Li et al. *Light: Science & Applications* (2022)11:153
https://doi.org/10.1038/s41377-022-00849-x

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

Polarization multiplexed diffractive computing: all-optical implementation of a group of linear transformations through a polarization-encoded diffractive network

Jingxi Li[1,2,3], Yi-Chun Hung[1], Onur Kulce[1,2,3], Deniz Mengu[1,2,3] and Aydogan Ozcan[1,2,3]✉

Abstract

Research on optical computing has recently attracted significant attention due to the transformative advances in machine learning. Among different approaches, diffractive optical networks composed of spatially-engineered transmissive surfaces have been demonstrated for all-optical statistical inference and performing arbitrary linear transformations using passive, free-space optical layers. Here, we introduce a polarization-multiplexed diffractive processor to all-optically perform multiple, arbitrarily-selected linear transformations through a single diffractive network trained using deep learning. In this framework, an array of pre-selected linear polarizers is positioned between trainable transmissive diffractive materials that are isotropic, and different target linear transformations (complex-valued) are uniquely assigned to different combinations of input/output polarization states. The transmission layers of this polarization-multiplexed diffractive network are trained and optimized via deep learning and error-backpropagation by using thousands of examples of the input/output fields corresponding to each one of the complex-valued linear transformations assigned to different input/output polarization combinations. Our results and analysis reveal that a single diffractive network can successfully approximate and all-optically implement a group of arbitrarily-selected target transformations with a negligible error when the number of trainable diffractive features/neurons (N) approaches $N_pN_iN_o$, where N_i and N_o represent the number of pixels at the input and output fields-of-view, respectively, and N_p refers to the number of unique linear transformations assigned to different input/output polarization combinations. This polarization-multiplexed all-optical diffractive processor can find various applications in optical computing and polarization-based machine vision tasks.

Introduction

With the increasing global demand for machine learning and computing in general, using light to perform computation has been a rapidly growing focus area of optics and photonics[1–5]. The research on optical computing has a long history spanning decades of exciting research and development efforts[6–31]. Motivated by the massive success of artificial intelligence and deep learning, in specific, a myriad of new hardware designs for optical computing have been reported recently, including, e.g., on-chip integrated photonic circuits[16–22], free-space optical platforms[23–28], and others[29–31]. Among these different optical computing systems, the integration of successive transmissive diffractive layers (forming an optical network) has been demonstrated for optical information processing, such as object classification[23,32–43], image reconstruction[38,44], all-optical phase recovery and quantitative phase imaging[45], and logic operations[46–48].

Correspondence: Aydogan Ozcan (ozcan@ucla.edu)
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA
[2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA
Full list of author information is available at the end of the article

© The Author(s) 2022
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

A diffractive network is trained using deep learning and error-backpropagation methods implemented in a digital computer, after which the resulting transmissive layers are fabricated to form a physical network that computes based on the diffraction of the input light through these spatially-engineered transmissive layers. Because the computational task is completed as the light passes through thin and passive optical elements, this approach is very fast, and the inference process does not consume power except for the illumination light. It is also scalable since an increase in the input field-of-view (FOV) can be handled by fabricating larger transmissive layers and/or deeper diffractive designs with more successive layers positioned one after another. Furthermore, both the phase and the amplitude information channels of the input scene/FOV can be processed by a diffractive optical network, without the need for phase retrieval or digitizing, vectorizing an image of the scene, which makes diffractive computing highly desirable for machine vision applications[38,44]. Harnessing light-matter interactions using engineered diffractive surfaces also enabled the inverse design of optical elements for e.g., spatially-controlled wavelength demultiplexing[49], pulse engineering[50], and orbital angular momentum multiplexing/demultiplexing[51,52]. It has also been shown that a diffractive network can be trained by optimizing its diffractive layers to perform an arbitrary complex-valued linear transformation between its input and output fields-of-view, demonstrating its computing capability for complex-valued matrix-vector operations at the speed of light propagation through a passive diffractive system.

All these results highlight the unique capabilities of diffractive networks to manipulate various physical properties of light, including e.g., its amplitude and phase distribution, spatial frequency, spectral bandwidth, orbital angular momentum, for performing specific computational tasks that are desired. As another important physical property of light, polarization specifies the geometrical orientation of electromagnetic wave oscillations. Utilizing the polarization state of light has played a pivotal role in numerous applications, including telecommunications[53–55], imaging[56–61], sensing[62–64], computing[65], and displays[66,67]. For example, polarization-division multiplexing (PDM) has been used in telecommunication systems to permit two channels of information to be simultaneously transmitted using orthogonal polarization states over a single wavelength[54,68].

Here, we report the design of polarization-multiplexed diffractive optical networks to perform a group of arbitrary linear transformations using a common set of diffractive layers that are jointly optimized to all-optically perform each one of the target complex-valued linear transformations at a different combination of input/output polarization states. In our earlier work[69], we showed that a diffractive optical network composed of spatially-engineered layers could all-optically perform an arbitrary complex-valued linear transformation between an input and output field-of-view with a negligible error when the number of trainable diffractive elements/neurons (N) approaches N_iN_o, where N_i and N_o represent the number of pixels at the input and output FOVs, respectively. In this work, we use polarization multiplexing between the input and output FOVs of a diffractive network to increase the capacity of diffractive computing and all-optically perform a group of arbitrary linear transformations that are complex-valued. These polarization-multiplexed diffractive network designs are *not* based on birefringent, anisotropic or polarization-sensitive materials; instead, our designs utilize standard diffractive surfaces where the phase and amplitude transmission coefficients of each trainable diffractive feature are independent of the polarization state of the input light. Using a network design solely based on standard isotropic diffractive materials makes our designs simpler in terms of material selection, fabrication and scale-up; however, it also makes the diffractive network insensitive to different polarization states, and therefore, polarization-multiplexed all-optical computation of different transformations becomes impossible. To overcome this challenge, we used a non-trainable, pre-determined array of linear polarizers (at 0°, 45°, 90° and 135°) within the diffractive network that acted as polarization seeds for the trainable isotropic diffractive layers to all-optically execute different linear transformations through input/output polarization multiplexing (see Fig. 1a). Stated differently, we used data-driven training and optimization of isotropic diffractive layers to encode different linear transformations into different input/output polarization combinations, and this encoding is made possible by the polarization mode diversity introduced by a non-trainable, pre-determined array of linear polarizers within the diffractive volume.

In our first implementation, we performed two different, arbitrarily selected linear transformations (i.e., $N_p = 2$) using a diffractive network composed of four transmissive layers that are jointly optimized using deep learning, where the first target linear transformation was assigned to x (0°) linear input and x linear output polarization combination, and the second target linear transformation was assigned to y (90°) linear input and y linear output polarization combination. For this case of $N_p = 2$, there are two different schemes (Fig. 1b) to all-optically access/implement the desired linear transformations: sequential (x and y input polarization states encode the input information sequentially, one after another) or simultaneous (x and y input polarizations encode the input information at the same time within the input FOV). Our numerical results

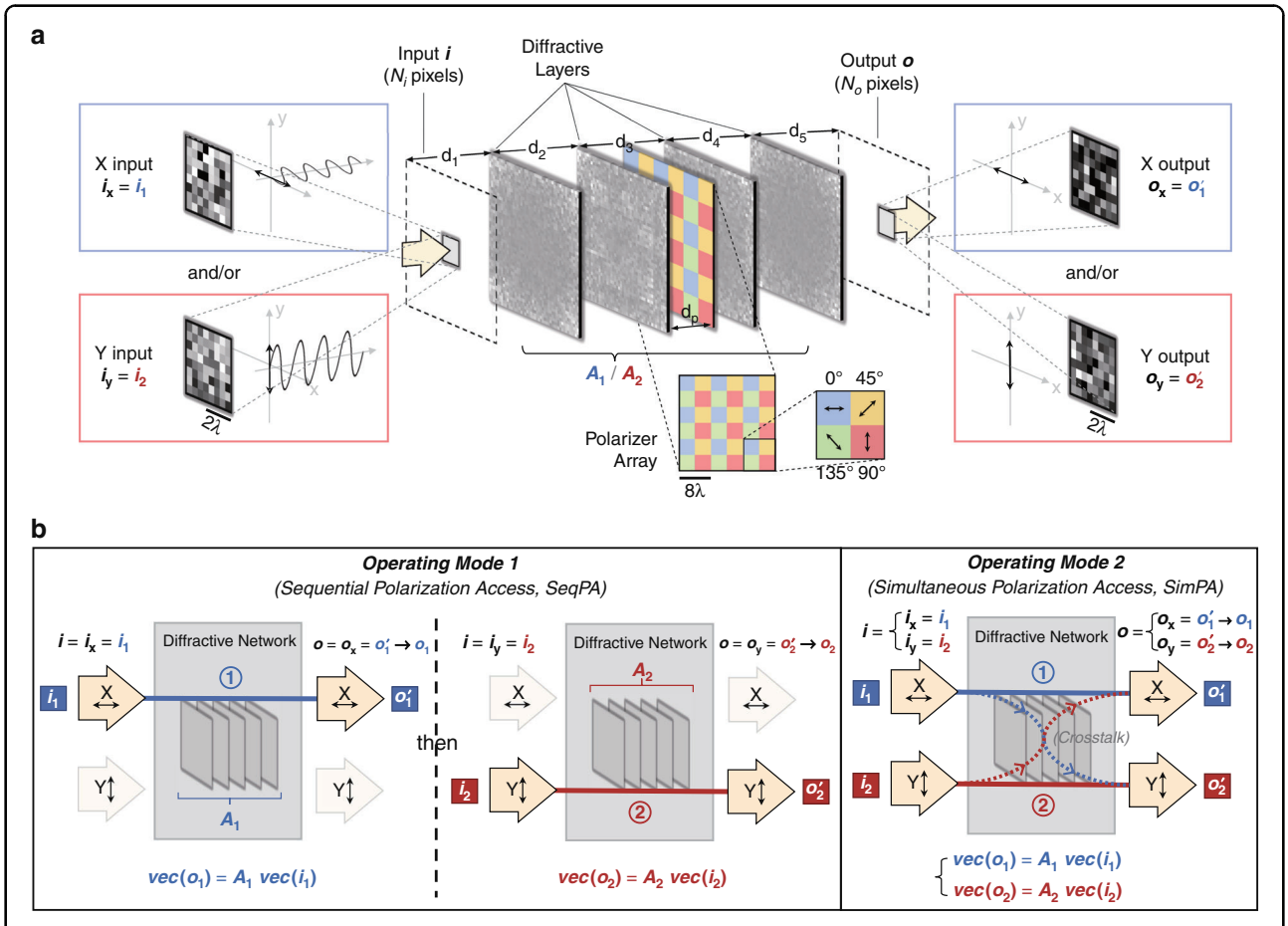

Fig. 1 Schematic of polarization-multiplexed all-optical diffractive computing. a Optical layout of the polarization-encoded diffractive network, where four isotropic diffractive layers and one array of linear polarizers are jointly used to perform two distinct, complex-valued linear transformations between the input field ***i*** and the output field ***o*** by using polarization encoding/decoding at the input/output FOVs. **b** Schematic for the sequential polarization access (SeqPA, left) mode and the simultaneous polarization access (SimPA, right) mode that can be used to operate the 2-channel polarization-multiplexed diffractive network

(Figs. 2–5) reveal that one can successfully train a diffractive network under each one of these operation modes (sequential vs. simultaneous) to approximate the two target, arbitrary-selected linear transformations with a negligible error when the number of trainable diffractive neurons N approaches $N_pN_iN_o = 2N_iN_o$.

In our second implementation (Fig. 6), we performed four different, arbitrary linear transformations (i.e., $N_p = 4$) using a diffractive network composed of eight transmissive layers that are jointly optimized using deep learning and examples of input/output fields corresponding to the selected complex-valued linear transformations (ground truth). In this case, the first target transformation was assigned to x linear input and 45° linear output polarization combination, the second target transformation was assigned to y linear input and 135° linear output polarization combination, the third target transformation was assigned to x linear input and 135° linear output polarization combination and finally the fourth target transformation was assigned to y linear input and 45° linear output polarization combination. Our analyses of this 4-channel polarization-multiplexed diffractive system show that when $N \geq N_pN_iN_o = 4N_iN_o$, all the target linear transformations can be successfully approximated, following a similar conclusion as in the first implementation case ($N_p = 2$).

Without the use of a non-trainable, pre-determined array of linear polarizers acting as polarization seeds within the network, none of these multiplexing results could be achieved using isotropic diffractive materials, no matter how they are trained or optimized, since they would normally perform the same transformation under different input polarization states.

Our results should *not* be confused with polarization-multiplexed (or wavelength/illumination multiplexed) projection of a set of desired complex fields at the output of a metamaterial design; such multiplexed metamaterial systems do not implement an arbitrary matrix

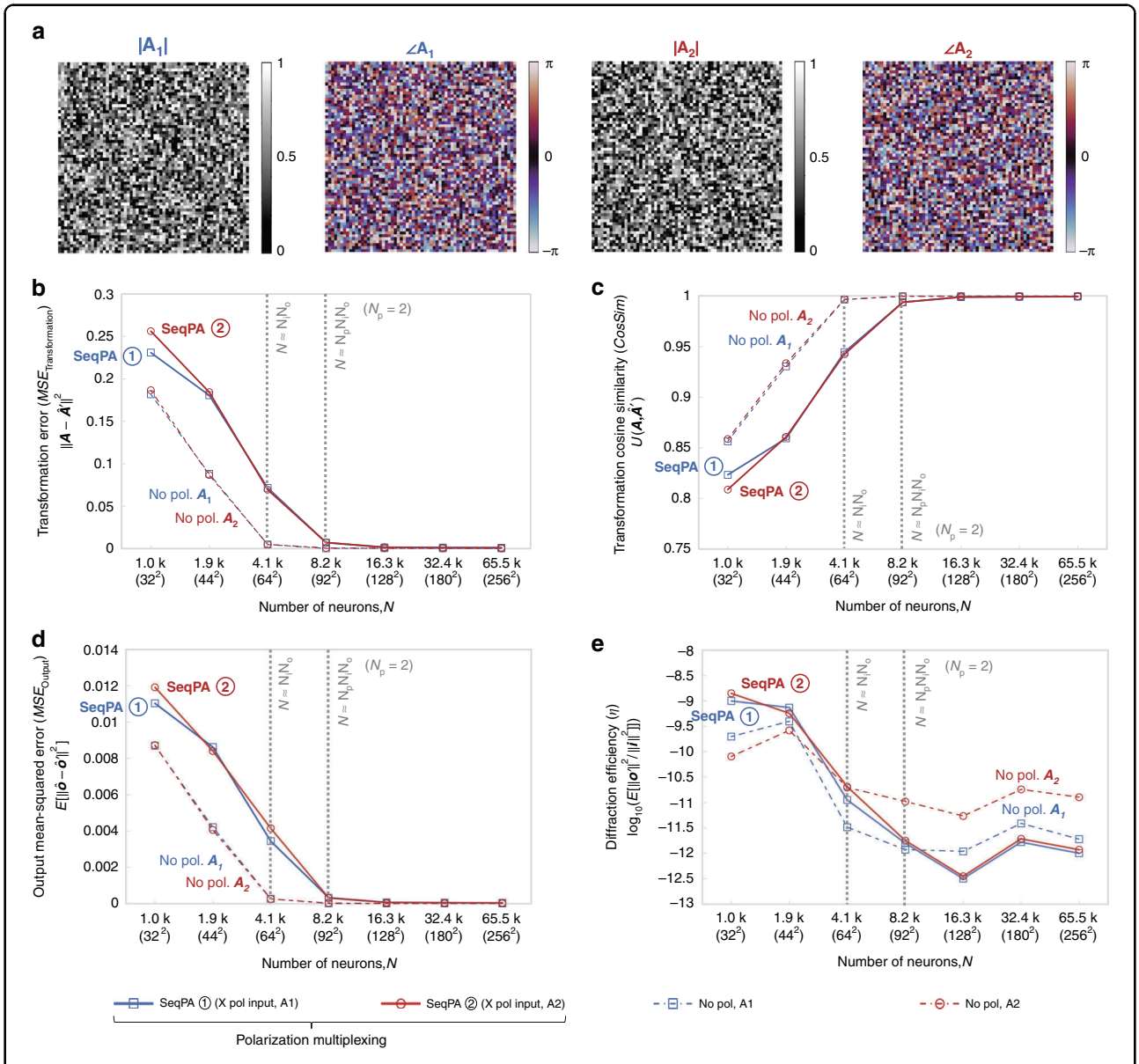

Fig. 2 Diffractive all-optical transformation results for 2-channel polarization multiplexing using the sequential polarization access (SeqPA) mode. a Amplitude and phase of the arbitrarily generated matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, which serve as the ground truth (target) for the diffractive all-optical transformations. **b** Curves representing the normalized mean-squared error between the ground truth transformation matrices ($\boldsymbol{A}_1$ and $\boldsymbol{A}_2$) and the all-optical transforms ($\boldsymbol{A}_1'$ and $\boldsymbol{A}_2'$) resulting from the trained diffractive networks as a function of the number of diffractive neurons N. The solid curves are achieved by the polarization-multiplexed diffractive networks trained using the SeqPA mode, which are compared with the dashed curves achieved by the regular diffractive networks trained with the same set of N but without any polarization multiplexing. For the polarization-multiplexed models, the results for the two polarization channels ① and ②, corresponding to transforms $\boldsymbol{A}'_1$ and $\boldsymbol{A}'_2$, are shown in separate curves that are labeled with "SeqPA ①" and "SeqPA ②", respectively. For the regular diffractive models without polarization multiplexing, the results for all-optical approximation of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are shown in separate curves labeled with "No pol. $\boldsymbol{A}_1$" and "No pol. $\boldsymbol{A}_2$", respectively. The space between the simulation data points is linearly interpolated. **c** Same as (**b**) but the cosine similarity between the all-optical transforms and their ground truth shown in (**a**) is reported. **d** Same as (**b**) but the mean-squared error between the diffractive network output fields and their ground truth is reported. **e** Diffraction efficiency of the presented diffractive networks

multiplication operation. Each input-output polarization combination in our diffractive design represents an all-optical implementation of a unique linear transformation between the input and output FOVs. Therefore, for each input-output polarization combination, infinitely many different target complex fields can be all-optically synthesized by the trained diffractive network in response to different input field distributions; and this capability

accurately defines the corresponding complex-valued linear transformation at the output FOV for all the possible and infinitely many combinations of phase and amplitude distributions at the input FOV.

A polarization-multiplexed diffractive network can perform an arbitrary set of target linear transformations using the same diffractive layers that all-optically implement a distinct complex-valued linear transformation at a selected input/output polarization combination. We believe that this unique framework will be valuable in developing high-throughput optical processors and polarization-based machine vision systems operating at different parts of the electromagnetic spectrum. Moreover, the presented diffractive computing platform and the underlying concepts can be used to develop polarization-aware optical information processing systems for e.g., detection, localization, and statistical inference of objects with unique polarization properties.

Results

Throughout this section, the terms "diffractive optical network," "diffractive network," and "diffractive processor" are interchangeably used. The schematic of our framework for 2-channel polarization-multiplexed all-optical computing ($N_p = 2$) is shown in Fig. 1a. A polarization-encoded diffractive neural network, composed of 4 trainable diffractive layers, is trained to all-optically perform 2 distinct, complex-valued linear transformations between the input and output FOVs through 2 orthogonal polarization channels. The predetermined polarizer array (which is treated as non-trainable) consists of multiple linear polarizer units with four different polarization directions: 0°, 45°, 90° and 135°. This non-trainable polarizer array is positioned close to the center of the diffractive volume (i.e., between the 2nd and 3rd trainable diffractive layers) so that the resulting polarization modulation does not directly dominate the output field; the former and latter diffractive layers are jointly optimized to effectively communicate with the polarizer array and all-optically implement the desired group of linear transformations. More details about the architecture, optical forward model and training details of the polarization diffractive network can be found in the Methods section.

We use $\boldsymbol{i}$ and $\boldsymbol{o}$ to denote the complex-valued, vectorized versions of the 2D input and output complex fields located at the input and output FOVs of the diffractive network, respectively, as presented in Fig. 1a. Based on the scalar diffraction theory, here $\boldsymbol{i_x}$ and $\boldsymbol{o_x}$ represent the column vectors of the complex fields generated by sampling the x-polarized optical fields within the input and output FOVs, respectively, and vectorizing the resulting 2D matrices in a column-major order. Similar to $\boldsymbol{i_x}$ and $\boldsymbol{o_x}$, $\boldsymbol{i_y}$ and $\boldsymbol{o_y}$ are their counterparts generated by sampling the y-polarized optical fields within the input and output FOVs, respectively. Based on this notation, ($\boldsymbol{i_x}$, $\boldsymbol{i_y}$) and ($\boldsymbol{o_x}$, $\boldsymbol{o_y}$) can be considered to represent the input and output channels of our polarization-multiplexed diffractive network, respectively. In our analyses, the number of pixels in the input and output FOVs are both taken as $N_i = N_o = 8^2 = 64$, such that each target linear transformation matrix has 64^2 complex-valued entries.

In this first implementation with $N_p = 2$, we randomly generated two complex-valued matrices $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, each with a size of $N_i \times N_o = 64^2$, to serve as two unique arbitrary linear transformations that we would like to all-optically implement using a single polarization diffractive network. Visualized in Fig. 2a with their amplitude and phase components, these two matrices are independently generated using different random seeds, and the difference between the two matrices can be found in Fig. S1. We also randomly generated two training sets of complex-valued vectors $\{\boldsymbol{i_1}\}$ and $\{\boldsymbol{i_2}\}$ with $N_i = 64$ as input fields, and constructed the corresponding sets of output field vectors $\{\boldsymbol{o_1}\}$ and $\{\boldsymbol{o_2}\}$ using $\boldsymbol{o}_1 = \boldsymbol{A}_1\boldsymbol{i}_1$ and $\boldsymbol{o}_2 = \boldsymbol{A}_2\boldsymbol{i}_2$, respectively. For each one of these training sets, $\{\boldsymbol{i_1}\}$ and $\{\boldsymbol{i_2}\}$, we used 55,000 randomly generated complex fields in our training process. A further increase in the size of this training dataset (to e.g., >100,000 randomly generated complex fields) could improve the transformation approximation accuracy of the trained diffractive networks, but would not change the general conclusions of this manuscript and therefore is left as future work.

Based on the given inputs of $\{\boldsymbol{i_1}\}$ and $\{\boldsymbol{i_2}\}$, the ultimate goal of training our polarization-multiplexed diffractive network is to simultaneously compute the all-optical output fields $\{\boldsymbol{o}'_1\}$ and $\{\boldsymbol{o}'_2\}$ to come close to the output ground truth (target) fields $\{\boldsymbol{o_1}\}$ and $\{\boldsymbol{o_2}\}$; this way, the all-optical transformations $\boldsymbol{A}'_1$ and $\boldsymbol{A}'_2$ performed by the trained single diffractive system represent an accurate approximation to their ground truth (target) transformation matrices $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$. It should be emphasized that we are not aiming to train the diffractive network to implement the correct linear transformations for only a few input-output field pairs. Instead, despite the limited number of input/output field patterns used during the training process, our goal is to generalize to *any* pairs of ($\boldsymbol{i_1}$, $\boldsymbol{o_1}$) and ($\boldsymbol{i_2}$, $\boldsymbol{o_2}$) that satisfy $\boldsymbol{o}_1 = \boldsymbol{A}_1\boldsymbol{i}_1$ and $\boldsymbol{o}_2 = \boldsymbol{A}_2\boldsymbol{i}_2$. More details about the training data generation can be found in "Methods".

To form two unique diffractive information processing pipelines in the same diffractive network for performing the linear transformations given by $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, as shown in Fig. 1a we matched the input fields and the diffractive output pairs, $\{(\boldsymbol{i_1}, \boldsymbol{o}'_1)\}$ and $\{(\boldsymbol{i_2}, \boldsymbol{o}'_2)\}$, with the input and output polarization channels of our diffractive system, i.e., $\boldsymbol{i_x} = \boldsymbol{i}_1$, $\boldsymbol{i_y} = \boldsymbol{i}_2$, $\boldsymbol{o_x} = \boldsymbol{o}'_1$ and $\boldsymbol{o_y} = \boldsymbol{o}'_2$. That is to say, the $\boldsymbol{A}'_1$ transformation is performed by encoding the

corresponding input field data $\boldsymbol{i_1}$ into the x-polarized optical field within the input FOV, using e.g., an x-aligned linear polarizer, and decoding (sampling) the x-polarized component of the field within the output FOV as the computed output field $\boldsymbol{o}'_1$ using e.g., an x-polarized analyzer. We denote this diffractive information processing channel as the channel ① in Fig. 1b. It is also a similar case for the $\boldsymbol{A}'_2$ transformation, except this time the y polarization is employed at the input and output FOVs, and this diffractive information processing channel is denoted as the channel ②. With this polarization encoding scheme, there are potentially two modes to perform the data inference through the same diffractive network: (1) in two sequential, successive accesses to the diffractive system, each time feeding the input data using its assigned polarization channel, and obtaining the corresponding output (see Fig. 1b, left); and (2) in single access to the diffractive system, by feeding the input data of both of the two polarization channels in parallel, and obtaining the two corresponding outputs simultaneously (see Fig. 1b, right). We term the former and latter approaches as the "sequential polarization access" (SeqPA) mode and the "simultaneous polarization access" (SimPA) mode, respectively. We should emphasize that the fundamental difference between these two modes of operation lies in the input information: the SimPA mode can simultaneously accept both of the input polarization states (e.g., x and y polarization) for encoding two different channels of input information, while the SeqPA mode can accept a single polarization state as its input so that only one channel of input information is encoded at a given time. Therefore, if the input FOV simultaneously encodes the data to be processed in two different polarization states, or if the time lag caused by switching between different input polarization states is unacceptable (such as e.g., an input FOV that includes a rapidly changing dynamic scene with specific polarization information), then only the SimPA mode would be suitable to process the input encoding. Conversely, if the system is only required to compute a single linear transformation at a given time, or if the time lag caused by switching back and forth between two different input polarization states is acceptable, then SeqPA mode can be used. Detailed analyses of these two modes of operation are presented in the following subsections.

2-channel polarization-multiplexed all-optical diffractive computing using the sequential polarization access (SeqPA) mode

As shown in Fig. 1b, left, with the input data $\boldsymbol{i_1}$ and $\boldsymbol{i_2}$ being separately and sequentially fed into the polarization channels ① and ②, respectively, the all-optical computed outputs $\boldsymbol{o}'_1$ and $\boldsymbol{o}'_2$ are also collected successively using the same diffractive network hardware. By employing this SeqPA strategy, we trained polarization-multiplexed diffractive networks with different numbers of trainable diffractive neurons, i.e., $N = \{32^2, 44^2, 64^2, 92^2, 128^2, 180^2, 256^2\}$, all using the same training datasets $\{(\boldsymbol{i_1}, \boldsymbol{o_1})\}$ and $\{(\boldsymbol{i_2}, \boldsymbol{o_2})\}$ and the same number of epochs. To benchmark the performances of these multiplexed diffractive networks, for each transformation dataset and N, we also trained regular diffractive networks without the polarizer array or any polarization encoding/decoding at the input/output FOVs, which constitute our baseline. These regular diffractive networks, denoted as "*No pol.*" in our analyses, are trained to approximate *only one* linear transformation (i.e., either $\boldsymbol{A_1}$ or $\boldsymbol{A_2}$), and therefore they are referred to as $N_p = 1$ (no polarization multiplexing).

Figure 2b–e present the quantitative comparison of the all-optical transformation results obtained using the trained diffractive networks described above. Three different metrics were used to quantify the transformation accuracy and generalization performance of these diffractive networks: (1) the normalized transformation mean-squared error ($\mathrm{MSE}_{\mathrm{Transformation}}$), (2) the cosine similarity (*CosSim*) between the all-optical transforms and the target transforms, and (3) the mean-squared error between the diffractive network output fields and their ground truth ($\mathrm{MSE}_{\mathrm{Output}}$). These performance metrics are reported in Fig. 2b-d, as a function of the number of diffractive neurons (N) used in each design. Note that the transformation error of the polarization-multiplexed diffractive systems is calculated per polarization channel. More details about the formulations of these performance metrics can be found in Methods. In Fig. 2b, it can be seen that the transformation errors of all the trained diffractive models monotonically decrease as N increases, which is expected due to the increased degrees of freedom in the diffractive processor. In the standard diffractive networks without polarization multiplexing (dash-dotted curves labeled with "No pol. $\boldsymbol{A_1}$" or "No pol. $\boldsymbol{A_2}$"), the transformation errors for implementing $\boldsymbol{A_1}$ or $\boldsymbol{A_2}$ are almost the same (which indicates that these randomly selected matrices, $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, represent similar computational complexity; also see Fig. S1). The approximation errors of these standard diffractive networks, No pol. $\boldsymbol{A_1}$ and No pol. $\boldsymbol{A_2}$, both approach to 0 as N approaches $N_iN_o = 64^2 \approx 4.1\,k$. In the polarization-multiplexed diffractive models (solid curves labeled with "SeqPA ①" or "SeqPA ②"), the transformation errors $\mathrm{MSE}_{\mathrm{Transformation}}$ for the two distinct transforms computed through the two polarization channels are also very close to each other for all values of N, demonstrating no bias toward any specific polarization channel or transform. The approximation errors of these polarization-multiplexed models approach to 0 as N approaches $N_pN_iN_o = 2N_iN_o = 92^2 \approx 8.2\,k$. This finding indicates that compared with the baseline diffractive models that can only perform a *single* transform,

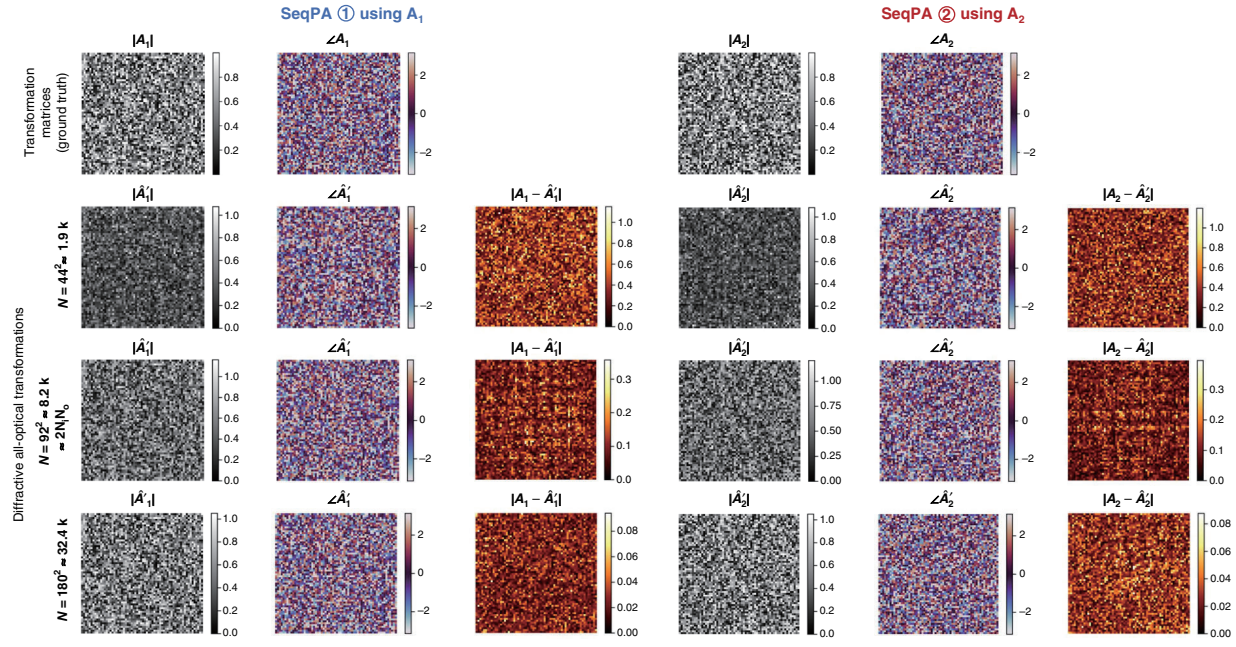

Fig. 3 All-optical transformation matrices estimated by the 2-channel polarization multiplexed diffractive designs trained using the SeqPA mode with $N = 44^2$, 92^2 and 180^2. The differences of these matrices from the ground truth matrices are also shown

performing two unique transforms using polarization multiplexing through the same diffractive model requires the number of trainable neurons N to double. This conclusion is further supported by the results of the other two performance metrics, *CosSim* (Fig. 2c) and $\mathrm{MSE}_{\mathrm{Output}}$ (Fig. 2d) that both show the same trends as in Fig. 2b: for the baseline diffractive models *CosSim* and $\mathrm{MSE}_{\mathrm{Output}}$ approach 1 and 0 as N approaches N_iN_o, respectively, while for the polarization-multiplexed diffractive models, the two metrics approach 1 and 0 as N approaches $N_pN_iN_o = 2N_iN_o$. Apart from the metrics that are used to evaluate the transformation performance, we also report the output diffraction efficiencies (η) of these diffractive models in Fig. 2e, which reveal that compared with the baseline diffractive networks (*No pol.*), the diffraction efficiencies of the polarization-multiplexed diffractive models trained using the SeqPA mode reach a similar level.

To further demonstrate the performance of our polarization-multiplexed diffractive networks, in Fig. 3 we show examples of the ground truth transformation matrices (i.e., $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$) and their counterparts (i.e., $\boldsymbol{A'}_1$ and $\boldsymbol{A'}_2$) resulting from the diffractive designs with $N = \{44^2, 92^2, 180^2\}$, along with the amplitude and phase absolute errors. Exemplary complex-valued input-output fields from the same set of diffractive designs are also presented in Fig. 4. Figures 3 and 4 reveal that for both of the polarization channels, when $N \geq N_pN_iN_o = 2N_iN_o$, the all-optical transformation matrices and the output complex fields very well match their ground truth targets with negligible absolute errors, which are also in line with the observations made in Fig. 2.

2-channel polarization-multiplexed all-optical diffractive computing using the simultaneous polarization access (SimPA) mode

As an alternative to the sequential polarization access (SeqPA) used earlier, we also explored the use of the simultaneous polarization access (SimPA) mode in our all-optical computing framework. As shown in Fig. 1b, right, in single access to the diffractive system, the input complex-valued data $\boldsymbol{i_1}$ and $\boldsymbol{i_2}$ are fed into the polarization channels ① and ②, respectively, and the all-optical diffractive outputs $\boldsymbol{o}'_1$ and $\boldsymbol{o}'_2$ are collected at the same time through two orthogonal polarization states at the output FOV. Before we trained a new polarization-multiplexed diffractive network from scratch using the SimPA mode, we first took our earlier diffractive designs trained using the SeqPA mode and tested them directly using the SimPA mode by inputting both polarization channels ① and ② *at the same time*, deviating from their training scheme, which only used SeqPA. The results of blindly testing the SeqPA-trained diffractive networks under the SimPA mode are shown in Fig. S2, which reveals inference results with significantly higher values of $\mathrm{MSE}_{\mathrm{Transformation}}$ and $\mathrm{MSE}_{\mathrm{Output}}$ and decreased values of *CosSim*, all of which indicate a performance degradation, when we operate a SeqPA-trained diffractive network using the

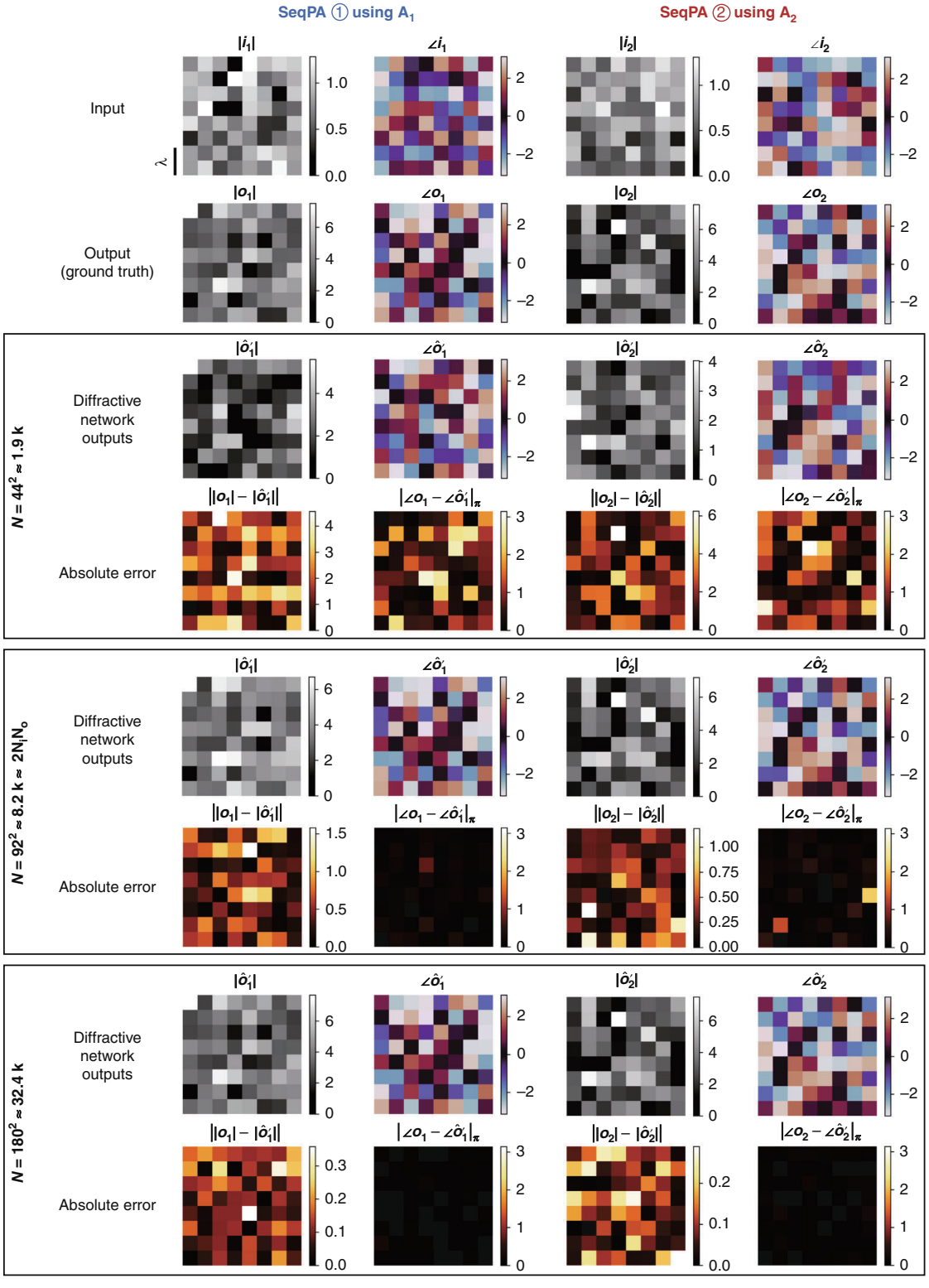

Fig. 4 Examples of input/output complex fields for the ground truth transformations presented in Figs. 2 and 3 along with the output fields computed by the 2-channel polarization-multiplexed diffractive designs trained with the SeqPA mode using $N = 44^2$, 92^2 and 180^2. Note that $\left|\angle \boldsymbol{o} - \angle \widehat{\boldsymbol{o}'}\right|_{\pi}$ indicates the wrapped phase difference between the ground truth output field $\boldsymbol{o}$ and the normalized diffractive network output field $\widehat{\boldsymbol{o}'}$

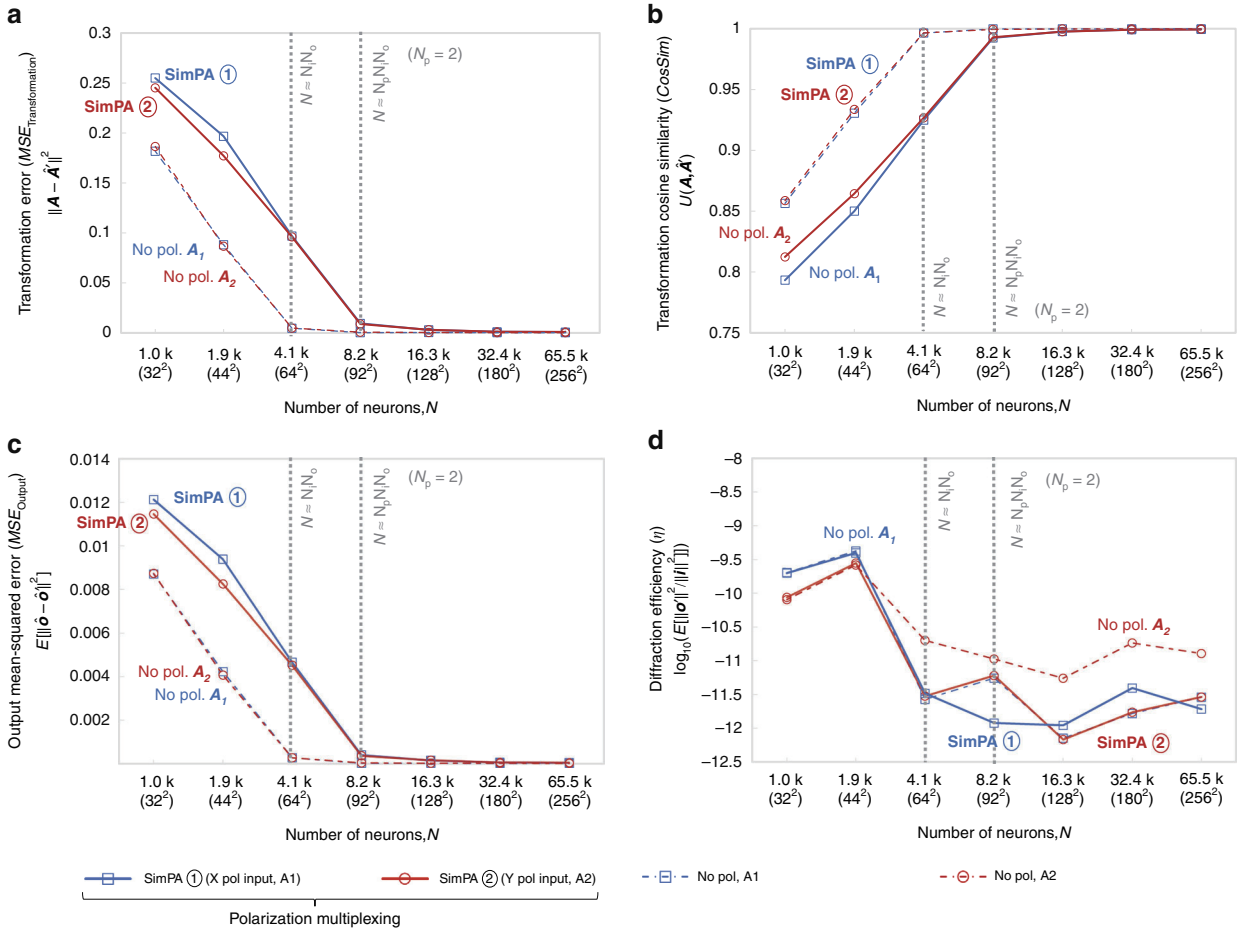

Fig. 5 Diffractive all-optical transformation results for 2-channel polarization multiplexing using the simultaneous polarization access (SimPA) mode. a Curves representing the normalized mean-squared error between the ground truth transformation matrices ($\boldsymbol{A}_1$ and $\boldsymbol{A}_2$) and the all-optical transforms ($\boldsymbol{A}'_1$ and $\boldsymbol{A}'_2$) resulting from the trained diffractive networks as a function of the number of diffractive neurons N. The solid curves are achieved by the polarization-multiplexed diffractive systems trained using the SimPA mode, which are compared with the dashed curves achieved by the regular diffractive networks trained with the same set of N but without any polarization multiplexing. For the polarization-multiplexed models, the results for the two polarization channels ① and ②, corresponding to transforms $\boldsymbol{A}'_1$ and $\boldsymbol{A}'_2$, are shown in separate curves that are labeled with "SimPA ①" and "SimPA ②", respectively. For the regular models without polarization multiplexing, the results for all-optical approximation of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are shown in separate curves labeled with "No pol. $\boldsymbol{A}_1$" and "No pol. $\boldsymbol{A}_2$", respectively. The space between the simulation data points is linearly interpolated. **b** Same as (**a**) but the cosine similarity between the all-optical transforms and their ground truth is reported. **c** Same as (**a**) but the mean-squared error between the diffractive network output fields and their ground truth is reported. **d** Diffraction efficiency of the presented diffractive networks

SimPA mode. As shown in Fig. S3, this performance degradation is due to the "cross-talk" between the two transformation channels when both of the input polarization states are at the same time present, which was not considered during the SeqPA-based training process. These results highlight the necessity of training the diffractive system from scratch under the SimPA mode, so that the impact of this cross-talk can be taken into account and minimized during the iterative design process. A related mathematical analysis that supports the same conclusion is reported in Supplementary Note 1.

After training our diffractive models from scratch using the SimPA mode, we report their blind testing results in Fig. 5 using the solid curves labeled with "SimPA ①" and "SimPA ②". The results of the new diffractive designs trained using the SimPA mode demonstrate the success of all-optically performing two different linear transformations in parallel using polarization multiplexing. Our analysis (Fig. 5) also reveals the same conclusions discussed earlier for the models trained using the SeqPA mode: the all-optical transformation performance of polarization-multiplexed diffractive networks very well match the ground truth, desired transformations as N approaches $N_pN_iN_o = 2N_iN_o$. Furthermore, as shown in Fig. 5d, the diffraction efficiencies achieved by the polarization-multiplexed diffractive networks reach a

similar level as their baseline counterparts that use the same number of diffractive layers, but without the linear polarizer array.

We further compared the blind testing results of these two different modes of operation (SeqPA vs. SimPA) and performed a cross-talk field analysis (see Fig. S3). We found out that the amount of transformation cross-talk in the diffractive models trained using the SimPA mode (shown in the right column of Fig. S3c, d), is ~300-fold lower when compared with the amplitude values of the cross-talk observed in the diffractive designs trained using the SeqPA mode (shown in the left column of Fig. S3c, d). During the diffractive model training, these cross-talk fields are gradually eliminated (penalized) using the SimPA mode of operation to better approximate the ground truth fields. However, for the diffractive models trained under the SeqPA mode, such cross-talk fields are ignored (i.e., remain non-penalized during the training phase) since the SeqPA operation assumes successive access to the diffractive network, one input polarization state at a time. Stated differently, SeqPA trained diffractive networks successfully approximate the target transformations only when they are tested under the same SeqPA mode of operation, and fail due to the field cross-talk when tested under the SimPA mode.

4-channel polarization-multiplexed all-optical diffractive computing

So far, we have demonstrated to perform all-optical computing with 2-channel polarization multiplexing through a single diffractive network. To further exploit the polarization multiplexing capability of this diffractive computing framework, next, we explored a 4-channel polarization-multiplexed design for performing 4 different arbitrarily-selected linear transformations through a single diffractive network (i.e., $N_p = 4$). Figure 6 illustrates the schematics of this framework. As depicted in Fig. 6b, by sequentially connecting one of the two input polarization states with one of the two output polarization states, four transformation channels, ①, ②, ③ and ④, can be formed to all-optically perform $N_p = 4$ distinct complex-valued transforms using the same diffractive processor. This 4-channel polarization-multiplexed design operates in a similar way as the SeqPA mode, where the different input data are separately and sequentially fed into different input polarization channels. Using this SeqPA operation mode, our diffractive system can accurately perform 4 different complex-valued linear transformations using the same passive diffractive layers, in a single optical network. For example, when only one polarization state (e.g., $\boldsymbol{i_x}$) is utilized to encode the input data (i.e., $\boldsymbol{i} = \boldsymbol{i_x} = \boldsymbol{i_1} = \boldsymbol{i_3}$), we can measure the output field at two orthogonal polarization states and simultaneously read out two computed outputs (i.e., $\boldsymbol{o_\alpha} = \boldsymbol{o_1}$ and $\boldsymbol{o_\beta} = \boldsymbol{o_3}$), each corresponding to the result of a uniquely different linear transformation (i.e. $\boldsymbol{A_1}$ or $\boldsymbol{A_3}$) computed based on the same input; this capability enables parallel optical information processing through the same polarization-encoded diffractive network. The overall design of this 4-channel diffractive system can be considered to utilize the remaining degrees of freedom in the cross-talk channels of the 2-channel system. Additional analysis that supports the same conclusions can be found in Supplementary Note 1.

It is also worth noting that, compared to the 2-channel polarization-multiplexed system reported earlier, the polarization states for the output field sampling in this 4-channel system are selected to be 45° and 135° linear polarization. This design choice is made to balance out the diffraction efficiencies of the resulting 4 different linear transformations that are all-optically performed by the diffractive network. Stated differently, this design choice introduces symmetry to all the input/output polarization combinations that are each assigned to a different linear transformation. In Fig. 6a, b, we denote the two output fields corresponding to the linear polarization directions at 45° and 135° as $\boldsymbol{o_\alpha}$ and $\boldsymbol{o_\beta}$, respectively.

In the light of our earlier findings that point to the need for more diffractive neurons in the case of $N_p = 2$ when compared to $N_p = 1$, here we employed 8 successive trainable diffractive layers to increase our degrees of freedom for $N_p = 4$ design (see Fig. 6a). Also, compared to the earlier 2-channel polarization-multiplexed design, we included an additional linear polarizer array with the same configuration as before (with polarization orientations of 0°, 45°, 90° and 135°) to further enhance the spatial diversity of polarization modes within the diffractive processor. These two linear polarizer arrays are positioned after the 3rd and 5th diffractive layers, respectively. Same as the $N_p = 2$ diffractive designs, these linear polarizer arrays are pre-determined (i.e., non-trainable) and act as "polarization seeds" within the trained diffractive network.

Next, we generated random data to train and test our diffractive networks under $N_p = 4$. In addition to the two randomly generated ground truth transforms $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ that were earlier used for the 2-channel models, we randomly generated two additional complex-valued transforms $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$ and accordingly constructed the training and testing dataset consisting of the input and ground truth output fields. These four ground truth (target) transforms are visualized in Fig. 7a, and their differences can be found in Fig. S1. Following the training of the polarization-multiplexed diffractive networks with different N, their transformation performance for $N_p = 4$ is analyzed in Fig. 7b–d based on the same set of performance metrics that were used earlier. These results reveal that, when N approaches $N_pN_iN_o = 4N_iN_o = 16.4\ k$, the $\mathrm{MSE}_{\mathrm{Transformation}}$ and $\mathrm{MSE}_{\mathrm{Output}}$ of all the four diffractive

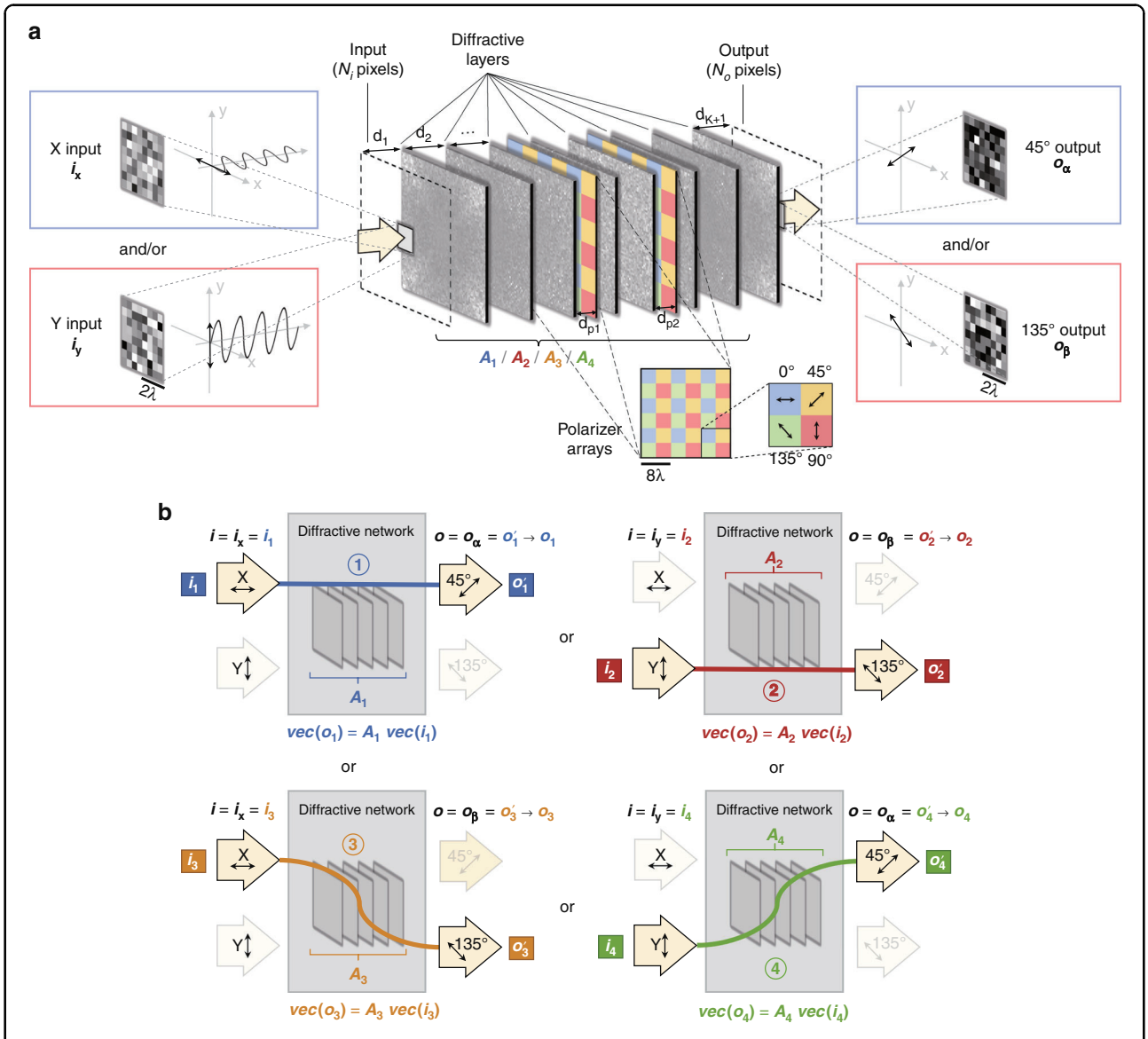

Fig. 6 Schematic of 4-channel polarization-multiplexed all-optical diffractive computing framework for performing four unique linear transformations through a single diffractive network. a Optical layout of the polarization-encoded diffractive network, where eight trained diffractive layers and two arrays of linear polarizers are jointly used to perform four distinct, complex-valued linear transformations between the input field ***i*** and the output field ***o*** by using polarization encoding/decoding at the input/output FOVs. **b** Schematic for the operation of the 4-channel polarization-multiplexed all-optical computing framework, where the four polarization channels, ①, ②, ③ and ④, are formed by sequentially connecting one of the two input polarization states with one of the two output polarization states

transformations approach 0, while the *CosSim* approaches 1, demonstrating that all the target linear transformations ($\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$) can be successfully approximated by a single diffractive processor with a negligible error if $N \geq N_p N_i N_o$. This is the same conclusion that was reached earlier for $N_p = 2$.

To further demonstrate the success of these 4-channel polarization-multiplexed diffractive systems, in Fig. S4 we present the ground truth transformation matrices (i.e., $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$) and their diffractive counterparts (i.e., $\boldsymbol{A'}_1$, $\boldsymbol{A'}_2$, $\boldsymbol{A_3}'$ and $\boldsymbol{A_4}'$) designed with $N = \{14.3\text{k}, 66.5\text{k}\}$, along with the amplitude and phase errors made in each case. Furthermore, exemplary complex-valued output fields achieved by these diffractive systems are also shown in Fig. S5, all of which confirm the success of the presented 4-channel polarization-multiplexed diffractive designs. Finally, we also analyzed the output diffraction efficiencies of these diffractive models, reported in Fig. 7e. The results show that, compared to their counterparts without polarization encoding ($N_p = 1$), the polarization-

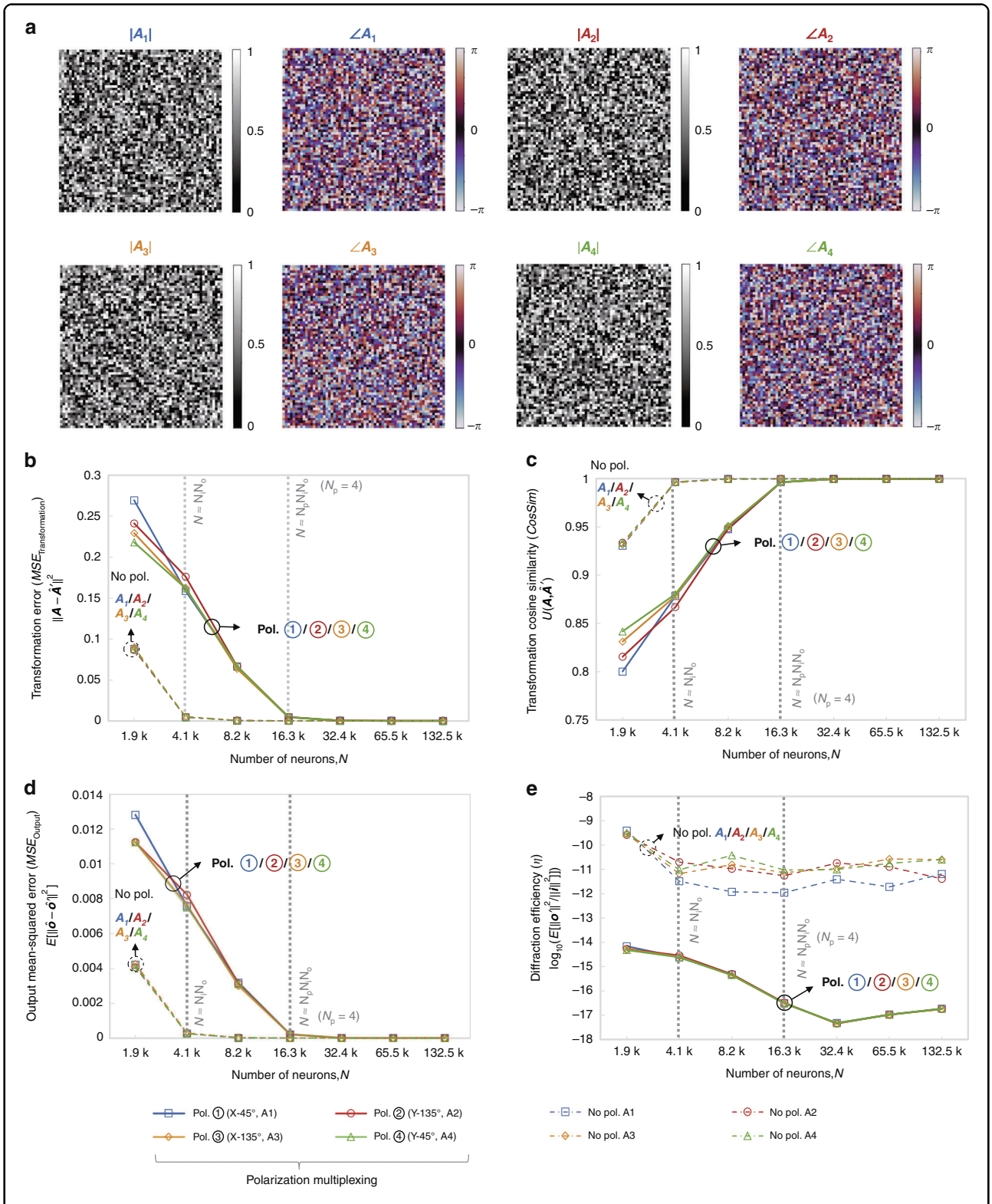

Fig. 7 (See legend on next page.)

(see figure on previous page)

Fig. 7 Diffractive all-optical transformation results for 4-channel polarization multiplexing of four distinct arbitrary linear transforms (depicted in Fig. 6). a Amplitude and phase of the arbitrarily generated matrices $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$, which serve as the ground truth (target) for the diffractive all-optical transformations. **b** Curves representing the normalized mean-squared error between the ground truth transformation matrices ($\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$) and the all-optical transforms ($\boldsymbol{A'}_1$, $\boldsymbol{A'}_2$, $\boldsymbol{A}_3'$and $\boldsymbol{A}_4'$, examples shown in Fig. S4) resulting from the trained diffractive networks as a function of the number of diffractive neurons *N*. The solid curves are achieved by the 4-channel polarization-multiplexed diffractive systems, which are compared with the dashed curves achieved by the regular diffractive networks trained with the same set of *N* but without polarization multiplexing. For the polarization-multiplexed models, the results for the four polarization channels ①, ②, ③ and ④ are shown in separate curves but jointly labeled with "Pol. ①/②/③/④" due to the spatial overlap of these curves. For the regular diffractive models without polarization multiplexing, the results for all-optical approximation of $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$ (individually) are shown in separate curves but jointly labeled with "No pol. $\boldsymbol{A_1}$/$\boldsymbol{A_2}$/$\boldsymbol{A_3}$/$\boldsymbol{A_4}$" due to the spatial overlap of these curves. The space between the simulation data points is linearly interpolated. **c** Same as (**b**) but cosine similarity between the all-optical transforms and their ground truth is reported. **d** Same as (**b**) but the mean-squared error between the diffractive network output fields and their ground truth is reported. **e** Diffraction efficiency of the presented diffractive networks

multiplexed diffractive models with $N_p = 4$ turn out to be less power efficient (per transformation), with an efficiency decrease of ~6 dB at the output FOV. This relatively small difference in the output diffraction efficiencies mainly stems from the different number of diffractive layers used in these two systems: the baseline diffractive systems without polarization encoding use 4 diffractive layers, whereas the 4-channel polarization-multiplexed systems are much deeper, utilizing 8 diffractive layers. Considering that the optical field within a deeper system with more diffractive layers propagates and spreads over a longer axial distance, it exhibits a relatively lower diffraction efficiency. Therefore, these results do not contradict our previous conclusion that the diffraction efficiency of the polarization-multiplexed diffractive network is similar to that of the baseline diffractive system when using the same number of diffractive layers.

Our results and analyses presented so far demonstrated that a single polarization-multiplexed diffractive network can all-optically compute four different complex-valued, arbitrarily-selected linear transformations between its input and output FOVs by using orthogonal linear polarization states. In addition to linear polarization, other polarization states can also be used, without loss of generality, to perform the same multiplexed computational tasks. To demonstrate this capability, we used two orthogonal circular polarization states (i.e., left- and right-hand circular polarization) at the input of a polarization-multiplexed diffractive network to encode the input information; the output channels in this case included x and y linear polarization states, i.e., the 4 different, arbitrarily-selected linear transformations were each assigned to one combination of circular-linear polarization. Our results, reported in Fig. S8, revealed that circular input polarization-multiplexed diffractive processors successfully approximated the target, complex-valued linear transformations, when N approaches $N_pN_iN_o = 4N_iN_o = 16.4\ k$, arriving at the same conclusion that we had for linear input polarization states. In this diffractive design, we used the same linear polarizer array (i.e., the seed) within the diffractive network volume to communicate between the circular polarization states at the input FOV and the linear polarization states at the output FOV, all-optically performing 4 different complex-valued transformations through the same diffractive network. A mathematical analysis of this design and its relationship to earlier diffractive designs with linear input/output polarization states is also provided in Supplementary Note 1. Since any arbitrary polarization state can be expressed through a superposition of orthogonal linear or circular polarization states, the same diffractive design can be extended to different input/output combinations of other polarization states. As detailed in Supplementary Note 1, a polarization-multiplexed diffractive processor with $N_p = 4$ can be designed by using input-output combinations of 2 orthogonal input polarization states (e.g., linear, circular or elliptical) and 2 orthogonal output polarization states (e.g., linear, circular or elliptical), where each input-output polarization combination all-optically performs one of the target complex-valued transformations ($\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$, $\boldsymbol{A_4}$). Supplementary Note 1 further proves that any additional transformation matrix $\boldsymbol{A_a}$ that can be assigned to a new combination of input-output polarization states of the diffractive network can be written as a linear combination of $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$.

Discussion

Our results and analysis demonstrated that, using polarization multiplexing in a single diffractive network, one can all-optically perform a group of complex-valued arbitrary linear transformations at the same output FOV of the diffractive network. In practical applications, these different transformations can cover, for example, various machine vision tasks, such as detection, classification, and localization of objects, which can be programmed into different input/output polarization states. These different tasks could potentially be also performed by employing multiple, separately-optimized diffractive networks, each of which is dedicated to performing a single computational task. However, such an approach would require the precise optical projection of an input FOV (while preserving its phase and amplitude distribution and

polarization information) onto separately positioned, individual diffractive networks, and would naturally suffer from additional optical losses and aberrations, misalignment issues, a much larger device footprint and higher manufacturing/alignment-related costs. In contrast, integrating multiple tasks to be all-optically performed within the same diffractive network and a common input FOV provides a much simpler and better design, offering unique advantages such as e.g., speed, compactness, resilience to misalignments and aberrations, power efficiency and cost-effectiveness.

Also note that, it is not practical to spatially superimpose multiple diffractive subsystems, each one separately designed for a unique transformation, using e.g., phase-composite metasurfaces or other metamaterials to create a polarization-multiplexed diffractive processor. First, in the design of each diffractive meta-unit, the cross-talk between the meta-atoms for the two orthogonal polarization states cannot be neglected. Therefore, the direct superposition of two or more different metasurface designs separately trained/designed for each one of the complex transformations would not work due to the cross-talk between the polarization channels of different metasurface designs. Stated differently, different metasurface designs, when put together in order to achieve multiplexed linear transformations in the same optical unit, will fail each other's transformation accuracy. In addition to this, there will be field cross-talk between the adjacent meta-units that are merged together on the same layer due to the in-plane propagating waves. Although increasing the lateral distance between two adjacent meta-units (from different designs, each targeting one transformation) can weaken the impact of this field cross-talk problem, it will then lead to lower diffraction efficiencies at the output and sacrifice the lateral density of the meta-units at each diffractive layer, thus degrading the computational performance and accuracy of the system. Furthermore, the desired phase response of such polarization-encoded meta-units in general covers a small angular range, leading to a low numerical aperture (NA) that fundamentally limits the connectivity between the diffractive layers. In our diffractive solutions, each isotropic feature of our diffractive network communicates with the following diffractive layer (s) with an NA of n ($n = 1$ in air). However, metasurface-based designs would fall short to offer such high numerical apertures, because the high spatial frequency components for the orthogonal input polarizations would deviate from the ideal phase response of the meta-unit, introducing errors to the multiplexed linear transformations that are targeted. Due to some of these challenges outlined above, metasurface or metamaterial-based diffractive surfaces have not yet been demonstrated as a solution to universal, all-optical implementation of an arbitrary linear transformation or a group of transformations.

In addition to polarization multiplexing, we should note that other degrees of freedom can be used to implement multiple computational tasks through a single diffractive network. For example, one can divide the input/output FOVs of the diffractive network into multiple regions, where each region is assigned to a unique computing task through spatial-division multiplexing. It is also possible to achieve wavelength-division multiplexing by assigning different wavelengths or spectral bands to independent computing tasks and employing dispersive elements in the diffractive computing system. In contrast to these other possible methods of information multiplexing, the polarization-based multiplexing that we reported here requires solely the addition of linear polarizers to a diffractive network without changing its architecture. Such polarizers are readily available (e.g., polarizing films), even integrated with the individual pixels of polarization-based imaging systems[60], and can be adapted to a wide range of wavelengths. Furthermore, polarization multiplexing can be flexibly coupled with other multiplexing methods (such as spectral and/or spatial multiplexing) to further increase the computing capacity of the diffractive network.

Unlike the diffractive layers, where the transmission coefficients are trained and optimized to all-optically perform the target transformations, the design and arrangement of the seed polarizer arrays between the diffractive layers are treated as hyperparameters that are pre-determined and non-trainable. Therefore, the parameters of the embedded polarizers including their number, size, and orientation are fixed during the training process. The polarization modulation induced by these polarizer arrays remains unchanged and was not used as learnable degrees of freedom for our diffractive computing system to approximate the target transformations. Furthermore, their total number is small, i.e., we only used $6 \times 6 = 36$ linear polarizers per array, which is negligible when compared to N. An increase in the number of linear polarizers per plane would not improve the approximation power of our diffractive network to perform arbitrary linear transformations. However, the topology of such polarizer seeds could potentially impact the performance of our polarization-multiplexed diffractive computing system. To explore this, we adjusted several key parameters of the linear polarizer array used in our diffractive processor designs including e.g., 1) the period of each polarizer unit, 2) the overall size of each polarizer array, and 3) the number and position of the polarizer arrays within the diffractive network. For this comparative analysis we used as our test-bed the 4-channel polarization-multiplexed diffractive system with $N = N_p N_i N_o = 16.3\ k$ and the same complex-valued target linear transforms (i.e., $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$), the results of which are summarized in Supplementary Note 2. Based on these analyses, we observe that: (1) a better

approximation accuracy can be achieved when the period of each linear polarization unit on the polarizer array is $\leq 4\lambda$, and a period of $\sim 4\lambda$ empirically appears as an optimal choice, also providing an improved output diffraction efficiency (see Supplementary Fig. SN2); (2) the linear transformation accuracy and the diffraction efficiency of the system can be optimized by using polarizer arrays with a sufficiently large size, i.e., at least matching the size of the neighboring diffractive layers; (3) using two polarizer arrays and placing them apart with an axial distance of $\sim 8\lambda$ within the diffractive volume can provide improved results for the all-optical transformation accuracy and diffraction efficiency of $N_p = 4$ designs; and (4) using too many (e.g., >6) polarizer arrays within a diffractive network can lead to severe degradation in the computational accuracy of the system (unless more diffractive layers are added to the design).

We would like to also emphasize that the reported polarization-multiplexed diffractive networks can be directly applied to 2D arrays of phase and amplitude input data. Compared to other optical computing systems operating based on e.g., integrated photonics, which requires 1D inputs and phase recovery if the information is represented in the phase channel, the capability to directly process and analyze raw 2D complex fields makes our framework highly advantageous for visual computing tasks. On the other hand, unless spatial light modulators (SLMs) are employed as part of the diffractive system (see e.g., the Supplementary Information of ref. [23]. for a discussion on reconfigurable networks), each physically fabricated diffractive network is fixed and would need to be retrained and fabricated again as the target transformations change, which is a limitation of *passive* diffractive systems.

There are additional limitations of the presented diffractive computing framework. First, polarization-multiplexed diffractive computing systems present lower diffraction efficiencies at their output FOV compared to regular diffractive networks without polarization multiplexing (see Figs. 2e and 5d). Several remedies can be used to improve the output diffraction efficiency such as e.g., adding a diffraction-efficiency-related penalty term to the training loss function, and/or restricting the diffractive layers to perform phase-only modulation. The efficacy of using these approaches in a regular diffractive network design (without polarization multiplexing) to improve the output diffraction efficiency has already been demonstrated in our earlier work[69]. To exemplify the performance of a phase-only diffractive design and how it can be used to improve the output diffraction efficiency, we trained phase-only diffractive networks from scratch for the 4-channel polarization multiplexing case ($N_p = 4$), the results of which are summarized in Fig. S6. This analysis revealed that phase-only diffractive designs can achieve significantly better output diffraction efficiencies (improved on average by ~12 dB), while still successfully approximating the target linear transformations ($\boldsymbol{A_1}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$ and $\boldsymbol{A_4}$). As a trade-off, however, these phase-only diffractive designs also exhibit reduced degrees of freedom compared to their complex-valued counterparts. As a result of this, we observed that all the target linear transformations were successfully approximated by a single phase-only diffractive processor when N approached $2N_pN_iN_o = 8N_iN_o$. This 2-fold "threshold increase" in the number of diffractive features (i.e., $2N_pN_iN_o$ vs. $N_pN_iN_o$) is a direct reflection of the reduced number of trainable transmission parameters per diffractive layer due to the phase-only operation, which is a limitation of phase-only diffractive networks, despite their enhanced output diffraction efficiency. To further validate this conclusion, we also selected another set of 4 target linear transformations by changing the matrix elements to be real-valued, and used them as ground truth to train phase-only polarization-multiplexed diffractive networks with $N_p = 4$. As shown in Fig. S7, our results reveal that these phase-only diffractive networks can successfully approximate the real-valued target linear transforms when $N \geq N_pN_iN_o = 4N_iN_o$, demonstrating a similar approximation performance, with significantly higher output diffraction efficiency compared to their complex-valued diffractive counterparts. These findings emphasize the value of phase-only diffractive network designs as a photon-efficient solution in polarization-multiplexed diffractive computing, also providing an important rationale for planning the diffractive neuron budget (N) for a given computational task.

Other practical concerns that need to be discussed include the potential fabrication and alignment errors, surface reflections, material absorption and non-ideal polarization modulation within the diffractive network, which may altogether limit the performance and accuracy of diffractive computing. Some of these errors can be mitigated by selecting appropriate fabrication methods, e.g., high-precision lithography, and using less absorptive materials. Moreover, our previous results[23,38,44,49,50] showed that some of these uncontrolled physical errors and imperfections did not lead to a significant discrepancy between the experimental and numerical, expected results, indicating the correctness of the assumptions involved in our optical forward model and training procedures. Even if these errors and imperfections become considerable, the performance degradation of a diffractive network caused by some of these experimental factors can be compensated by incorporating them as random variables into the physical forward model of the diffractive network during the training process. One example of this has been demonstrated in previous work[36] where the destructive impact of the lateral and axial misalignments

of diffractive layers was mitigated by randomly misaligning the diffractive network during its training process. Following a similar strategy, the imperfect polarization extinction ratio (PER) of the polarizer arrays/seeds can also be included as part of our physical forward model using a modified form of the Jones matrices for linear polarizers. This modeling of imperfect PER of linear polarizers during the training phase can mitigate a potential performance degradation in the computational power of a polarization-multiplexed diffractive processor. Supporting this conclusion, Supplementary Note 3 and Supplementary Fig. SN3 report our mathematical analysis and simulation results for using imperfect linear polarizer arrays/seeds in our diffractive network designs. In the same Supplementary Note 3, we also quantified the overall PER of SimPA-based polarization-multiplexed diffractive designs, considering each diffractive network as a monolithic polarization optical element. Our analysis reveals that the SimPA-based 2-channel polarization-multiplexed diffractive design exhibits a very high PER of >51,000. In fact, such a high PER is expected since the SimPA mode is designed to *simultaneously* perform two different linear transformations using two orthogonal polarization states, and therefore undesired polarization cross-talk at the output field-of-view was penalized during the training phase, successfully leading to a high PER per diffractive network. For the SeqPA mode of operation, however, PER is not a meaningful figure-of-merit since only one orthogonal polarization state is read/measured at a given time due to the sequential access of each target transformation through the diffractive network; stated differently, the SeqPA mode of operation does not penalize the leakage of power into an orthogonal polarization state at the output as it does not impact at all the accuracy of each all-optical transformation that is sequentially performed.

In addition to performing multiple arbitrarily-selected linear transformations through polarization encoding, the presented framework can also be used for polarization-aware optical imaging and sensing tasks. Polarization-based optical imaging has been used in many biomedical applications, such as performing diagnoses of diseases, including gout[59,60,70], malaria infection[71], squamous cell carcinoma[72], and cerebral amyloid[73]. We believe that the presented polarization-multiplexed diffractive computing framework exhibits translational potential for some of these biomedical applications including e.g., the all-optical detection and classification of birefringent crystals in bodily fluids for diagnosing various forms of crystal arthropathy[74].

In conclusion, we introduced a diffractive network-based all-optical computing framework that can perform multiple complex-valued, arbitrary linear transformations using polarization multiplexing. This framework is very compact; for instance, the system depicted in Fig. 1 has a total length of only 20λ in depth, where λ is the illumination wavelength. Our results show that when the number of diffraction elements/neurons, N, in a given diffractive network design approaches $N_pN_iN_o$, a group of N_p arbitrarily-selected linear transforms can be all-optically computed at the output FOV of the network with negligible error. We believe that this polarization-multiplexed diffractive computing framework can be used to build all-optical, passive processors that can execute multiple inference tasks in parallel. We further envision that artificially engineered materials with polarization manipulation capabilities[75–79] can also be combined with advanced diffractive surface fabrication techniques (e.g., high-precision 3D additive manufacturing and photolithography) to allow the use of our diffractive computing framework in different parts of the electromagnetic spectrum.

Materials and methods

Forward model of the polarization-multiplexed diffractive optical network

Using Jones calculus[80], the complex-valued, polarization-multiplexed electrical field $\boldsymbol{E}$ at a spatial location (x_m, y_m, z_m) can be represented as:

$$\boldsymbol{E}(x_m, y_m, z_m) = \begin{bmatrix} E_x(x_m, y_m, z_m) \\ E_y(x_m, y_m, z_m) \end{bmatrix} \tag{1}$$

In our implementation, E_x and E_y are computed in parallel throughout the entire diffractive system. Since the trainable diffractive layers are *not* polarization-sensitive, the complex-valued modulation generated by these thin diffractive layers is the same for the two orthogonal polarization states. The diffractive layers are assumed to be thin optical modulation elements, where the mth feature on the kth diffractive layer at location (x_m, y_m, z_m) represents a complex-valued transmission coefficient, t^k, given by:

$$t^k(x_m, y_m, z_m) = a^k(x_m, y_m, z_m)\exp\left(j\phi^k(x_m, y_m, z_m)\right) \tag{2}$$

In Eq. 2, a and ϕ denote the amplitude and phase coefficients, respectively. The amplitude and phase coefficients of the diffractive neurons, a^k and ϕ^k ($k \in \{1, 2, \cdots, K\}$), are both trainable, with a permitted range of 0 to 1 and 0 to 2π, respectively. Before the training starts, a^k and ϕ^k are randomly initialized with a uniform ($\mathcal{U}$) distribution of $\mathcal{U}[0, 1]$and $\mathcal{U}[0, 2\pi)$, respectively. For a phase-only diffractive design $a^k = 1$. The size of each diffractive neuron on the transmissive layers and the width of the pixels of the input/output fields are both chosen as $\lambda/2$.

The diffractive layers are connected to each other by free-space wave propagation, which is modeled through the Rayleigh-Sommerfeld diffraction equation:[23,32]

$$w_m^k(x,y,z) = \frac{z-z_i}{r^2}\left(\frac{1}{2\pi r}+\frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right) \tag{3}$$

where $w_m^k(x,y,z,\lambda)$ is the complex-valued field on the mth neuron of the kth layer at (x, y, z) with a wavelength of λ, which can be viewed as a secondary wave generated from the source at (x_m, y_m, z_m); and $r = \sqrt{(x-x_m)^2+(y-y_m)^2+(z-z_m)^2}$ and $j=\sqrt{-1}$. For the kth layer ($k \geq 1$, treating the input plane as the 0th layer), the modulated optical field E_p^k at location (x_m, y_m, z_m) with a polarization state of p ($p \in \{\text{x}, \text{y}\}$) is given by:

$$E_p^k(x_m,y_m,z_m) = t^k(x_m,y_m,z_m)\cdot\sum_{n\in S}E_p^{k-1}(x_n,y_n,z_n) \cdot w_m^{k-1}(x_m,y_m,z_m) \tag{4}$$

where S denotes all the pixels on the previous diffractive layer. For all the diffractive networks trained in this paper, the axial distances $d_0, d_1, \ldots, d_K$ are all chosen as 4λ.

When modeling the polarizer elements in our diffractive system, we used Jones matrices to represent the modulation of the complex field brought by the input polarizer, output analyzer, or the polarizer array at location (x, y, z), the process of which can be written as:

$$\boldsymbol{E}_{\text{out}}(x,y,z) = \boldsymbol{J}_{\text{linear}}(x,y,z)\boldsymbol{E}_{\text{in}}(x,y,z) \tag{5}$$

where E_{in} and E_{out} are the vectors denoting the input and output complex field before and after the polarization modulation, each containing two orthogonal components along the x and y directions, i.e., $\boldsymbol{E}_{\text{out}}(x,y,z) = \begin{bmatrix} E_{\text{out,x}}(x,y,z) \\ E_{\text{out,y}}(x,y,z)\end{bmatrix}$ and $\boldsymbol{E}_{\text{in}}(x,y,z) = \begin{bmatrix} E_{\text{in,x}}(x,y,z) \\ E_{\text{in,y}}(x,y,z)\end{bmatrix}$. $\boldsymbol{J}_{\text{linear}}(x,y,z)$ represents the Jones matrix of a linear polarizer element, which is given by:

$$\boldsymbol{J}_{\text{linear}}(x,y,z) = \begin{bmatrix} \cos^2\theta(x,y,z) & \cos\theta(x,y,z)\sin\theta(x,y,z) \\ \sin\theta(x,y,z)\cos\theta(x,y,z) & \sin^2\theta(x,y,z)\end{bmatrix} \tag{6}$$

where $\theta(x,y,z)$ is the angle between the x-axis and the polarizing axis of the linear polarizer located at (x, y, z). For the non-trainable, pre-determined polarizer array that is composed of multiple square-shaped linear polarizers, we used in total 4 types of linear polarizer units with 4 different polarizing axis directions, $\theta = \{0, 0.25\pi, 0.5\pi, \text{and } 0.75\pi\}$. As illustrated in Fig. 1a, these 4 different types of linear polarizers are spatially binned to have a 2 × 2 period and repeated with 3 periods in each direction, extending into a square region. The side length of each linear polarizer array is 24λ. The residual space surrounding the polarizer array is filled with air, without any polarization modulation. For all the diffractive network designs presented in this paper, the axial distances (i.e., d_p, d_{p1} and d_{p2}) between the pre-determined polarizer arrays and the adjacent diffractive layers in front of them are all empirically chosen as 0; stated differently, each linear polarizer array is attached to the isotropic diffractive layer in front of it.

Preparation of the linear transformation datasets

In our diffractive network designs, the input and output FOVs have the same size of 8 × 8 pixels, i.e., $\boldsymbol{i}_c, \boldsymbol{o}_c \in \mathbb{C}^{8\times 8}$ ($c \in \{1,2,3,4\}$). The size of the transformation matrices is equal to 64 × 64, i.e., $\boldsymbol{A}_c \in \mathbb{C}^{64\times 64}$ ($c \in \{1,2,3,4\}$). The amplitude and phase components of the complex-valued transformation matrices $\boldsymbol{A}_c$ used in this paper were generated with a uniform (U) distribution of $U[0,1]$and $U[0,2\pi)$, respectively, using the pseudo-random number generation function *random.uniform()* built-in NumPy. Different random seeds were used to generate these transformation matrices to ensure they were uniquely different (see Fig. S1). Next, the amplitude and phase components of the input fields $\boldsymbol{i}_c$ ($c \in \{1,2,3,4\}$) were also randomly generated with a uniform (U) distribution of $U[0,1]$ and $U[0,2\pi)$, respectively. The ground truth (target) fields $\boldsymbol{o}_c$ ($c \in \{1,2,3,4\}$) were generated by calculating $\boldsymbol{o}_c = \boldsymbol{A}_c\boldsymbol{i}_c$. For each $\boldsymbol{A}_c$ ($c \in \{1,2,3,4\}$) we generated a total of 70,000 input/output complex fields to form a dataset, divided into three parts: training, validation, and testing, each containing 55,000, 5,000, and 10,000 complex-valued field pairs, respectively.

Training loss function

For training of our diffractive networks, we used the mean-squared-error (MSE) loss function, which is defined as:

$$\begin{aligned}\mathcal{L}_{\text{MSE},c} &= E\left[\frac{1}{N_o}\sum_{n=1}^{N_o}\left|\widehat{\boldsymbol{o}_1}[n]-\widehat{\boldsymbol{o}'_1}[n]\right|^2\right] \\ &= E\left[\frac{1}{N_o}\sum_{n=1}^{N_o}\left|\sigma_c\boldsymbol{o}_c[n]-\sigma'_c\boldsymbol{o}'_c[n]\right|^2\right]\end{aligned} \tag{7}$$

where $E[\cdot]$ denotes the average across the current batch, c stands for the cth polarization channel that is being accessed, and $[n]$ indexes the nth element of the vector. σ_c and σ'_c are the coefficients used to normalize the energy of the ground truth (target) field $\boldsymbol{o}_c$ and the diffractive

network output field $\boldsymbol{o}'_c$, respectively, which are given by:

$$\sigma_c = \frac{1}{\sqrt{\sum_{n=1}^{N_o} |\boldsymbol{o}_c[n]|^2}} \tag{8}$$

$$\sigma'_c = \frac{\sum_{n=1}^{N_o} \sigma_c \boldsymbol{o}_c[n] \boldsymbol{o}'^*_c[n]}{\sum_{n=1}^{N_o} |\boldsymbol{o}'_c[n]|^2} \tag{9}$$

During the training of the diffractive networks using the SeqPA mode, each polarization channel of the diffractive network is accessed and evaluated cyclically based on the order of the channel number. For instance, for the 2-channel polarization-multiplexed design illustrated in Fig. 1b, left, the access sequence during the training is set to be {①, ②, ①, ②, ...}; for the 4-channel polarization-multiplexed design illustrated in Fig. 6, the access sequence is {①, ②, ③, ④, ①, ②, ③, ④, ...}. During the access of a certain polarization channel, the diffractive network is fed with one batch of the training input/output complex fields corresponding to the transformation matrix assigned to this channel, and then trained based on the average loss across this batch. Thus, the loss function for training the diffractive designs through the cth polarization channel using the SeqPA mode, $\mathcal{L}_{\mathrm{Seq},c}$, can be simply written as:

$$\mathcal{L}_{\mathrm{Seq},c} = \mathcal{L}_{\mathrm{MSE},c} \tag{10}$$

During the training of the diffractive networks using the SimPA mode, as illustrated in Fig. 1b, right, all the polarization channels of the diffractive network are accessed simultaneously, and the training data are fed into the channels at the same time. For this SimPA mode, the diffractive network is trained based on the loss averaged across the different polarization channels and complex-valued fields in the current batch, where the loss function $\mathcal{L}_{\mathrm{Sim}}$ can be written as:

$$\mathcal{L}_{\mathrm{Sim}} = \frac{1}{N_p} \sum_{c=1}^{N_p} \mathcal{L}_{\mathrm{MSE},c} \tag{11}$$

Performance metrics used for the quantification of all-optical transformation errors

To quantitatively evaluate the transformation results of the polarization-multiplexed diffractive networks, four performance metrics were calculated per polarization channel of the diffractive designs using the testing dataset: (1) the normalized transformation mean-squared error ($\mathrm{MSE}_{\mathrm{Transformation}}$), (2) the cosine similarity (*CosSim*) between the all-optical transforms and the target transforms, (3) the normalized mean-squared error between the diffractive network output fields and their ground truth ($\mathrm{MSE}_{\mathrm{Output}}$), and (4) the output diffraction efficiency (η). The transformation error for the cth polarization channel of the diffractive network, $\mathrm{MSE}_{\mathrm{Transformation},c}$, is defined as:

$$\begin{aligned} \mathrm{MSE}_{\mathrm{Transformation},c} &= \frac{1}{N_i N_o} \sum_{n=1}^{N_i N_o} \left| \boldsymbol{a}_c[n] - m_c \boldsymbol{a}'_c[n] \right|^2 \\ &= \frac{1}{N_i N_o} \sum_{n=1}^{N_i N_o} \left| \boldsymbol{a}_c[n] - \widehat{\boldsymbol{a}}'_c[n] \right|^2 \end{aligned} \tag{12}$$

where $\boldsymbol{a}_c$ is the vectorized version of the ground truth transformation matrix assigned to the cth polarization channel $\boldsymbol{A}_c$, i.e., $\boldsymbol{a}_c = \mathrm{vec}(\boldsymbol{A}_c)$. $\boldsymbol{a}'_c$ are the vectorized version of $\boldsymbol{A}'_c$, which is the all-optical transformation matrix computed using the optimized diffractive transmission coefficients. m_c is a scalar normalization coefficient used to eliminate the effect of diffraction-efficiency related scaling mismatch between $\boldsymbol{A}_c$ and $\boldsymbol{A}'_c$, i.e.,

$$m_c = \frac{\sum_{n=1}^{N_i N_o} \boldsymbol{a}_c[n] \boldsymbol{a}'^*_c[n]}{\sum_{n=1}^{N_i N_o} |\boldsymbol{a}'_c[n]|^2} \tag{13}$$

The cosine similarity between the all-optical transform and their target transform for the cth polarization channel, $CosSim_c$, is defined as:

$$CosSim_c = \frac{\left| \boldsymbol{a}_c^H \widehat{\boldsymbol{a}}'_c \right|}{\sqrt{\sum_{n=1}^{N_i N_o} |\boldsymbol{a}_c[n]|^2} \sqrt{\sum_{n=1}^{N_i N_o} \left| \widehat{\boldsymbol{a}}'_c[n] \right|^2}} \tag{14}$$

The normalized mean-squared error between the diffractive network outputs and their ground truth for the cth polarization channel, $\mathrm{MSE}_{Output,c}$, is defined using the same formula as in Eq. 7 (the loss function used during the training process), except for that $E[\cdot]$ is calculated across the entire testing set.

The mean diffraction efficiency η_c for the cth polarization channel of the diffractive system is defined as:

$$\eta_c = E\left[\frac{\sum_{n=1}^{N_o} |\boldsymbol{o}'_c[n]|^2}{\sum_{n=1}^{N_i} |\boldsymbol{i}_c[n]|^2} \right] \tag{15}$$

Training-related details

All the diffractive optical networks used in this work were simulated and trained using Python (v3.8.11) and TensorFlow (v2.6.0, Google Inc.). We selected Adam optimizer[81] for training all the models, and its parameters were taken as the default values in TensorFlow and kept identical in each model. The batch size and learning rate were set as 8 and 0.001, respectively. The training of the diffractive network models using the SimPA mode was performed with 50 epochs. For training the diffractive

models using the SeqPA mode, the 2-channel and 4-channel polarization-multiplexed designs were trained for 100 and 200 epochs, respectively, so that equivalently 50 epochs are dedicated for training each polarization channel of these designs. The best models were selected based on the MSE loss calculated on the validation dataset. For the training of our diffractive models, we used a desktop computer with a GeForce GTX 1080Ti graphical processing unit (GPU, NVidia Inc.) and Intel® Core^{TM} i7-8700 central processing unit (CPU, Intel Inc.) and 64 GB of RAM, running Windows 10 operating system (Microsoft Inc.). The typical time to train a diffractive network model using the SeqPA mode with 2 and 4 polarization channels is ~7 and ~14 h, respectively. The training time for a diffractive model using the SimPA mode with 2 polarization channels is ~4 h.

Acknowledgements
The Ozcan Research Group at UCLA acknowledges the support of US Air Force Office of Scientific Research (AFOSR), Materials with Extreme Properties Program funding (FA9550-21-1-0324). The authors also thank Dr. Jingtian Hu and Tairan Liu for their useful discussion during the preparation of the manuscript.

Author details
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA. [2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute (CNSI), University of California, Los Angeles, CA 90095, USA

Author contributions
A.O. conceived the research and initiated the project, J.L. and Y.H. conducted the numerical experiments, J.L. and Y.H. processed the data. O.K. and D.M. helped with the all-optical transformation performance metrics. All the authors contributed to the preparation of the manuscript. A.O. supervised the research.

Data availability
The deep learning models reported in this work used standard libraries and scripts that are publicly available in TensorFlow. All the data and methods needed to evaluate the conclusions of this work are presented in the main text and Supplementary Information. Additional data can be requested from the corresponding author.

Conflict of interest
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41377-022-00849-x.

Received: 14 March 2022 Revised: 10 May 2022 Accepted: 11 May 2022
Published online: 26 May 2022

References

1. Solli, D. R. & Jalali, B. Analog optical computing. *Nat. Photonics* **9**, 704–706 (2015).
2. Athale, R. & Psaltis, D. Optical computing: past and future. *Opt. Photonics News* **27**, 32–39 (2016).
3. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
4. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).
5. Zhou, H. et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci. Appl.* **11**, 30 (2022).
6. Cutrona, L., Leith, E., Palermo, C. & Porcello, L. Optical data processing and filtering systems. *IRE Trans. Inf. Theory* **6**, 386–400 (1960).
7. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci.* **79**, 2554–2558 (1982).
8. Psaltis, D. & Farhat, N. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Opt. Lett.* **10**, 98–100 (1985).
9. Farhat, N. H., Psaltis, D., Prata, A. & Paek, E. Optical implementation of the Hopfield model. *Appl. Opt.* **24**, 1469–1475 (1985).
10. Wagner, K. & Psaltis, D. Multilayer optical learning networks. *Appl. Opt.* **26**, 5061–5076 (1987).
11. Psaltis, D., Brady, D., Gu, X.-G. & Lin, S. Holography in artificial neural networks. *Nature* **343**, 325 (1990).
12. Vandoorne, K., Dambre, J., Verstraeten, D., Schrauwen, B. & Bienstman, P. Parallel reservoir computing using optical amplifiers. *IEEE Trans. Neural Netw.* **22**, 1469–1481 (2011).
13. Silva, A. et al. Performing mathematical operations with metamaterials. *Science* **343**, 160–163 (2014).
14. Vandoorne, K. et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **5**, 3541 (2014).
15. Carolan, J. et al. Universal linear optics. *Science* **349**, 711–716 (2015).
16. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
17. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
18. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208 (2019).
19. Miscuglio, M. & Sorger, V. J. Photonic tensor cores for machine learning. *Appl Phys. Rev.* **7**, 031404 (2020).
20. Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).
21. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
22. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
23. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
24. Bueno, J. et al. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **5**, 756–760 (2018).
25. Chang J., Sitzmann V., Dun X., Heidrich W., Wetzstein G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci Rep.* **8**, https://doi.org/10.1038/s41598-018-30619-y (2018).
26. Zuo, Y. et al. All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).
27. Estakhri, N. M., Edwards, B. & Engheta, N. Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).
28. Wright, L. G. et al. Deep physical neural networks trained with back-propagation. *Nature* **601**, 549–555 (2022).
29. Hughes, T. W., Williamson, I. A. D., Minkov, M. & Fan, S. Wave physics as an analog recurrent neural network. *Sci. Adv.* **5**, eaay6946 (2019).
30. Dong, J., Rafayelyan, M., Krzakala, F. & Gigan, S. Optical reservoir computing using multiple light scattering for chaotic systems prediction. *IEEE J. Sel. Top. Quantum Electron* **26**, 1–12 (2020).
31. Teğin, U., Yıldırım, M., Oğuz, İ., Moser, C. & Psaltis, D. Scalable optical learning operator. *Nat. Comput Sci.* **1**, 542–549 (2021).
32. Mengu, D., Luo, Y., Rivenson, Y. & Ozcan, A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron* **26**, 1–14 (2020).
33. Li, J., Mengu, D., Luo, Y., Rivenson, Y. & Ozcan, A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv. Photonics* **1**, 046001 (2019).
34. Yan, T. et al. Fourier-space diffractive deep neural network. *Phys. Rev. Lett.* **123**, 023901 (2019).
35. Mengu D., Rivenson Y., Ozcan A. Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics*. https://doi.org/10.1021/acsphotonics.0c01583 (2020).

36. Mengu, D. et al. Misalignment resilient diffractive optical networks. *Nanophotonics* **9**, 4207–4219 (2020).
37. Rahman, M. S. S., Li, J., Mengu, D., Rivenson, Y. & Ozcan, A. Ensemble learning of diffractive optical networks. *Light Sci. Appl* **10**, 14 (2021).
38. Li, J. et al. Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).
39. Mengu D., Veli M., Rivenson Y. & Ozcan A. Classification and reconstruction of spatially overlapping phase images using diffractive optical networks. *Sci. Rep.* **12**, 8446 (2022).
40. Kulce, O., Mengu, D., Rivenson, Y. & Ozcan, A. All-optical information-processing capacity of diffractive surfaces. *Light Sci. Appl* **10**, 25 (2021).
41. Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
42. Chen, H. et al. Diffractive deep neural networks at visible wavelengths. *Engineering* **7**, 1483–1491 (2021).
43. Liu C., et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat Electron* 1–10 (2022).
44. Luo, Y. et al. Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight* **2**, 4 (2022).
45. Mengu, D. & Ozcan, A. All-optical phase recovery: diffractive computing for quantitative phase imaging. *Adv. Opt. Mat.* 2200281 (2022).
46. Qian, C. et al. Performing optical logic operations by a diffractive neural network. *Light Sci. Appl.* **9**, 1–7 (2020).
47. Wang, P. et al. Orbital angular momentum mode logical operation using optical diffractive neural network. *Photonics Res.* **9**, 2116–2124 (2021).
48. Luo, Y., Mengu, D. & Ozcan, A. Cascadable all-optical NAND gates using diffractive networks. *Sci. Rep.* **12**, 7121 (2022).
49. Luo, Y. et al. Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci. Appl.* **8**, 1–14. (2019).
50. Veli, M. et al. Terahertz pulse shaping using diffractive surfaces. *Nat. Commun.* **12**, 37 (2021).
51. Huang, Z. et al. All-optical signal processing of vortex beams with diffractive deep neural networks. *Phys. Rev. Appl* **15**, 014037 (2021).
52. Wang, P. et al. Diffractive deep neural network for optical orbital angular momentum multiplexing and demultiplexing. *IEEE J. Sel. Top. Quantum Electron* **28**, 1–11 (2022).
53. Han, Y. & Li, G. Coherent optical communication using polarization multiple-input-multiple-output. *Opt. Express* **13**, 7527–7534 (2005).
54. Chen, Z.-Y. et al. Use of polarization freedom beyond polarization-division multiplexing to support high-speed and spectral-efficient data transmission. *Light Sci. Appl.* **6**, e16207–e16207 (2017).
55. Oshima, N., Hashimoto, K., Suzuki, S. & Asada, M. Terahertz wireless data transmission with frequency and polarization division multiplexing using resonant-tunneling-diode oscillators. *IEEE Trans. Terahertz Sci. Technol.* **7**, 593–598 (2017).
56. Kadambi, A., Taamazyan, V., Shi, B., Raskar, R. Polarized 3D: High-quality depth sensing with polarization cues. In *2015 IEEE International Conference on Computer Vision (ICCV)* 3370–3378 (IEEE: Santiago, Chile, 2015).
57. Dongfeng, S. et al. Polarization-multiplexing ghost imaging. *Opt. Lasers Eng.* **102**, 100–105 (2018).
58. Liu, F. et al. Deeply seeing through highly turbid water by active polarization imaging. *Opt. Lett.* **43**, 4903–4906 (2018).
59. Liu, T. et al. Deep learning-based holographic polarization microscopy. *ACS Photonics* **7**, 3023–3034 (2020).
60. Bai, B. et al. Pathological crystal imaging with single-shot computational polarized light microscopy. *J. Biophotonics* **13**, e201960036 (2020).
61. Deschaintre, V., Lin, Y., Ghosh, A. Deep polarization imaging for 3D shape and SVBRDF acquisition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 15562–15571 (IEEE, Nashville, TN, USA, 2021).
62. Wolff, L. B. Polarization-based material classification from specular reflection. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 1059–1071 (1990).
63. Zhan, Z. et al. Optical polarization–based seismic and water wave sensing on transoceanic cables. *Science* **371**, 931–936 (2021).
64. Mecozzi, A. et al. Polarization sensing using submarine optical cables. *Optica* **8**, 788–795 (2021).
65. Zhou, S., Campbell, S., Yeh, P. & Liu, H.-K. Two-stage modified signed-digit optical computing by spatial data encoding and polarization multiplexing. *Appl Opt.* **34**, 793–802 (1995).
66. Tan, G., Zhan, T., Lee, Y.-H., Xiong, J. & Wu, S.-T. Polarization-multiplexed multiplane display. *Opt. Lett.* **43**, 5651–5654 (2018).
67. Zhan, T. et al. Improving near-eye display resolution by polarization multiplexing. *Opt. Express* **27**, 15327–15334 (2019).
68. Evangelides, S. G., Mollenauer, L. F., Gordon, J. P. & Bergano, N. S. Polarization multiplexing with solitons. *J. Light Technol.* **10**, 28–35 (1992).
69. Kulce, O., Mengu, D., Rivenson, Y. & Ozcan, A. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. *Light Sci. Appl* **10**, 196 (2021).
70. Kohn, N. N., Hughes, R. E., McCarty, D. J. & Faires, J. S. The significance of calcium phosphate crystals in the synovial fluid of arthritic patients: the 'pseudogout syndrome'. II. Identification of crystals. *Ann. Intern Med.* **56**, 738–745 (1962).
71. Lawrence, C. & Olson, J. A. Birefringent hemozoin identifies malaria. *Am. J. Clin. Pathol.* **86**, 360–363 (1986).
72. Arun Gopinathan, P. et al. Study of collagen birefringence in different grades of oral squamous cell carcinoma using picrosirius red and polarized light microscopy. *Scientifica* **2015**, e802980 (2015).
73. Jin, L.-W. et al. Imaging linear birefringence and dichroism in cerebral amyloid pathologies. *Proc. Natl Acad. Sci. USA* **100**, 15294–15298 (2003).
74. Zhang, Y. et al. Wide-field imaging of birefringent synovial fluid crystals using lens-free polarized microscopy for gout diagnosis. *Sci. Rep.* **6**, 28793 (2016).
75. Solli, D. R., McCormick, C. F., Chiao, R. Y. & Hickmann, J. M. Photonic crystal polarizers and polarizing beam splitters. *J. Appl Phys.* **93**, 9429–9431 (2003).
76. Liu, T., Zakharian, A. R., Fallahi, M., Moloney, J. V. & Mansuripur, M. Design of a compact photonic-crystal-based polarizing beam splitter. *IEEE Photonics Technol. Lett.* **17**, 1435–1437 (2005).
77. Zabelin, V. et al. Self-collimating photonic crystal polarization beam splitter. *Opt. Lett.* **32**, 530–532 (2007).
78. Hao, J. et al. Optical metamaterial for polarization control. *Phys. Rev. A* **80**, 023807 (2009).
79. Cong, L. et al. A perfect metamaterial polarization rotator. *Appl Phys. Lett.* **103**, 171107 (2013).
80. Jones, R. C. A new calculus for the treatment of optical systemsI. Description and discussion of the calculus. *JOSA* **31**, 488–493 (1941).
81. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc. 3rd International Conference on Learning Representations* (ICLR, 2014).