ADVANCED PHOTONICS RESEARCH
Open Access
www.adpr-journal.com

Digitized Phase-Change Material Heterostack for Transmissive Diffractive Optical Neural Network

*Ruiyang Chen, Jichao Fan, Cunxi Yu, and Weilu Gao**

All-optical and fully reconfigurable transmissive diffractive optical neural network (DONN) architectures emerge as high-throughput and energy-efficient machine learning (ML) hardware accelerators in broad applications. However, current device and system implementations have limited performance. In this work, a novel transmissive diffractive device architecture, a digitized phase-change material (PCM) heterostack, which consists of multiple nonvolatile PCM layers with different thicknesses, is demonstrated. Through this architecture, the advantages of PCM electrical and optical properties can be leveraged and challenges associated with multilevel operations in a single PCM layer can be mitigated. Through proof-of-concept experiments, the electrical tuning of one PCM layer is demonstrated in a transmissive spatial light modulation device, and thermal analysis guides the design of multilayer devices and DONN systems to avoid thermal cross talk if individual heterostacks are assembled into an array. Further, a heterostack containing three PCM layers is designed based on experimental results to produce a large-phase modulation range and uniform coverage, and the ML performance of DONN systems with the designed heterostack is evaluated. The developed device architecture is practically feasible and scalable for future energy-efficient, fast-reconfigured, and compact transmissive DONN systems.

1. Introduction

Machine learning (ML) has transformed a broad range of applications, including imaging and sensing,[1–3] chip design,[4] and scientific discovery.[5,6] With the explosively increasing size of ML models, such as ChatGPT, their execution on hardware requires unsustainably large computational resources and energy consumption. Alternative to electronic platforms, optical computing platforms have recently gained much interest as new high-throughput and energy-efficient hardware ML accelerators thanks to the extreme parallelism and low static energy consumption of photons.[7] For example, 2D integrated photonics-based processors with different types of on-chip optoelectronic modulators can perform various ML tasks and mathematical operations.[8–11] In addition, 3D free-space optical systems,[12–23] which spatially manipulate the amplitude, phase, and polarization of light with engineered 2D surfaces, advance optical computing capabilities. In particular, diffractive optical neural network (DONN) systems operating based on the physics of spatial light modulation and optical diffraction enable ML functionalities.[23]

DONN systems can be constructed in either transmission or reflection modes; see Figure S1, Supporting Information for illustrations. Transmissive DONNs are advantageous over reflective ones due to multiple reasons. First, the normal light incidence in transmissive systems facilitates optical alignment compared to reflective systems with oblique light incidence. Second, the conventional diffraction theory and model assume normal incidence so that they match well with practical transmissive systems. In contrast, when the oblique incident angle is large in reflective systems, the diffraction calculations require substantial modifications with new numerical procedures to reduce the discrepancy with practical reflective DONN systems.[24] Third, transmissive DONNs can achieve a large number of diffractive neurons and deep network structures in a compact footprint, while reflective systems need more space because of inevitable oblique incidence and layer separation. Further, most current DONN systems cannot be reconfigured for different ML tasks on demand once diffractive arrays are manufactured. Note that the reconfiguration of diffractive arrays occurs only when ML tasks change, which is not frequent and is distinct from encoding data streams using fast electro-optic modulators. Limited reconfigurability was implemented in a hybrid optoelectronic reflective DONN system,[17] where electrical-to-optical (E/O) and optical-to-electrical (O/E) conversions occurred between diffractive arrays. However, these intermediate conversions increased energy consumption and processing latency, and the induced deployment errors required additional iterative adaptive tuning processes. Hence, the most desirable DONN systems are transmissive

R. Chen, J. Fan, W. Gao
Department of Electrical and Computer Engineering
The University of Utah
Salt Lake City, UT 84112, USA
E-mail: weilu.gao@utah.edu

C. Yu
Department of Electrical and Computer Engineering
The University of Maryland College Park
College Park, MD 20742, USA

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adpr.202400201.

DOI: 10.1002/adpr.202400201

and fully reconfigurable in optical domains. We recently made a major advance toward this ultimate goal by constructing a fully reconfigurable transmissive DONN system based on cascaded liquid crystal spatial light modulators in the visible wavelength range.[20] This system is all-optical without E/O and O/E conversions. However, the employed liquid-crystal-based diffractive devices need external voltage and current to maintain the rotation states of liquid crystal molecules when performing the inference of ML tasks, leading to static energy consumption. Further, the reconfiguration speed is limited to hundreds of milliseconds. The lateral size of each liquid crystal pixel is also much larger than the operation wavelength. Hence, it is crucial to create new transmissive diffractive devices and arrays based on new materials and architectures, which are energy efficient without maintaining external stimulus during ML inference, fast reconfigured, and of compact footprint, toward the most desirable systems.

Here, we present a novel near-infrared transmissive diffractive device based on nonvolatile chalcogenide phase-change materials (PCMs), which can be beneficial for DONN systems from multiple perspectives. First, the material states of PCMs can be fast and reversibly reconfigured between crystalline or amorphous phases with an electrical or optical pulse.[25–27] Second, the material states can be preserved after reconfiguration without external stimulus for >10 years[28,29] such that static energy consumption is minimal.[30] Third, the change of optical refractive indices can be >1 over a wide spectral range,[25] enabling compact and broadband operations for integrated[9,31–34] and free-space photonic components and systems.[35–40] Fourth, PCMs are scalable[41] and compatible with other materials for constructing complex architectures. The capability of handling complex ML tasks in DONN systems is dependent on the number of modulation levels achieved in each diffractive pixel. Multilevel operations in practical PCM-based devices are implemented with multiple intermediate states, which are a mixture of crystalline and amorphous phases and typically implemented by carefully designing electrical pulse profiles to deliver various heating energy.[31,38,39,42,43] However, uniformly and reliably achieving many levels (e.g., >10) in a single PCM layer is still challenging and the formation of intermediate states is intrinsically stochastic,[44] which can potentially lead to nonuniform and unreliable operations in large-area devices. Hence, to mitigate these challenges associated with multilevel operations in a single PCM layer, we demonstrate a digitized PCM heterostack architecture containing multiple PCM films with different thicknesses and other dielectrics. Each PCM film is only under the crystalline and amorphous phases or with only a few intermediate states. We fabricated and experimentally characterized a proof-of-concept device for spatial light modulation and performed thermal analysis to determine the needed spacing between PCM films in the heterostack and spacing between heterostack pixels to avoid thermal cross talk if individual diffractive devices are assembled into an array for a transmissive DONN system. Together with the obtained proof-of-concept experimental results, we further designed layer thicknesses in a heterostack containing three PCM layers to have a large-phase modulation range ($>0.5\pi$) and uniform coverage and finally evaluated the performance of the DONN system employing designed heterostacks.

2. Results

Figure 1a illustrates the general architecture of a transmissive DONN system consisting of multiple cascaded diffractive arrays. The input laser beam is incident on the images, diffracted by cascaded diffractive arrays, and captured on a camera that is essentially a 2D photodetector array. On the camera, there are multiple predefined localized regions to represent the meaning of input images. For example, we can define 10 regions to represent 10 handwritten digits in the Modified National Institute of Standards and Technology (MNIST) database. Each pixel on a diffractive array can modulate the amplitude and phase of transmitted light. By training (i.e., optimizing) excitation and thus responses of each pixel on all diffractive arrays, the light wavefront after diffraction can converge into a predefined localized region and hence the DONN system can perform classification tasks on input images.

To create an energy-efficient, compact reconfigurable diffractive array, we developed a digitized PCM heterostack for each pixel as illustrated in Figure 1b. The specific PCM we utilized is the widely demonstrated $Ge_2Se_2Te_5$ (GST) material. There are multiple stacks of GST films with different thicknesses, such as three thicknesses $\{d_1, d_2, d_3\}$ in Figure 1b. Each GST layer has an independent heater and electrodes to control GST phases in fully crystalline or amorphous phases, which is called a 2-level GST operation. Since layer thicknesses are different, the

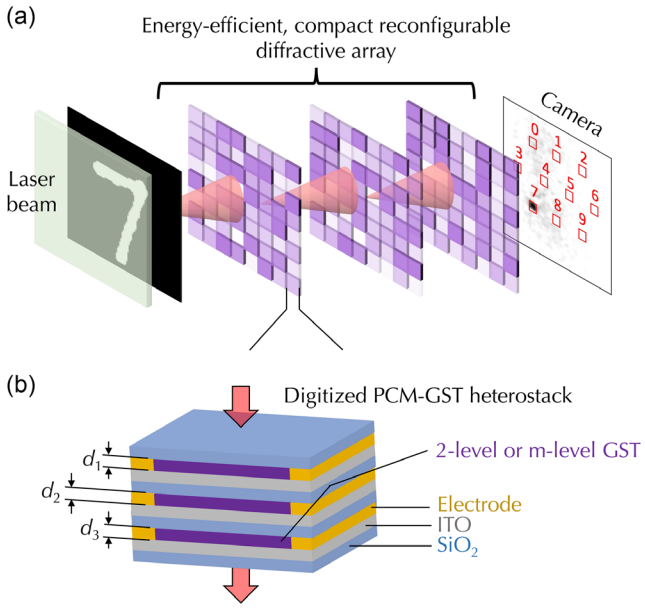

Figure 1. a) Schematic of a transmissive DONN system with cascaded energy-efficient, compact reconfigurable diffractive arrays. A laser is incident on images, diffracted by cascaded diffractive arrays, and captured on a camera that consists of multiple predefined regions. Each diffractive pixel modulates the amplitude and phase of the transmitted light under external stimulus. By optimizing the stimulus of each pixel on diffractive arrays, the diffracted light converges into a predefined region based on input images for performing classification tasks. b) Illustration of each diffractive pixel implemented using a digitized GST heterostack consisting of multiple GST layers with different thicknesses. Each GST layer can be addressed electrically and individually.

amplitude and phase modulation from the phase transition in each GST layer have different weights in the total modulation of transmitted light, which is the principle of implementing multilevel operations. This is analogous to the conversion from binary to decimal numeral systems. For example, each 1 in a binary number $(111)_2$ has a different weight in its corresponding decimal number. The leading 1 represents 4, the second 1 represents 2, and the trailing 1 represents 1. The weight of each 1 is similar to the thickness of each GST layer, and the 0 (1) on each digit is similar to the fully amorphous (crystalline) phase. Because of multiple reflections between layers, the thicknesses d_1, d_2, and d_3 in a 3-layer architecture may not follow the ratio 1:2:4 as in binary numbers. Instead, these thicknesses can be optimized freely to have a large range and uniformly distributed multilevel modulation. Further, the 2-level GST operation can be extended to a few levels (e.g., 5 levels) with some relatively stable and easy-to-achieve intermediate states, named the *m*-level operation. Practically, the as-fabricated heterostack with as-deposited amorphous GST films is first globally heated through hot plates or ovens to have crystalline GST films as the initial state, and then each GST layer is reconfigured sequentially using electrical pulses. The heterostack can be reset to its initial state through global heating or large electrical pulses.

For a proof-of-concept demonstration of the electrical tuning of one GST film in a transmissive spatial light modulation device, as illustrated in **Figure 2**a, we experimentally fabricated the device containing one nanopatterned GST film that can be electrically driven by an indium tin oxide (ITO)-based heater through gold electrodes. The operation wavelength was chosen in the near-infrared range because of the relatively low loss of GST and ITO. The glass substrate and the ITO heater are both transparent for transmission operation. Figure 2b displays the flowchart of the nanofabrication processes, including multiple steps of photolithography, deposition, liftoff, and annealing. More details can be found in Section 4. Figure 2c shows a photo of a manufactured device. Current demonstrations of PCM-based free-space optoelectronic devices for spatial light modulation cannot be employed or have limitations for transmissive DONN systems. For example, a reconfigurable visible display has been demonstrated by tuning individual pixels through electrical nanotips while the display operates in the reflection mode and the electrical excitation approach is not scalable as electrical addressing using wires and electrodes;[36] reconfigurable metasurfaces have been demonstrated through optical[35] and electrical[38] excitation while they still operate in the reflection mode. Very recently, a transmissive reconfigurable metasurface has been demonstrated through direct electrical excitation,[40] but the requirement of defining lateral structures with fine resolutions in the metasurface using sophisticated electron beam lithography leads to challenging scalable manufacturing. In contrast, our heterostack architecture has a relatively simple lateral structure and the fabrication process does not require sophisticated electron beam lithography. Moreover, the heterostack is scalable in the vertical or thickness dimension through large-scale thin-film deposition. Hence, our heterostack can be practically feasible for large-scale arrays and be utilized for DONN systems with reasonably good performance as shown later.

Figure 3a illustrates the optoelectronic characterization setup to characterize the manufactured devices. Specifically, we focused a 1550 nm laser onto the manufactured chip using a pair of lenses, and the focused spot size was ≈90 × 90 μm; see Figure S2a, Supporting Information, for a photo of the characterization setup. Because of the relatively large laser spot size and small GST area (≈30 × 30 μm), we created a metallic titanium hole (≈10 × 10 μm) to block all other light so that the transmitted light passing through the device is from the GST film; see Figure 2b,c

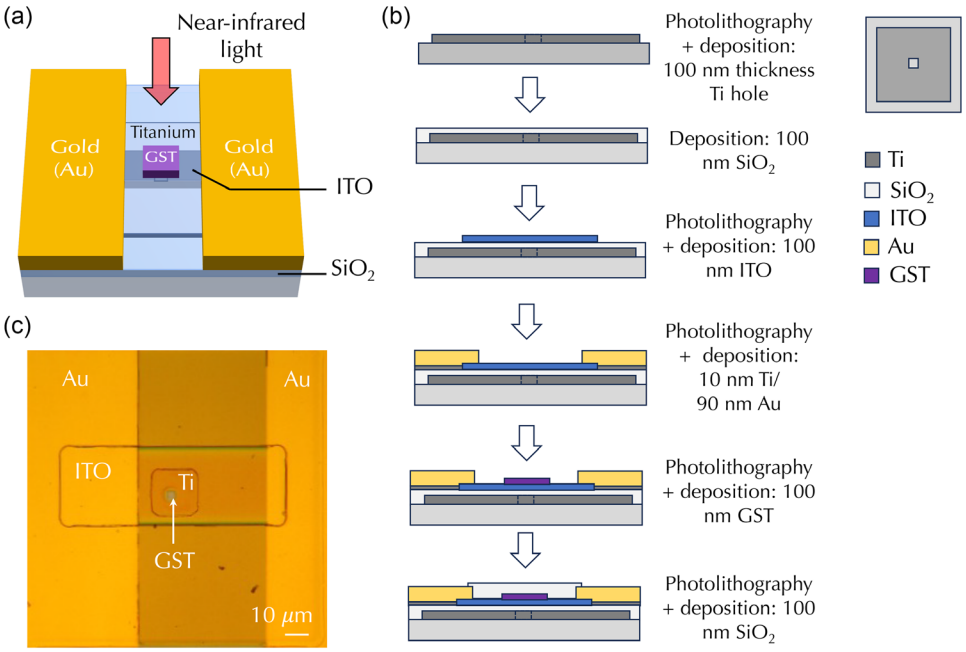

Figure 2. a) Illustration of the spatial light modulation device demonstrating the electrical tuning of one GST film using an ITO heater. b) Flow chart of nanofabrication process including multiple steps of photolithography, deposition, liftoff, and annealing. c) Photo of a manufactured device.

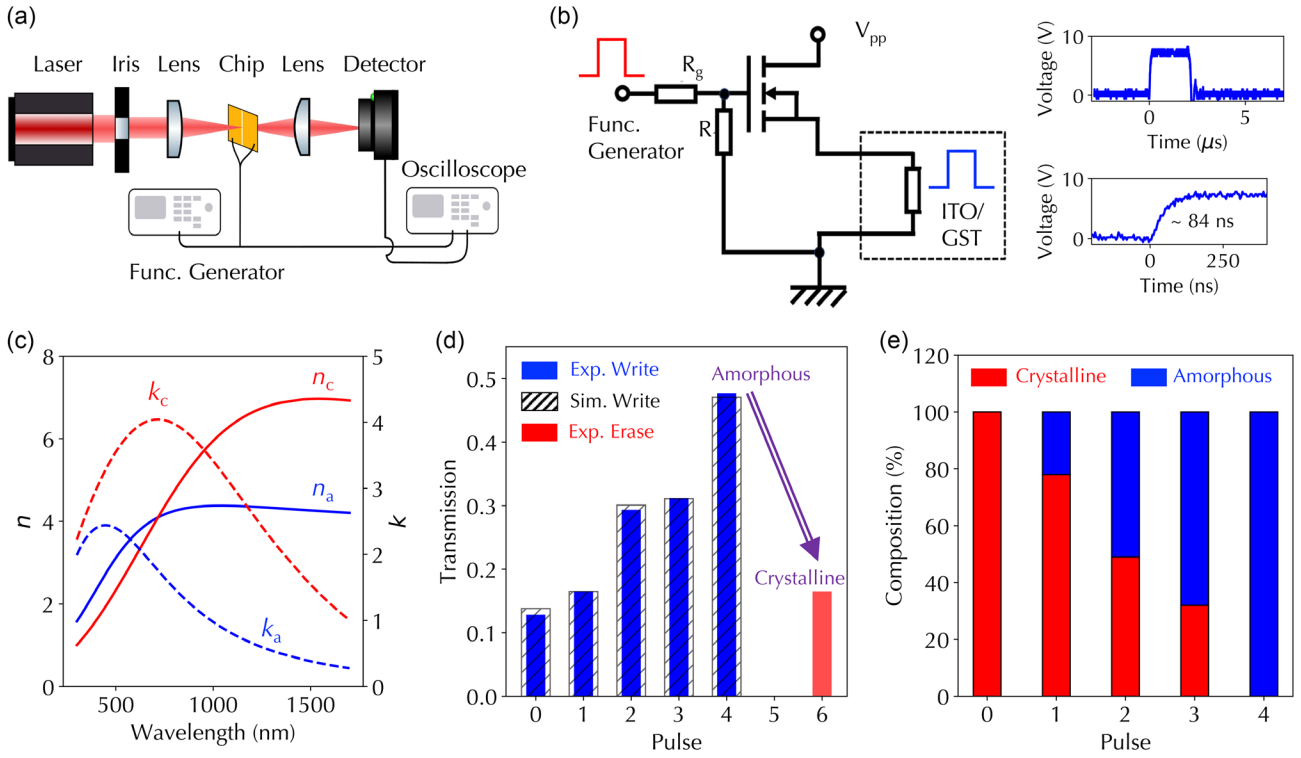

Figure 3. a) Illustration of the laser characterization setup consisting of a near-infrared laser, an iris, a pair of lenses to focus the beam on the chip, and a fast detector connected to an oscilloscope. b) Diagram of the electrical driving circuit. The base of the power transistor is connected to a functional generator supplying square waves with various amplitude and pulse widths to switch on and off. The source of the transistor is connected to a constant voltage source to provide driving voltage pulses to GST devices. c) Experimentally measured refractive index (*n*, solid lines) and extinction coefficient (*k*, dashed lines) of the GST film under crystalline (red lines) and amorphous (blue lines) phases using ellipsometry. d) Experimentally measured (solid rectangles) and simulated transmittance using transfer matrix (black-striped rectangles) under multiple writing (blue color) and erasing pulses (red color). e) Composition ratios of crystalline (red rectangles) and amorphous (blue rectangles) phases under different pulses extracted from transfer matrix simulations.

as well. A function generator was used to drive a circuit to apply electrical pulses to the ITO heater and the oscilloscope was used to measure the time-dependent response from the driving circuit and a fast InGaAs detector. Figure 3b displays a diagram of the driving electrical circuit; see Figure S2b, Supporting Information, for a photo of the driving electrical circuit. A fixed square voltage wave from the function generator was the input to the gate of a power transistor, whose source was connected to a DC voltage source with a controllable voltage V_{pp}. More details can be found in Section 4. Hence, a voltage pulse was applied to the ITO heater to tune the phase of the GST film (Figure 3b). The obtained pulse width on the ITO heater was determined by the input square wave and the circuit response, and the pulse height was determined by V_{pp}. By tuning V_{pp}, we changed the pulse height. Figure 3b also displays time-domain waveforms of one representative applied voltage pulse. The pulse width, rising time, and peak voltage were ≈2 µs, ≈84 ns, and ≈7.5 V, respectively.

The as-deposited GST film was in the amorphous phase and then heated up on a hot plate to be converted to the fully crystalline phase. We measured the refractive index (*n*) and extinction coefficient (*k*) in a wavelength range of 300–1700 nm using ellipsometry; see Figure 3c and more details in Section 4. Although there are other low-loss PCMs such as $Ge_2Sb_2Se_4Te$[38,40] and Sb_2Se_3,[45] the GST material is the front runner for large-scale devices, which will be discussed later in detail. We set V_{pp} at 19 V and applied multiple pulses to drive devices and switch the GST film from the crystalline phase to the amorphous phase. Blue bars in Figure 3d display the measured transmittance, which is defined as the ratio of the transmitted power passing through the GST film over the incident power passing through other layers except the GST film, after several writing pulses. "0" pulse means the initial state, which is the crystalline phase. Note that the measured transmittance was calibrated based on the laser spot and Ti pinhole sizes. We observed a 5-level GST operation. Further, we applied two additional pulses as erasing pulses to switch the amorphous phase back to the crystalline phase. The repeatability of writing and erasing cycles is a few times and Figure S3, Supporting Information, displays a photo of a burned ITO heater after multiple cycles. This poor repeatability could be due to the poor quality of the fabricated ITO heater and its interface with the electrodes. However, the ITO heater has been demonstrated to sustain >100 cycles.[46] Moreover, nanomaterials have become emerging candidates as electrodes for controlling PCMs. For example, graphene has been demonstrated as an energy efficient and robust electrical heater in a PCM-based reconfigurable integrated photonic circuit with repeatability of >1000 cycles. Further, graphene can be

manufactured on a large scale[47] and possesses broadband optical transparency,[48] making it promising for reconfigurable transmissive diffractive devices.

Moreover, with experimentally obtained n and k of the GST film, available optical constants of other materials,[49] and experimental film thicknesses in the stack, we employed a standard 2×2 transfer matrix method[50] to simulate the transmittance of the stack at 1550 nm wavelength. For intermediate states, we assumed a linear combination of crystalline and amorphous phases and corresponding n and k. Specifically, the composition ratio of the crystalline phase was denoted as α_c and the amorphous ratio α_a is $1 - \alpha_c$. Hence, n and k for intermediate states are $\alpha_c n_c + \alpha_a n_a$ and $\alpha_c k_c + \alpha_a k_a$, respectively, where n_c (n_a) and k_c (k_a) are n and k for the crystalline (amorphous) phase shown in Figure 3c. α_c was a fitting parameter during simulations. Black-striped bars in Figure 3d show simulation results under writing pulses, agreeing well with experimental results. Figure 3e displays the composition ratio of the amorphous and crystalline phases at intermediate states and we observed a gradual and nearly linear phase change under writing pulses.

In the digitized heterostack shown in Figure 1b, the SiO_2 spacing layers between GST layers are necessary to minimize the thermal cross talk that can induce the phase transition from amorphous to crystalline phases and have independent heater control for each GST layer. In addition, when heterostack pixels are grouped into a 2D diffractive array (Figure 1a) for a DONN system, the lateral spacing between pixels is also needed for minimal cross talk between pixels. To analyze the temperature profiles of the GST film under experimental electrical pulses, we performed COMSOL Multiphysics simulations. **Figure 4**a displays the diagram and dimensions of the COMSOL simulations. The device was at the center of the simulation region, which was filled with SiO_2. The top panel of Figure 4b shows a schematic of the simulated device consisting of one GST layer and one ITO heater. The size was set as 30×30 μm. More details can be found in Section 4. The bottom panels of Figure 4c display the cross section and top view of temperature distribution in the device under experimental writing pulses. Figure 4c shows the temperature profile across the green line in the cross section view in Figure 4b and the highest temperature achieved in the GST layer is ≈780 °C. This temperature is sufficient to induce the phase transition from crystalline to amorphous phases, typically requiring a melting temperature of >600 °C. The one-side temperature decay length from the highest temperature to a temperature close to ≈100 °C is ≈4 μm and the transition temperature from amorphous to crystalline phases is >150 °C, meaning that within the heterostack, the thickness of SiO_2 spacing layers needs to ≥4 μm to minimize thermal cross talk. Similarly, Figure 4d shows the temperature profile across the green line in the top view of Figure 4b. At a distance of 5 μm away from the device boundary (purple dot and arrow in Figure 4d), the temperature becomes close to room temperature (≈30 °C). Hence, in lateral dimensions, the pixel spacing ≥4 μm can avoid thermal cross talk. Figure 4e shows the time-dependent response of

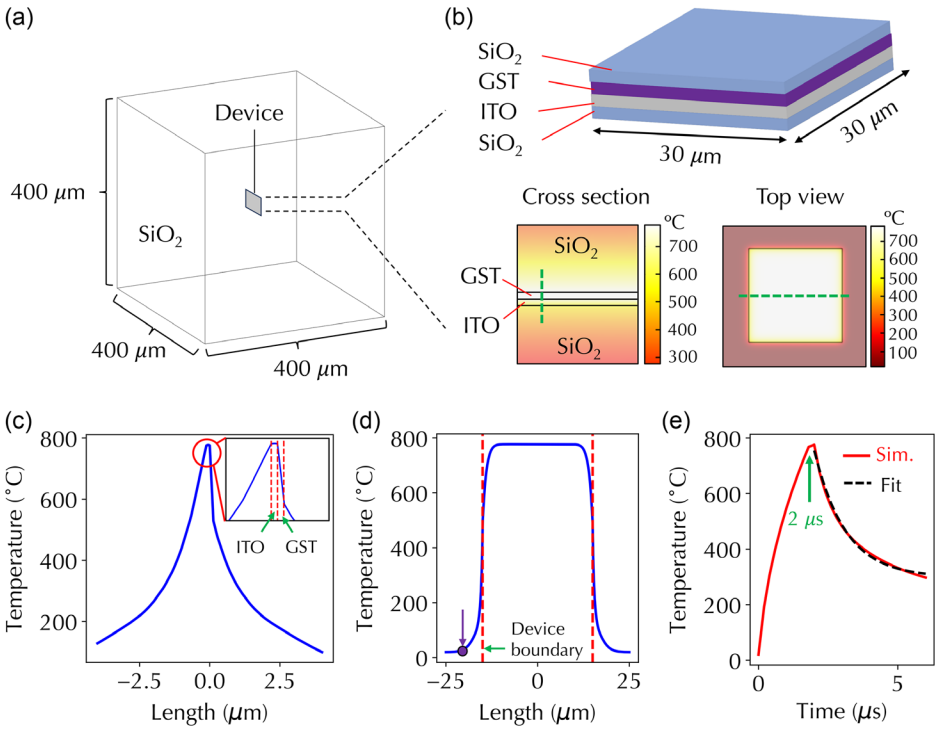

Figure 4. a) Schematic of COMSOL simulation setup. The device is located at the center of a $400 \times 400 \times 400$ μm cube filled with SiO_2. b) Illustration of a heterostack with one GST film, and the cross section and top view of temperature distribution under experimental pulses. c) Temperature profile across the green line in the cross section of temperature distribution. The inset is a zoomed-in temperature profile near the peak temperature. d) Temperature profile across the green in the top view of temperature distribution. Purple dot and arrow indicate a location 5 μm away from the device boundary. e) Simulated (red line) time-dependent response of the highest temperature in the GST film and an exponential fitting curve (blue line) for the decay process.

the highest temperature observed in Figure 4c,d. The peak temperature occurs at the end of pulse 2 μs. After the excitation, the temperature drops due to heat dissipation in the surrounding environment. We fit the temperature decay using an exponential function and obtained a time constant ≈1.12 μs.

With experimentally measured transmission properties of one GST film, we further designed thicknesses of SiO_2, ITO, and GST films in a heterostack containing three layers of GST films using the transfer matrix method, as illustrated in **Figure 5**a. As mentioned before, the heterostack can be scaled up in the thickness dimension by repeating the large-scale thin-film deposition process. All thicknesses $\{d_1, d_2, \ldots, d_9\}$ were designed to have large and uniform phase responses. Each GST film can perform either under a 2-level operation with only crystalline and amorphous phases or under a 5-level operation with intermediate states as the experimentally obtained composition ratios shown in Figure 3e. The thickness of SiO_2 capping layer (d_1) was fixed as 100 nm and the thicknesses of SiO_2 spacing layer (d_4 and d_7) were required to be $\geq$4 μm to avoid thermal cross talks based on the COMSOL simulations in Figure 4. The optimized ITO thickness was 122 nm and the thicknesses of three GST layers (d_2, d_5, and d_8) were 10, 60, and 80 nm, respectively. Figure 5b,c displays the phase and amplitude response of transmitted light under different combinations of levels when the GST film was under the 2-level operation. There are in total $8 = 2^3$ different combinations for 3 GST films. The phase modulation range is $\approx 0.5\pi$. Further, Figure 5d,e displays the phase and amplitude response of transmitted light under different combinations of levels when the GST film was under the 5-level operation. In total, there are $125 = 5^3$ combinations for three GST films and hence there are more intermediate phase and amplitude response points with 5-level-operation GST films compared to 2-level-operation GST films.

Finally, we evaluated the performance of DONN systems for classifying handwritten digits from the MNIST dataset, if the individual heterostack diffractive device designed in Figure 5 can be assembled into a 2D diffractive array. Although not experimentally demonstrated here, such an integrated 2D array is practically feasible because of the simple lateral structure and scalable vertical dimension in the heterostack, as discussed before. Further, the mature scalable thin-film transistor techniques used in modern displays with the individual control of $> 1000 \times 1000$ pixels can be utilized to replace discrete electronic components in driving circuits shown in Figure S2b, Supporting Information, for large-scale integration. For example, recent work has demonstrated a spatial light modulator consisting of 128×128 array of driving transistors and simple GST PCM thin-film devices with 5×5 μm square patterns. Each pixel can be addressed electrically and individually.[46] Although the demonstrated modulator operates in a reflective manner, the scalable manufacturing of devices confirms that our heterostack is also practically feasible and scalable. In addition, the industry has developed mature scalable PCM-based memory technologies.[41] In particular, the architecture of developed commercial 3D XPoint memory technology consists of multiple stacked PCM layers, which is similar to our proposed device architecture and further confirms the feasibility and scalability of our heterostack architecture. To train diffractive arrays, meaning finding the best combination of levels for all heterostack diffractive pixels, through the backpropagation algorithm in ML software frameworks, we need to incorporate the irregular amplitude and phase device responses into DONN calculations in a differentiable manner. However, the device responses cannot be fitted with analytical equations and we utilized a Gumbel–Softmax reparameterization approach to approximate discrete distributions with a differentiable continuous function.[20] Each pixel size was set as

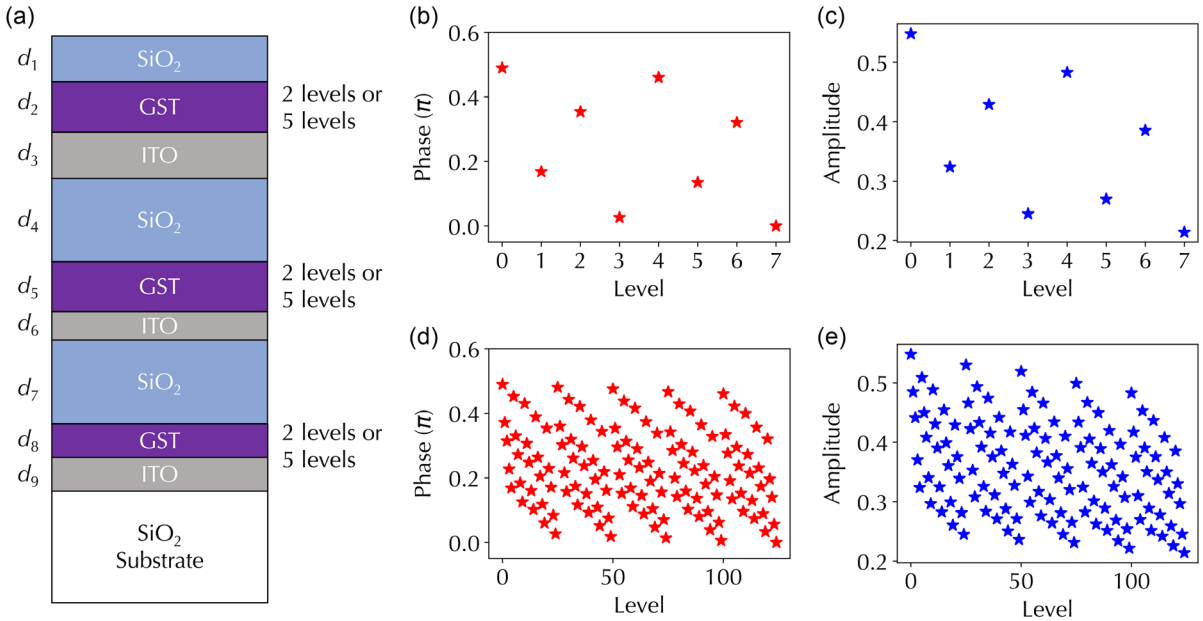

Figure 5. a) Diagram of a digitized GST heterostack containing 3 GST layers with each layer having different thicknesses and operating under 2-level or 5-level operations. b) Phase and c) amplitude response of transmitted light under different combinations of levels when the GST film can achieve 2 states. d) Phase and e) amplitude response of transmitted light under different combinations of levels when the GST film can achieve 5 states.

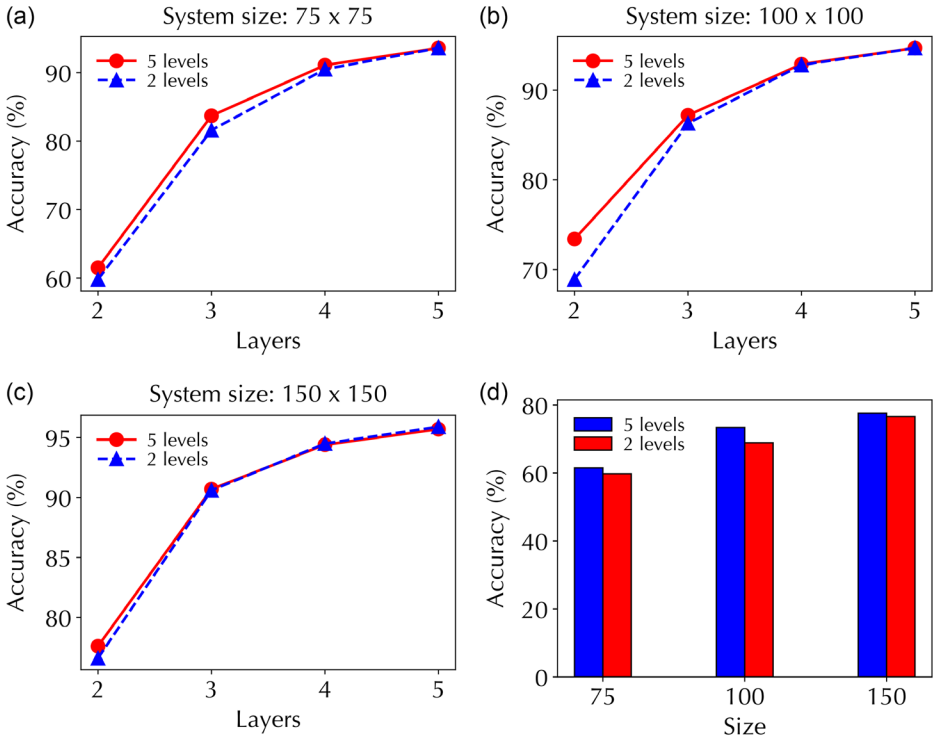

Figure 6. a–c) MNIST classification accuracies of DONNs of different system sizes and layers with digitalized GST heterostack operating under 2 or 5 levels for each GST film. d) Comparison of different system sizes for 2-level and 5-level GST films in 2-layer DONN systems.

30×30 μm to be consistent with experiments in Figure 2 and a 5 μm spacing between pixels was set to avoid thermal cross talk. Hence, the pixel filling factor was 56.25%. We evaluated DONN systems with different array sizes, the number of layers, and heterostacks containing 2-level and 5-level GST films, as summarized in **Figure 6**. As expected, classification accuracies increase with increasing array size and layer number (Figure 6a–c). Even though the phase modulation range of the designed heterostack is $\approx 0.5\pi$, the MNIST classification accuracy is reasonably good when the number of diffractive layers is 5. Our previous work[51] demonstrated that a larger phase modulation range of 2π can achieve high classification accuracy with a smaller number of diffractive layers, and a smaller phase modulation range of 0.1π leads to low accuracies even if the number of diffractive layers increases to 6. However, achieving a 2π phase modulation range is challenging for transmissive diffractive devices. Hence, our designed heterostack with a 0.5π range demonstrates a good balance considering the practical feasibility of device implementation and system performance. Further, for 2-layer DONN systems, the heterostack with 5-level GST films improves the DONN system classification performance compared to 2-level GST films, as shown in Figure 6d. Their difference in 5-layer DONN systems becomes negligible. Hence, for deep DONNs, transmissive diffractive devices under 2-level operations with only crystalline and amorphous phases of GST films are sufficient for good system performance with high classification accuracies and robust operations. Further, we evaluated the performance of DONN systems for classifying the Fashion MNIST dataset. Figure S4a–c, Supporting Information, displays classification accuracies for DONN systems with different array sizes, the number of layers, and heterostack containing 2-level GST films. For the 5-layer DONN system with 150×150 size, transmissive diffractive devices under 2-level operations with only crystalline and amorphous phases of GST films can already provide good classification accuracy of ≈82%, which is comparable to previously demonstrated DONN systems.[12] The corresponding confusion matrix is shown in Figure S4d, Supporting Information.

3. Conclusion and Discussion

We demonstrated a feasible design of a novel diffractive device for DONN systems based on a PCM heterostack, which can leverage the advantages of PCM optical properties and mitigate challenges associated with implementing multilevel operations. We performed not only proof-of-concept experiments but also detailed analyses on both device and system levels. The cyclability of current devices is not good because of the poor quality of ITO heaters. More thermally stable, efficient, robust, and transparent heaters, such as carbon nanomaterials including graphene[34] and aligned carbon nanotube films,[52] can substantially improve the efficiency and repeatability of heaters. Further, in addition to GST materials, the utilized PCM in heterostacks can be extended to other materials with lower loss for broadband operations, such as $Ge_2Sb_2Se_4Te$[38,40] and Sb_2Se_3.[45]

4. Experimental Section

Device Fabrication: The overall nanofabrication flow is illustrated in Figure 2b. Specifically, a 10 × 10 μm hole was defined by photolithography using a mask aligner Suss MA1006 and a 100 nm thick Ti film was deposited using a Denton Discovery 18 sputtering system. The Ti hole structure was then formed after a liftoff process. A 100 nm thick SiO_2 film was then deposited on top using a Denton 635 sputtering system. A 150 × 50 μm area was defined by photolithography and a 100 nm thick ITO film was deposited using the Denton Discovery 18 system to form a heater after liftoff. ITO was annealed at 300 °C in a furnace (Lindberg/Blue M Moldatherm 1100 °C box furnace) under N_2 atmosphere for 1 h. The ITO heater was connected by two electrodes, which contained 10 nm thick Ti and 90 nm Au defined by photolithography and deposited by the Denton 635 system. Afterward, a 30 × 30 μm area was defined by photolithography and a 100 nm thick GST film was deposited using a Denton Discovery 18 system at an argon pressure of 4.5 mTorr and a power setting of 35 W.

Optoelectronic Characterization Setup: A schematic diagram of the experimental setup is illustrated in Figure 3a. The employed laser diode was LQC1550-05E from Newport Corporation with a center wavelength of 1550 nm. The elliptical output beam from the laser diode was reshaped into a round beam through lenses and an iris. Manufactured chips were driven by an electrical circuit connected to a function generator (Tektronix AFG2020). In the driving circuit, the model of the employed power transistor was Infineon IRLZ34N and the model of the employed DC source was Tektronix PS281. The output light power was measured by an InGaAs detector with a 5 GHz speed bandwidth and an operation wavelength range of 800–1700 nm (DET08C from Thorlabs, Inc.). The time responses of the driving electrical pulses and the detector reading were recorded by an oscilloscope with a bandwidth of 300 MHz (Tektronix TDS3034).

Ellipsometry Measurement: The film thickness and optical refractive index n and extinction coefficient k of GST films were measured using a J. A. Woollam variable angle spectroscopic ellipsometer over a wavelength range of 300–1700 nm. In the measurement, polarized light was reflected off the sample surface, and the change in polarization was measured as two quantities: amplitude ratio Ψ and phase difference Δ. By fitting measurements with a model describing materials and sample structures, the optimal film thickness and optical indices were obtained to minimize the error between experiments and calculations.

COMSOL Simulations: A 3D finite-element simulation using COMSOL Multiphysics was conducted to analyze the temperature distribution in Figure 4. The ITO thickness was 120 nm and the GST thickness was 130 nm. The thermal conductivity of ITO was set as 11 $W\,m^{-1}K^{-1}$ and the GST thermal conductivity was set as 0.27 $W\,m^{-1}K^{-1}$.[53] The heat source was a square wave with a duration of 2 μs and was calculated based on a 19 V pulse height, 30 Ω ITO resistance, and 150 × 150 μm × 120 nm ITO volume.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

R.C., J.F., C.Y., and W.G. acknowledge the support from the National Science Foundation through grant nos. 2230727, 2316627, and 2428520.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ruiyang Chen: conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); software (lead); validation (lead); visualization (lead); writing—original draft (equal); and writing—review and editing (equal). **Jichao Fan**: investigation (supporting) and methodology (supporting). **Cunxi Yu**: software (supporting). **Weilu Gao**: supervision (lead); writing—original draft (lead); and writing—review and editing (lead).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

diffractive optical neural networks, heterostacks, phase-change materials

Received: December 3, 2024
Published online:

[1] Y. Lecun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[2] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, in *Deep Learning*, Vol. *1*, MIT Press, Cambridge **2016**.
[3] S. P. Rodrigues, Z. Yu, P. Schmalenberg, J. Lee, H. Iizuka, E. M. Dede, *Nat. Photonics* **2021**, *15*, 66.
[4] A. Mirhoseini, A. Goldie, M. Yazgan, J. W. Jiang, E. Songhori, S. Wang, Y.-J. Lee, E. Johnson, O. Pathak, A. Nova, J. Pak, A. Tong, K. Srinivasa, W. Hang, E. Tuncer, Q. V. Le, J. Laudon, R. Ho, R. Carpenter, J. Dean, *Nature* **2021**, *594*, 207.
[5] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
[6] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, *Nature* **2020**, *577*, 706.
[7] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, D. Psaltis, *Nature* **2020**, *588*, 39.
[8] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, *Nat. Photonics* **2017**, *11*, 441.
[9] C. Ríos, N. Youngblood, Z. Cheng, M. Le Gallo, W. H. Pernice, C. D. Wright, A. Sebastian, H. Bhaskaran, *Sci. Adv.* **2019**, *5*, eaau5759.
[10] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, H. Bhaskaran, *Nature* **2021**, *589*, 52.
[11] H. Feng, T. Ge, X. Guo, B. Wang, Y. Zhang, Z. Chen, S. Zhu, K. Zhang, W. Sun, C. Huang, Y. Yuan, C. Wang, *Nature* **2024**, *627*, 80.
[12] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, *Science* **2018**, *361*, 1004.
[13] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, D. Englund, *Phys. Rev. X* **2019**, *9*, 021032.
[14] L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, *Nature* **2020**, *579*, 62.
[15] J. Spall, X. Guo, T. D. Barrett, A. I. Lvovsky, *Opt. Lett.* **2020**, *45*, 5752.
[16] M. Miscuglio, Z. Hu, S. Li, J. K. George, R. Capanna, H. Dalir, P. M. Bardet, P. Gupta, V. J. Sorger, *Optica* **2020**, *7*, 1812.
[17] T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, Q. Dai, *Nat. Photonics* **2021**, *15*, 367.

[18] W. Gao, C. Yu, R. Chen, *Adv. Photonics Res.* **2021**, 2100048.
[19] T. Wang, S.-Y. Ma, L. G. Wright, T. Onodera, B. C. Richard, P. L. McMahon, *Nat. Commun.* **2022**, *13*, 123.
[20] R. Chen, Y. Li, M. Lou, J. Fan, Y. Tang, B. Sensale-Rodriguez, C. Yu, W. Gao, *Laser Photonics Rev.* **2022**, 2200348.
[21] R. Chen, Y. Tang, J. Ma, W. Gao, *Adv. Intell. Syst.* **2023**, *5*, 2300536.
[22] J. Fan, Y. Tang, W. Gao, *Adv. Intell. Syst.* **2023**, *5*, 2300147.
[23] J. Hu, D. Mengu, D. C. Tzarouchis, B. Edwards, N. Engheta, A. Ozcan, *Nat. Commun.* **2024**, *15*, 1525.
[24] K. Matsushima, H. Schimmel, F. Wyrowski, *JOSA A* **2003**, *20*, 1755.
[25] M. Wuttig, H. Bhaskaran, T. Taubner, *Nat. Photonics* **2017**, *11*, 465.
[26] S. Abdollahramezani, O. Hemmatyar, H. Taghinejad, A. Krasnok, Y. Kiarashinejad, M. Zandehshahvar, A. Alù, A. Adibi, *Nanophotonics* **2020**, *9*, 1189.
[27] Y. Zhang, C. Ríos, M. Y. Shalaginov, M. Li, A. Majumdar, T. Gu, J. Hu, *Appl. Phys. Lett.* **2021**, *118*, 210501.
[28] M. Wuttig, N. Yamada, *Nat. Mater.* **2007**, *6*, 824.
[29] H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, K. E. Goodson, *Proc. IEEE* **2010**, *98*, 2201.
[30] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, E. Eleftheriou, *Nat. Nanotechnol.* **2020**, *15*, 529.
[31] C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, *Nat. Photonics* **2015**, *9*, 725.
[32] Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, H. Bhaskaran, *Sci. Adv.* **2017**, *3*, e1700160.
[33] Y. Zhang, J. B. Chou, J. Li, H. Li, Q. Du, A. Yadav, S. Zhou, M. Y. Shalaginov, Z. Fang, H. Zhong, *Nat. Commun.* **2019**, *10*, 4279.
[34] Z. Fang, R. Chen, J. Zheng, A. I. Khan, K. M. Neilson, S. J. Geiger, D. M. Callahan, M. G. Moebius, A. Saxena, M. E. Chen, C. Rios, J. Hu, E. Pop, A. Majumdar, *Nat. Nanotechnol.* **2022**, *17*, 842.
[35] Q. Wang, E. T. F. Rogers, B. Gholipour, C.-M. Wang, G. Yuan, J. Teng, N. I. Zheludev, *Nat. Photonics* **2016**, *10*, 60.
[36] C. Ríos, P. Hosseini, R. A. Taylor, H. Bhaskaran, *Adv. Mater.* **2016**, *28*, 4720.
[37] Y. Wang, P. Landreman, D. Schoen, K. Okabe, A. Marshall, U. Celano, H.-S. P. Wong, J. Park, M. L. Brongersma, *Nat. Nanotechnol.* **2021**, *16*, 667.
[38] Y. Zhang, C. Fowler, J. Liang, B. Azhar, M. Y. Shalaginov, S. Deckoff-Jones, S. An, J. B. Chou, C. M. Roberts, V. Liberman, *Nat. Nanotechnol.* **2021**, *16*, 661.
[39] S. Abdollahramezani, O. Hemmatyar, M. Taghinejad, H. Taghinejad, A. Krasnok, A. A. Eftekhar, C. Teichrib, S. Deshmukh, M. A. El-Sayed, E. Pop, M. Wuttig, A. Alù, W. Cai, A. Adibi, *Nat. Commun.* **2022**, *13*, 1.
[40] C. C. Popescu, K. Aryana, P. Garud, K. P. Dao, S. Vitale, V. Liberman, H.-B. Bae, T.-W. Lee, M. Kang, K. A. Richardson, *Adv. Mater.* **2024**, 2400627.
[41] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y.-C. Chen, R. M. Shelby, M. Salinga, D. Krebs, S.-H. Chen, H.-L. Lung, C. H. Lam, *IBM J. Res. Dev.* **2008**, *52*, 465.
[42] H. Zhang, L. Zhou, L. Lu, J. Xu, N. Wang, H. Hu, B. M. A. Rahman, Z. Zhou, J. Chen, *ACS Photonics* **2019**, *6*, 2205.
[43] X. Li, N. Youngblood, C. Ríos, Z. Cheng, C. D. Wright, W. H. P. Pernice, H. Bhaskaran, *Optica* **2019**, *6*, 1.
[44] Q.-J. Hong, J. Schroers, D. Hofmann, S. Curtarolo, M. Asta, A. Van De Walle, *NPJ Comput. Mater.* **2021**, *7*, 1.
[45] M. Delaney, I. Zeimpekis, D. Lawson, D. W. Hewak, O. L. Muskens, *Adv. Funct. Mater.* **2020**, *30*, 2002447.
[46] Y.-H. Kim, S. M. Cho, K. Choi, C. Y. Hwang, G. H. Kim, S. Cheon, C.-S. Hwang, *JOSA A* **2019**, *36*, D23.
[47] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. Banerjee, L. Colombo, R. Ruoff, *Science* **2009**, *324*, 1312.
[48] R. R. Nair, P. Blake, A. N. Grigorenko, K. S. Novoselov, T. J. Booth, T. Stauber, N. M. R. Peres, A. K. Geim, *Science* **2008**, *320*, 1308.
[49] Y. Tang, P. T. Zamani, R. Chen, J. Ma, M. Qi, C. Yu, W. Gao, *Laser Photonics Rev.* **2022**, 2200381.
[50] P. Yeh, in *Optical Waves in Layered Media*, 2nd ed., Wiley-Interscience, New York **2005**.
[51] H. Zeng, J. Fan, Y. Zhang, Y. Su, C. Qiu, W. Gao, *Opt. Express* **2022**, *30*, 12712.
[52] X. He, W. Gao, L. Xie, B. Li, Q. Zhang, S. Lei, J. M. Robinson, E. H. Hároz, S. K. Doorn, W. Wang, R. Vajtai, P. M. Ajayan, W. W. Adams, R. H. R. H. Hauge, J. Kono, *Nat. Nanotechnol.* **2016**, *11*, 633.
[53] H. Taghinejad, S. Abdollahramezani, A. A. Eftekhar, T. Fan, A. H. Hosseinnia, O. Hemmatyar, A. Eshaghian Dorche, A. Gallmon, A. Adibi, *Opt. Express* **2021**, *29*, 20449.