

# Electrically programmable phase-change photonic memory for optical neural networks with nanoseconds *in situ* training capability

Maoliang Wei,<sup>a,†</sup> Junying Li,<sup>a,†</sup> Zequn Chen,<sup>b,c</sup> Bo Tang,<sup>d</sup> Zhiqi Jia,<sup>a</sup> Peng Zhang,<sup>d</sup> Kunhao Lei,<sup>a</sup> Kai Xu,<sup>a</sup> Jianghong Wu,<sup>b,c</sup> Chuyu Zhong,<sup>a</sup> Hui Ma,<sup>a</sup> Yuting Ye,<sup>b,c</sup> Jialing Jian,<sup>b,c</sup> Chunlei Sun,<sup>b,c</sup> Ruonan Liu,<sup>d</sup> Ying Sun,<sup>a</sup> Wei. E. I. Sha,<sup>e</sup> Xiaoyong Hu,<sup>e</sup> Jianyi Yang,<sup>a</sup> Lan Li,<sup>b,c</sup> and Hongtao Lin<sup>a,\*</sup>

<sup>a</sup>Zhejiang University, College of Information Science and Electronic Engineering, State Key Laboratory of Modern Optical Instrumentation, Key Laboratory of Micro-Nano Electronics and Smart System of Zhejiang Province, Hangzhou, China

<sup>b</sup>Westlake University, School of Engineering, Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, Hangzhou, China

<sup>c</sup>Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, China

<sup>d</sup>Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China

<sup>e</sup>Peking University, School of Physics, Frontiers Science Center for Nano-optoelectronics, State Key Laboratory for Mesoscopic Physics, Beijing, China

**Abstract** Optical neural networks (ONNs), enabling low latency and high parallel data processing without electromagnetic interference, have become a viable player for fast and energy-efficient processing and calculation to meet the increasing demand for hash rate. Photonic memories employing nonvolatile phase-change materials could achieve zero static power consumption, low thermal cross talk, large-scale, and high-energy-efficient photonic neural networks. Nevertheless, the switching speed and dynamic energy consumption of phase-change material-based photonic memories make them inapplicable for *in situ* training. Here, by integrating a patch of phase change thin film with a PIN-diode-embedded microring resonator, a bifunctional photonic memory enabling both 5-bit storage and nanoseconds volatile modulation was demonstrated. For the first time, a concept is presented for electrically programmable phase-change material-driven photonic memory integrated with nanosecond modulation to allow fast *in situ* training and zero static power consumption data processing in ONNs. ONNs with an optical convolution kernel constructed by our photonic memory theoretically achieved an accuracy of predictions higher than 95% when tested by the MNIST handwritten digit database. This provides a feasible solution to constructing large-scale nonvolatile ONNs with high-speed *in situ* training capability.

Keywords: phase-change materials; optical neural networks; photonic memory; silicon photonics; reconfigurable photonics.

Received Dec. 7, 2022; revised manuscript received Jun. 2, 2023; accepted for publication Jun. 25, 2023; published online Jul. 18, 2023.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.5.4.046004](https://doi.org/10.1117/1.AP.5.4.046004)]

## 1 Introduction

In recent years, neural networks based on central processing units (CPUs) have been used in mobile phones for speech recognition and image classification,<sup>1</sup> but they are still in their infancy in more sophisticated and expansive application fields where massive amounts of data should be processed in real time,

such as autonomous driving<sup>2</sup> and computer vision.<sup>3</sup> Optical neural networks (ONNs) based on photonic integrated circuits (PICs)<sup>4-9</sup> have the potential to meet this demand as a consequence of their low latency, high parallel (e.g., wavelength/spatial division multiplexing), and strong anti-electromagnetic interference capability of PICs, as well as the low cost and high yield provided by a complementary metal-oxide-semiconductor (CMOS) fabrication process.<sup>10-13</sup> Recently, a series of ONNs have been demonstrated for artificial intelligence, including vowel recognition,<sup>14</sup> perceptron,<sup>15,16</sup> pattern recognition,<sup>17</sup> and

\*Address all correspondence to Hongtao Lin, [hometown@zju.edu.cn](mailto:hometown@zju.edu.cn)

<sup>†</sup>These authors contributed equally to this work.

image classification.<sup>18,19</sup> However, for real-world applications, more efforts are needed to improve the energy efficiency, scalability, and algorithm accuracy of ONNs.

In on-chip ONNs, weights are determined by basic units of PICs altering their optical phase<sup>20</sup> or intensity.<sup>21</sup> These basic units commonly employ the thermo-optic (TO) effect, free-carrier dispersion effect, or nano-opto-electromechanical systems,<sup>22,23</sup> suffering from severe heat accumulation, high static power consumption or/and large footprint, which constrains the scalability of programmable photonic networks. On-chip integrated photonic memories, which can retain specific optical states after training (referring to all types of training), are anticipated to be embedded in programmable PICs to reduce or even eliminate static power consumption. Chalcogenide phase-change materials (PCMs) are promising candidates for zero static power-consumption photonic memories due to their reversible amorphous-crystalline phase transition,<sup>24–26</sup> and exceptional long-term, self-sustaining capability.<sup>27</sup> Moreover, the high optical contrast ( $\Delta n$ ) of PCMs between their covalent-bonded amorphous and resonant-bonded crystalline states makes ultracompact photonic memories achievable. Compared with photonic memories based on charge trapping,<sup>28</sup> and ferroelectric domain configuration,<sup>29</sup> or programmable nodes of PICs based on latched micromechanical systems,<sup>30</sup> photonic memories and nonvolatile PICs based on PCMs have the advantages of high stability, low loss, and especially small footprint. In the past decade, PCM-based integrated photonic memory (PM) has been demonstrated by adopting GeSbTe,<sup>31–36</sup> GeSbSeTe,<sup>37</sup> SbS,<sup>38</sup> SbSe,<sup>39,40</sup> etc. On-chip light-induced reconfigurable GST-based PM and its application in an ONN have been demonstrated.<sup>41</sup> However, for low-loss PCMs such as SbSe, optically induced reprogramming is inapplicable for scalable networks due to the negligible absorption loss at the telecom C-band. Electrothermal control of PCM not only addresses this issue but also has the potential for constructing large-scale nonvolatile programmable PICs. This makes electrically programmable PCM-based PICs much coveted in the future of high-efficiency and large-scale ONNs.<sup>42</sup>

On the other hand, *in situ* training (referring to training the ONN directly in the optical domain) is a potent remedy for enhancing the accuracy of algorithm execution in integrated ONNs,<sup>43–45</sup> which can not only improve the training speed but also reduce the influence of manufacturing errors and electrical/thermal cross talk.<sup>46</sup> However, although PCM-integrated photonic memories can make PICs highly energy-efficient after training, their long switching time and high switching energy consumption make them unsuitable for *in situ* training of ONNs, which hampers more accurate algorithm operation. Hence, an energy-efficient PM that could achieve high-speed volatile modulation at the same time is not only necessary but also pivotal, especially for *in situ* training of sporadic reprogramming ONNs exemplified by convolutional neural networks (CNNs).

Wavelength division multiplexing (WDM)-based computing is a potential arena for implementing optical CNNs.<sup>47</sup> Combined with the nonvolatile modulation of PCM, zero-static power consumption optical CNNs can be achievable.<sup>48</sup> Moreover, the combination of WDM and frequency comb makes ONNs with more complex functionality achievable.<sup>49</sup> Increasing the number of WDM channels can increase the amount of parallel computation of optical computing. The 2- $\mu\text{m}$  waveband is a promising candidate for expanding the number of channels thanks to the ignorable two-photon absorption at the 2- $\mu\text{m}$  waveband of

silicon<sup>50</sup> and the higher free-carrier dispersion effect of silicon at 2  $\mu\text{m}$ .<sup>51</sup>

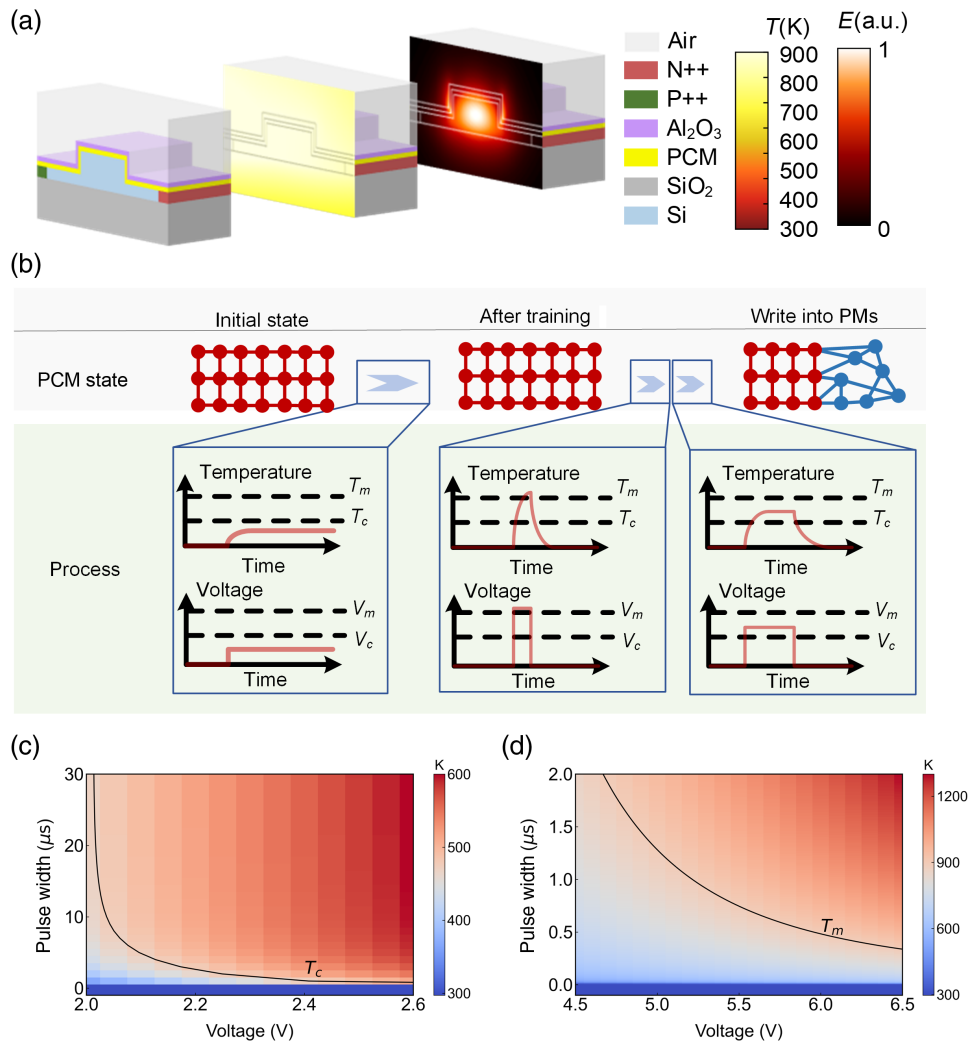
To date, to the best of our knowledge, nanosecond *in situ* training-compatible multilevel PM has not yet been studied. Here, we address these challenges by demonstrating an electrically programmable phase-change PM for ONNs. In this work, by integrating a low-loss PCM  $\text{Sb}_2\text{Se}_3$  with a p-i-n (PIN)-diode-embedded micro-ring resonator (MRR), a 2- $\mu\text{m}$  multilevel PM with more than 5 bits was demonstrated, and any specific intermediate optical state can be configured from an unknown state by applying certain electrical pulses. Meanwhile, volatile modulation with a speed of 15.2 MHz was enabled by keeping the driving voltage of the waveguide-integrated PIN diode under the threshold for triggering the phase change of the PCM. Such photonic memories can simultaneously realize *in situ* training and data storage in PICs for ONNs. In addition, this work provides a new paradigm for constructing CMOS-compatible, electrically programmable, nonvolatile on-chip photonic accelerators with high-speed *in situ* training capability, which we believe would contribute to the further development of energy-efficient, large-scale, high-yield ONNs.

## 2 Device Design

Figure 1(a) shows a schematic diagram of the PM enabling *in situ* training of ONNs. A patch of  $\text{Sb}_2\text{Se}_3$  phase-change thin film with a thickness of 30 nm was covered on a 600 nm-wide 150 nm-etched silicon waveguide, forming a low-loss  $\text{Sb}_2\text{Se}_3$ /silicon hybrid waveguide configuration similar to what we previously demonstrated in Ref. 52. When the phase transition of  $\text{Sb}_2\text{Se}_3$  occurs, it modifies the refractive index of the PCM patch and the effective refractive index ( $n_{\text{eff}}$ ) of the hybrid waveguide, which alters the resonant peak of the microring, thus changing the optical output of the PM. A 30 nm-thick  $\text{Al}_2\text{O}_3$  film was capped on the top to avoid oxidization of  $\text{Sb}_2\text{Se}_3$  during phase switching. A PIN diode was embedded in the silicon waveguide to not only support fast volatile modulation but also induce phase transition of the PCM above the waveguide by resistive heating.

Figure 1(b) depicts how PCM-integrated photonic memories operate in on-chip ONNs. Before the *in situ* training began, PCM patches of the photonic memories in an ONN were all initialized to the crystalline state. This was achieved by heating the PCM up via the PIN diode to a temperature higher than its crystallization temperature ( $T_c$ ) and holding for a period of time, for instance, 1 ms. During *in situ* training, the PIN diode in each PM was driven by a relatively low driving voltage, realizing the free-carrier dispersion effect-based volatile modulating, thus updating the weight in nanoseconds while keeping the temperature of the PCM below its crystallization temperature ( $T_c$ ). After *in situ* training, the trained weight information from volatile modulation was written into PCM-integrated memories by the ohmic heating effect of the PIN diode. To realize multibit memory, PCM was melted and then rapidly quenched, further heated to various temperatures between  $T_c$  and  $T_m$  (melting temperature) to partially crystallize to a certain optical state. After weights are written into PM, the on-chip ONN can compute passively, i.e., maintaining the weight info without power consumption.

The design of the PIN microheater is the key to the PCM-integrated PM. Since we employed standard concentrations of ion implantation in a multiproject wafer (MPW) run offered by the Institute of Microelectronics of the Chinese Academy of



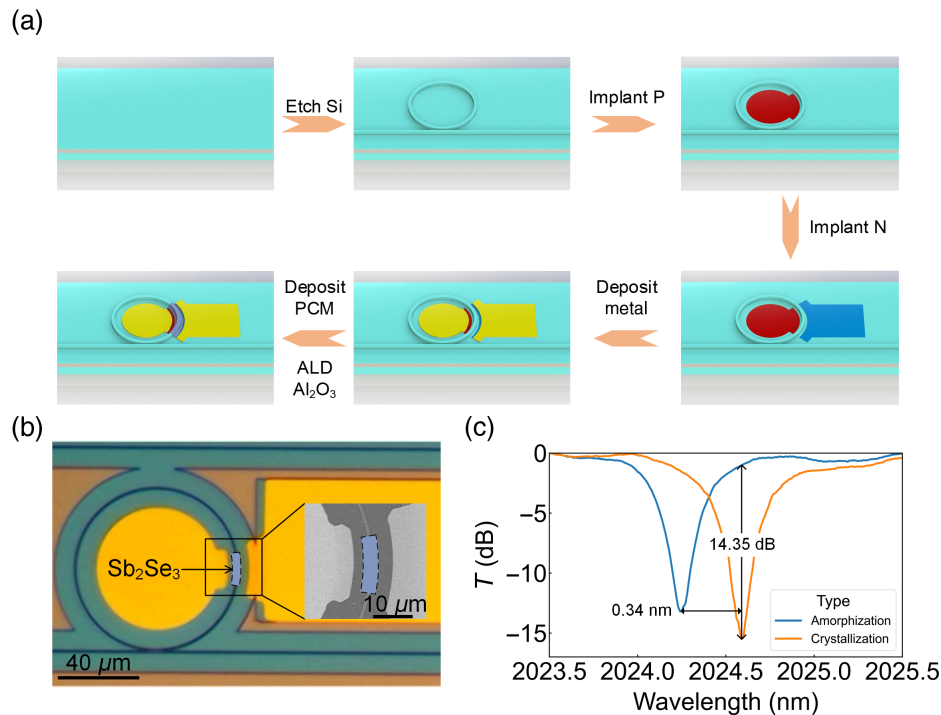
**Fig. 1** Design and operation principle of our PM. (a) Schematic diagram of the PM's structure, thermal distribution at a 6 V/500 ns voltage pulse, and optical mode profile at 2025 nm, respectively. (b) Operation principle of our PM. Simulated temperature variation of Sb<sub>2</sub>Se<sub>3</sub> patch with applied single pulses of different voltages and durations for (c) crystallization and (d) amorphization.

Sciences (IMCAS), the distance between the P<sup>++</sup>/N<sup>++</sup> heavily doping area and waveguide core was designed to balance the insertion loss and heating efficiency. The propagation loss of our PIN-diode-embedded waveguide is simulated to be 0.0042 dB/μm and experimentally measured to be 0.0065 dB/μm (see Sec. S1 in the [Supplementary Material](#)). Figure 1(a) shows the distribution of the thermal field in the PM when a 6 V/500 ns voltage pulse is applied. It could be seen that the PIN diode can effectively heat the PCM up to a certain temperature and induce a corresponding phase change by applying specific electrical pulses.

To separately manipulate volatile modulation and nonvolatile storage, electric pulses needed to be studied. According to our simulation, the bias current applied for fast volatile modulation based on the free-carrier dispersion effect should be lower than 5.84 mA to avoid the TO effect (see Sec. S2 in the [Supplementary Material](#)). At this point, the temperature of the whole waveguide region was simulated to be lower than 355 K, far below the crystallization temperature of Sb<sub>2</sub>Se<sub>3</sub>.

To write data to the PM, the driving voltage and pulse duration are the main parameters that need to be carefully designed and optimized. The longest pulse duration (or switching speed) of a PCM-based PM is limited by the crystallization process. Figure 1(c) shows the crystallization temperature of an Sb<sub>2</sub>Se<sub>3</sub> patch on the PIN diode with applied single pulses of different voltages and durations. It could be seen that the pulse duration needed for crystallization could be shortened by appropriately increasing the driving voltage, considering that the driving voltage required for crystallization is relatively low. In contrast, the highest driving voltage needed for a PCM-based PM depends on the amorphization process due to higher  $T_m$  than  $T_c$ , as shown in Fig. 1(d). However, the voltage of the amorphization pulse cannot be arbitrarily lowered by prolonging the pulse duration. On the one hand, the thermal decay rate of the system has to be larger than the critical cooling rate<sup>53</sup> to avoid recrystallization, yet the thermal decay time of the system is simulated to increase with the prolonged pulse duration. On the other hand, continuous increasing of the pulse duration with a certain voltage





**Fig. 2** Device fabrication and switching performance of our PM. (a) Fabrication flowchart of the device. (b) Microscope image of an  $\text{Sb}_2\text{Se}_3$  MRR PM. The inset shows an SEM image of the  $\text{Sb}_2\text{Se}_3$  (the shaded region) on top of the PIN diode. (c) Normalized transmittance spectra of the PM after the phase switching between two states of  $\text{Sb}_2\text{Se}_3$ .

amplitude ultimately leads to thermal saturation, and an over-long pulse duration brings about limited benefits. Hence, the duration of amorphization pulses is limited to within  $2 \mu\text{s}$  in our design. It could be seen from Fig. 1(d) that the driving voltage could be optimized down to 5 V theoretically. This driving voltage could be supplied by integrated circuits in standard CMOS technologies.<sup>54</sup>

Therefore, this PCM-integrated PM could potentially achieve a nonvolatile write speed of microseconds and write voltage lower than 5 V, as well as volatile modulation with nanoseconds for *in situ* training for ONNs. Although the optical loss in volatile phase modulation of a PIN diode is higher than that of a p-i-p (PIP) or n-i-n (NIN) doping waveguide,<sup>55,56</sup> it has prominent advantages of higher speed for volatile modulation due to the usage of the free-carrier dispersion effect of silicon rather than the TO effect of silicon. Moreover, the optical loss induced during volatile modulation becomes exploitable by integrating such a design with an MRR. Finally, the PIN diode microheater can reduce the driving voltage needed for phase switching of PCM compared to the PIP or NIN doping profile.<sup>31,40</sup>

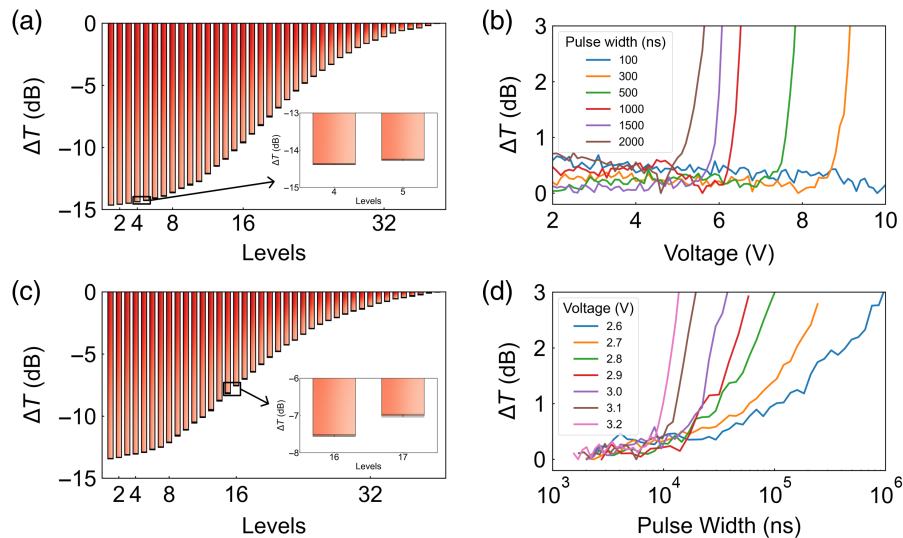
### 3 Multibit Low-Loss Photonic Memory

We experimentally demonstrated the  $\text{Sb}_2\text{Se}_3$ -integrated PM in the form of an all-pass MRR ( $\text{Sb}_2\text{Se}_3$  MRR). Figure 2(a) shows a schematic diagram of the fabrication process. The waveguide patterning and ion implantation were performed in an MPW run offered by IMCAS. The doping concentrations of p-type and n-type were  $2.0 \times 10^{20} \text{ cm}^{-3}$  and  $5.0 \times 10^{20} \text{ cm}^{-3}$ , respectively. Then, metallic electrodes (5 nm Cr/100 nm Au) and  $\text{Sb}_2\text{Se}_3$  patches were fabricated by UV lithography followed by a

lift-off process. Finally, a 30 nm  $\text{Al}_2\text{O}_3$  was deposited, and the metal contact window was opened by etching.

Figure 2(b) shows an optical microscope image of the fabricated  $\text{Sb}_2\text{Se}_3$  MRR with a radius of  $40 \mu\text{m}$ . A  $15\text{-}\mu\text{m}$ -long  $\text{Sb}_2\text{Se}_3$  patch was covered on a  $20\text{-}\mu\text{m}$ -long PIN diode embedded in the resonator. A home-built integrated photonic measurement setup (see Sec. S3 in the [Supplementary Material](#)) was used to characterize the PM. To eliminate the temperature perturbation derived from ambient temperature variation, the temperature of the substrate of the photonic chip is held to  $30^\circ\text{C}$  throughout the test via a temperature control system. Figure 2(c) shows the change of normalized transmittance ( $T$ ) spectra of the  $\text{Sb}_2\text{Se}_3$  MRR when the phase transition of  $\text{Sb}_2\text{Se}_3$  occurs. When  $\text{Sb}_2\text{Se}_3$  was crystallized by a 3.0 V/1 ms voltage pulse or amorphized by an 8.2 V/500 ns pulse, a resonance peak shift of 0.34 nm and an extinction ratio over 14 dB were realized.

Here, we systematically characterized the effect of amplitude and duration of voltage pulses on the multilevel switching response of photonic memories. The  $\text{Sb}_2\text{Se}_3$  patch was gradually amorphized and generated 38 levels in the PM by applying an electric pulse with a duration of 500 ns and voltage amplitudes not exceeding 8.2 V. The transmission change ( $\Delta T$ ) and storage levels are shown in Fig. 3(a). Each optical storage level is the average value of 50 measurements in the same state to avoid test errors due to systematic noise. The lowest resolution of these memory states is 0.07 dB. Among them, 28 levels were distinguishable after the transmission change is converted to the linear region, which can be used for info storage for optical computing. As our simulations confirmed, prolonging the pulse width can reduce the driving voltage for the melt quenching of  $\text{Sb}_2\text{Se}_3$



**Fig. 3** The change in transmittance of the PM under multilevel states. (a) Amorphization (at 2024.59 nm) and (c) crystallization (at 2024.25 nm). The inset shows the enlarged error bar of two randomly chosen storage levels. Change in the transmittance of the PM with different voltages and pulse widths for (b) amorphization and (d) crystallization.

during amorphization [see Fig. 3(b)]. By employing a pulse duration of 2  $\mu$ s, the driving voltage needed for partial amorphization of  $\text{Sb}_2\text{Se}_3$  to generate a transmittance change could be reduced to 5.3 V. The device would be damaged once the pulse duration of the relatively high-voltage amorphization pulse exceeded 2  $\mu$ s; hence, the pulse duration should be kept within 2  $\mu$ s. The amorphization driving voltage could be reduced to 4.4 V by narrowing the gap between the waveguide and the metal contact (see Sec. S4 in the [Supplementary Material](#)), suggesting good scaling potential with improved energy efficiency.

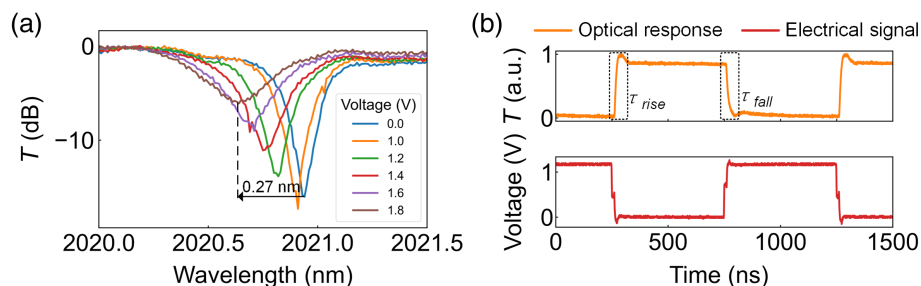
As for multilevel crystallization, by applying fixed voltage amplitude at 3 V and various pulse durations of no more than 50  $\mu$ s, 40 memory states were demonstrated with a resolution higher than 0.07 dB, as shown in Fig. 3(c). After conversion to the linear domain, there are still 34 different states (more than 5 bits). Each level was also averaged by 50 measurements. The standard deviation in Fig. 3(c) confirms that the states are separable even with noise in the measurement system. The write speed of the PM could be further improved by increasing the driving voltage for crystallization, as shown in Fig. 3(d), consistent with our design.

Hence, a 5-bit PCM-integrated PM was demonstrated, with a driving voltage lower than 10 V and a switching time within

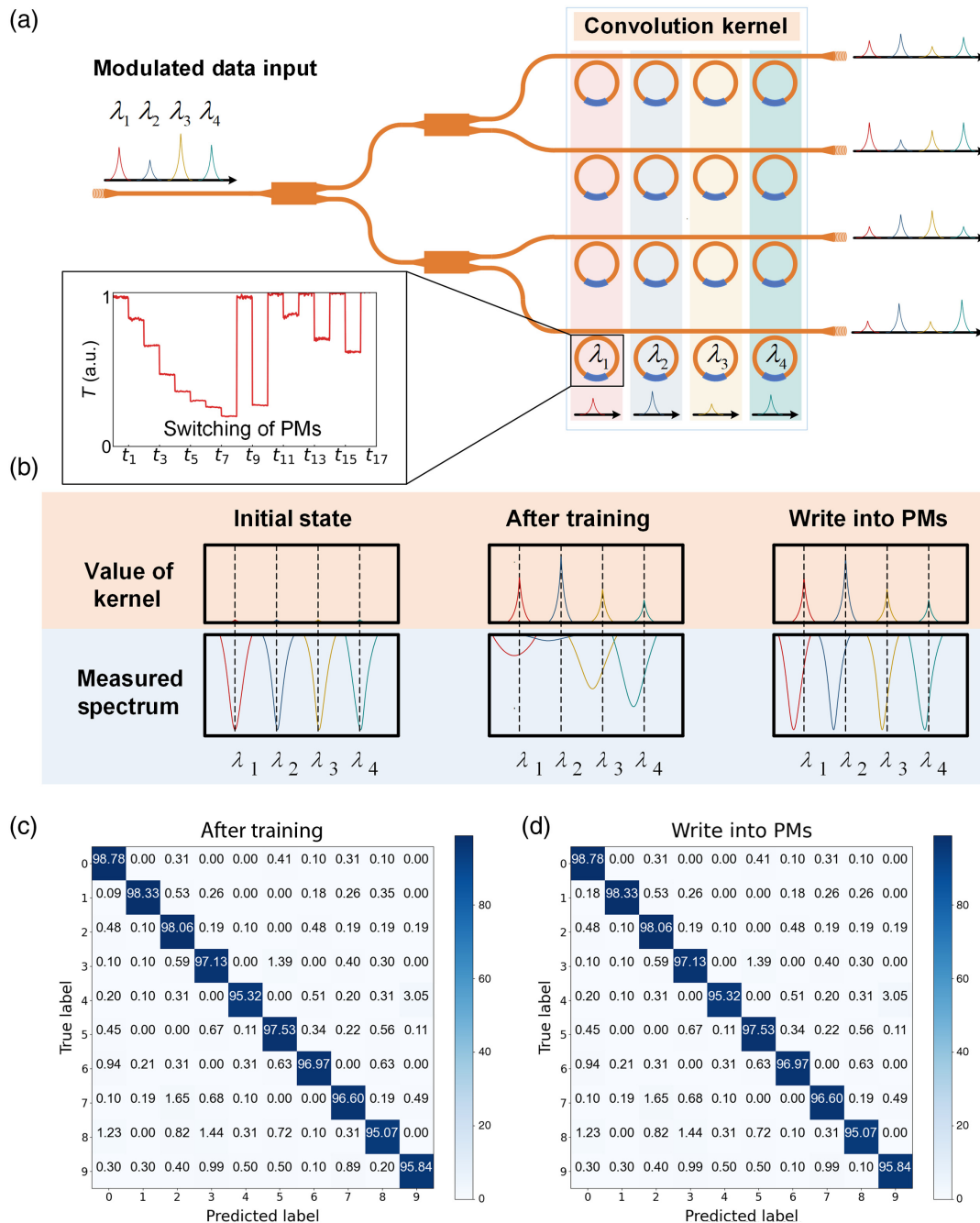
tens of microseconds. The experimental driving voltage is not as low as the simulated one, which may result from nonideal ion implantation and activation in the device fabrication.

#### 4 Volatile Modulation-Compatible Photonic Memory for ONNs

A photonic neural network with PCM-integrated memory is of zero static power consumption, but *in situ* training via continually and intensively switching the phase of PCM is neither energy-efficient nor fast enough. Here, we address this issue by embedding a volatile modulation function into nonvolatile PM. Figure 4(a) shows the change of normalized transmittance spectra of the PM during volatile modulation used in the *in situ* training process. Note that the  $\text{Sb}_2\text{Se}_3$  patch on the PIN diode is now amorphized. The ripples of the measured spectra resulted from the Fabry–Perot resonance due to the reflection of the grating coupler. A peak shift efficiency of 0.15 nm/V was realized. Figure 4(b) shows the dynamic response of the PM when a 1.3 V, 1 MHz square-wave signal was applied. The 10%-to-90% rising time ( $\tau_{\text{rise}}$ ) and 90%-to-10% falling time ( $\tau_{\text{fall}}$ ) are characterized to be 13.4 and 23.0 ns, respectively, corresponding to a 3 dB bandwidth of 15.2 MHz.



**Fig. 4** Volatile modulation of an  $\text{Sb}_2\text{Se}_3$  MRR. (a) Normalized transmittance spectra with different forward biases. (b) Dynamic response of the  $\text{Sb}_2\text{Se}_3$  MRR.



**Fig. 5** OCK based on the volatile-modulation-compatible PM. (a) Schematic architecture of a  $4 \times 4$  OCK. The inset is the solidifying method of the output value of each basic unit in ONNs after on-chip training of OCK. (b) Schematic diagram of the on-chip training and writing operation of the OCK. The accuracy of predictions (c) after the simulated on-chip training of OCK and (d) after simulated writing into PMs.

Here, we simulated electrically programmable ONNs by Python exemplified by a  $4 \times 4$  optical convolution kernel (OCK) constructed by the PM, as shown in Fig. 5(a). Since the PCM-integrated PM was demonstrated in the form of an MRR, the convolution operation was implemented through a WDM scheme. Modulated optical signals with four different wavelengths were equally sent to the OCK in four equal channels. After the optical convolution operation, optical signals were

converted to electrical signals, amplified by the transimpedance amplifier, and then processed by the CPU. Any intermediate storage state could be configured from an unknown state by employing two electrical pulses (one for amorphization and the other for crystallization), and the measured transmission change ( $T$ ) is shown in the inset of Fig. 5(a). Thus, our proposed OCK is capable of both fast on-chip training and computing with near-zero power consumption.



The PM-embedded OCK was theoretically verified by the MNIST handwritten digit database. Before the on-chip training of OCK execution, the states of all SbSe patches are initialized to their crystalline state. After that, the on-chip training of OCK was implemented by exploiting the volatile modulation of our PM. Then, the trained weights were written to the PM by applying a reset (amorphization) pulse followed by a fractional-crystallization pulse after the on-chip training of OCK. Figure 5(b) shows a schematic diagram of the evolution of measured transmittance spectra and kernel value. The trained and stored MRR arrays have different transmittance spectra, since the on-chip training of OCK and writing were conducted through different principles and approaches. Yet the value of weights after the on-chip training of OCK and writing should be as close as possible (and ideally the same). The question naturally arises over whether the discrete storage states of PCM-based PMs may lead to performance deterioration of the OCK. To verify this, the accuracy of predictions after the simulation of the on-chip training of OCK via PIN diodes (>95%) is shown in Fig. 5(c). After the trained parameters were written into the PMs, the implementation of the network reached minimal deviation in accuracy, as shown in Fig. 5(d). Note that the scale of the MRR array could be easily expanded. Considering there are  $M$  channels for data processing, the OCK could be scaled up to  $M \times 21$  by simply decreasing the radius of the Sb<sub>2</sub>Se<sub>3</sub> MRR to 8  $\mu\text{m}$  in theory (see Sec. S5 in the [Supplementary Material](#)).

The PM-based convolution core benefits both on-chip training of OCK and low-static power computing. The on-chip training of OCK based on the volatile-compatible PM provides a training speed typically 1000 times faster than the commonly used TO scheme.<sup>20</sup> After on-chip training of OCK, the computing is done passively without static power consumption. With this scheme, the saved power consumption of an  $M \times 21$  OCK is  $M \times 210$  mW, compared with the typical TO modulator array with 10 mW of each discrete device on average.<sup>20,57</sup> Therefore, the ONNs with PM are attractive in sporadic programming applications, and the power efficiency would increase with the scaling up of PICs.

In large-scale ONNs where PMs are expected to be used in the whole linear network, multibit storage of PMs can play a significant role. For instance, constructing an ONN (with a  $16 \times 4$  OCK) from PMs where the *in situ* training results showed an averaged prediction accuracy rate of 94.64% identifying the MNIST data set, PMs need at least 4 bits to achieve comparable prediction accuracy (averaged accuracy rate >94%), as shown in Sec. S6 in the [Supplementary Material](#). This indicates that multibit PMs are necessary for high-performance ONNs, and higher bits are expected for more complicated applications.

## 5 Conclusion

In this work, we proposed an electrically programmable phase-change PM for energy-efficient *in situ* training ONNs with CMOS compatibility and scalability. By integrating an Sb<sub>2</sub>Se<sub>3</sub> phase-change patch onto a PIN diode, we designed and experimentally validated the PCM-driven 5-bit PM using an MRR. The PM exhibits a transmittance contrast of 14.63 dB/13.42 dB, creating 28/34 storage levels during amorphization/crystallization, and the corresponding pulse voltages (pulse durations) are 7.4 to 8.2 V (0.5  $\mu\text{s}$ )/3 V (10 to 50  $\mu\text{s}$ ). Furthermore, theoretically, complete amorphization of

Sb<sub>2</sub>Se<sub>3</sub> can be induced by a 500-ns electrical pulse with an actuation voltage as low as 3.3 V, which can be provided by an integrated circuit with standard CMOS technology. In our experiment, fractional amorphization was achieved by applying a 4.4 V/2  $\mu\text{s}$  voltage pulse. Volatile modulation with a bandwidth of >15 MHz was also achieved in this PM when electric pulses with voltages lower than 2 V were applied, enabling a 1000 times faster training in theory for nonvolatile ONNs composed of such PMs than the commonly used TO switches. After training, PMs are configured to specific states via PIN-microheater-assisted multilevel switching (i.e., partial phase transition) of Sb<sub>2</sub>Se<sub>3</sub> to match the target weight values in the ONNs. According to our simulations, at least 4 bits are needed for PMs to maintain the accuracy of predictions of ONNs after the simulated *in situ* training when tested by the MNIST handwritten data set. This study on volatile modulation-compatible PM provides a feasible solution for constructing nonvolatile ONNs with high-speed and energy-efficient on-chip training capability.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (2019YFB2203002 and 2021YFB2801300), National Natural Science Foundation of China (62105287, 91950204, and 61975179), and Zhejiang Provincial Natural Science Foundation (LD22F040002). The authors would like to acknowledge the fabrication support from the Institute of Microelectronics of the Chinese Academy of Sciences, ZJU Micro-Nano Fabrication Center at Zhejiang University, and Westlake Center for Micro/Nano Fabrication at Westlake University. The authors would also like to thank Qing Zhao and Liming Shan for their help in thin-film depositions and Xingjie Li for his help in developing the test program. The authors declare no conflicts of interest.

## References

1. C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: a survey," *IEEE Commun. Surv. Tutor.* **21**(3), 2224–2287 (2019).
2. L. Chen et al., "Deep neural network based vehicle and pedestrian detection for autonomous driving: a survey," *IEEE Trans. Intell. Transp. Syst.* **22**(6), 3234–3246 (2021).
3. J. Chai et al., "Deep learning in computer vision: a critical review of emerging techniques and application scenarios," *Mach. Learn. Appl.* **6**, 100134 (2021).
4. P. Xu and Z. Zhou, "Silicon-based optoelectronics for general-purpose matrix computation: a review," *Adv. Photonics* **4**(4), 044001 (2022).
5. H. Zhou et al., "Photonic matrix multiplication lights up photonic accelerator and beyond," *Light Sci. Appl.* **11**(1), 30 (2022).
6. B. J. Shastri et al., "Photonics for artificial intelligence and neuro-morphic computing," *Nat. Photonics* **15**(2), 102–114 (2021).
7. C. Li et al., "The challenges of modern computing and new opportunities for optics," *Photonix* **2**(1), 20 (2021).
8. J. Liu et al., "Research progress in optical neural networks: theory, applications and developments," *Photonix* **2**(1), 5 (2021).
9. X. Xu et al., "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* **589**(7840), 44–51 (2021).
10. T. J. Seok et al., "Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers," *Optica* **3**(1), 64–70 (2016).
11. W. Bogaerts et al., "Programmable photonic circuits," *Nature* **586**(7828), 207–216 (2020).

12. S. Y. Siew et al., "Review of silicon photonics technology and platform development," *J. Lightwave Technol.* **39**(13), 4374–4389 (2021).
13. H. Shu et al., "Microcomb-driven silicon photonic systems," *Nature* **605**(7910), 457–463 (2022).
14. S. Bandyopadhyay et al., "Single chip photonic deep neural network with accelerated training," arXiv:2208.01623 (2022).
15. S. Pai et al., "Experimentally realized *in situ* backpropagation for deep learning in photonic neural networks," *Science* **380**, 398–404 (2023).
16. H. Zhang et al., "An optical neural chip for implementing complex-valued neural network," *Nat. Commun.* **12**(1), 457 (2021).
17. J. Y. S. Tan et al., "Monadic Pavlovian associative learning in a backpropagation-free photonic network," *Optica* **9**(7), 792–802 (2022).
18. J. Feldmann et al., "Parallel convolutional processing using an integrated photonic tensor core," *Nature* **589**(7840), 52–58 (2021).
19. F. Ashtiani, A. J. Geers, and F. Aflatouni, "An on-chip photonic deep neural network for image classification," *Nature* **606**(7914), 501–506 (2022).
20. Y. Shen et al., "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**(7), 441–446 (2017).
21. Z. G. Cheng et al., "On-chip photonic synapse," *Sci. Adv.* **3**(9), e1700160 (2017).
22. P. Edinger et al., "Silicon photonic microelectromechanical phase shifters for scalable programmable photonics," *Opt. Lett.* **46**(22), 5671–5674 (2021).
23. D. Pérez, I. Gasulla, and J. Capmany, "Programmable multifunctional integrated nanophotonics," *Nanophotonics* **7**(8), 1351–1371 (2018).
24. K. Shportko et al., "Resonant bonding in crystalline phase-change materials," *Nat. Mater.* **7**(8), 653–658 (2008).
25. A.-K. U. Michel et al., "Using low-loss phase-change materials for mid-infrared antenna resonance tuning," *Nano Lett.* **13**(8), 3470–3475 (2013).
26. L. Mao et al., "Reversible switching of electromagnetically induced transparency in phase change metasurfaces," *Adv. Photonics* **2**(5), 056004 (2020).
27. M. Wuttig and N. Yamada, "Phase-change materials for rewritable data storage," *Nat. Mater.* **6**(11), 824–832 (2007).
28. J.-F. Song et al., "Integrated photonics with programmable non-volatile memory," *Sci. Rep.* **6**(1), 22616 (2016).
29. J. Geler-Kremer et al., "A ferroelectric multilevel non-volatile photonic phase shifter," *Nat. Photonics* **16**(7), 491–497 (2022).
30. S. Abe and K. Hane, "A silicon microring resonator with a nanolatch mechanism," *Microsyst. Technol.* **21**(9), 2019–2024 (2015).
31. J. Zheng et al., "Nonvolatile electrically reconfigurable integrated photonic switch enabled by a silicon PIN diode heater," *Adv. Mater.* **32**(31), 2001218 (2020).
32. C. Ríos et al., "In-memory computing on a photonic platform," *Sci. Adv.* **5**(2), eaau5759 (2019).
33. D. Wu et al., "Resonant multilevel optical switching with phase change material GST," *Nanophotonics* **11**(15), 3437–3446 (2022).
34. N. Farmakidis et al., "Electronically reconfigurable photonic switches incorporating plasmonic structures and phase change materials," *Adv. Sci.* **9**(20), 2200383 (2022).
35. C. Wu et al., "Low-loss integrated photonic switch using subwavelength patterned phase change material," *ACS Photonics* **6**(1), 87–92 (2019).
36. H. Zhang et al., "Miniature multilevel optical memristive switch using phase change material," *ACS Photonics* **6**(9), 2205–2212 (2019).
37. Y. Zhang et al., "Broadband transparent optical phase change materials for high-performance nonvolatile photonics," *Nat. Commun.* **10**(1), 4279 (2019).
38. Z. Fang et al., "Non-volatile reconfigurable integrated photonics enabled by broadband low-loss phase change material," *Adv. Opt. Mater.* **9**(9), 2002049 (2021).
39. Z. Fang et al., "Ultra-low-energy programmable non-volatile silicon photonics based on phase-change materials with graphene heaters," *Nat. Nanotechnol.* **17**(8), 842–848 (2022).
40. C. Ríos et al., "Ultra-compact nonvolatile phase shifter based on electrically reprogrammable transparent phase change materials," *Photonix* **3**(1), 26 (2022).
41. J. Feldmann et al., "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature* **569**(7755), 208–214 (2019).
42. X. Ma et al., "Photonic tensor core with photonic compute-in-memory," in *Opt. Fiber Commun. Conf. and Exhibit. (OFC)*, pp. 1–3 (2022).
43. T. W. Hughes et al., "Training of photonic neural networks through *in situ* backpropagation and gradient measurement," *Optica* **5**(7), 864–871 (2018).
44. H. Zhou et al., "All-in-one silicon photonic polarization processor," *Nanophotonics* **8**(12), 2257–2267 (2019).
45. H. Zhou et al., "Chip-scale optical matrix computation for PageRank algorithm," *IEEE J. Sel. Top. Quantum Electron.* **26**(2), 8300910 (2020).
46. S. M. Buckley et al., "Photonic online learning: a perspective," *Nanophotonics* **12**, 833–845 (2023).
47. S. Xu, J. Wang, and W. Zou, "Optical convolutional neural network with WDM-based optical patching and microring weighting banks," *IEEE Photonics Technol. Lett.* **33**(2), 89–92 (2021).
48. F. Brücknerhoff-Plückelmann et al., "Broadband photonic tensor core with integrated ultra-low crosstalk wavelength multiplexers," *Nanophotonics* **11**(17), 4063–4072 (2022).
49. B. Bai et al., "Microcomb-based integrated photonic processing unit," *Nat. Commun.* **14**(1), 66 (2023).
50. R. Soref, "Mid-infrared photonics in silicon and germanium," *Nat. Photonics* **4**(8), 495–497 (2010).
51. M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, "Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1–14  $\mu\text{m}$  infrared wavelength range," *IEEE Photonics J.* **3**(6), 1171–1180 (2011).
52. K. Lei et al., "Magnetron-sputtered and thermal-evaporated low-loss Sb-Se phase-change films in non-volatile integrated photonics," *Opt. Mater. Express* **12**(7), 2815–2823 (2022).
53. Y. Zhang et al., "Myths and truths about optical phase change materials: a perspective," *Appl. Phys. Lett.* **118**(21), 210501 (2021).
54. H. Ballan and M. Declercq, *High Voltage Devices and Circuits in Standard CMOS Technologies*, pp. 268–269, Springer Science & Business Media, Boston (2013).
55. C. Zhong et al., "Fast thermo-optical modulators with doped-silicon heaters operating at 2  $\mu\text{m}$ ," *Opt. Express* **29**(15), 23508–23516 (2021).
56. M. Wei et al., "TDFA-band silicon optical variable attenuator," *Prog. Electromagn. Res.* **174**, 33–42 (2022).
57. S. Liu et al., "Thermo-optic phase shifters based on silicon-on-insulator platform: state-of-the-art and a review," *Front. Optoelectron.* **15**(1), 9 (2022).
58. J. R. Erickson et al., "Designing fast and efficient electrically driven phase change photonics using foundry compatible waveguide-integrated microheaters," *Opt. Express* **30**(8), 13673–13689 (2022).
59. H. Ma et al., "Passive devices at 2  $\mu\text{m}$  wavelength on 200 mm CMOS-compatible silicon photonics platform [Invited]," *Chin. Opt. Lett.* **19**(7), 071301 (2021).



**Maoliang Wei** is a doctoral candidate student of Professor Hongtao Lin at Zhejiang University. He received his BS degree in electronic information science and technology from Xiamen University. His current research focuses on the study of micronano devices and systems.

**Junying Li** is an associate professor at the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang, China. She received her BS and PhD degrees from Chongqing University, Chongqing, China. Her research is focused on chalcogenide phase-change materials, reconfigurable photonic devices, and their applications.

**Hongtao Lin** is a ZJU100 Young Professor at the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang, China. He received his BS degree from the University of Science and Technology of China, Hefei, China, and his PhD from the University of Delaware, Newark, Delaware, USA. His research interests are focused on chalcogenide integrated nanophotonics and their applications.

Biographies of the other authors are not available.