Article

Inverse-designed low-index-contrast structures on a silicon photonics platform for vector–matrix multiplication

Vahid Nikkhah [1], **Ali Pirmoradi** [1], **Farshid Ashtiani** [1,2], **Brian Edwards** [1], **Firooz Aflatouni** [1] & **Nader Engheta** [1] ✉

Inverse-designed silicon photonic metastructures offer an efficient platform to perform analogue computations with electromagnetic waves. However, due to computational difficulties, scaling up these metastructures to handle a large number of data channels is not trivial. Furthermore, a typical inverse-design procedure is limited to a small computational domain and therefore tends to employ resonant features to achieve its objectives. This results in structures that are narrow-bandwidth and highly sensitive to fabrication errors. Here we employ a two-dimensional (2D) inverse-design method based on the effective index approximation with a low-index contrast constraint. This results in compact amorphous lens systems that are generally feed-forward and low-resonance. We designed and experimentally demonstrated a vector–matrix product for a 2 × 2 matrix and a 3 × 3 matrix. We also designed a 10 × 10 matrix using the proposed 2D computational method. These examples demonstrate that these techniques have the potential to enable larger-scale wave-based analogue computing platforms.

Photonic structures that can perform mathematical operations and solve equations are becoming increasingly popular due to the resurgence of optical analogue computing with the promise of low-power, high-speed, parallel computations enabled by light[1–11]. Specifically, designing a photonic structure for performing robust linear operations has central importance in a range of applications, including optical imaging[12], optical signal processing[13], optical analogue computing and equation solving[4], and linear optical quantum computing[14]. Equipped with proper reconfigurable/tuning elements, linear optical structures, along with nonlinear elements[15], can also be applicable in optical neuromorphic architectures. These architectures can benefit from vector–matrix multiplication rendered as optical analogue computing modules, with improved energy consumption per arithmetic operation, while also achieving substantially higher speeds[16–20]. Physically defining a $N \times N$ matrix for vector–matrix multiplication requires the realization of N^2 different objectives within a single device.

Performing vector–matrix multiplication with electromagnetic waves can be formulated either as free-space Fourier optics in k-space[1,3,21,22] or as spatial modes as the basis for performing in real space[4,23]. This can be done in one of two ways. The n modes can be expanded into n^2 modes, utilizing an additional dimension. These are directly acted upon and then summed[24]. Within a three-dimensional (3D) device, such an operation can be achieved using metasurfaces defined under the Born approximation considering only their transmission coefficients[24,25]. Alternatively, the n modes can be operated through a series of mixing operations in the form of a mesh architecture, generally 'forward' only[26–29]. Finally, the modes may be operated by an inverse-designed structure, potentially containing internal resonances[1,4].

Using a combination of topological and shape optimizations, inverse design[1,4] is particularly well suited to the design of such structures. Specifically, density-based topology optimization is a technique in which the material within a design region is suitably discretized into

[1]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. [2]Nokia Bell Labs, Murray Hill, NJ, USA. ✉e-mail: engheta@seas.upenn.edu
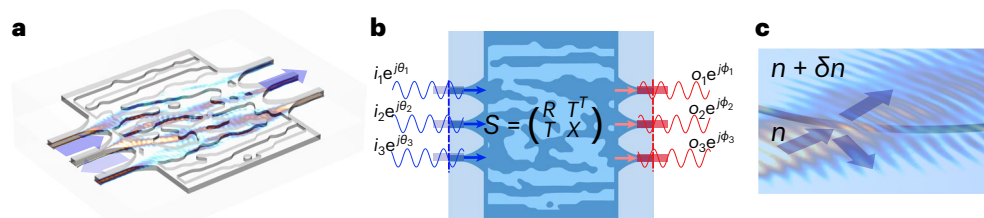
Fig. 1 | Schematic of our inverse-designed structures in a SiPh platform. a, Vector–matrix multiplication consisting of the transmissive part of the scattering matrix, *S*. **b**, Generic structure, optimized to achieve this transmissive part. **c**, Feed-forward due to the low-index contrast, leading to a small reflection and primarily refractive behaviour.

a large number of free parameters without any preconceived idea of the nature of the structure[30]. These parameters are then optimized utilizing a suitable mathematical technique such as the adjoint method, which can compute the gradient utilizing comparatively few simple forward simulations[31,32]. Following topological optimization, the material can then be binarized and vectorized into highly parameterized shapes, which are similarly optimized using the adjoint method.

Although a structure in a silicon photonics (SiPh) platform is planar in nature, the fields do not possess 2D symmetry, forcing designers to often utilize 3D simulations[33]. Due to the computational difficulties of 3D simulations and the potentially large number of iterations needed to reach optimal performance, inverse design in 3D is limited to optically small structures (that is, tens of cubic wavelengths). This effectively limits the number of information channels and therefore the matrix size that may be considered. There is thus a need for an efficient computational platform that reduces the computational costs of full-wave 3D simulations in SiPh.

A 'reduced-order' approximation for photonic structures that can perform analogue computation between a large number of input and output information channels is vital. In this Article we employ a propagation-based 2D effective index approximation (p2DEIA)[34–37] for 3D planar structures that can greatly reduce the computational effort. By designing SiPh structures with relatively small variations in silicon thickness, the effective index difference between these regions will be small. This low-index contrast reduces out-of-plane scattering, increases the accuracy of the p2DEIA approximation, and reduces reflections (that is, it reduces resonances and therefore increases the bandwidth). We validate the inverse-design method based on the low-index-contrast p2DEIA by designing and fabricating SiPh metastructures that perform vector–matrix multiplication.

Results

Problem description

In this Article we perform vector–matrix multiplication for a given matrix *M* on a SiPh platform using the complex amplitude (that is, magnitude and phase) of a series of spatial modes to represent our vectors. For that purpose, an array of single-mode silicon waveguides is considered as the input channels and a second array represents the output channels. Both the input and output waveguides are connected to a design region in which the silicon thickness is parameterized, varying between two values (Fig. 1a). In general, such a structure can be fully described using a complex scattering matrix, in which the incoming spatial modes on any waveguide (designated input or output) will map to all others (Fig. 1b). The waves injected through the designated input ports can be mathematically described as a complex-valued vector. Despite being secondary to the design, an incoming vector from the designated output ports can be similarly defined. By representing the incoming waves as two concatenated $N \times 1$ vectors, the scattering matrix ($2N \times 2N$) can be defined using four block matrices (each $N \times N$). As we are interested in the propagation from the input to the output ports as the physical process for realizing a vector–matrix product,

we are primarily concerned with the transmission block matrix ($T = M$). Additionally, we would like to minimize the reflections back towards the designated input ports (that is, $R = 0$). Our goal, therefore, is to compute a spatially varying silicon thickness distribution within the 2D design region, which achieves these objectives.

Effective index approximation

It is known that the simulation and design of a 3D structure is a computationally costly process, especially in any iterative design method. The confined wave within the silicon slab propagates through the in-plane geometry in the form of a guided slab mode. The effective index of a slab mode depends on the thickness of the slab and the refractive index of the substrate, silicon core and cladding. For example, a 220-nm-thick silicon layer with $n_{si} = 3.48$ immersed in SiO_2 with $n_{SiO_2} = 1.44$ supports a fundamental slab mode with $n_{eff, 220 nm} = 2.86$ at $\lambda_0 = 1.525$ μm. The effective index decreases as the thickness of the silicon core decreases. For example, for a similarly defined 150-nm-thick silicon layer, the effective index of the fundamental slab mode reduces to $n_{eff, 150 nm} = 2.56$. Within certain constraints, one can construct a 2D computational model that approximates the in-plane wave propagation (that is, the guided slab modes) through a 3D planar structure by using the effective indices of the slab modes to represent the various silicon core thicknesses. For instance, a region in the 2D model with a refractive index of $n = n_{eff, 220 nm} = 2.86$ fully and accurately models the in-plane propagation of the fundamental slab mode in the 3D structure with a silicon thickness of 220 nm. This model is here referred to as the p2DEIA.

Suppose we have two regions, one with a silicon layer thickness of 220 nm and the other that has been uniformly etched to achieve a smaller thickness (Fig. 1c). When the fundamental slab mode hits the interface between these two regions, various phenomena can occur. The p2DEIA model will effectively capture the angle of the reflection and transmission of the fundamental slab modes through Snell's law (Fig. 1c). However, a 3D height discontinuity will experience other effects. For instance, some portion of the wave is scattered away from the silicon slab and into the SiO_2 cladding. Additionally, excitation of higher-order slab modes inside the silicon core may take place. With these additional scattering channels, a 3D height discontinuity may only be approximated by a 2D model. The accuracy of the 2D model depends on the strength of the coupling to these other scattering channels. When the silicon thicknesses are similar, the modal overlap between the fundamental slab modes is very large, and thus all other couplings must be small (Supplementary Fig. 1). Therefore, for shallow-enough etch depths, one can assume that the coupling to the unmodelled open channels is negligible and safely use the approximate p2DEIA method with good accuracy, even at interfaces.

The p2DEIA method based on the index of the fundamental slab mode defined previously accurately captures the propagation within a uniformly etched region, angle of refraction and angle of reflection. However, it may not capture well the magnitude and phase of the reflection and transmission at an interface, compared to other

models. For example, in models based on variational methods, one chooses a thickness where one expects most of the energy of the wave to reside[33]. The effective index of this region is calculated similarly to above and accurately models the propagation. However, all other effective indices are defined in reference to this chosen region and are designed to accurately model the interface behaviour, at the expense of accurately modelling propagation in these other regions[33]. These variational methods may be preferred for waveguiding structures based on total internal reflection (TIR) and grating couplers, where accurate modelling of the interfaces between regions is crucial. However, they will be less accurate when substantial energy is propagating across etched regions for extended distances, where phase error due to compromised effective indices will accumulate. We will refer to this as the interface-based effective index approximation (i2DEIA).

As mentioned, the accuracy of the p2DEIA requires that the relative difference in silicon thickness across an interface is not large, and we will limit ourselves to designs in which this statement is true. As a result, the effective indices associated with each region will have low-index contrast, rendering the interface reflection magnitude small and transmission magnitude large. Therefore, one can expect that the wave mostly travels forward from input to output ports, refracting at interfaces as it moves from region to region. The lack of substantial reflections will result in a low-resonance feed-forward optical system that behaves similarly to a series of irregular-shaped lenses. This type of system will naturally present broadband optical responses and will be less sensitive to small perturbations in the shape of the interfaces, as long as the angles of the interfaces are generally maintained. These are useful features from an experimental point of view, as a metastructure that is robust against fabrication errors such as 'over etching' is highly desired.

As for the design degrees of freedom (DOFs), the low-index-contrast interfaces will provide only weak diffraction and will offer only weak control over wave manipulation compared to highly contrasted interfaces. More precisely, the DOFs of design depend on the in-plane area of the design domain and the effective index contrast (dn) between etched and non-etched regions. Within the low-contrast regime where the out-of-plane scattering is much smaller than the in-plane scattering, the larger the (dn) becomes, the more design DOFs are provided by the etched interface due to the stronger in-plane scattering. However, as the (dn) increases further, there is a point where the in-plane diffraction into the etched regions is accompanied by substantial out-of-plane scattering. At that point, the in-plane diffraction channels become lossy and therefore less useful for in-plane wave manipulation. Therefore, increasing the effective index contrast beyond some point removes the in-plane diffraction channels from the design DOFs.

Effects of passivity on geometry definition

Any photonic structure that is enclosed by a boundary S can be described as a unitary scattering matrix, provided that it exhibits only lossless propagation. The modes on such a boundary can be enumerated. Some of these modes will be designated input (I), output (O) and the remainder will be considered simply 'absorbing' (A). In all cases, any outgoing waves on these modes (I_{out}, O_{out}, A_{out}) will not return to the system. In addition to block matrices T and R, the addition of the vector **A** allows us to conceptualize other block matrices as well (χ_1, and so on).

$$\begin{pmatrix} I_{out} \\ O_{out} \\ A_{out} \end{pmatrix} = S \begin{pmatrix} I_{in} \\ O_{in} \\ A_{in} \end{pmatrix} \quad S = \begin{pmatrix} R & T^T & \chi_1 \\ T & X & \chi_2 \\ \chi_3 & \chi_4 & \chi_5 \end{pmatrix} \quad (1)$$

These absorbing modes will never be illuminated (that is $A_{in} = 0$), and the precise shape of these modes will not be characterized. However, these modes do affect the system.

Because the scattering matrix is unitary, all singular values of S are 1. As a block matrix within a unitary matrix, the singular values of T must be less than 1. This condition is known as passivity. When the number of unconstrained absorbing modes is very large, then S being unitary ceases to be an important constraint, and the condition of passivity on T is sufficient. On the other hand, if the number of absorbing modes is small (for example, 1), there is no guarantee that even a scattering matrix defined with a passive T, $R = 0$ and unconstrained X belongs to the space of unitary matrices. Therefore, within this Article, we assume that the unconstrained A represents a sufficiently large number of absorbing modes. In that case, we may use the passivity condition to select the target transmission matrices. Within the subspace described by the passivity constraints on the transmission matrix, we arbitrarily choose the following matrices for a 2 × 2 and 3 × 3 optical network:

$$T_{2\times2} = \begin{pmatrix} 0.43 + j0.43 & -0.47 + j0.12 \\ 0.43 + j0.22 & 0.51 - j0.33 \end{pmatrix} \quad R_{2\times2} = 0 \quad (2)$$

$$T_{3\times3} = \begin{pmatrix} 0.49 + j0.24 & -0.53 + j0.44 & 0.44 - j0.16 \\ 0.60 + j0.41 & 0.20 - j0.46 & -0.32 - j0.36 \\ 0.41 + j0.11 & 0.5 + j0.15 & 0.20 + j0.71 \end{pmatrix} \quad R_{3\times3} = 0$$

$$(3)$$

Metastructure configuration and inverse design

The proposed optical metastructures for realizing the target transmission matrices are designed based on a silicon photonics platform at $\lambda_0 = 1.525$ μm. The 2D schematics of the optimized 3 × 3 and 2 × 2 structures using the p2DEIA model are shown in Fig. 2a,d, respectively. The width of each silicon waveguide is 500 nm with a thickness of 220 nm. For the 2 × 2 case, the design region, which is distinguished by the domain inside the dashed black boundary in Fig. 2a, is a rectangular domain of width 11 μm and length 10.3 μm. For the 3 × 3 case, it is an ellipse with a width of 34 μm and length 23 μm (Fig. 2d). In both structures, the absorbers' boundaries, illustrated by dashed red lines, span many wavelengths across the sides, providing a large enough number of absorbing modes A required as part of the design process.

The optimization problem is mathematically described by

$$\min_{n(x,y)} \| T - T_{targ} \|_F + \| R - 0 \|_F \quad (4)$$

$$\text{s.t.} \quad n_{eff, 150\,nm} \le n(x, y) \le n_{eff, 220\,nm}$$

where $\|\cdot\|_F$ is the Frobenius norm. The objective scalar defined for inverse design is the 'distance' between the complex-valued target transmission matrix and the transmission matrix that is realized by the current structure in each iteration. A similar expression for minimizing reflections is also added to the objective scalar. Using p2DEIA, the design region is parameterized by an effective index distribution that is bound between $n_{eff, 150\,nm}$ and $n_{eff, 220\,nm}$. By employing the inverse-design method in COMSOL Multiphysics[38] using the density-based topology optimization based on the adjoint method, the effective index distribution within the p2DEIA model is optimized to minimize the scalar objective.

During the inverse-design process, the effective index distribution is encouraged to the extremum values of $n_{eff, 150\,nm}$ and $n_{eff, 220\,nm}$ using a sigmoidal projection function that becomes successively steeper. The optimized distribution is then vectorized to define two classes of domain associated with $n_{eff, 220\,nm}$ and $n_{eff, 150\,nm}$, shown as the highlighted and complementary unhighlighted regions, respectively, within the design regions in Fig. 2a,d.
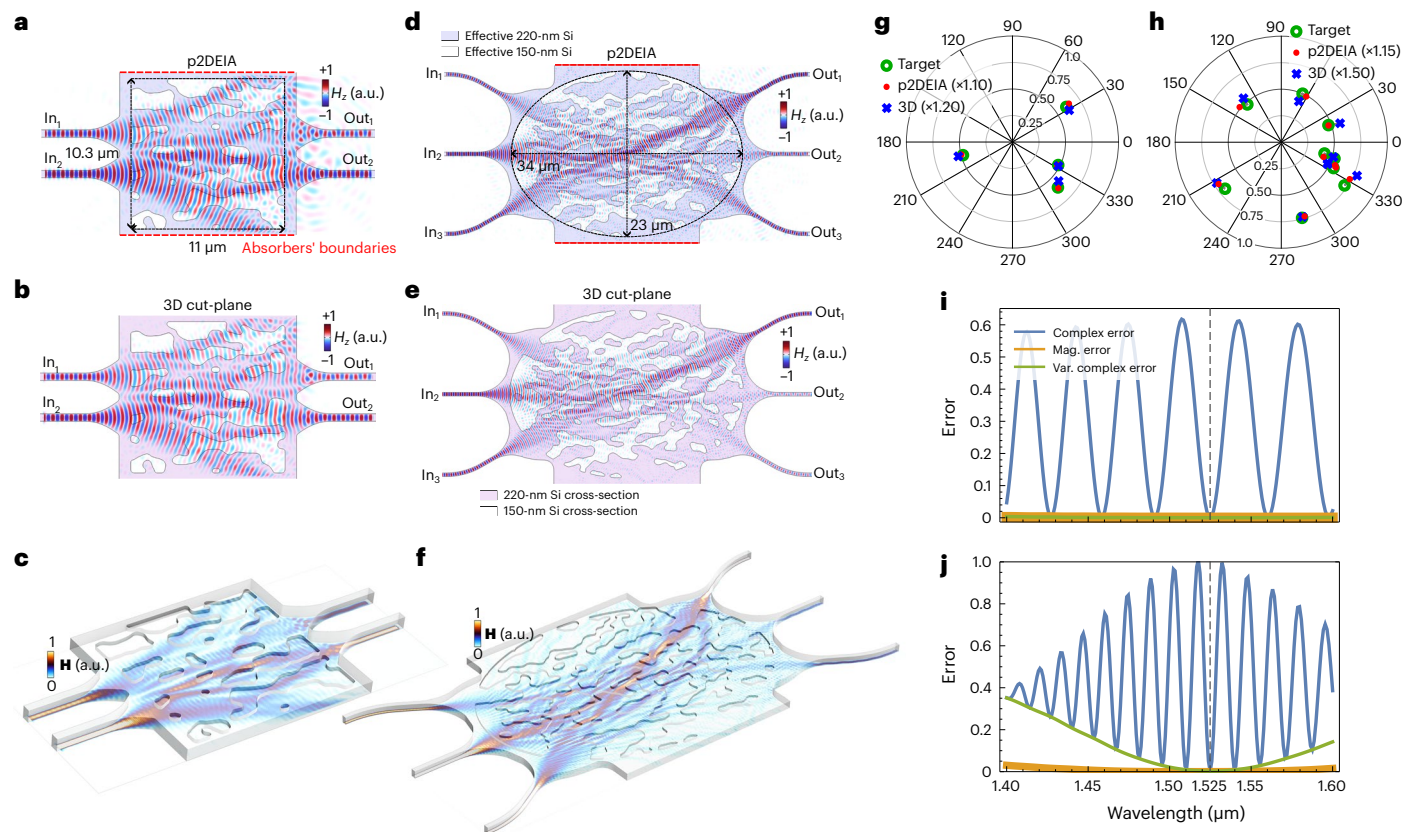
Fig. 2 | Inverse-designed metastructures based on silicon photonics for performing 2 × 2 and 3 × 3 vector–matrix multiplications. The designs and simulations are carried out at free-space wavelengths of $\lambda_0 = 1.525$ μm. **a,d**, Optimized structures in p2DEIA and time snapshots of out-of-plane magnetic-field distributions for arbitrary input vector excitations for 2 × 2 (**a**) and 3 × 3 (**d**) examples. **b,e**, Time snapshots of the magnetic-field distributions on cut-planes through silicon in the full-wave 3D-rendered structures for the same excitations as in p2DEIA. Here we choose a vertical view for easy comparison to the results in the p2DEIA case (**a,d**). **c,f**, Absolute values of the magnetic-field distributions through the silicon in the 3D view of the designed structures (**a,d**). **g,h**, Target transmission matrix elements (green circles), p2DEIA values (after application of the proper adjustment factors, shown in parentheses) (red circles) and 3D values (after application of the proper adjustment factors, shown in parentheses) (blue circles) for 2 × 2 (**g**) and 3 × 3 (**h**) structures. **i,j**, Normalized Frobenius norm of the error squared (that is, $\| T - T_{targ} \|_F^2 / 4N$) between the target and 3D transmission values versus wavelength for 2 × 2 (**i**) and 3 × 3 (**j**) structures.

These domains ultimately define the corresponding thicknesses in the 3D structure both for simulation and fabrication, as shown in Fig. 2c,f.

To assess the advantages of utilizing p2DEIA instead of the full-wave 3D model, we point out that there are two primary reasons for preferring p2DEIA over full-wave 3D simulations. First, in p2DEIA, only a scalar field for the out-of-plane magnetic-field component is solved, whereas in full-wave 3D, all three vector components must be taken into account, that is, $\{E_x, E_y, E_z\}$, resulting in a greater number of DOFs to be solved. Second, in 3D simulations, the electromagnetic fields along the out-of-plane dimension must also be accurately estimated, which requires additional DOFs compared to the p2DEIA, which merely discretizes the in-plane geometry. As a result, in the 3D model, in addition to discretizing the fields over the silicon thickness, it is essential to ensure that the perfectly matched layer boundaries perpendicular to the out-of-plane direction are positioned at least one wavelength (λ_0) away from each side of the slab to prevent unwanted interaction with the near-fields along the etched interfaces. Therefore, a cladding height of at least $2\lambda_0$ (where λ_0 is the wavelength in SiO$_2$) is necessary. The above-mentioned factors contribute to the noticeable larger DOFs in the 3D model compared to the p2DEIA. We have compared the amount of computational resources required for each scenario and present them in Supplementary Table 1. It is worth noting that the number of required DOFs needed to calculate the EM fields in the 3D model is a constant scale factor times the DOFs in 2D, and both

scale quadratically with the length scale of the in-plane dimension, that is, DOF $\sim \mathcal{O}(d^2)$. However, this scale factor turns out to be a large number (~105) due to the vectorial nature of the fields in the 3D model and due to the discretization of the device's thickness and cladding height to capture the out-of-plane scattering modes and the high-k near-fields around the etched interfaces (Supplementary Fig. 6).

The p2DEIA model is judged based on how well it compares to full 3D simulations of the structure, named here as the ground truth. Figure 2a,d shows time snapshots of the out-of-plane magnetic-field distribution in the vectorized p2DEIA model, and Fig. 2b,e shows the corresponding field distributions on a cut-plane through the silicon slab in the 3D-rendered structure for the same excitation as in the p2DEIA model. Visually comparing the simulated field distributions of the p2DEIA model to the 3D structure reveals similar features, indicating the success of the p2DEIA in approximating the in-plane wave propagation through the 3D structure. Finally, Fig. 2c,f illustrates the absolute value of the magnetic-field distributions on cut-planes through the silicon in the 3D view of the designed metastructures.

In Fig. 2g,h we compare the complex values for the target transmission matrix (T_{targ}), realized for the p2DEIA model (T_{p2DEIA}) and for the 3D model (T_{3D}). There are two classes of error that we allow ourselves to normalize out. First, a constant magnitude adjustment is applied to all values in T_{p2DEIA} to account for the unmodelled energy loss due to out-of-plane scattering. Second, a constant phase adjustment for each class of waveguide is applied to all values in T_{p2DEIA}. This is due to
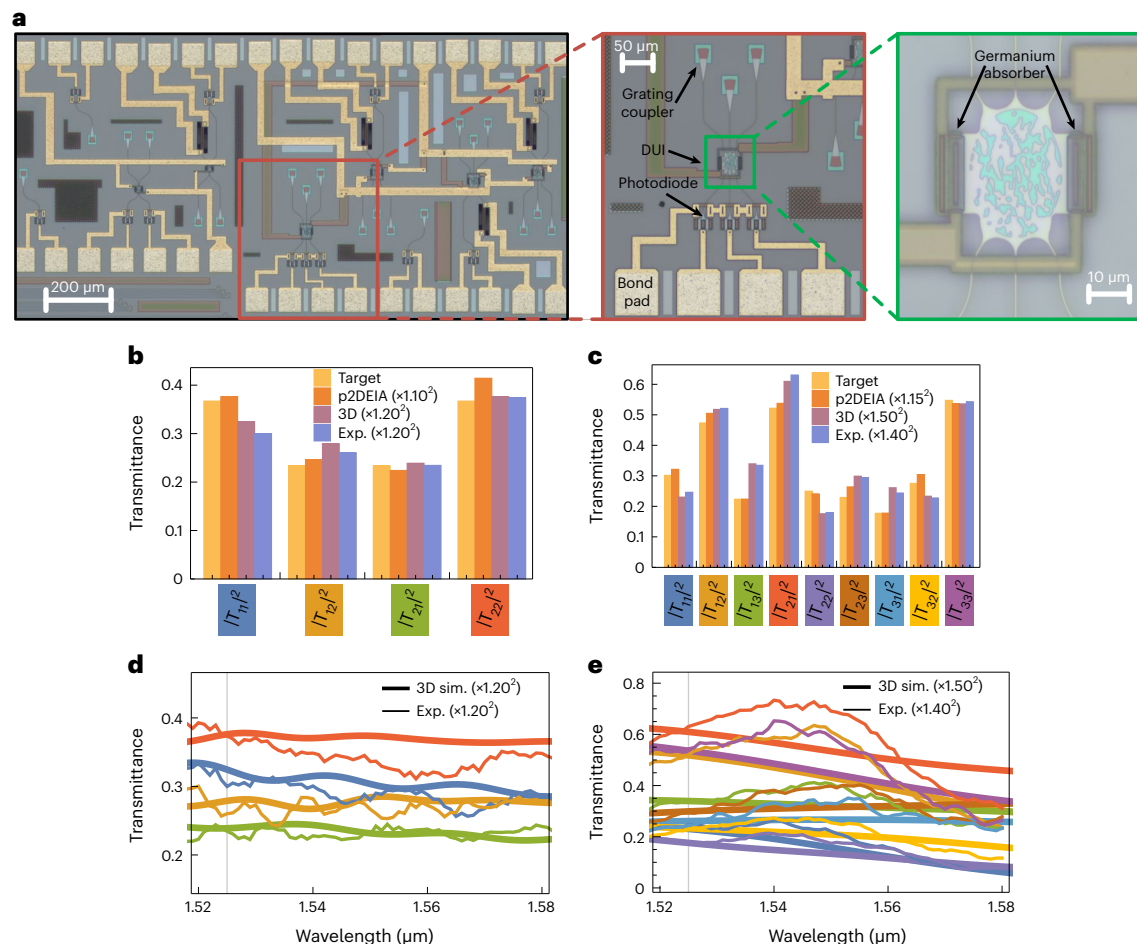
Fig. 3 | Comparison between simulations and experimental results.
a, Micrographs of the experiment, showing several 2 × 2 and 3 × 3 kernels and various calibration structures. The 3 × 3 kernel is highlighted. **b,c**, Comparison of the measured transmittance values illustrated as blue Exp. bars and the simulated transmittance values of the p2DEIA and 3D models and the target values at $\lambda_0 = 1.525\,\mu m$ for 2 × 2 (**b**) and 3 × 3 (**c**) metastructures, respectively. **d,e**, Measured transmittance values (thin solid lines) and the simulation results in 3D (thick solid lines) versus wavelength for 2 × 2 (**d**) and 3 × 3 (**e**) metastructures, respectively. The vertical grey lines mark the design wavelength at $\lambda_0 = 1.525\,\mu m$.

the ineffectiveness of the p2DEIA in modelling waveguides compared to the i2DEIA model. In other words, the guided modes in waveguides present a slightly different wavenumber, which introduces a phase error between p2DEIA and both the i2DEIA and 3D models. Once these trivial adjustments are made, we can see the close agreement between the target, p2DEIA and 3D values, indicating the success in achieving the desired transmission coefficients and the good accuracy of the p2DEIA in approximating 3D structures.

Next, in Fig. 2i,j we examine the spectral response of the realized 3D design by examining three types of error. In all cases, we normalize the squared error by 4N, which renders a unitary matrix of size N with 180° phase error to 100% (Supplementary Information). The orange curve is the Frobenius distance between the absolute values of the transmission parameters ($\| \, |T_{3D}(\lambda)| - |T_{targ}| \, \|_F^2/4N$). The small distance between the target values and the simulated ones indicates that the magnitude of the transmission parameters remains fairly invariant and close to the target as the wavelength changes. However, as a device that is intended to perform complex matrix multiplication, this is an inadequate indicator of success. The blue curve is the Frobenius distance between the transmission parameters over a range of wavelengths and the target values $\| \, T_{3D}(\lambda) - T_{targ} \, \|_F^2/4N$. This distance is quite small at the design wavelength, indicating the high accuracy of the 3D structure in realizing the target transmission parameters once the aforementioned amplitude and phase corrections are applied. However, the error presents an oscillatory behaviour as the wavelength changes.

This is due to the propagation delay between the input and output ports, which introduces a global phase error on the transmission parameters. This is conceptually related to the rotation of phase that occurs upon changing wavelength when light propagates in the forward direction through any medium (that is, $e^{j(k(\lambda)-k(\lambda_0))l}$). This global phase is immaterial as far as the phase differences are concerned, and these phase differences are typically what matters. Therefore, we consider an error in which we apply a global phase correction (green curve), $\phi(\lambda)$, to all values such that $T'_{3D}(\lambda) = T_{3D}(\lambda)e^{-j\phi(\lambda)}$, which minimizes $\| \, T'_{3D}(\lambda) - T_{targ} \, \|_F^2/4N$. As expected, this removes the oscillatory nature of the error and we may conclude that the designed metastructure is capable of performing vector–matrix operations over an extended range of wavelengths, hence being a broadband optical structure for such analogue computation.

Experimental results
The designed metastructures were fabricated using Advanced Micro Foundry (AMF), a commercial foundry, using their 180-nm process. This process provides two levels of etch (70 nm and 130 nm), and we used the most shallow. Each of the inputs of the inverse-designed devices is coupled to a grating coupler (GC), and each of the outputs is also coupled to a photodiode (PD). In addition to the devices under test, a calibration structure consisting of a directly connected GC and PD was fabricated, which was used to normalize the effect of the GC efficiency and the PD responsivity so that power transmission coefficients could
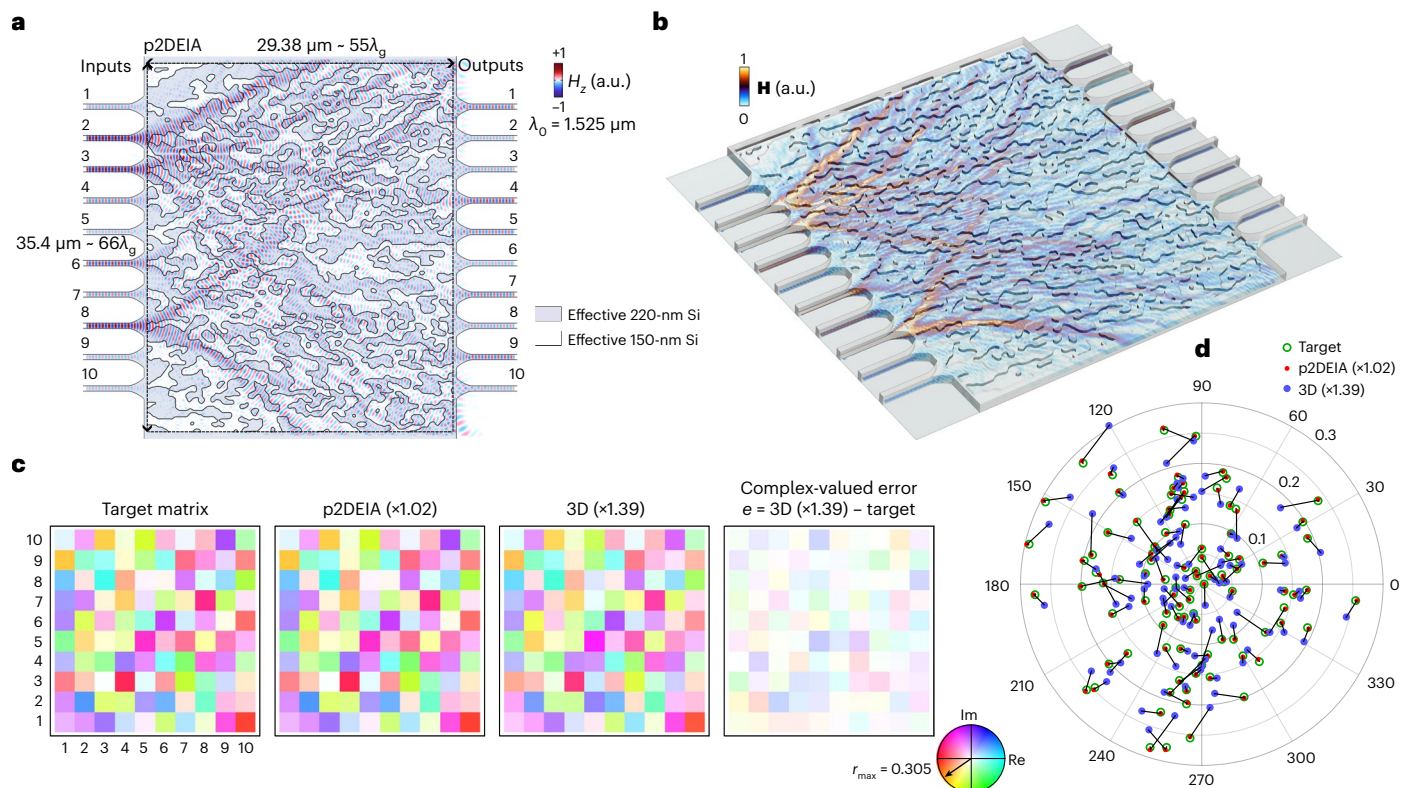
Fig. 4 | Inverse design of a SiPh metastructure for performing a 10 × 10 vector–matrix multiplication. a, Optimized geometry in p2DEIA and the distribution of the out-of-plane-polarized magnetic field for an arbitrary input excitation vector. **b**, Absolute value of magnetic-field distribution on a cut-plane through silicon in the 3D-rendered structure for the same excitation as in **a**. **c**,

Target transmission matrix, the simulated transmission parameters in p2DEIA and 3D models, and the complex-valued error between the target and 3D transmission parameters. **d**, Corresponding transmission values on a complex plane polar plot.

be measured. Our goal is to demonstrate that these devices perform complex matrix multiplication, meaning that both the magnitude and the phase of each element of matrix T are important. The magnitude of $|T_{ij}|$ can be directly inferred from the power measurements. However, the phase, $\angle T_{ij}$, is difficult to measure, as small variations in single-mode waveguides will result in substantial phase variation for the mode propagated over any significant length. Indeed, these variations are what we attempt to circumvent by using the low-index-contrast design with relatively few scattering interfaces. As such, a reliable reference is difficult to design and fabricate. Similar to other optical measurement techniques such as reflectometry and variable-angle spectral ellipsometry, we rely on the spectral output from a series of measurements to produce a large number of coupled measured values and then check these values against a model. However, unlike traditional techniques, our model is not a simple material stack, but rather the full 3D simulation of the device. Regardless, from simulation, the fields used by various elements T_{ij} heavily overlap, indicating strong coupling between measurements. A good agreement of a sufficiently large number of measurements serves to validate the model, which includes phase. Further discussion is provided in Supplementary Section Fabrication and measurement methodology. The results of the experiment are illustrated in Fig. 3.

In Fig. 3b,c, the target transmission parameters, the simulated transmittance in p2DEIA, the simulated values in the 3D model, and the measured transmittance at $\lambda_0 = 1.525\,\mu m$, are compared with each other for the 2 × 2 and 3 × 3 structures, respectively. As mentioned before, because the p2DEIA does not model the out-of-plane scattering, a small amount of energy will scatter away from the silicon slab in the 3D structure, and there is thus a slight reduction in transmittance. Therefore, to visually compare the transmittance in the 3D structure

to those of the p2DEIA, the 3D values of the 2 × 2 structure are multiplied by a power scale factor of 1.2^2 and those of the 3 × 3 structure by 1.5^2. After applying the adjustment factors, the target, p2DEIA and 3D values agree with each other. To visually compare the measured transmittances to those of the target, similar scale factors of 1.2^2 and 1.4^2 are needed for the 2 × 2 and 3 × 3 structures, respectively. Therefore, the measured transmittances closely follow the values from the 3D model, indicating the fidelity of the model. In conclusion, despite the constraints present in the p2DEIA model and the imperfection in the measurement, the good agreement between the target, p2DEIA, 3D and measured transmission values at $\lambda_0 = 1.525\,\mu m$ is an indication of the success of the p2DEIA as a computational tool for the inverse design of 3D planar structures.

In Fig. 3d,e, the results of the wavelength sweep for the transmittances are illustrated. The thick solid curves present the simulation data from the 3D model and the thin solid lines are the measurement data. The agreement between the simulation and measurement data is excellent at the design wavelength (measurement details are provided in Supplementary Information). As mentioned before, the magnitudes of the transmission parameters show a broadband response, as one can see from the slow variation of the simulation and experimental data versus wavelength.

Metastructure for 10 × 10 vector–matrix multiplication
As mentioned already, p2DEIA enables the simulation and design of an optically large structure via tractable computational efforts. This allows one to increase the number of input and output data channels by incorporating more waveguides. For that purpose, we used the same inverse-design technique, utilizing the p2DEIA to explore and design a structure to perform a 10 × 10 vector–matrix multiplication

(Fig. 4a). This was done within a 35.4 µm × 29.4 µm design region using the same SiPh platform and wavelength as in the previous examples. The design region spans many guided wavelengths, that is, $66\lambda_g \times 55\lambda_g$, where $\lambda_g = \lambda_0/n_{\text{eff, 220 nm}}$. The choice of the target matrix $M_{10 \times 10}$ was arbitrary within the space of passive transmission matrices (Supplementary Information).

Figure 4a illustrates the optimized geometry and a time snapshot of the out-of-plane magnetic-field distribution for an arbitrary vector excitation in the p2DEIA model. Figure 4b shows the absolute value of the magnetic-field distribution for the same excitation through the 3D-rendered structure. The full-wave 3D simulation of such an exceedingly large optical structure is arduous using either frequency- or time-domain simulations; however, with p2DEIA we were able to simulate the structure for hundreds of iterations during the inverse-design procedure. Figure 4c illustrates the complex values of the target transmission matrix elements, the values from the optimized p2DEIA model, and the values from the 3D-rendered structure. Similar to the 2 × 2 and 3 × 3 results, the p2DEIA and 3D values are offset by appropriate adjustment factors. For a more clear comparison, the corresponding transmission values are represented in the complex plane polar plot shown in Fig. 4d. (The proper adjustment factors are shown in parentheses.)

The agreement between the target and p2DEIA values is excellent, signifying that the inverse design converged nicely. The 3D values show an acceptable agreement with the p2DEIA values, despite the use of a coarser meshing and reduced-order absorbing boundary conditions forced by the huge size of the 3D simulation. However, the p2DEIA model introduces some errors on each interface (Supplementary Information). Larger structures might require more sophisticated interface modelling to take into account out-of-plane scattering and phase discontinuities due to the stored energy in non-propagating modes. Regardless, the optimized design performs well judging by the close agreement between the target, the p2DEIA and the 3D transmission parameters. Therefore, with this design, p2DEIA proved to be successful in designing an optically large structure for manipulating a large number of input and output data channels.

Summary and conclusion

In this Article we have employed the p2DEIA using the effective index of the slab mode as an efficient computational platform for the inverse design of 3D planar structures in silicon photonics. Assuming a low-index contrast, we have demonstrated inverse-designed metastructures for performing 2 × 2 and 3 × 3 vector–matrix products. The results of the simulation and experiment illustrate the high accuracy of the p2DIEA for approximating 3D planar structures. Furthermore, the essential assumption of low-index contrast for p2DEIA leads to feed-forward, low-resonant structures in which the response is slowly varying as a function of wavelength. Therefore, the resulting structures using the p2DEIA present broadband optical responses for performing analogue computations and are also less sensitive to fabrication errors. Because the p2DEIA is a reduced-dimension approximation of a 3D structure, scaling to larger structures is much less of a computational burden than full-wave 3D simulation, thus enabling the inverse design of metastructures for larger-scale vector–matrix multiplications. We demonstrated this by designing a structure for performing a 10 × 10 vector–matrix product.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-024-01394-2.

References

1. Cordaro, A. et al. Solving integral equations in free-space with inverse-designed ultrathin optical metagratings. *Nat. Nanotechnol.* **18**, 365–372 (2023).
2. Nikkhah, V., Tzarouchis, D. C., Hoorfar, A. & Engheta, N. Inverse-designed metastructures together with reconfigurable couplers to compute forward scattering. *ACS Photonics* **10**, 977–985 (2022).
3. Silva, A. et al. Performing mathematical operations with metamaterials. *Science* **343**, 160–163 (2014).
4. Mohammadi Estakhri, N., Edwards, B. & Engheta, N. Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).
5. Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).
6. Camacho, M., Edwards, B. & Engheta, N. A single inverse-designed photonic structure that performs parallel computing. *Nat. Commun.* **12**, 1466 (2021).
7. Fu, W. et al. Ultracompact meta-imagers for arbitrary all-optical convolution. *Light Sci. Appl.* **11**, 62 (2022).
8. Pan, D. et al. Laplace metasurfaces for optical analog computing based on quasi-bound states in the continuum. *Photon. Res.* **9**, 1758–1766 (2021).
9. Zangeneh-Nejad, F., Sounas, D. L., Alù, A. & Fleury, R. Analogue computing with metamaterials. *Nat. Rev. Mater.* **6**, 207–225 (2021).
10. Zhu, T. et al. Plasmonic computing of spatial differentiation. *Nat. Commun.* **8**, 15391 (2017).
11. Pors, A., Nielsen, M. G. & Bozhevolnyi, S. I. Analog computing using reflective plasmonic metasurfaces. *Nano Lett.* **15**, 791–797 (2015).
12. Wang, H. et al. Design of compact meta-crystal slab for general optical convolution. *ACS Photonics* **9**, 1358–1365 (2022).
13. Choutagunta, K., Roberts, I., Miller, D. A. & Kahn, J. M. Adapting Mach–Zehnder mesh equalizers in direct-detection mode-division-multiplexed links. *J. Light. Technol.* **38**, 723–735 (2019).
14. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
15. Nikkhah, V., Mencagli, M. J. & Engheta, N. Reconfigurable nonlinear optical element using tunable couplers and inverse-designed structure. *Nanophotonics* **12**, 3019–3027 (2023).
16. Wang, T. et al. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **13**, 123 (2022).
17. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).
18. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
19. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
20. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
21. Wang, H., Guo, C., Zhao, Z. & Fan, S. Compact incoherent image differentiation with nanophotonic structures. *ACS Photonics* **7**, 338–343 (2020).
22. Zhu, T. et al. Topological optical differentiator. *Nat. Commun.* **12**, 680 (2021).
23. Goh, H. & Alù, A. Nonlocal scatterer for compact wave-based analog computing. *Phys. Rev. Lett.* **128**, 073201 (2022).
24. Farhat, N. H. Photonic neural networks and learning machines. *IEEE Expert* **7**, 63–72 (1992).
25. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
26. Miller, D. A. Self-configuring universal linear optical component. *Photon. Res.* **1**, 1–15 (2013).

27. Bogaerts, W. et al. Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
28. Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walmsley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016).
29. Tzarouchis, D. C., Mencagli, M. J., Edwards, B. & Engheta, N. Mathematical operations and equation solving with reconfigurable metadevices. *Light Sci. Appl.* **11**, 263 (2022).
30. Bendsoe, M. P. & Sigmund, O. *Topology Optimization: Theory, Methods and Applications* (Springer, 2004).
31. Molesky, S. et al. Inverse design in nanophotonics. *Nat. Photon.* **12**, 659–670 (2018).
32. Hughes, T. W., Minkov, M., Williamson, I. A. & Fan, S. Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics* **5**, 4781–4787 (2018).
33. Hammer, M. & Ivanova, O. V. Effective index approximations of photonic crystal slabs: a 2-to-1-D assessment. *Opt. Quantum Electron.* **41**, 267–283 (2009).
34. Knox, R. & Toulios, P. Integrated circuits for the millimeter through optical frequency range. *Proc. Symp. Submillimeter Waves* **20**, 497–515 (1970).
35. Chiang, K. S. Analysis of optical fibers by the effective-index method. *Appl. Opt.* **25**, 348–354 (1986).
36. Hocker, G. B. & Burns, W. K. Mode dispersion in diffused channel waveguides by the effective index method. *Appl. Opt.* **16**, 113–118 (1977).
37. Van De Velde, K., Thienpont, H. & Van Geen, R. Extending the effective index method for arbitrarily shaped inhomogeneous optical waveguides. *J. Light. Technol.* **6**, 1153–1159 (1988).
38. COMSOL Multiphysics v. 5.6 (COMSOL; 2023); www.comsol.com
39. Nikkhah, V. et al. Replication data for 'Inverse-designed low-index-contrast structures on silicon photonics platform for vector-matrix multiplication' (Zenodo, 2023); https://doi.org/10.5281/zenodo.10083901

Author contributions

N.E., F. Aflatouni. and B.E. conceived the idea, envisioned the experiments and supervised the project. V.N. conducted theoretical analysis, numerical simulations and inverse-design of the structures. A.P., F. Ashtiani. and B.E. prepared the designs for nanofabrication by the Advanced Micro Foundry (AMF) and designed the experiments. A.P. performed the experiments and collected data. All authors reviewed, studied and discussed the experimental and numerical simulation results, and discussed the main outcomes of the project. V.N. prepared the first draft of the main text and the Supplementary Information. All authors subsequently worked on the manuscript up to its completion for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-024-01394-2.

Correspondence and requests for materials should be addressed to Nader Engheta.

Peer review information *Nature Photonics* thanks Yannick Salamin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.