An Integrated All-Optical Multimodal Learning Engine Built by Reconfigurable Phase-Change Meta-Atoms

Yuhao Wang[1,2], Jingkai Song[1,2], Penghui Shen[1,2], Qisheng Yang[1,2], Yi Yang[1,2], and Tian-ling Ren[1,2(✉)]

[1] School of Integrated Circuits, Tsinghua University, Beijing 100084, China
RenTL@tsinghua.edu.cn

[2] Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

Abstract. Optical computing is regarded as one of the most promising computing paradigms for solving the computational bottleneck and accelerating artificial intelligence in the post-Moore age. While reconfigurable optical processors make artificial general intelligence (AGI) possible, they often cannot process multimodal signals. Here, we propose an integrated all-optical multimodal learning engine (AOMLE) built by reconfigurable phase-change meta-atoms. The engine architecture can be mapped to different optical neural networks by laser direct writing for phase-change materials, enabling more efficient processing of visual and auditory information at the speed of light. The AOMLE provides a cutting-edge idea for reconfigurable optical processors with increasing demands for complicated AI models.

Keywords: All-Optical Computing · Reconfigurable Chip · Multimodal Learning

1 Introduction

The thriving development of photonics has paved the way for faster and more energy-efficient AI computing. Optical processors are considered one of the most promising solutions for accelerating AI [1–7], leveraging the unique advantages of light speed, ultralow power consumption, and multiplexing. As the complexity of AI models continues to increase, the development of reconfigurable optical processors becomes increasingly important. There is a need to design new devices and explore suitable materials to make progress [8–12]. Chalcogenide phase-change materials play a crucial role in the field of reconfigurable photonics [13–16]. The non-volatile materials can transition between crystalline and amorphous phases under external excitation, exhibiting significant differences in optical properties [17], which find widespread applications in light field modulation. Undeniably, reconfigurable optical computing hardware enabled by phase-change materials has fruitful achievements [18–20]. However, current advanced optical processors can only demonstrate particular types of information, such as visual

L. Fang et al. (Eds.): CICAI 2023, LNAI 14474, pp. 442–451, 2024.
https://doi.org/10.1007/978-981-99-9119-8_40

signals like images and videos or sequential signals represented by audio. Due to the monotony of optical computing architecture and coupling signals, there are still limitations in multimodal signals processing, which prevents optical processors from progressing toward general computers.

Herein, we propose an integrated all-optical multimode learning engine (AOMLE) built by reconfigurable phase-change meta-atoms. By arranging phase-change meta-atoms covering an individual optical waveguide, we map them to different optical neural networks, enabling light-speed multimodal learning. We successfully reconstruct the all-optical computing architecture by taking advantage of the excellent properties of chalcogenide phase-change materials: disordered metasurface corresponds to the optical scattered neural network suitable for auditory signals, whereas layer-by-layer metalines corresponds to the optical diffractive neural network that is better for visual signals. We unify the training models for both optical neural network architectures by solving Maxwell's equations, and the adjoint method is used for backpropagation to update the medium gradient. Finally, we obtain 95.83% accuracy in vowel recognition and 96.34% accuracy in handwritten digit recognition, both of which are comparable to state-of-the-art electronic platforms and with a boost in energy efficiency. In conclusion, our proposed all-optical computing engine can efficiently perform multimodal learning, providing promise for general AI processors.

2 Architecture of AOMLE

Figure 1 depicts the architecture of AOMLE, which is actually a physical neural network built by phase-change materials. The red and purple marks represent the directions of data propagation in the feed-forward neural network and recurrent neural network, respectively, in the artificial neural network model shown in Fig. 1(a), and AOMLE is mapped to two neural networks by programming the pattern of phase-change materials covered on the waveguide, as illustrated in Fig. 1(b). The amplitude of light is pre-coded at the left input port of AOMLE. The optical path is obviously changed as light propagates through the intermediate training region due to the modulation of phase-change materials at the top of the waveguide, and the light is eventually coupled out at the right port, and the intensity is detected by photodetectors to obtain the classification result. This is the entire inference procedure of AOMLE.

It should be noted that the chalcogenide phase-change material used in AOMLE is Sb_2Se_3, and its extinction coefficient in the telecommunication wavelength tends to be negligible, meaning that intrinsic loss to the propagating light will be minimal. The refractive index of crystalline Sb_2Se_3 is around 4.0, it has a stronger effect on the phase modulation of light than that of amorphous Sb_2Se_3, so we regard the crystalline phase-change meta-atom as an effective neuron, which function is to sum the input light and then transmit them to the next effective neuron. In the nonlinear activation function of the neural network, we use the Kerr effect of silicon itself to establish a nonlinear connection between output and input optical power.

Considering the wave characteristics of light itself, when it propagates through the disordered dielectric layer, it will continue to scatter in all directions, which will introduce a feedback loop. This process is equivalent to the recurrent neural network, which is more suitable for processing data with time series information. Previous work has also proved this in theory [21]. Therefore, we use the metasurface formed by the random distribution of meta-atoms in different crystal phases to map the optical scattered neural network to AOMLE. It should be pointed out that our preprocessing of time series information only involves basic operations such as windowing and sampling, and we will not use spectrogram and other methods to make it into a matrix for subsequent calculation so that the time step of data will be preserved and the recurrent neural network will be driven to compute when scattered light propagates backward.

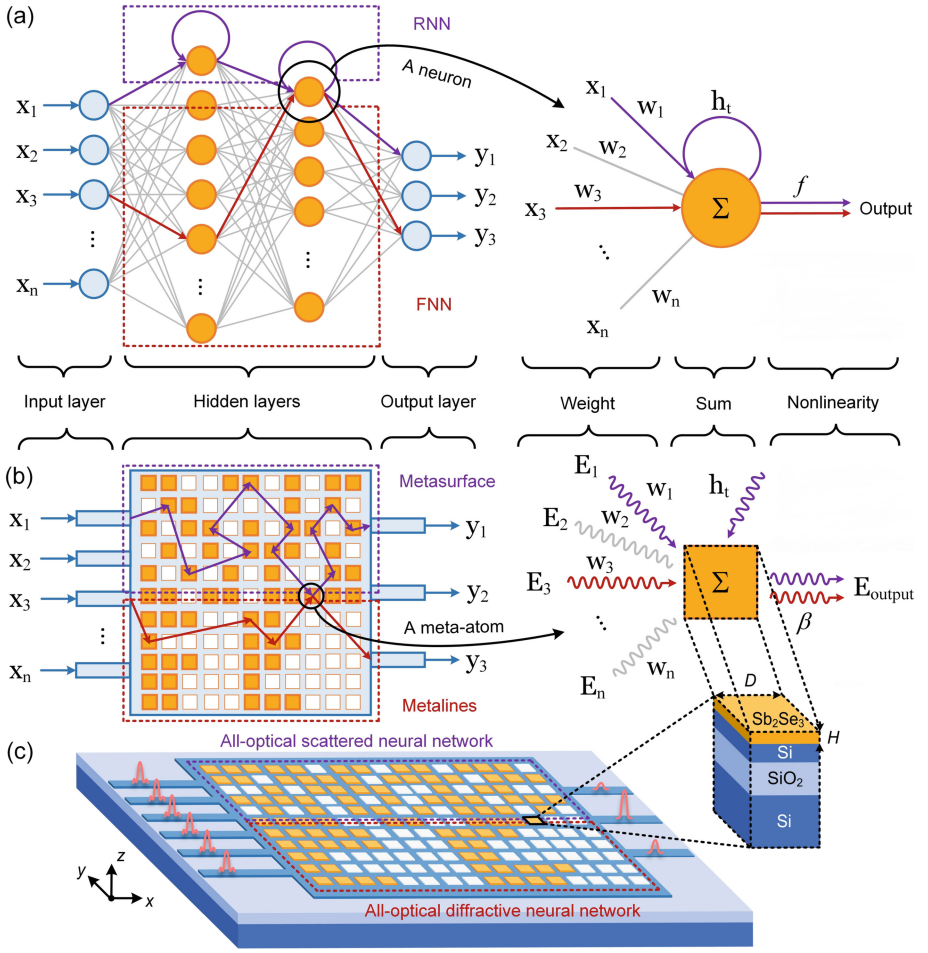

Fig. 1. Architecture of AMOLE. (a) Artificial neural network model. (b) Reconfigurable principle of AOMLE. (c) A phase-change meta-atom and hardware implementation of AOMLE.

For the on-chip diffractive neural network, its mathematical model is based on the Huygens-Fresnel diffraction principle, which reveals that diffractive neurons are essentially secondary wave sources, so their diffractive characteristics are more similar to convolution operation, even though this is a one-dimensional situation. In this way, we use crystalline effective neurons to form layers of diffractive metalines as hidden layers of feedforward neural networks. Furthermore, while the layered optical diffractive architecture is a subset of the bulk scattered architecture, the modulation mechanism of different architectures for the propagation light field determines which artificial neural network model they correspond to and which modal data computing scenarios are better suitable for.

In the experiment, the waveguide pattern of AOMLE architecture is realized by optically programming the phase-change materials, and its experimental platform is shown in Fig. 2. The experimental platform is mainly divided into two types of optical paths, propagation computation part and laser programming part. For the image input of the first kind of optical path, the 1550nm CW laser passes through the beam expander, and the image information is programmed by the digital micro-mirror. It is input to a single-mode fiber through the objective lens after passing through a 4f system. The acoustic-optical modulator provides waveform information of voice signals to the light, which is subsequently fed to AOMLE through the fiber. Finally, the photodetector receives the classified optical signal. The second type of optical path is used to implement the programming of AOMLE. We reconstruct the device using optical pulses generated by a 638nm laser diode, and the piezo stage accomplishes the movement required for array programming.

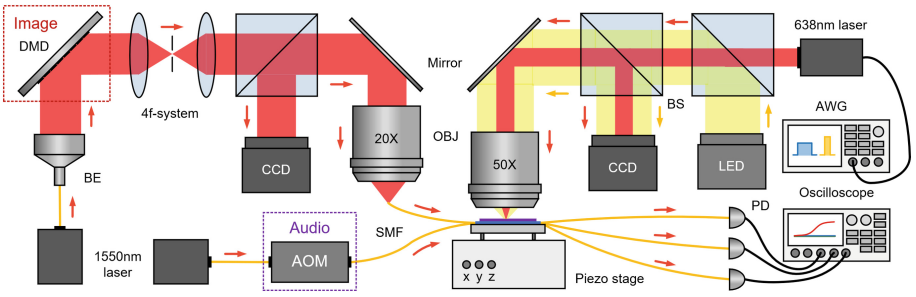

Fig. 2. Reconfigurable experimental schematic diagram of AOMLE

3 Training Algorithm of AOMLE

Light propagation in the AOMLE training region follows Maxwell's electromagnetic theory, and the primary light field distribution can be obtained by solving Maxwell's equations in the frequency domain. We describe their training methods in relation to the differences between the all-optical scattered and diffractive neural networks, respectively.

The following loss function Γ is defined.

$$\Gamma = \frac{1}{2}\sum_{i=1}^{N}(I_i - y_i)^2 \tag{1}$$

where I_i denotes the light intensity detected at the i th output port, and y_i is the ground truth as one-hot encoding.

First, we discuss the training process of the AOMLE scattered neural network model. At this moment, the structure of all phase-change meta-atoms in the training region is noticed. We use the FDTD method in the frequency domain to solve the primary light field $\overrightarrow{E}_{pri}(r)$ of any point r.

$$\left(\nabla^2 - \omega^2\mu_0\varepsilon(r)\right)\overrightarrow{E}_{pri}(r) = -i\omega\mu_0\overrightarrow{J}_s \tag{2}$$

where $\varepsilon(r)$ is the complex relative dielectric constant at r, μ_0 is the permeability of vacuum, and $\overrightarrow{J}_s$ is the current source density of the input light field distribution. We then determine the derivative $\partial\Gamma/\partial\overrightarrow{E}_{pri}(r)$ and use it as the excitation source of the adjoint field $\overrightarrow{E}_{adj}(r)$. Consequently, $\overrightarrow{E}_{adj}(r)$ can correspond to any r in the training region.

$$\left(\nabla^2 - \omega^2\mu_0\varepsilon(r)\right)\overrightarrow{E}_{adj}(r) = -\frac{\partial\Gamma}{\partial\overrightarrow{E}_{pri}(r)} \tag{3}$$

This is the solution process of two electromagnetic fields in the all-optical scattered neural network. The structural parameter in the diffractive neural network we are concerned about is a certain layer $\overrightarrow{m}$. According to the previous work [22], we can express the original field $\overrightarrow{E}_{pri}(\overrightarrow{m})$ and adjoint field $\overrightarrow{E}_{adj}(\overrightarrow{m})$ in each diffractive layer.

$$\overrightarrow{E}_{pri}(\overrightarrow{m}) = \left(\prod_{m=1}^{M} F^{-1}P_m F\Phi_m\right)\overrightarrow{E}_s \tag{4}$$

$$\overrightarrow{E}_{adj}(\overrightarrow{m}) = \overrightarrow{E}_{pri}(\overrightarrow{m}) \otimes (I_i - y_i) \tag{5}$$

where F and F^{-1} denote the discrete Fourier transform and inverse form, P_m and Φ_m express the diagonal matrix including the light propagation from m th layer to $m+1$ th layer, and phase shifts of m th layer, respectively.

Combining the adjoint field $\overrightarrow{\mathrm{E}}_{adj}$ with the original field $\overrightarrow{\mathrm{E}}_{pri}$, we get the gradient of AOMLE's structural parameters.

$$\frac{\partial\Gamma}{\partial\varpi} \propto \mathrm{Re}\left\{\overrightarrow{\mathrm{E}}_{adj}\cdot\overrightarrow{\mathrm{E}}_{pri}\right\} \tag{6}$$

where $\partial\Gamma/\partial\varpi$ denotes the gradient value, and ϖ here represents the structural parameter r and $\overrightarrow{m}$ corresponding scattered and diffractive neural network, respectively, and the

gradient has a linear relationship with the real component of $\left\{\vec{E}_{adj} \cdot \vec{E}_{pri}\right\}$. We can update the state $\Delta\varpi$ of phase-change meta-atoms.

$$\Delta\varpi = \varpi_{t+1} - \varpi_t \propto \frac{\partial\Gamma}{\partial\varpi} \tag{7}$$

The above process represents the training algorithm of AOMLE. It is crucial to acknowledge that training a scattered neural network requires an inverse design method rooted in photonics. Genetic algorithms, the adjoint method, generative adversarial networks, and reinforcement learning are all common methods for inverse design. For the proposed on-chip diffractive neural network, the primary modeling approach is based on the Rayleigh-Sommerfeld diffraction equation, although with substantial computational complexity. We build a unified model for all-optical scattered and diffractive neural networks in this work by solving the original and adjoint electromagnetic fields within the training domain for forward propagation and using the adjoint method for backpropagation, which enables the realization of a more efficient training algorithm. The training algorithm flow chart of AOMLE is presented in Fig. 3.

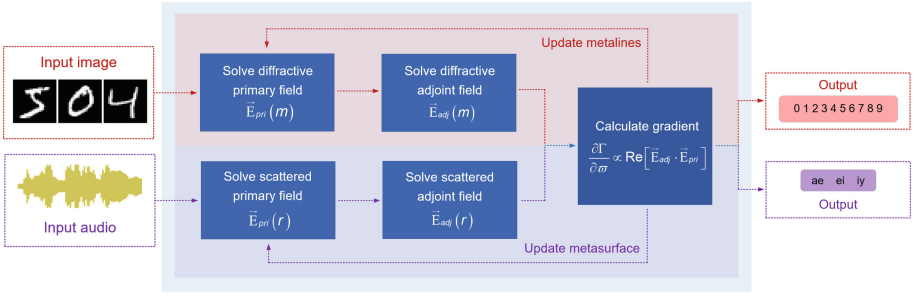

Fig. 3. Training algorithm flow chart of AOMLE

4 Multimodal Learning of AOMLE

According to the reconfigurable properties of AOMLE, we separate multimodal learning into scattered computing mode (SCM) and diffractive computing mode (DCM). In this work, we classify vowels and handwritten digital images using disctinct strategies.

In the face of SCM, we conduct vowel recognition tasks using the dataset [23]. This dataset consists of 270 audio messages from individuals of different genders and covers a variety of pronunciations, including ae, ei and ow. The training epoch of SCM is set to 30. As shown in Fig. 4(a-b), AOMLE achieves rapid convergence in the vowel recognition, with the training dataset reaching 96%, and the testing dataset likewise reaching 95.83%. Figure 4(c) shows the confusion matrix for both the training and testing dataset. Additionally, the effect of varying the length of the training region on recognition accuracy is also investigated as depicted in Fig. 4(d). By keeping the width of

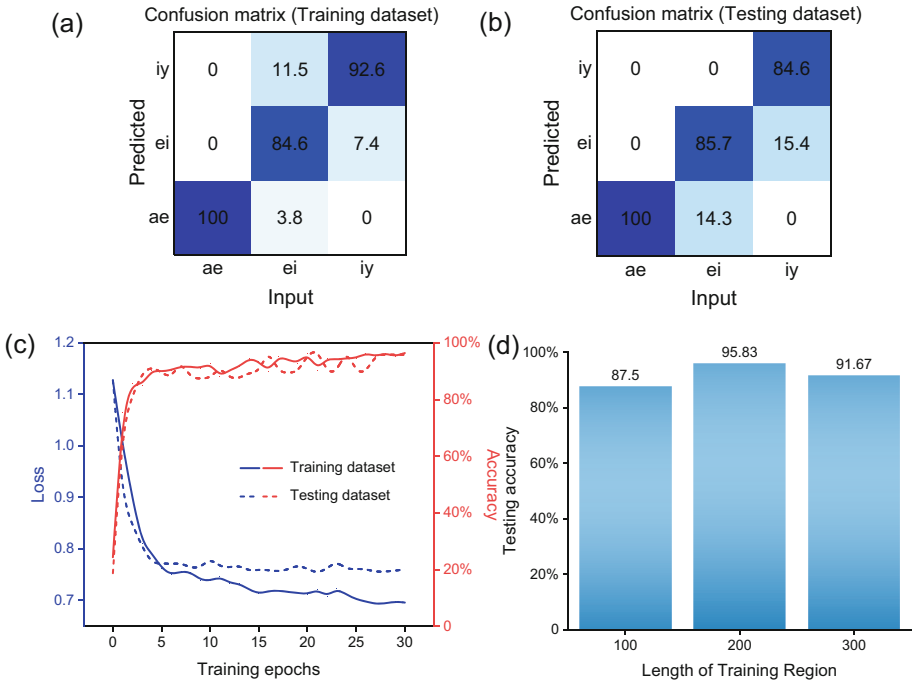

Fig. 4. Vowel recognition of SCM. (a-b) The confusion matrix of training dataset and testing dataset. (c) The loss and accuracy of SCM with training epochs. (d) The relationship between recognition accuracy and training region length.

the training area constant in AOMLE at 100, we identify that the testing dataset achieves the greatest accuracy (95.83%) when the training area length is set to 200.

We employ the classical MNIST dataset to assess its performance for DCM. Following 30 epochs of training, we achieve a recognition accuracy of 96.82% on the training dataset and 96.34% on the testing dataset. The confusion matrix is shown in Fig. 5(a-b), while Fig. 5(c) illustrates the performance of loss function and accuracy in handwritten digital classification. It is evident that AOMLE and other models have comparable accuracy. We further explore the scalability of DCM by changing the number of layers in the diffractive neural network, as presented in Fig. 5(d). By maintaining a fixed interval between metalines, we show that increasing the number of diffractive layers improves accuracy. The rate of improvement, however, becomes limited as the number of layers increases, indicating some redundancy within the neural network. We have a total of 2000 effective neurons per diffractive layer, which can be further optimized through pruning. Nevertheless, we decide to use five diffractive layers of meta lines for handwritten digital image classification, reaching a remarkable accuracy of 96.34% on the testing dataset.

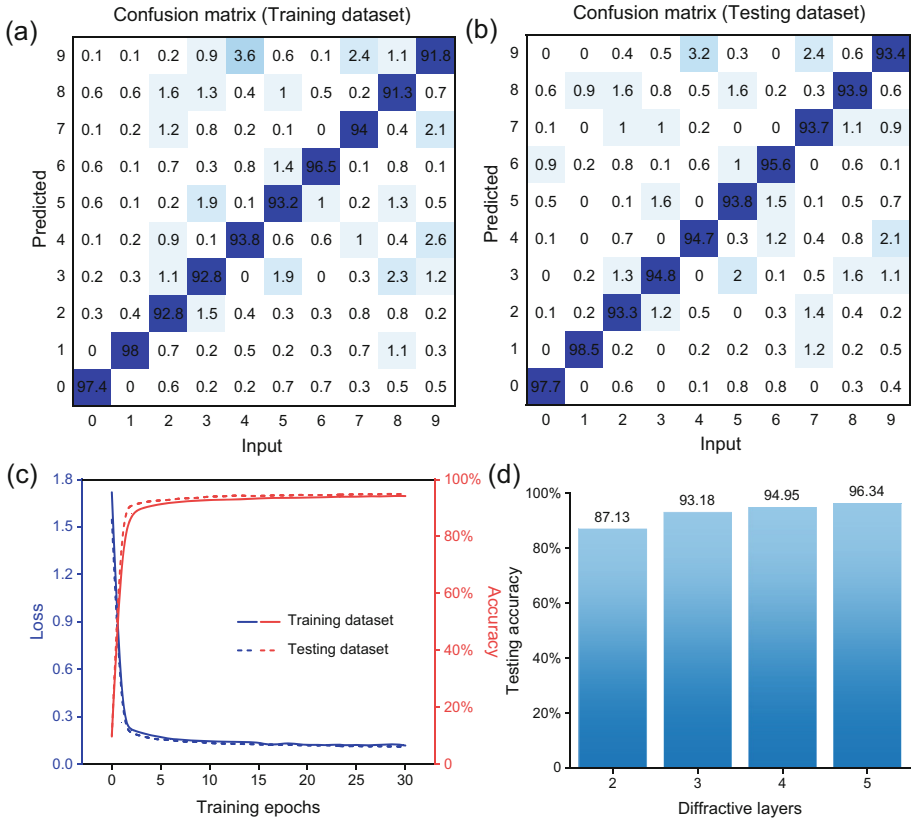

Fig. 5. Handwritten digital recognition of DCM. (a-b) The confusion matrix of training dataset and testing dataset. (c) The loss and accuracy of DCM with training epochs. (d) The scalablity of diffractive layers.

We conduct a comparative analysis between our proposed AOMLE and several processor chips used for intelligent classification tasks in the fields of natural language processing and machine vision, as shown in Table 1. The current optical processors primarily rely on devices or architectures such as Mach-Zehnder interferometer, optical diffraction, wavelength-division multiplexing, and optical scattering. Upon comparing AOMLE with other advanced optical processors, it becomes evident that AOMLE outperforms them in various indicators, including programmability, processible modality, among others. Furthermore, AOMLE achieves a remarkable increase in computing energy-efficiency, surpassing commercial electric processors by several orders of magnitude, while retaining outstanding recognition accuracy. These results highlight AOMLE's exceptional competitive edge over both optical and electric processors.

Table 1. Comparison of state-of-the-art integrated optical and electronic AI chips.

	Processor	Programmability	Modality	Energy	Latency
Optical processor	[1]	Electrical	Audios	30fJ/MAC	< 100ps
	[3]	Optical	Images	–	< 1ns
	[4]	Electrical	Images	1.58pJ/MAC	110ns
	[5]	Optical	Images	17fJ/MAC	250ps
	[7]	Electrical	Images	345fJ/MAC	< 60ps
	[9]	Electrical	Images & Videos	0.82fJ/MAC	–
	[12]	Optical	Audios	20pJ/MAC	40ps
Electronic processor	Google TPU	Electrical	–	0.43pJ/MAC	1.4ns
	Flash	Electrical	–	7fJ/MAC	15ns
Our work	AOMLE	Optical	Images & Audios	< 5fJ/MAC	< 200ps

5 Conclusion and Discussion

We propose a highly integrated all-optical multimodal learning engine called AOMLE, which effectively switches tasks based on input data modality and achieve array programming using externally modulated laser pulses. By leveraging the tunable property of phase-change materials, we successfully implement the reconfigurability of all-optical scattered and diffractive neural network on a single chip. We update the neural network's parameters using the unified form of the adjoint method, resulting in exceptional performance in both vowel recognition and handwritten digit recognition multimodal tasks.

It is worth mentioning the adjoint method has been widely employed in the inverse design of photonic devices. However, practical device fabrication often necessitates the binarization of trained medium parameters. The obtained device parameters of AOMLE can be quasi-continuous, with the level of discretization depending on the programming ability of laser pulses. This approach effectively overcomes the constraints imposed by binarization during fabrication, thereby further enhancing the computing performance of the optical processor.

Additionally, we utilize externally modulated laser pulses to program the phase-change materials, enabling precise control and alteration of the refractive index of the phase-change meta-atoms. Consequently, written laser pulses directly define the device pattern, eliminating the need for top-down lithography processes. This not only significantly enhances the flexibility of the silicon photonic device but also reduces fabrication errors and phase noise caused by lithography and etching. Collectively, these advantages demonstrate that our proposed AOMLE paves the way for more energy-efficient and flexible optical artificial intelligence processors.

References

1. Shen, Y., Harris, N., Skirlo, S., et al.: Deep learning with coherent nanophotonic circuits. Nat. Photonics **11**(7), 441–446 (2017)
2. Lin, X., Rivenson, Y., Yardimci, N.T., et al.: All-optical machine learning using diffrac-tive deep neural networks. Science **361**(6406), 1004–1008 (2018)
3. Feldmann, J., Youngblood, N., Wright, C.D., et al.: All-optical spiking neurosynaptic networks with self-learning capabilities. Nature **569**(7755), 208–214 (2019)
4. Xu, X., Tan, M., Corcoran, B., et al.: 11 TOPS photonic convolutional accelerator for optical neural networks. Nature **589**(7840), 44–51 (2021)
5. Feldmann, J., Youngblood, N., Karpov, M., et al.: Parallel convolutional processing using an integrated photonic tensor core. Nature **589**(7840), 52–58 (2021)
6. Zhang, H., Gu, M., Jiang, X.D., et al.: An optical neural chip for implementing complex-valued neural network. Nat. Commun. **12**(1), 457 (2021)
7. Ashtiani, F., Geers, A.J., Aflatouni, F.: An on-chip photonic deep neural network for image classification. Nature **606**(7914), 501–506 (2022)
8. Liu, W., Li, M., Guzzon, R., et al.: A fully reconfigurable photonic integrated signal processor. Nat. Photonics **10**(3), 190–195 (2016)
9. Zhou, T., Lin, X., Wu, J., et al.: Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. Nat. Photonics **15**(5), 367–373 (2021)
10. Xu, Z., Tang, B., Zhang, X., et al.: Reconfigurable nonlinear photonic activation function for photonic neural network based on non-volatile opto-resistive RAM switch. Light Sci. Appl. **11**(1), 288 (2022)
11. Zhou, W., Dong, B., Farmakidis, N., et al.: In-memory photonic dot-product engine with electrically programmable weight banks. Nat. Commun. **14**(1), 2887 (2023)
12. Wu, T., Menarini, M., Gao, Z., et al.: Lithography-free reconfigurable integrated photonic processor. Nat. Photonics (2023)
13. Wang, Q., Rogers, E., Gholipour, B., et al.: Optically reconfigurable metasurfaces and photonic devices based on phase change materials. Nat. Photonics **10**(1), 60–65 (2016)
14. Zhang, Y., Fowler, C., Liang, J., et al.: Electrically reconfigurable non-volatile metasurface using low-loss optical phase-change material. Nat. Nanotechnol. **16**(8), 661–666 (2021)
15. Fang, Z., Chen, R., Zheng, J., et al.: Ultra-low-energy programmable non-volatile silicon photonics based on phase-change materials with graphene heaters. Nat. Nanotechnol. **17**(9), 842–848 (2022)
16. Chen, R., Fang, Z., Perez, C., et al.: Non-volatile electrically programmable integrated photonics with a 5-bit operation. Nat. Commun. **14**(1), 3465 (2023)
17. Wuttig, M., Bhaskaran, H., Taubner, T., et al.: Phase-change materials for non-volatile photonic applications. Nat. Photonics **11**(8), 465–476 (2017)
18. Feldmann, J., Stegmaier, M., Gruhler, N., et al.: Calculating with light using a chip-scale all-optical abacus. Nat. Commun. **8**(1), 1256 (2017)
19. Ríos, C., Youngblood, N., Cheng Z., et al.: In-memory computing on a photonic platform. Sci. Adv. **5**(11), eaau5759 (2019)
20. Wu, C., Yu, H., Lee, S., et al.: Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. Nat. Commun. **12**(1), 96 (2021)
21. Hughes, T.W., Williamson, I.A.D., Minkov, M., et al.: Wave physics as an analog recurrent neural network. Sci. Adv. **5**(12), eaay6946 (2019)
22. Backer, A.S.: Computational inverse design for cascaded systems of metasurface optics. Opt. Express **27**(21), 30308–30331 (2019)
23. Hillenbrand, J., Getty, L.A., Clark, M.J., et al.: Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. **97**, 3099–3111 (1995)