



All-optical neural network with nonlinear activation functions

YING ZUO,^{1,†} BOHAN LI,^{1,†} YUJUN ZHAO,^{1,†} YUE JIANG,¹ YOU-CHIUAN CHEN,¹ PENG CHEN,¹ GYU-BOONG JO,¹ JUNWEI LIU,^{1,2} AND SHENGWANG DU^{1,3} 

¹Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

²e-mail: liuj@ust.hk

³e-mail: dusw@ust.hk

Received 28 May 2019; revised 30 July 2019; accepted 5 August 2019 (Doc. ID 368705); published 29 August 2019

Artificial neural networks (ANNs) have been widely used for industrial applications and have played a more important role in fundamental research. Although most ANN hardware systems are electronic-based, their optical implementation is particularly attractive because of its intrinsic parallelism and low energy consumption. Here, we demonstrate a fully functioning all-optical neural network (AONN), in which linear operations are programmed by spatial light modulators and Fourier lenses, while nonlinear optical activation functions are realized in laser-cooled atoms with electromagnetically induced transparency. Because all errors from different optical neurons are independent, it is possible to scale up the size of such an AONN. Moreover, our hardware system is reconfigurable for different applications without the need to modify the physical structure. We confirm its capability and feasibility in machine-learning application by successfully classifying order and disorder phases of a statistical Ising model. The demonstrated AONN scheme can be used to construct various ANN architectures with intrinsic optical parallel computation. © 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](https://doi.org/10.1364/OPTICA.6.001132)

<https://doi.org/10.1364/OPTICA.6.001132>

1. INTRODUCTION

Machine-learning techniques, especially artificial neural networks (ANNs), have grown significantly in the past decade and have been demonstrated to be powerful or even to surpass human intelligence in various fields, such as image recognition, medical diagnosis, and machine translation [1,2]. ANNs also show a great potential in scientific research [3–5], especially in discovering new materials [6], classifying phases of matter [7], representing variational wave functions [8], and accelerating Monte Carlo (MC) simulations [9,10]. They may be used to solve problems that are intractable in conventional approaches [11–14]. The power of an ANN comes from its extensive interconnections among a large number of neurons, requiring huge computational resources (time and energy) when they are implemented electronically [1].

Unlike electrons in a digital computer, photons, as noninteracting bosons, could naturally be used to realize multiple interconnections and simultaneous parallel calculations at the speed of light [15–20]. The key ingredients of an ANN are the artificial neurons, which perform both linear and nonlinear transformations for the input signals. In most hybrid optical neural networks (ONNs), optics is mainly used for linear operations, and nonlinear functions are usually implemented electronically [21–24]. Recently, ONNs based on nanophotonic circuits [25] and light-wave linear diffraction and interference [26] have been demonstrated for efficient machine learning, but nonlinear optical

activation functions are still absent from deep networks [24,27]. Although there have been proposals for implementing nonlinear optical activation functions [28,29], their experimental realizations have become the bottleneck for further extension of ONNs in practical applications.

In this work, we demonstrate an all-optical neural network (AONN) with both tunable linear operations and nonlinear activation functions in optics. We use spatial light modulators (SLMs) and Fourier lenses to implement the linear operations. The nonlinear optical activation functions are realized based on electromagnetically induced transparency (EIT) [30,31]—a light-induced quantum interference effect among atomic transitions. To verify the capability and feasibility of the AONN scheme, we implement a dense (fully connected) two-layer AONN and use it to successfully classify different phases for a prototypical Ising model.

2. GENERAL AONN STRUCTURE

In a typical ANN, as illustrated in Fig. 1(a), neurons are usually arranged in layered structures without connections between different neurons in the same layer, and the output from the neurons in one layer serves as the input for the neurons in the next layer. The working principle of an artificial neuron can be abstracted into the following two steps: (1) receiving multiple weighted W_{ij} input signals v_j from neurons in the preceding layer through a linear operation with some bias b_i , i.e., $z_i = b_i + \sum_j W_{ij}v_j$, and

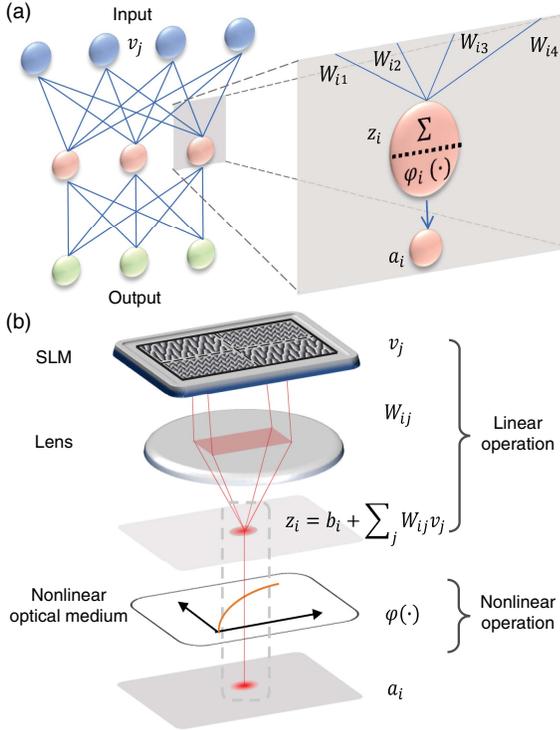


Fig. 1. General AONN. (a) A typical two-layer neural network; (b) schematic of experimental realization of an optical neuron including linear and nonlinear operations. The linear operation $z_i = b_i + \sum_j W_{ij} v_j$ is achieved by combining a programmable SLM and a Fourier lens. ϕ is the nonlinear activation function.

(2) generating new output signal a_i processing all input signals through nonlinear activation functions $a_i = \phi(z_i)$. In our optical configuration, the linear operation is implemented by an SLM (see Supplement 1 S1 [32]) followed by a Fourier lens, and the nonlinear optical activation function is realized with EIT, as shown in Fig. 1(b). Differentiating from the conventional diffractive ONNs, where the electric-field neurons are complex [26], in our AONN the signals are encoded in light power; thus, $v_i, z_i, a_i \geq 0$, and the real matrix elements satisfy $1 \geq W_{ij} \geq 0$.

3. LINEAR OPTICAL OPERATION

In the linear operation process, the incident light powers at different areas in the SLM [33] represent the input layer nodes v_j . By superposing multiple phase gratings, the incident light beam v_j can be split into different directions j with weight W_{ij} . The SLM is placed at the back focal plane of the lens, which performs Fourier transform and sums all diffracted beams in the same direction onto a spot at its front focal plane as the linear summation $z_i = \sum_j W_{ij} v_j$, as shown in Fig. 1(b). The linear bias b_i could be realized similarly from additional inputs. We obtain given matrix elements W_{ij} following the Gerchberg–Saxton iterative feedback algorithm [34–36], in which high accuracy (>95%) could be achieved in fewer than 10 iterations (see Supplement 1 S2 and S3 [32]). Moreover, it is worth mentioning that one great advantage of this method is that the error for a given spot does not depend on the total number of spots as long as the resolution of SLM is high enough, which is qualitatively different from the previous implementations [25,26].

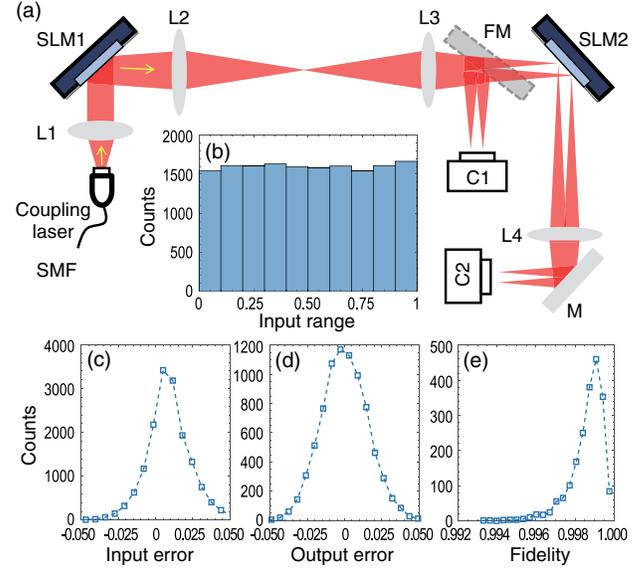


Fig. 2. Linear optical operation and characterization. (a) Optical setup. FM, flip mirror; M, mirror; L1–L4, optical lens; SMF, single-mode fiber; C1–C2, camera. The coupling laser beam emitted from an SMF is collimated by lens L1 ($f = 10$ cm) and illuminates the surface of SLM1 (Leto, Holoeye), where it is selectively reflected to eight separate light spots on the surface of SLM2 (Pluto-2, Holoeye) by a 4- f imaging system with lenses L2 ($f = 30$ cm) and L3 ($f = 25$ cm). An FM is inserted in the optical path to optionally reflect the beam to camera C1, which is located at the equivalent position of SLM2. SLM2 performs linear operation to transform the eight inputs to the four output beams, which are recorded by camera C2. (b) Histogram of 2000 random input vectors with elements equally distributing from 0 to 1; (c) error distribution of the input vectors. The standard deviation (STD) is 0.012. (d) Error distribution of the output vectors for the Hankel matrix operation. The STD is 0.014. (e) Fidelity distribution of the output vectors for the Hankel matrix operation.

Figure 2(a) shows the optical layout for performing the linear operation. Without losing generality, we take an 8-to-4 linear operation as an example. The coupling laser beam output from a single-mode fiber (SMF) is collimated and incident onto the first SLM (SLM1), which selectively reflects eight separate beam spots. These eight spots are then imaged onto the second SLM (SLM2) as the input v_j through a 4- f optical lens system (L2 and L3). The flip mirror (FM) and the first camera (C1) are used to monitor and measure v_j . The stray coupling light is blocked at the Fourier plane of lens L2. After SLM2, each laser beam is divided into four beams. The Fourier lens L4 performs a summation operation and the four output spots are recorded by the second camera (C2). To characterize the accuracy of the input vectors v_j , we measured error distribution of 2000 random eight-dimension input vectors with elements equally sampled from 0 to 1 [Fig. 2(b)]. As depicted in Fig. 2(c), we obtain very accurate input vectors with a standard deviation of only 0.017. The mean of the error is slightly off zero due to a possible laser power drift during the measurement.

Next, we confirm that an arbitrary positively valued 4×8 matrix ($1 \geq W_{ij} \geq 0$) can be realized by programming SLM2. We take two types of linear operations as examples (see Supplement 1 S4 [32]). The first is the Hankel matrix A , a typical symmetric matrix in mathematics whose elements satisfy

$A_{ij} = A_{i+k,j-k}$ ($k = 0, \dots, j - i$) (see Eq. S2 in Supplement 1 [32]). Because we cannot directly measure the matrix elements, we take the error distribution of output vectors using 2000 random input vectors previously described. As shown in Fig. 2(d), the errors $z_i - z'_i$ are very small, with a standard deviation of 0.014, where z_i and z'_i are the exact and measured vector components, respectively. Impressively, they are almost the same as the errors of input vectors, even though many more operations are involved here. This result further indicates that the error could maintain at a small level even for multiple linear operations, which is crucial for large-scale AONNs. For matrix production, the directions of output vectors are more useful than the exact value of different elements, the accuracy of which could be captured by the fidelity $\frac{\vec{z} \cdot \vec{z}'}{|\vec{z}| |\vec{z}'|}$. As shown in Fig. 2(e), the fidelity distribution is narrower than the error, and the mean value of fidelity is around 0.998 for the Hankel matrix. The high fidelity suggests that although there are some uncertainties for single elements, the output vectors are actually insensitive to these fluctuations. We also perform the same measurements for a random matrix and obtain similar results (see Eq. S3 and Fig. S3 in Supplement 1 [32]). Thus we verify that different matrixes can be implemented by reconfiguring SLM2 without changing the physical layout.

4. NONLINEAR OPTICAL ACTIVATION FUNCTION

The EIT nonlinear optical activation functions are implemented by laser-cooled ^{85}Rb atoms in a dark-line two-dimensional magneto-optical trap (MOT) [37–39] with a longitudinal length of 1.5 cm and an aspect ratio of 25:1, as shown in Fig. 3(a)

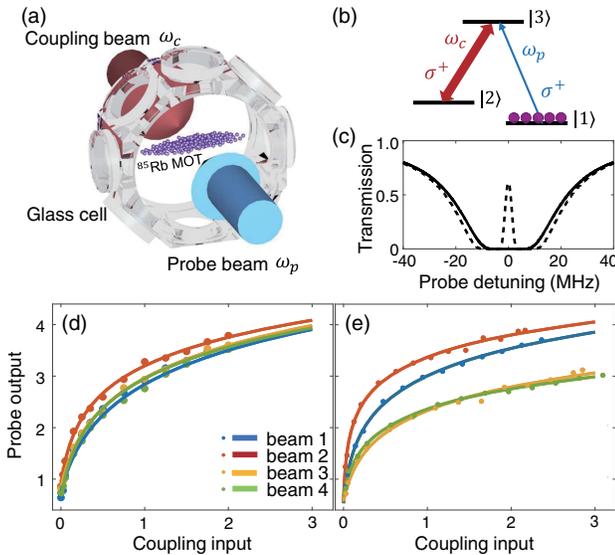


Fig. 3. Realization of EIT nonlinear optical activation functions. (a) EIT experimental configuration with cold ^{85}Rb atoms in the MOT; (b) three-level Λ -type EIT energy level diagram, where the ^{85}Rb atomic states are $|1\rangle = |5S_{1/2}, F = 2\rangle$, $|2\rangle = |5P_{1/2}, F = 3\rangle$, and $|3\rangle = |5S_{1/2}, F = 3\rangle$. Both the circularly polarized (σ^+) coupling (ω_c) laser and probe (ω_p) laser are on resonance with the transitions $|2\rangle \rightarrow |3\rangle$ and $|1\rangle \rightarrow |3\rangle$, respectively. (c) The EIT transmission spectrum of the probe beam. The solid (dashed) line is obtained without (with) the coupling beam. (d) and (e) are the nonlinear transmission functions for four on-resonance probe beams placed at different positions of the MOT. The coupling input and probe output are scaled for fitting the neural network input-output range [32].

(see Supplement 1 S5 [32]). The atoms are prepared in the ground state $|1\rangle$, as shown in the atomic energy level diagram in Fig. 3(b). The circularly polarized (σ^+) coupling laser (ω_c) beams, which are from the outputs of the linear operation, are on resonance to the atomic transmission $|2\rangle \leftrightarrow |3\rangle$ and incident to the atomic cloud along its transverse direction. A counterpropagating probe laser beam (ω_p, σ^+) is on resonance to $|1\rangle \rightarrow |3\rangle$. In the absence of the coupling beam, the atomic medium is opaque to the resonant probe beam, which is maximally absorbed by the atoms, as shown as the solid curve in the transmission spectrum of Fig. 3(c). Contrarily, in the presence of the coupling beam, the quantum interference between the transition paths leads to an EIT [30,31] spectral window, as shown as the dashed curve in Fig. 3(c), where the on-resonance peak transmission and the bandwidth are controlled by the coupling laser intensity. The on-resonance probe laser beam output can be expressed as

$$I_{p,\text{out}} = I_{p,\text{in}} e^{-\text{OD} \frac{4\gamma_{12}\gamma_{13}}{\Omega_c^2 + 4\gamma_{12}\gamma_{13}}} = \varphi(\Omega_c^2), \quad (1)$$

where $I_{p,\text{in}}$ and $I_{p,\text{out}}$ are the input and output probe beam intensity, OD is the atomic optical depth on the $|1\rangle \rightarrow |3\rangle$ transition, and γ_{ij} is the dephasing rate between the states $|i\rangle$ and $|j\rangle$. For ^{85}Rb atoms, $\gamma_{13} = 2\pi \times 3$ MHz, and the nonzero ground-state dephasing rate γ_{12} can be tuned by the stray background magnetic field. Ω_c is the coupling field Rabi frequency, and its square is proportional to the coupling laser intensity ($\Omega_c^2 \propto I_c$). As shown in Eq. (1), the probe beam intensity is nonlinearly controlled by the coupling beam intensity. The nonlinear activation function φ is achieved by taking the coupling intensity as the input and the transmitted probe intensity as the output. In the experiment, the input probe beam is collimated, and its beam size is large enough to cover the entire coupling beam profile. Moreover, Eq. (1) also indicates that the nonlinear activation function is determined by OD and γ_{12} , whose values vary at different positions of the MOT. Therefore, by placing the counterpropagating coupling-probe beams at different positions of the MOT, we can achieve different nonlinear activation functions for different neurons. Figure 3(d) shows nearly identical nonlinear activation functions are obtained by judiciously positioning the four input coupling beams. We can also assign these four neurons with different nonlinear activation functions, as shown in Fig. 3(e). Clearly, errors from different nonlinear activation functions are also independent. Together with the same advantages of linear operations realized by SLMs and lenses, the AONN scheme is expected to scale up to large size with error maintained at a small level.

5. TWO-LAYER AONN FOR APPLICATION

After the demonstration of linear and nonlinear operations, we are ready to assemble a fully functional AONN using SLMs, lenses, a MOT, and the coupling and probe laser beams. Here we show that we can actually apply such an AONN to classify different phases in condensed matter physics. It has been demonstrated recently that neural networks have great potential to identify different phases, including both symmetry-breaking phases [7] and topological phases [7,13]. Here we take the prototypical two-dimensional Ising model on a square lattice as an example for the demonstration. The Ising model could be written as

$$H(\sigma) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j, \quad (2)$$

where $\sigma_i = \pm 1$ represents localized spin on site i , and $\langle ij \rangle$ takes the summation over all the nearest neighbors. It is well known that there will be a continuous phase transition with the critical temperature as $T_c = \frac{2|J|}{K_B \ln(1+\sqrt{2})}$. In the following simulations and experiments, the interaction strength is set to be $J = 1$, and we take lattice size $L = 4$ with periodic boundary conditions as an example. We find that a dense (fully connected) two-layer neural network could work better when using the experimentally measured nonlinear activation functions in Fig. 3(e). In the same way as in a conventional electronic computer, our two-layer AONN consists of one input layer, one hidden layer, and one output layer. An optical signal propagates from one layer to the next layer through optical operation units. At the hidden layer, optical information is processed by optical nonlinear activation functions before propagating to the output layer.

The detailed optical implementation of the two-layer AONN is illustrated in Fig. 4(a). The input layer contains L^2 neurons with L as the linear system size. For the hidden and output layers, there are four and two neurons, respectively. In this particular configuration, because the input values are binary (0 or 1), the coupling beam input vector and the first linear operation can be realized by a single SLM, as shown as SLM1 in Fig. 4(a) (see Supplement 1 S6 [32]). The horizontally polarized output coupling beam passes through a polarizing beam splitter (PBS) and is incident to the cold atoms in MOT after a quarter-wave ($\lambda/4$) plate. The four transmitted counterpropagating and vertically polarized probe beams are reflected by a PBS and enter SLM2 for a second linear operation, which reduces the four

inputs to two outputs recorded by camera C3. The FMs and cameras C1 and C2 are used for configuring the network parameters.

We obtain the optimized linear operation matrix in our two-layer dense neural network by performing supervised learning in a computer with the measured EIT nonlinear optical activation functions [Fig. 3(e)]. The labeled raw training configurations are generated by MC simulations. By using only the configurations generated at low and high temperatures, the neural network learns to label them as ordered or disordered to determine the optimized linear matrix elements (see details in Supplement 1 S7 [32]). Then, for optical implementation, we configure AONN following the Gerchberg–Saxton iterative feedback algorithm described in Section 3 and Supplement 1 S2 [32]. Afterwards, we apply the two-layer AONN to identify different phases from the configurations sampled by MC simulation at intermediate temperatures and find the critical phase transition temperature (see Supplement 1 S7 and S8 for details [32]). The fraction of ordered or disordered configurations among all the samples can be regarded as the probability of being in order or disorder states. The crossing temperature with both probabilities at 50% indicates the phase transition point.

Figures 4(b) and 4(c) plot the mean probability of configuration sets generated at different temperatures for $L^2 = 4^2 = 16$ inputs. The experimental results reproduce the whole phase diagram, even though we only train the AONN at the temperatures far away from the critical temperature. The experimental phase transition temperature is close to the analytical thermodynamic limit represented by the vertical dashed line as the number of sites goes to infinite. It suggests that the AONN can successfully capture the essential features that distinguish the order and disorder phases. To clearly show performance of our AONN, we first

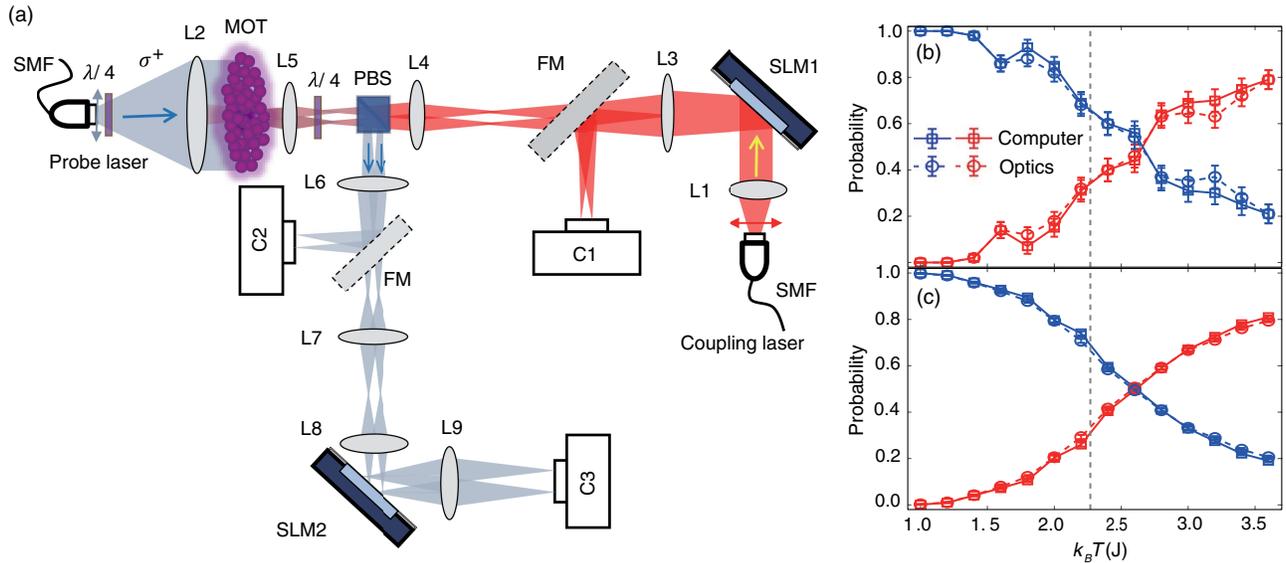


Fig. 4. Fully functioning two-layer AONN. (a) Experimental configuration of the two-layer AONN. The input layer is the pattern encoded on SLM1, whose area is divided into multiple subareas. The first layer consists of a linear operation done by SLM1 and EIT nonlinear optical activation functions at MOT. The second layer contains SLM2, which converts four beams into two output beams at camera C3. The collimated coupling laser beam passing lens L1 ($f = 10$ cm) is incident on the SLM1, which generates four beams at the focal plane of L3 ($f = 75$ cm). A FM and camera C1 are used to monitor this linear operation. The four beams are imaged on the MOT through a $4 - f$ system consisting of L4 ($f = 75$ cm) and L5 ($f = 5$ cm). A collimated probe laser beam propagates along the opposite direction of the coupling beam, which is imaged on camera C2 through L5 and L6 ($f = 45$ cm). With further magnification by L7 ($f = 7.5$ cm) and L8 ($f = 45$ cm), four beams are incident on SLM2 and generate two beams and then are focused on camera C3. (b) and (c) are the mean probability of the order (blue) and disorder (red) phases as functions of temperature for 100 and 4000 configurations, respectively. The comparison is made between the results from the AONN (circle) and the computer NN (square).

intentionally perform an experiment by using only 100 configurations; the results are shown in Fig. 4(b). Regarding phase classification, the results are slightly away from the thermodynamic limit. It is reasonable to have a large fluctuation, because we use very few configurations in a very large configured space (100 out of $2^{4 \times 4} \sim 0.15\%$). However, the results from our AONN and the computer dense neural network for the same configurations are nearly identical for all the temperatures, which clearly show that our AONN has the same accuracy as a well-trained computer-based ANN. To further demonstrate the capability of our AONN, we repeat the experiment with 4000 configurations. As expected, the phase transition curves become smoother because the random errors from statistical fluctuations are strongly reduced, and the optical results are almost the same as the computer data, as shown in Fig. 4(c). All results confirm that our implementation of AONN is successful and capable of classifying different phases for the Ising model.

6. DISCUSSION

In summary, we demonstrate an AONN scheme with both tunable linear optical operations and nonlinear optical activation functions. The linear interconnections are realized using SLMs and optical lenses. The EIT nonlinear optical activation functions are based on quantum interference. Although in this demonstration we work with cold atoms that allow us to have the EIT nonlinear optical function analytically as Eq. (1), a hot atomic vapor cell would also work well. As a proof-of-principle demonstration, we constructed a two-layer AONN for classifying the phases of a prototypical Ising model. In principle, it is possible to build a self-learning AONN with high-speed SLM feedback. However, the operating speed of most commercially available SLMs is not as fast as that of a computer. Therefore, in this work, the supervised training of the neural network has been done beforehand by a conventional computer, and then we follow the Gerchberg-Saxton iterative feedback algorithm to configure the ONN hardware. We focused mainly on the feasibility of the linear operations and nonlinear activation functions, which are the key ingredients of an ANN; the AONN is scalable to a larger system size with more SLMs and EIT nonlinear channels. The reasons are twofold: (1) As the computational power of ANNs comes from extensive interconnections between a large amount of neurons, ANNs are error-tolerant and robust against small local random errors, which means that even though the local parameters are not precise, we can still get very good results as long as the number of neurons is large enough. For the majority of problems, more neurons in ANNs usually give better performance; (2) as clearly demonstrated in our experiments, the final error of our AONN is insensitive to the total number of neurons, and the error could maintain at a similar level as a single neuron, even for large scale AONNs. Such a big advantage derives from the fact that all linear and nonlinear optical activation functions in our AONNs are independent, and errors from different optical neurons will not accumulate but may cancel each other out. Moreover, in our system the linear matrix elements and nonlinear functions can be independently programmed for realizing different AONN architectures and applications. Implementing a large-scale AONN requires more engineering resources, which is possible due to the recent developmental efforts in the miniaturization of cold atom devices [40,41] and EIT on chip [42].

Note Added. We became aware that during the manuscript preparation process [43], a work on an all-optical spiking neurosynaptic network based on phase-change nonlinear materials was published [44], which demonstrated only a single-layer system. Here we demonstrate a two-layer AONN with 16 inputs, four intermediate neurons with nonlinear optical activation functions, and two outputs. In addition, we use EIT quantum interference to realize the nonlinear optical activation functions, which is completely different from the approach in that just-published paper. With EIT quantum memory capacity [45], our system may be extended to realize a quantum neural network.

Funding. Hong Kong Research Grants Council (C6005-17G, ECS26302118).

Acknowledgment. B. L. and Y. C. C. acknowledge the support from the Undergraduate Research Opportunities Program at the Hong Kong University of Science and Technology.

See Supplement 1 for supporting content.

[†]These authors contributed equally to this work.

REFERENCES AND NOTES

1. A. J. Maren, C. T. Harston, and R. M. Pap, *Handbook of Neural Computing Applications* (Academic, 2014).
2. M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science* **349**, 255–260 (2015).
3. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature* **559**, 547–555 (2018).
4. G. Carleo, C. Ignacio, C. Kyle, D. Laurent, S. Maria, T. Naftali, V.-M. Leslie, and Z. Lenka, "Machine learning and the physical sciences," arXiv:1903.10563 (2019).
5. S. D. Sarma, D.-L. Deng, and L.-M. Duan, "Machine learning meets quantum physics," *Phys. Today* **72**(3), 48–54 (2019).
6. P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature* **533**, 73–76 (2016).
7. J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nat. Phys.* **13**, 431–434 (2017).
8. G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science* **355**, 602–606 (2017).
9. J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, "Self-learning Monte Carlo method," *Phys. Rev. B* **95**, 041101 (2017).
10. L. Huang and L. Wang, "Accelerated Monte Carlo simulations with restricted Boltzmann machines," *Phys. Rev. B* **95**, 035105 (2017).
11. D.-L. Deng, X. Li, and S. D. Sarma, "Machine learning topological states," *Phys. Rev. B* **96**, 195145 (2017).
12. J. Liu, H. Shen, Y. Qi, Z. Y. Meng, and L. Fu, "Self-learning Monte Carlo method and cumulative update in fermion systems," *Phys. Rev. B* **95**, 241104 (2017).
13. Y. Zhang and E.-A. Kim, "Quantum loop topography for machine learning," *Phys. Rev. Lett.* **118**, 216401 (2017).
14. C. Wang and H. Zhai, "Machine learning of frustrated classical spin models. I. Principal component analysis," *Phys. Rev. B* **96**, 144432 (2017).
15. Y. Abu-Mostafa and D. Psaltis, "Optical neural computers," *Sci. Am.* **256**, 88–95 (1987).
16. G. Zhou and D. Z. Anderson, "Acoustic signal recognition with a photo-refractive time-delay neural network," *Opt. Lett.* **19**, 655–657 (1994).
17. D. Z. Anderson, "Optical resonators and neural networks," *AIP Conf. Proc.* **151**, 12 (1986).
18. H. J. Caulfield, J. Kinser, and S. K. Rogers, "Optical neural networks," *Proc. IEEE* **77**, 1573–1583 (1989).
19. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutiérrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information process-

- ing beyond Turing: an optoelectronic implementation of reservoir computing," *Opt. Express* **20**, 3241–3249 (2012).
20. F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express* **20**, 22783–22795 (2012).
 21. S. Jutamulia and F. Yu, "Overview of hybrid optical neural networks," *Opt. Laser Technol.* **28**, 59–72 (1996).
 22. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Sci. Rep.* **2**, 287 (2012).
 23. D. Woods and T. J. Naughton, "Photonic neural networks," *Nat. Phys.* **8**, 257–259 (2012).
 24. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through in situ backpropagation and gradient measurement," *Optica* **5**, 864–871 (2018).
 25. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**, 441–446 (2017).
 26. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science* **361**, 1004–1008 (2018).
 27. J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," *Optica* **5**, 756–760 (2018).
 28. J. George, R. Amin, A. Mehrabian, J. Khurgin, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks," in *Signal Processing in Photonic Communications* (Optical Society of America, 2018), paper SpW4G-3.
 29. M. Miscuglio, A. Mehrabian, Z. Hu, S. I. Azzam, J. George, A. V. Kildishev, M. Pelton, and V. J. Sorger, "All-optical nonlinear activation function for photonic neural networks," *Opt. Mater. Express* **8**, 3851–3863 (2018).
 30. S. E. Harris, "Electromagnetically induced transparency," *Phys. Today* **50(7)**, 36–42 (1997).
 31. M. Fleischhauer, A. Imamoglu, and J. P. Marangos, "Electromagnetically induced transparency: Optics in coherent media," *Rev. Mod. Phys.* **77**, 633–673 (2005).
 32. See the supplemental material and Refs. [33–39] for (S1) the technical information of SLM; (S2) the Gerchberg–Saxton algorithm and feedback iteration process; (S3) the principle of linear optical power summation; (S4) the two matrices used for testing the linear operation; (S5) the operation of 2D MOT; (S6) the two-layer AONN implementation; (S7) the training of two-layer AONN; and (S8) Ising model related data processing.
 33. K. Lu and B. E. A. Saleh, "Theory and design of the liquid crystal TV as an optical spatial phase modulator," *Opt. Eng.* **29**, 240–246 (1990).
 34. G.-Z. Yang, B.-Z. Dong, B.-Y. Gu, J.-Y. Zhuang, and O. K. Ersoy, "Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison," *Appl. Opt.* **33**, 209–218 (1994).
 35. L. R. Di, F. Ianni, and G. Ruocco, "Computer generation of optimal holograms for optical trap arrays," *Opt. Express* **15**, 1913–1922 (2007).
 36. F. Nogrette, H. Labuhn, S. Ravets, D. Barredo, L. Beguin, A. Vernier, T. Lahaye, and A. Browaeys, "Single-atom trapping in holographic 2D arrays of microtraps with arbitrary geometries," *Phys. Rev. X* **4**, 021034 (2014).
 37. E. L. Raab, M. Prentiss, A. Cable, S. Chu, and D. E. Pritchard, "Trapping of neutral sodium atoms with radiation pressure," *Phys. Rev. Lett.* **59**, 2631 (1987).
 38. H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping* (Springer-Verlag, 1999).
 39. S. Zhang, J. F. Chen, C. Liu, S. Zhou, M. M. T. Loy, G. K. L. Wong, and S. Du, "A dark-line two-dimensional magneto-optical trap of ⁸⁵Rb atoms with high optical depth," *Rev. Sci. Instrum.* **83**, 073102 (2012).
 40. S. Du, M. B. Squires, Y. Imai, L. Czaia, R. A. Saravanan, V. Bright, J. Reichel, T. W. Hansch, and D. Z. Anderson, "Atom-chip Bose-Einstein condensation in a portable vacuum cell," *Phys. Rev. A* **70**, 053606 (2004).
 41. D. M. Farkas, K. M. Hudek, E. A. Salim, S. R. Segal, M. B. Squires, and D. Z. Anderson, "A compact, transportable, microchip-based system for high repetition rate production of Bose-Einstein condensates," *Appl. Phys. Lett.* **96**, 093102 (2010).
 42. B. Wu, J. F. Hulbert, E. J. Lunt, K. Hurd, A. R. Hawkins, and H. Schmidt, "Slow light on a chip via atomic quantum state control," *Nat. Photonics* **4**, 776–779 (2010).
 43. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, "All optical neural network with nonlinear activation functions," arXiv: 1904.10819 (2019).
 44. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature* **569**, 208–214 (2019).
 45. Y. Wang, J. Li, S. Zhang, K. Su, Y. Zhou, K. Liao, S. Du, H. Yan, and S.-L. Zhu, "Efficient quantum memory for single-photon polarization qubits," *Nat. Photonics* **13**, 346–351 (2019).