 3 (multiple layers in the linear neural network is to provide enough degrees of freedom to help complete the learning task). An SEM image of this network structure is shown in Fig. 1(c). The fixed network weights were pre-determined prior to the fabrication via an iterative training algorithm detailed in Supplementary information Section 3, and implemented using the silicon weight modulator structures in this optical CNN. The network weights are optimized such that the solution to the equation is given by the position of the waveguide yielding the maximum output intensity:

$$x = 3 + 8 \times i' / N', \quad (2)$$

where i' is an integer, and N' is the total number of output waveguides. To validate the performance of the optical CNN equation solver, solution of Eq. (1) is performed by inputting a series of waveforms with the parameter k equaling to 1.67, 1.84 and 2.35, respectively. Fig. 3(a) shows the optical CNN output for $k = 1.67$, and the results pertaining to the other two k values are presented in Supplementary information Section 3. Here, we define the deviation of the CNN output solution from the true solution as:

$$(x_{\text{exp}} - x_{\text{theor}}) / X_{\text{out}}, \quad (3)$$

where x_{exp} and x_{theor} correspond to the encoded x values associated with the maximum intensity waveguide (following Eq. (2)), and the subscripts exp and theor denote the experimentally measured and theoretically predicted results, respectively. X_{out} represents the entire range of x over which the solution is sought (3 to $3 + 8 \times 26/27$ in this case), which is dictated by the total number of output waveguides N' (27 in our device). Fig. 3(b) summarizes the solution deviations for the three k values. The test result shows that our transcendental equation solver has achieved high accuracy with a maximum deviation less than 5%, and in most cases the deviations are less than 3%. The deviation results from the finite number of output waveguides and imperfect sample fabrication. Thus, it should be emphasized that the accuracy of the solution can be improved by increasing the number of output waveguides in theory.

Besides excellent solution accuracy, the all-optical equation solver also features ultrafast and energy-efficient computation. The total computing time, characterized by the time-of-flight of light through the entire structure (including the waveform discretization section), is 9.4 ps, and the effective operation time of the

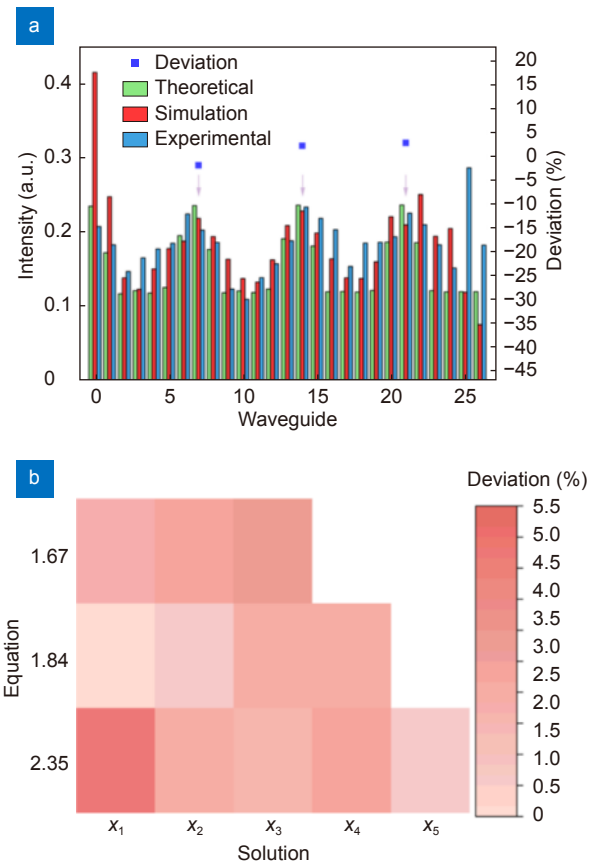


Fig. 3 | All-optical transcendental equation solver. (a) Output light intensity distribution in the output waveguides ($k = 1.67$). The arrows in the figure correspond to the locations of the solutions. The horizontal axis is the number of discrete waveguides, the vertical axis on the left represents the output signal intensity, and the vertical axis on the right gives the deviation between the experimental output signal and the theoretical value. **(b)** A graphic representation of solution deviation. The horizontal axis labels the individual solutions, and the vertical axis represents three values of the parameter k . The shade of the color indicates the magnitude of the deviation.

3-layer CNN is as short as 1.3 ps. The optical solver can also claim ultralow energy consumption. In our experiments, the computation energy overhead is 92 fJ/bit based on the laser pulse power we used. Our analysis further demonstrated that the shot-noise-limited mean error converges to a limit bounded by the discreteness of the network output at pulse energies above a few aJ/bit. (Supplementary information Section 6).

The optical CNN architecture presented here also offers the unique potential of crosstalk elimination. Crosstalk in optical analog computing is generally caused by light backscattering between adjacent layers in a densely integrated platform. Based on our device design, the crosstalk is expected to be naturally eliminated by means of the error back propagation optimization

process. Stability analysis of our network further demonstrates its high fault tolerance to defects such as weight deviation and waveguide damage (Supplementary information Section 5).

Multifarious logic gate operators

All-optical logic gates constitute the basic building blocks for ultra-high-speed all-optical chips, as any complex optical logic circuit can be composed of these logic gates. In addition, logic operation sets the foundation for more complex optical signal processing functions, such as addressing²¹, data coding²², parity checking²³ and signal extraction²⁴. However, current all-optical logic device designs based on linear coherence of signal light or non-linear interactions still face challenges in realizing reconfigurability and multifunctional operation (implementing multiple logic functions in a single chip) with high speed and low power consumption.

We leverage the scalability of our network to optimize on-chip all-optical multifarious logic devices. The design optimizes 6 input ports, including 2 signal input terminals and 4 control bits with a total of 5 layers (as shown in Fig. 4(a, b)). Similar to the all-optical equation solver, the fixed network weights were optimized using the iterative algorithm. Sixteen logic functions (representing exhaustive combinations of output results corresponding to all four possible input signals 11, 10, 01, and 00 is $2^4 = 16$) can be realized through seven different CNN structures, each with different network weights and responds to a different set of control bits. Each structure can perform 3 to 4 logic functions. Here we illustrate one of the optical CNN structures in Fig. 4(a), and characterization results of the other six structures are elaborated in Supplementary information Section 4. As we can see from Fig. 4(c), when the control bits are 1001, 0110 and 1010, the optical CNN performs “ $A + B(\text{OR})$ ”, “ $A \odot B(\text{XNOR})$ ”,

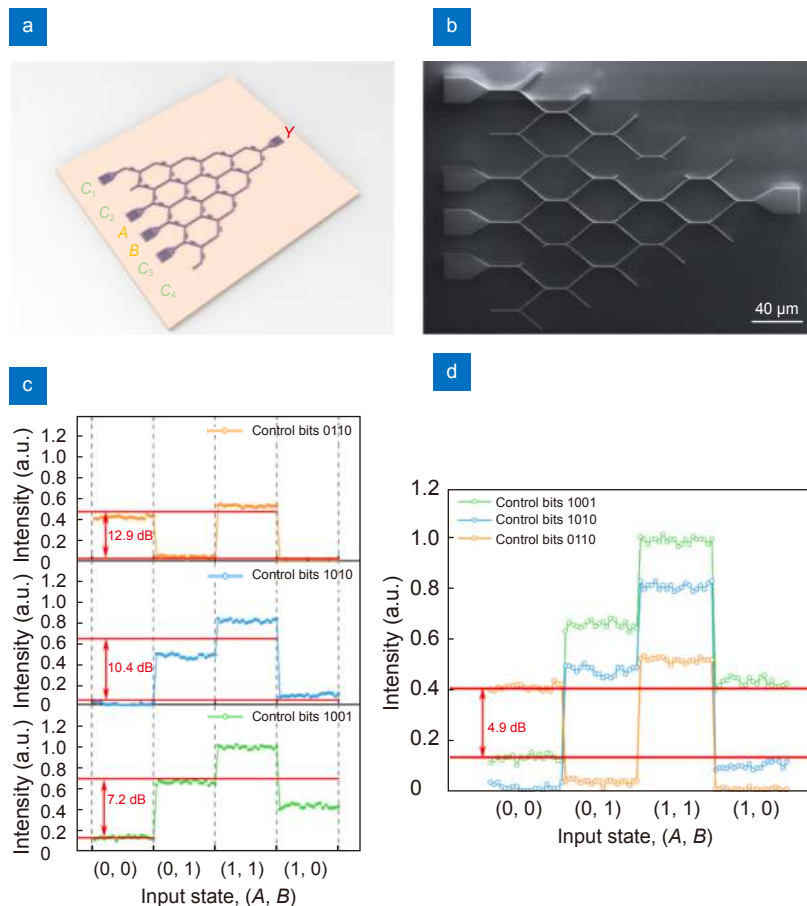


Fig. 4 | Multifarious logic gates. (a) Schematic diagram of the multifarious logic gate operator. Ports A, B are the signal inputs, and ports C₁, C₂, C₃, C₄, together constitute the control bits, and Y represents the signal output. (b) Top-view SEM image of the multifarious logic gate operator. (c) 0–1 intensity distribution when the optical CNN device acts as three different types of logic gates. (d) Overlay of three logic function responses in the optical CNN structure. The top red line corresponds to the minimum intensity of “1”, and the bottom red line shows the maximum intensity of “0”.

and “ $(A + \bar{A})B$ ” functions respectively. The intensity contrast of logic states 0 and 1 are experimentally measured as 7.2 dB, 10.4 dB, and 12.9 dB respectively for the three functions. The time-of-flight computing time is 3.3 ps with an energy consumption of 71 fJ/bit. Our analysis further demonstrates that energy consumption down to 10.4 aJ/bit can be achieved while maintaining a low error rate of 10^{-9} (Supplementary information Section 6). In Fig. 4(d), the optical CNN responses when the performance of the three logic functions are overlaid in one plot, showing a minimal output optical intensity contrast between the logic states 0 and 1 of 4.9 dB. The result shows that the optical CNN’s output logic states are readily distinguishable while performing multiple logic functions. That is to say, more cascades to scale up in the future may still work.

Half-adder

All-optical half-adder can perform the calculation task of adding two input data bits and yielding a Sum bit and a Carry bit in an all-optical implementation (Fig. 5(a)).

Half-adder is a basic unit of arithmetic logic operation optical circuits: for example, a full-adder can be realized by cascading two half-adders. Here we demonstrate an all-optical half-adder based on our optical CNN platform. We use 2D convolutional layers to train our CNNs for half-adder as well as multifarious logic gate operators, because shared weights cannot meet the demands in these two scenarios. After the training process, only the weights corresponding to the non-zero positions are extracted (Supplementary information Section 1). Here, 12 network weights are determined through the algorithm optimization, and an SEM image of half-adder is shown in Fig. 5(b). The arithmetic logic operations of “1” + “1” = (Sum “0”, Carry “1”), “0” + “1” = (Sum “1”, Carry “0”), and “1” + “0” = (Sum “1”, Carry “0”) are realized. The average optical intensity contrast between logic states 0 and 1 is 14.2 dB (Fig. 5(c)). The time-of-flight computing time is 2.7 ps with an energy consumption of 50.8 fJ/bit. Similarly, our analysis further demonstrates that energy consumption down to 23.8 aJ/bit can be achieved while maintaining a low error rate of 10^{-9}

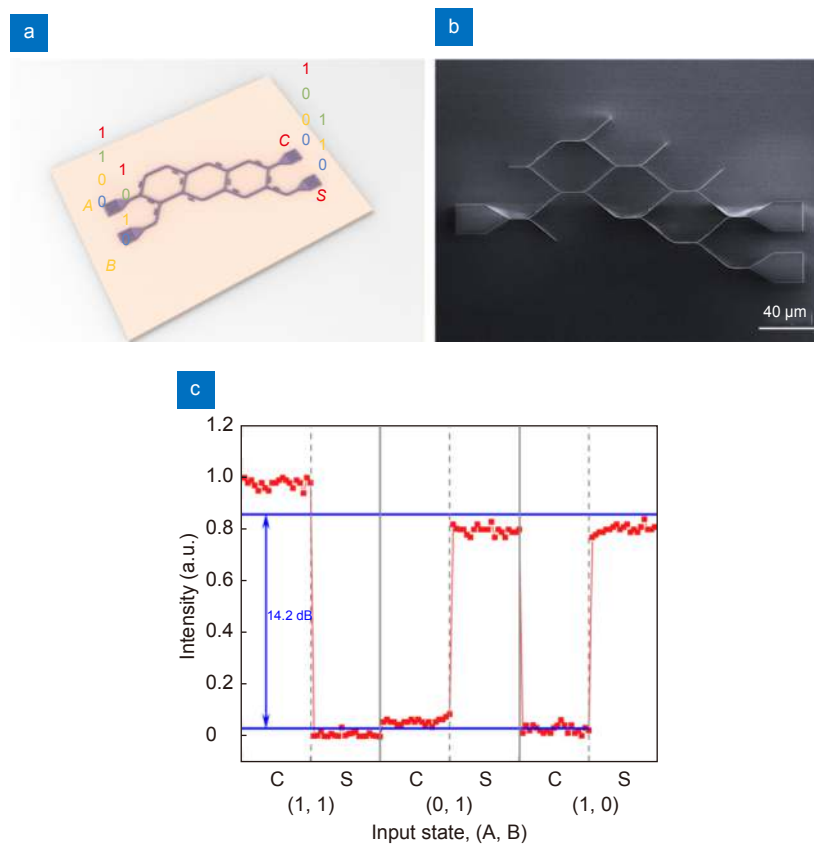


Fig. 5 | Half-adder. (a) Schematic diagram of the half-adder. Ports *A*, *B* are the signal inputs, and *C* and *S* represent the Carry and the Sum bit, respectively. (b) Top-view SEM image of the half-adder. (c) Intensity distribution of Sum bit and Carry bit corresponding to three different input signals in the half-adder. The blue lines give the average intensity values of the 0 and 1 logic states.

(Supplementary information Section 6). The function of the half-adder is successfully demonstrated while achieving high intensity contrast, which further validates that the CNN design is highly scalable and broadly applicable to a wide variety of all-optical processing functions.

Moreover, based on this element structure, the desired phase distribution can be obtained at output ports by adjusting weights of network, then the spatial filtering system can be constructed to realize the Fourier transformation of the input signal. Similarly, the input function can be expressed as a linear combination of multiple monomials at a given point to achieve series expansion. In addition, by defining the input-output relationship in advance to realize the network training, the output signals corresponding to different input signals are specified to represent specific code groups. Thus the encoder can be implemented. From above, a number of signal processing functions are allowed to be implemented on the proposed platform, which promotes the whole field of nanophotonics. The performance benchmark and significance of this work are presented in Supplementary information Section 7.

Conclusion

In this paper, we experimentally demonstrated the first physically-fixed CNN for all-optical computing based on silicon waveguides. Our optical CNN is formed by cascading a simple, universal element structure comprising Y-shaped silicon waveguides side-coupled with silicon weight modulators. We implemented the design to realize all-optical transcendental equation solvers, multifarious logic gate operators, and half-adders, all of which exhibit picosecond-scale ultrafast operation and ultralow energy consumption of the order of tens of femtojoules per bit. This optical network architecture is readily scalable which has the potential to be further extended to execute other complex computing tasks simply by cascading the basic element structures. Furthermore, this platform offers the possibility of parallel computing using wavelength multiplexing. Our work therefore points to a promising direction for next-generation all-optical computing systems.

Methods

Theoretical analysis and numerical simulation.

PyTorch, a custom package in Python which is used popularly for machine learning, was used to construct the

theoretical modeling of our optical neural networks. The calculations were based on 1D CNN used for the equation solver and 2D CNN used for logic devices and half-adder, respectively. Some optimizers were then used in PyTorch, applying stochastic gradient descent (SGD) in the learning process, to compute the parameters in our networks and minimize the loss function related to the model's performance as possible. The simulation results were conducted from finite element method (via the COMSOL Multiphysics commercial software).

Device fabrication.

Devices were fabricated leveraging standard silicon microfabrication technologies. A 6% hydrogen silsesquioxane (HSQ) electron beam resist was spun onto a double-side polished silicon-on-insulator (SOI) wafer and was patterned by an Elionix ELS-F125 electron beam lithography (EBL) tool. Development of the resist was performed by immersing the chip into 25% tetramethylammonium hydroxide solution for 150 seconds. The chip was subsequently etched in an RIE tool (PlasmaTherm Inc.) with chlorine gas at a power of 200 W and a pressure of 5 mTorr (1 Torr = 133.322 Pa). After stripping the electron beam resist in HF, an additional EBL step was conducted to pattern the waveguide grating couplers with ZEP resist on the same EBL tool (etching depth of grating couplers is different from that of transmission waveguide to obtain higher coupling efficiency). The chip was developed in ZED-N50 developer and etched in the same RIE tool under identical conditions. Finally, the resist was stripped by soaking in N-Methyl-2-Pyrrolidone (NMP) overnight.

Optical measurement.

Devices were tested on a microspectroscopy measurement system. Laser beam from a home-built femtosecond pulse fiber laser system was used as the light source. The laser central wavelength was 1560 nm with a repetition rate of 100 MHz and a pulse width of 80 fs (The results are stable in the range of femtosecond pulse wavelength broadening). The signal light with adjustable spot size was focused to the input-coupling port of the sample. The output signal was collected with a long working distance objective lens (Mitutoyo 20, NA = 0.58) and imaged onto a charge coupled device (CCD) camera (Xenics, XS-4407, Belgium).

References

- Salem R, Foster MA, Gaeta AL. Application of space-time duality to ultrahigh-speed optical signal processing. *Adv Opt Photonics* **5**, 274–317 (2013).
- Marpaung D, Yao JP, Capmany J. Integrated microwave photonics. *Nat Photonics* **13**, 80–90 (2019).
- Willner AE, Khaleghi S, Chitgarha MR, Yilmaz OF. All-optical signal processing. *J Lightwave Technol* **32**, 660–680 (2014).
- Rajaei R, Mamaghani SB. Ultra-low power, highly reliable, and nonvolatile hybrid MTJ/CMOS based full-adder for future VLSI design. *IEEE Trans Device Mater Reliab* **17**, 213–220 (2017).
- Salahuddin S, Ni K, Datta S. The era of hyper-scaling in electronics. *Nat Electron* **1**, 442–450 (2018).
- Smitha GS, Aradhya HV. mGDI based parallel adder for low power applications. *Microsyst Technol* **25**, 1653–1658 (2019).
- Sun C, Wade MT, Lee Y, Orcutt JS, Alloatti L et al. Single-chip microprocessor that communicates directly using light. *Nature* **528**, 534–538 (2015).
- Ambis P. Optical computing: a 60-year adventure. *Adv Opt Technol* **2010**, 372652 (2010).
- Yan H, Choe HS, Nam S, Hu YJ, Das S et al. Programmable nanowire circuits for nanoprocessors. *Nature* **470**, 240–244 (2011).
- Ando M, Kadono K, Haruta M, Sakaguchi T, Miya M. Large third-order optical nonlinearities in transition-metal oxides. *Nature* **374**, 625–627 (1995).
- Leuthold J, Koos C, Freude W. Nonlinear silicon photonics. *Nat Photonics* **4**, 535–544 (2010).
- Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**, 607–617 (2019).
- Xia QF, Yang JJ. Memristive crossbar arrays for brain-inspired computing. *Nat Mater* **18**, 309–323 (2019).
- Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- Yan T, Wu JM, Zhou TK, Xie H, Xu F et al. Fourier-space diffractive deep neural network. *Phys Rev Lett* **123**, 023901 (2019).
- Shen YC, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).
- Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WHP. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
- Brunner D, Soriano MC, Mirasso CR, Fischer I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat Commun* **4**, 1364 (2013).
- Antonik P, Marsal N, Brunner D, Rontani D. Human action recognition with a large-scale brain-inspired photonic computer. *Nat Mach Intell* **1**, 530–537 (2019).
- Estakhri NM, Edwards B, Engheta N. Inverse-designed meta-structures that solve equations. *Science* **363**, 1333–1338 (2019).
- Ballarini D, De Giorgi M, Cancellieri E, Houdré R, Giacobino E et al. All-optical polariton transistor. *Nat Commun* **4**, 1778 (2013).
- Lu CC, Hu XY, Yang H, Gong QH. All-optical logic binary encoder based on asymmetric plasmonic nanogrooves. *Appl Phys Lett* **103**, 121107 (2013).
- Wang FF, Gong ZB, Hu XY, Yang XY, Yang H et al. Nano-scale on-chip all-optical logic parity checker in integrated plasmonic circuits in optical communication range. *Sci Rep* **6**, 24433 (2016).
- Liu CS, Chen HW, Hou X, Zhang H, Han J et al. Small footprint transistor architecture for photoswitching logic and in situ memory. *Nat Nanotechnol* **14**, 662–667 (2019).

Acknowledgements

We are grateful for financial supports from the National Key Research and Development Program of China (2018YFB2200403); National Natural Science Foundation of China (NSFC) (61775003, 11734001, 91950204, 11527901, 11604378, 91850117).

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at <https://doi.org/10.29026/oea.2021.200060>