scientific reports

OPEN

The Goldilocks principle of learning unitaries by interlacing fixed operators with programmable phase shifters on a photonic chip

Kevin Zelaya[1,3], Matthew Markowitz[1,2,3] & Mohammad-Ali Miri[1,2,3✉]

Programmable photonic integrated circuits represent an emerging technology that amalgamates photonics and electronics, paving the way for light-based information processing at high speeds and low power consumption. Programmable photonics provides a flexible platform that can be reconfigured to perform multiple tasks, thereby holding great promise for revolutionizing future optical networks and quantum computing systems. Over the past decade, there has been constant progress in developing several different architectures for realizing programmable photonic circuits that allow for realizing arbitrary discrete unitary operations with light. Here, we systematically investigate a general family of photonic circuits for realizing arbitrary unitaries based on a simple architecture that interlaces a fixed intervening layer with programmable phase shifter layers. We introduce a criterion for the intervening operator that guarantees the universality of this architecture for representing arbitrary $N \times N$ unitary operators with $N + 1$ phase layers. We explore this criterion for different photonic components, including photonic waveguide lattices and meshes of directional couplers, which allows the identification of several families of photonic components that can serve as the intervening layers in the interlacing architecture. Our findings pave the way for efficiently designing and realizing novel families of programmable photonic integrated circuits for multipurpose analog information processing.

Levering the unique properties of light to perform computations in novel ways is a subject with a long history[1,2]. Although an all-optical processor for universal computing seems to be a far reach goal, photonics can provide exciting opportunities for unconventional computing building on analog logic and with novel information processing configurations. What makes photonics an intriguing option for unconventional computing are exotic potentials such as the intrinsically high speed and low energy consumption, the capabilities for massive parallelization, and long-range interactions. Nevertheless, a significant challenge in optical computing lies in the absence of appropriate computing paradigms, methods, and algorithms that harness the unique capabilities of this technology to develop efficient and application-specific photonic processors. In particular, matrix-by-vector multiplication is one of the most basic mathematical operations that lies at the core of various tasks ranging from optical convolution schemes[3,4] and matrix eigenvalue solver[5] to novel optical memristors[6,7] and optical artificial neural networks[8,9]. In the past decade, with rapid technological progress, there has been a resurrection in efforts devoted to developing such programmable photonic integrated circuit that performs matrix-vector multiplication[10–12]. The utility of such a device as an energy-efficient photonic accelerator in conjunction with electronic processors appears to be a distant possibility, considering the inherent difficulties associated with scaling and precision. However, there is no doubt that an on-chip programmable photonic matrix-vector multiplier can create exciting opportunities in classical and quantum computing through various applications that range from quantum information and quantum transport simulations[13–15], to optical signal processing[16], neuromorphic

[1]Department of Physics, Queens College of the City University of New York, Queens, NY 11367, USA. [2]Physics Program, The Graduate Center, City University of New York, New York, NY 10016, USA. [3]These authors contributed equally: Kevin Zelaya, Matthew Markowitz and Mohammad-Ali Miri. ✉email: mmiri@qc.cuny.edu

computing[17], and optical neural networks[3,18,19], as well as putting forward a platform for rapid prototyping of linear multiport photonic devices[20,21].

Indeed, the optical realization of arbitrary unitary operations has been known since the seminal paper of Reck *et al.*[22], which originally concerned free space optics but successfully translated to photonic integrated circuits by Miller[23–25]. This architecture builds on breaking down unitary matrices of any order into lower-dimension unitary matrices, which ensures the existence of an optical realization through two fundamental building blocks that are beam splitters (couplers) and phase shifters. Despite its generality, this method uses a pyramid-shaped array of Mach-Zehnder interferometers (MZI), which is impractical for larger implementations because the number of beam splitters grows quadratically with the number of ports. In turn, Clements *et al.*[26] introduced an alternative and symmetric rectangular-shaped array, resulting in a device with half the total optical depth and, consequently, more loss-tolerant. Such a rectangular array has been proved robust enough to create photonic realizations of Haar-random matrices[27]. Further unitary realizations related to other mesh geometries have been explored in[28,29], as well as topological photonic lattices with hexagonal-shaped arrays of MZI[30,31]. In turn, free-space propagation setups have been devised based on plane-light conversion[32,33] and using diffractive surface layers[34].

While the latter devices originally consisted of bulky optical components, the principle has recently been applied to on-chip structures as well. Recent studies have explored the use of particular transfer matrices (henceforth called *F*) alternating with phase mask layers to obtain an arbitrary unitary transformation[10,12,35–37]. Pastor et al.[12] considers wave propagation in multimode slab waveguides to implement a Discrete Fourier transform (DFT) as their transformation *F*. They showed that an arbitrary transformation could be performed when $6N + 1$ phase layers and $6N$ DFT elements, where N is the number of ports. Tanomura *et al.*[35] interleave the phase masks with multimode interference couplers connected with single-mode waveguides and use simulated annealing optimization to argue for well-approximated conversions when $M \approx N$. Fully functional unitary four-, eight-, ten-, and twelve-port devices have been proposed and manufactured[38–41]. Moreover, an alternative device using polarization and multiple wavelength degrees of freedom has been considered in[42]. Markowitz and Miri[37] have explored similar structures and have found rigorous numerical evidence that interleaved phase arrays and discrete fractional Fourier transform (DFrFT) are indeed universal, whereas the use of Haar-random unitary matrices has been proved to lead to the desired universality[10]. A further waveguide array with varying propagation constants with step-like profiles has been reported[43]. The interlacing architecture appears to exhibit interesting auto-calibrating properties, which makes it resilient to fabrication errors[44]. Furthermore, we recently have shown that the intervening structure can go beyond implementing unitaries to directly implementing arbitrary non-unitary operations when the diagonal matrices are relaxed to leave the unitary circle in the complex domain[45]. In this sense, by utilizing both amplitude and phase modulations, one can realize a fully programmable device for arbitrary matrix operations[45]. This shows an important generalization of the previous results that show by itself an advantage of the interlacing architecture over the mesh geometries.

This manuscript discusses a broad class of universal on-chip photonic architectures based on a layered configuration of phase mask layers as programmable units interlaced with a passive random matrix *F*. It is shown that the proposed interlaced architecture is far more flexible by showing that broad families of matrices *F* can serve as the fixed intervening operator. The phases are steered to reconstruct a unitary target matrix, provided that *F* has well-posed properties. Numerical evidence based on rigorous optimization algorithms reveals that universality is reached for dense matrices *F*, while a phase transition in the accuracy of reconstructed $N \times N$ targets occurs at $M = N + 1$, with *M* the total number of phase mask layers. Tests using the discrete Fourier transform (DFT) and discrete fractional Fourier transform (DFrFT) confirm the latter claim, and Haar-random matrices also show outstanding convergence. To generalize the domain of the valid intervening operators, a density criterion is derived so that matrices *F* can be classified according to their elements to ensure universality. To demonstrate this result, photonic lattices with uniform, nonuniform, and disordered coupling coefficients are considered as photonic realizations for the matrix *F*, while using the proposed density criterion, it is shown that universality is reached for specific length intervals. Furthermore, we explore waveguide coupler meshes as an alternative intervening unit *F* and determine the minimum number of coupler layers required to guarantee the universality of the interlacing architecture.

Results
Architecture and mathematical foundation
Let us consider an arbitrary unitary matrix $\mathscr{U} \in U(N)$, with $U(N)$ the group of $N \times N$ unitary matrices. Our goal is to implement a proper factorization of $\mathscr{U}$ in terms of another unitary matrix F to be defined and a set $\{P_k\}_{k=1}^M$ composed of phase matrices $P_k = e^{iD_k}$, with $D_k = diag(\phi_1^{(k)}, \ldots, \phi_N^{(k)})$ a diagonal matrix and $\phi_n^{(k)} \in (0, 2\pi]$, where $n \in \{1 \ldots, N\}$ and $k \in \{1, \ldots, M \in \mathbb{N}\}$. The factorization proposed here is such that it intercalates F with a phase matrix P_k through the relation

$$\mathscr{U} = FP_M F \ldots FP_1 F, \tag{1}$$

which is resourceful as it allows for optical implementations. The phase matrices can be implemented through layers of phase shifters (active optical elements), and the matrix F (passive optical element) has to be selected so that arbitrary unitary target matrices $\mathscr{U}_t$ can be reconstructed with minimal error by adequately tuning the phase shifters $\phi_n^{(k)}$. If the latter is achieved, it is said that the *universality property* has been met. An arbitrary matrix $\mathscr{U} \in U(N)$ requires N^2 real parameters to be fully defined. Therefore, while performing the factorization, it is vital to consider at least the same number of parameters. For the device proposed in (1), we have MN free parameters in total and expect that $M \geq N$ in order to achieve the desired universality. Although there are some cases where $\mathscr{U}$ has a particular symmetry that reduces the number of parameters, we aim for the general
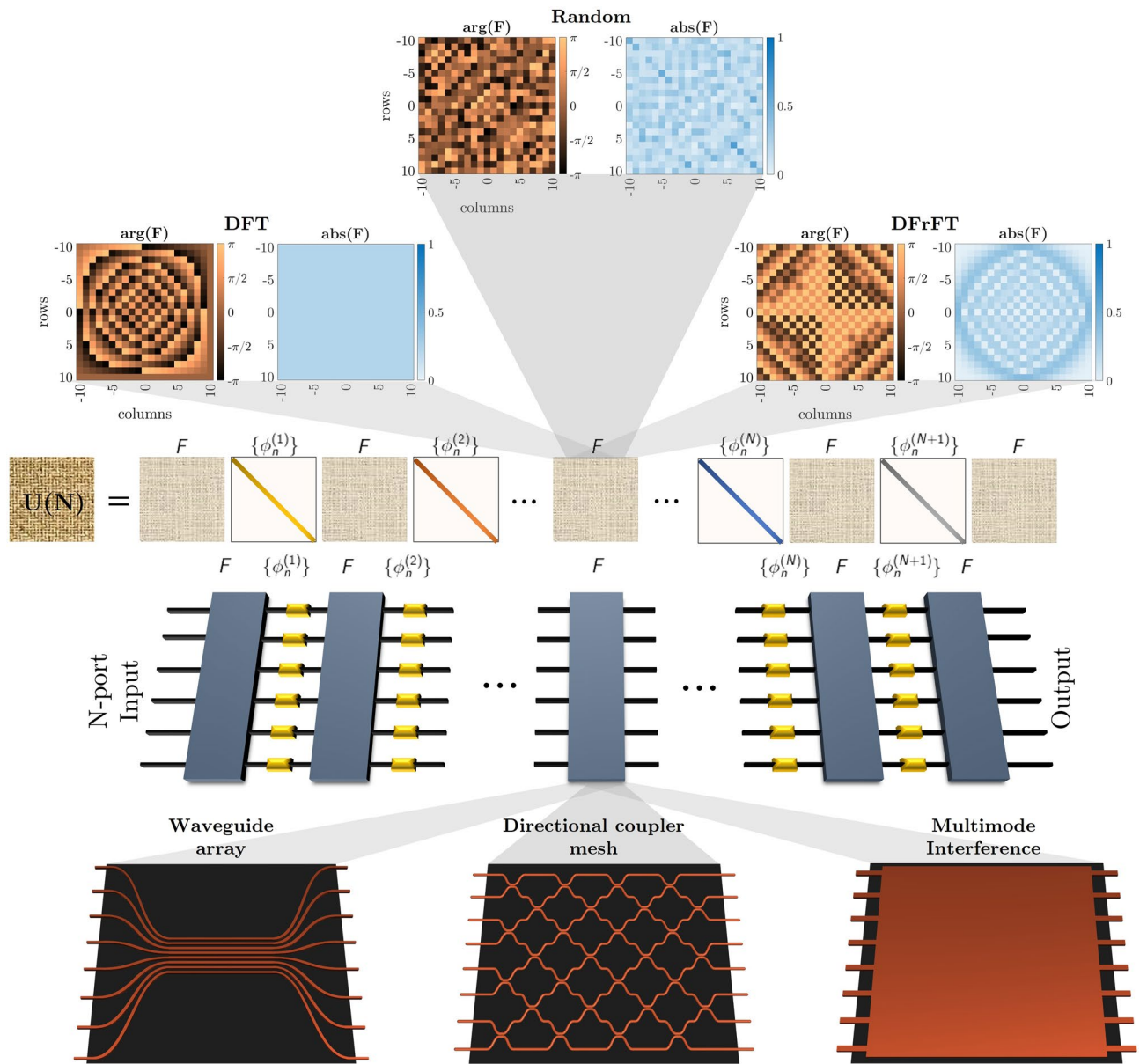
Figure 1. Universal architecture scheme. The proposed architecture involving alternating layers of random unitary matrices F and diagonal phase shifts layers (PL) $\{\phi_n^{(p)}\}$, with $p = 1, \ldots, N + 1$. The upper insets depict the modulus and argument of the potential candidates for the unitary matrix F, which have been selected as the DFT, DFrFT, and a random unitary matrix. The lower insets illustrate potential photonic implementations to perform the unitary matrix F.

case. The proposed architecture (1) and its optical implementation are sketched in Fig. 1. On the one hand, the mathematical structure of the passive unitary F (see top panels of Fig. 1) can be that of the discrete Fourier transform (DFT), discrete fractional Fourier transform[37], or simply a Haar-random matrix. On the other hand, the photonic implementation of F can be performed through different optical realizations involving a unitary wave evolution (see bottom panels of Fig. 1), such as waveguide arrays[45,46], meshes of directional coupler[26,28,29], and multimode interference (MMI)[12,47]. Particularly, MMI couplers have been shown to be suitable to represent the DFT matrix[12,48], whereas waveguide arrays lead to simple representations of the DFrFT[46].

Here, we focus on architectures based on the first two solutions, the numerical analysis and universality of which are discussed below. Layered architectures akin to Eq. (1) have been numerically validated in previous works using different optical arrays and MZI meshes, where different optimization algorithms such as gradient-descent, stochastic gradient descent, simulated annealing, and basin-hopping have been implemented. See for instance[10,40,43,49]. In order to demonstrate the universality of this device, we optimize the NM phases for a variety of randomly chosen target matrices $\mathscr{U}_t$ generated in accordance with the Haar measure[50]. The objective function to be minimized, also called *error norm*, is defined by

$$L = \frac{1}{N^2} ||\mathscr{U} - \mathscr{U}_t||^2 \qquad (2)$$

where $\| \cdot \|$ stands for the Frobenius definition of the norm, $\mathscr{U}_t$ is the target matrix being tested, and $\mathscr{U}$ is the reconstructed matrix using the factorization (1). The Levenberg-Marquardt (LM) algorithm[51,52] is used to find the minimum of this function. For a given target $\mathscr{U}_t$, the phases are randomly initialized between 0 and 2π. The optimization was performed in MATLAB. In Fig. 2a, the norm for 100 target unitary matrices is shown for various cases, fixing the default tolerance values to 10^{-10}. A phase transition occurs between the $M = N$ and $M = N + 1$ layers, which is unsurprising given the system becomes over-determined by N parameters. These jumps are larger than reported by Tang et al.[53] due to their usage of a probabilistic algorithm (Simulated Annealing) rather than a gradient-based one such as LMA. A downside of gradient-based approaches is they may require many runs with different starting conditions. We can decrease the overhead by using a stopping criteria for the norm along with a maximum iteration for each run of LMA. Using a maximum iteration of 50, we find that we rarely need more than 100 runs in the case $N = 8$, $M = 9$ to achieve norms less than 10^{-10} (Fig. 2b). For systems with a lower number of ports N, the distribution skews towards lower values. To more confidently label choices of F which are not Haar-random generated as "bad" mixing layers, we set the maximum number of runs somewhat higher to 250 or 500 as found appropriate.

The density criterion and the Goldilocks principle

Our preliminary numerical results suggest that Haar random matrices F also possess the required properties to render a universal architecture from (1). Random matrices drawn from the Haar measure are typically dense matrices whose sparsity is low, suggesting that density might be a criterion to classify F matrices as good candidates. To test this idea, it is thus desirable to find a proper measure to quantify the density for unitary matrices. Indeed, random matrix theory establishes robust criteria for studying random complex-valued matrices at the limit of large size N based on the singular value decomposition (SVD)[54]. On the one hand, in the current setup, we focus on unitary matrices of relatively small size, as the number of ports in our architecture is not necessarily long enough to take into account the asymptotic analysis of random matrix theory. On the other hand, the singular value decomposition is not particularly helpful when dealing with unitary matrices, as for a complex-valued matrix A, the SVD requires the computation of the spectral properties of $AA^\dagger$ and $A^\dagger A$, which for unitary matrices is always equal to the identity matrix. For these reasons, the density criterion discussed below suits better for the interlaced architectures here constructed.

To better understand the importance of dense matrices in our universal architecture, let us further inspect the factorization in (1). By fixing $F = diag(e^{i\xi_1}, \ldots, e^{i\xi_N})$ as a diagonal unitary matrix, for $\xi_j \in (0, 2\pi]$ for $j = 1, \ldots, N$, it is straightforward to notice that (1) reduces to a diagonal unitary matrix as well, which is far from representing a universal device. That is, diagonal F matrices can only reconstruct diagonal unitary matrices. This behavior extends to any of the $N!^2$ different permutations allowed to the diagonal matrix F, as the
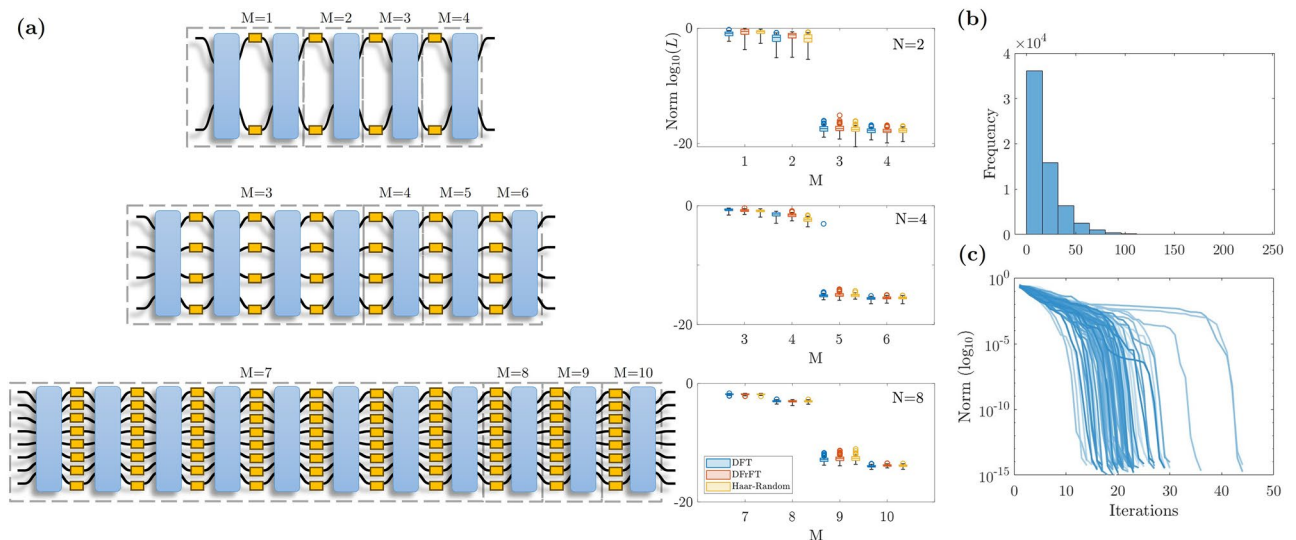

Figure 2. Numerical universality test. (**a**) Architecture depiction (left column) and optimization objective function (right column) for 100 target matrices at various values of M and N. Black boxes denote any possible realization for the F matrix. (**b**) Multiple trials for $N = 8$ and $M = 9$ using 250 random F matrices were considered; 250 targets were used for each matrix F. Shown is the distribution of the number of LMA runs to achieve a norm lower than the stopping norm of 10^{-10}, with a maximum of 50 iterations per run. (**c**) Norm ($\log_{10}$) in terms of the number of iterations for the run with the best norm. Using 100 random matrices F, each with a single target matrix.

factorized matrix $\mathcal{U}$ acquires the same structure as that of the permuted F matrix. Now, let us consider the set of unitary matrices $\{\mathcal{V}_{n_j}\}_{j=1}^k$, where $\mathcal{V}_{n_j} \in U(n_j)$ and $\sum_{j=1}^k n_j = N$, such that $F = diag(\mathcal{V}_{n_1}, \ldots, \mathcal{V}_{n_k})$ is an N-dimensional block-diagonal unitary matrix. This particular selection for F leads to a unitary operator with the same structure, lacking the required universality property. Although block-diagonal matrices F lose their structure when randomly rearranged, anti-diagonal block matrices will always result in either block-diagonal or anti-block-diagonal matrices, neither of which are universal.

Thus, we have identified a particular class of *bad-performing* interlacing matrices F, which are useful to trace a suitable Goldilocks region where the passive F matrices have the required density property. To this end, let us first remark that any N-dimensional unitary matrix can be written as $\mathcal{V} = (\vec{v}_1, \ldots, \vec{v}_N)$, where $\vec{v}_j \in \mathbb{C}^N$ are complex-valued column vectors that form an orthogonal set through the Euclidean inner product in $\mathbb{C}^N$, i.e., $\vec{v}_{j_1} \cdot \vec{v}_{j_2} \equiv \vec{v}_{j_1}^\dagger \vec{v}_{j_2} = \delta_{j_1,j_2}$. From the orthogonality condition, we can simply focus on the columns (or equivalently the rows) of $\mathcal{V}$, whereas the normalization imposes a constraint on elements across the columns (rows). Since the elements of $\mathcal{V}$ are complex numbers, we alternative work with the matrix $\widetilde{\mathcal{V}}$, composed of the modulus of the elements of $\mathcal{V}$. Let us define $v_{p;q}$ as the q-th element of the p-th column vector $\vec{v}_p$, with $p, q = 1, \ldots, N$, so that $\widetilde{\mathcal{V}}_{p;q} = |v_{p;q}|$. From the unitarity of $\mathcal{V}$, it follows that $\sum_{q=1}^N |v_{p;q}|^2 = \sum_{p=1}^N |v_{p;q}|^2 = 1$ for all p, q, so that we can focus on the density of either the columns or rows of $\widetilde{\mathcal{V}}$. Without loss of generality, we work with the columns. Since we are interested in how the elements are sparse across each column, we compute the corresponding variance

$$S_p = \frac{1}{N} - \mu_p^2, \quad \mu_p = \frac{\sum_{q=1}^N |v_{p;q}|}{N} \geq \frac{1}{N},$$

for each column p.

Following the normalization condition, it is straightforward to prove that the variance is a bounded quantity in the closed interval $S_p = \left[0, \frac{N-1}{N^2}\right]$, where the lower bound corresponds to the case where all the elements of $\vec{\tilde{v}}_p$ are equal, i.e., $\vec{\tilde{u}}_p$ is maximally spread. The upper bound corresponds to the case where $\vec{\tilde{v}}_p$ is one of the canonical basis vectors, $(0, \ldots, 1, \ldots, 0)^T$. Thus, a given column p of $\mathcal{V}$ is said to be denser if its corresponding variance S_p approaches $(N-1)/N^2$. The more sparse the elements of the column p, the more S_p approaches 0. These are the key ideas we use henceforth to characterize density across the full matrix $\mathcal{V}$. Let $\mathcal{S} = \{S_p\}_{p=1}^N$ be the set of variances associated with each column of $\widetilde{\mathcal{V}}$. We define the mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$ associated with the elements of $\mathcal{S}$, so that the density of a given unitary matrix $\mathcal{V}$ can be characterized by defining the point

$$\vec{R} := (N\tilde{\mu}, N\tilde{\sigma})). \tag{3}$$

Since row permutation leaves S_p invariant and column permutation only permutes the index p, the quantities $\tilde{\mu}$ and $\tilde{\sigma}$, and consequently the point $\vec{R}$, are permutation invariant.

There are two note-worthy extremal cases, namely the maximally dense and the diagonal cases (sparsest cases) unitary matrices. In the former case, $\mathcal{V}$ is composed of column vectors so that the variances vanish, $S_p = 0$, for all $p = 0, \ldots, N$. The DFT matrix of dimension N is such an example, leading to $\tilde{\mu} = \sigma = 0$, which we consider as the ideal case. For the second case, the variances are maximal per each column, $S_p = (N-1)/N^2$, for all $p = 1, \ldots, N$, and the statistical information of the matrix reduces to $\tilde{\mu} = (N-1)/N^2$ and $\tilde{\sigma} = 0$. We thus have two comparison points, from which we find the bounded interval $N\tilde{\mu} \in [0, (N-1)/N]$. Additional reference points can be traced out if we take the block-diagonal matrices $F = diag(\mathcal{V}_{n_1}, \ldots, \mathcal{V}_{n_k})$, with $\mathcal{V}_{n_j} \in U(n_j)$ unitary and maximally dense (DFT) matrices of dimension n_j for $j = 1, \ldots, k, 1 \leq k \leq N$, and $\sum_{j=1}^k n_j = N$, so that bad-performing matrices are generated (see discussion above). Particularly, let us consider the case $k = 2$ so that $F = diag(\mathcal{V}_{n_1}, \mathcal{V}_{n_2})$, with $n_1 + n_2 = N$. One can assign the indexes $n_1 = \ell$ and $n_2 = N - \ell$, with $\ell = 1, \ldots, \lfloor N/2 \rfloor$ nonequivalent ways to define the two block-diagonal matrices F that leads to the k_2 reference points

$$\vec{R}_{k=2} = \left(\frac{2\ell(N-\ell)}{N^2}, \frac{\sqrt{\ell(N-\ell)}(N-2\ell)}{N^2}\right), \quad \ell = 1, \ldots, \lfloor \frac{N}{2} \rfloor.$$

Although further reference points exist for $k \geq 3$, those points are farther from the ideal (maximally dense) case $\vec{R}_0 \equiv \vec{R}_{k=1} = (0, 0)$ than those marked with $k = 2$, and are thus disregarded.

In this form, we can just focus on the area spanned by the points marked between $k = 1$ and $k = 2$. Interestingly, for $k = 2$ and $\ell = \lfloor N/2 \rfloor$, one obtains the reference points with smaller standard deviation, which are $\vec{R}_{k=2, \ell=\frac{N}{2}} = \left(\frac{1}{2}, 0\right)$ and $\vec{R}_{k=2, \ell=\frac{N-1}{2}} = \left(\frac{1}{2} - \frac{1}{2N^2}, \frac{\sqrt{N^2-1}}{2N^2}\right)$ for even and odd N, respectively. Note that in the limit $N \to \infty$, the mean converges to the non-vanishing value $1/2$. In turn, the maximum standard deviation is determined by minimizing $N\tilde{\sigma}$ in terms of ℓ, from which one obtains the critical value $\ell_c = \frac{N}{2\sqrt{2}}(\sqrt{2} - 1)$ and the maximum standard deviation $N\tilde{\sigma}|_{\ell_c} = 1/4$. That is, the standard deviation is bounded to the interval $N\tilde{\sigma} \in [0, 1/4]$, where the upper bound is independent of N and is given by $max\left(N\tilde{\sigma}|_{\lfloor \ell_c \rfloor}, N\tilde{\sigma}|_{\lceil \ell_c \rceil}\right)$. The latter allows us forming a polygon with vertices at $\vec{R}_{k=1} = (0, 0)$ and $\vec{R}_{k=2}$, the area of which is non-null and finite even for $N \to \infty$ (see Fig. 3). We focus on unitary matrices whose vector $\vec{R}$ lies inside the latter polygon while
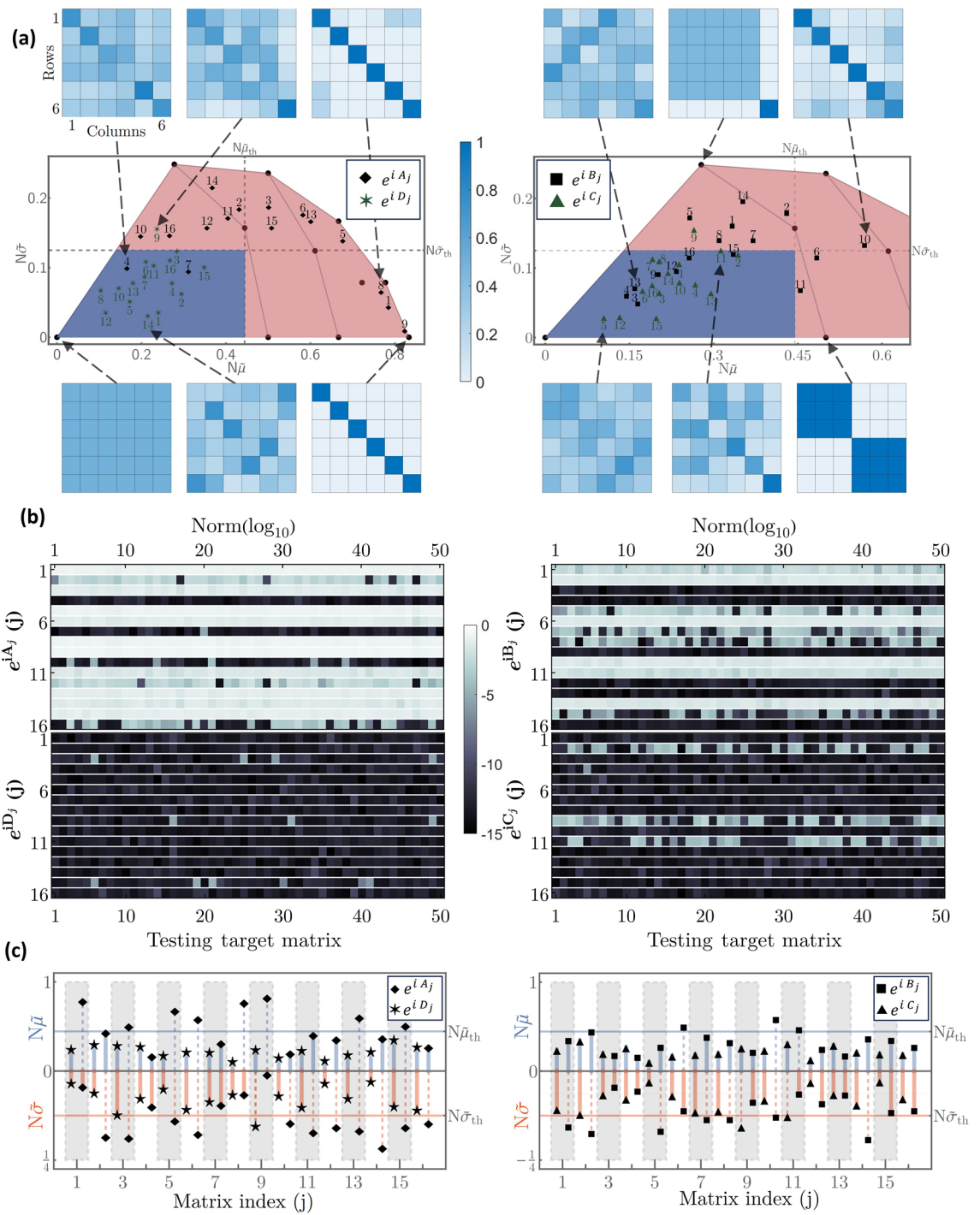
Figure 3. Density estimation and performance test. (**a**) Points $\vec{R}$ associated with density criterion for the set of unitary matrices $\{e^{A_j}, e^{D_j}\}_{j=1}^{50}$ (left column) and $\{e^{B_j}, e^{C_j}\}_{j=1}^{50}$ (right column). The shaded blue area denotes the Goldilocks region where universality is expected for $N = 6$. In turn, the blue heat maps denote the modulus of some particular choices of the unitary $F = e^{iX_j}$ matrices, with $X \in \{A, B, C, D\}$. (**b**) Error norm ($\log_{10}$) L in (2) for each unitary matrix under consideration with fifty testing targets per matrix. (**c**) Mean and standard deviation $N\widetilde{\mu}$ and $N\widetilde{\sigma}$, respectively, related to the density estimation for each unitary matrix in (**a**). The horizontal blue and red lines denote the universality threshold for $N\widetilde{\mu}$ and $N\widetilde{\sigma}$, respectively.

avoiding the vertices, as the latter are the well-known bad-performing cases (with the exception of $\vec{R}_0$). We can go a step further and make a better prediction of unitary matrices by reducing the area of the polygon and imposing a threshold to $N\widetilde{\sigma}$ and $N\vec{\mu}$. For the former, we already know that the standard deviation reaches its maximum

value at $N\widetilde{\sigma}|_{k=2,\ell_\sigma} = 1/4$ for all N. We thus implement the threshold at half of the maximum allowed standard deviation, i.e., $N\widetilde{\sigma}_{th} = 1/8$. For even N, the reference points $\vec{R}_{k=2,\ell\geq\ell_\sigma}$ are above such a threshold for $\ell_\sigma = \frac{N}{2} - \lceil\frac{N}{4}\sqrt{2-\sqrt{3}}\rceil$. Likewise, we fix the threshold for the mean at $N\widetilde{\mu}_{th} = N\widetilde{\mu}|_{k=2,\ell=\ell_\sigma} = 2\ell_\sigma(N-\ell_\sigma)/N^2$.

Therefore, the Goldilocks region is defined as the region spanned by the interception of the polygon spanned by the set of points $\{\vec{R}_0\} \cup \{\vec{R}_{k=2}\}_{\ell=1}^{\lfloor\frac{N}{2}\rfloor}$ and the thresholds $N\widetilde{\mu}_{th}$ and $N\widetilde{\sigma}_{th}$. This region is illustrated in Fig. 3a by the blue-shaded area. Any unitary matrix F whose associated vector $\vec{R}$ lies inside the Goldilocks region is said to fulfill the Goldilocks principle; i.e., the components of F have the statistical properties to be deemed as dense enough to render a universal factorization in (1).

Random matrices and performance test

To test the performance of the Goldilocks principle, we generate sets of random 6×6 unitary matrices (not necessarily Haar random) and determine the corresponding $\vec{R}$ in the plane $(N\widetilde{\mu}, N\widetilde{\sigma})$. To ensure control over the testing matrices, we generate random matrices using the decomposition $F_X = e^{iX}$, with $X^\dagger = X$ a Hermitian matrix in $\mathbb{C}^{N\times N}$ to be defined. Furthermore, we introduce the four non-symmetric matrices $X_A = (\vec{\chi}_{1;A}, \vec{0}, \vec{0}, \vec{0}, \vec{0}, \vec{0})$, $X_B = (\vec{\chi}_{1;B}, \vec{\chi}_{2;B}, \vec{0}, \vec{0}, \vec{0}, \vec{0})$, $X_C = (\vec{\chi}_{1;C}, \vec{\chi}_{2;c}, \vec{\chi}_{3;C}, \vec{0}, \vec{0}, \vec{0})$, and $D = (\vec{\chi}_{1;D}, \vec{\chi}_{2;D}, \vec{\chi}_{3,D}, \vec{\chi}_{4;D}, \vec{0}, \vec{0})$, with $\vec{0}$ the column null-vector in $\mathbb{C}^N$. The column vectors $\vec{\chi}_{j;\wp}$ are composed of zeros in the first j inputs and random numbers elsewhere, with $\wp \in \{A, B, C, D\}$. We thus consider the Hermitian construction as $X = X_\wp + X_\wp^\dagger$, with $\wp \in \{A, B, C, D\}$. In this form, the number of random parameters increases in each case as additional non-null columns are included, rendering random unitary matrices defined by higher number of random parameters. In other words, the density of the random matrices is expected to be higher for D than A, B, C.

In this form, we establish a controlled benchmark for the Goldilocks principle and the subsequent universality of the matrix under consideration. Here, 50 target unitary matrices are considered for each testing unitary matrix F_{X_j} so that the relative error of the optimized targets from (1) and the corresponding vector $\vec{R}_{X_j}$ can be analyzed for a broad number of cases. Particularly, Fig. 3a depicts the reference points $\vec{R}_k$ (filled-circles) and the points $\vec{R}_{X_j}$ associated with the sets of random unitary matrices F_{X_j} for $X \in \{A, B, C, D\}$ and $j \in \{1, \ldots, 50\}$. This allows determining which unitary matrices have points $\vec{R}_{j;\wp}$ lying inside the Goldilocks region. One may notice that, for the random unitary matrices F_A, only the points A_4 and A_7 are expected to fulfill the Goldilocks principle. In turn, we expect more well-behaved matrices F_D, with only D_9 outside of the Goldilocks region. This is corroborated in Fig. 3b, where the optimization routine using the LMA has been implemented for each testing target matrix. Indeed, the error norm for the matrices A_4 and A_7 render values within the preestablished tolerance values, as predicted in Fig. 3a. On the other hand, the numerical optimization also reveals that A_{10} and D_9 should be good candidates, whereas the density criterion has ruled them out. This is an example of false-negative outcomes. As seen from both Fig. 3b,c, this is usually the case for matrices F with $\vec{R}_F$ lying in the vicinity of the Goldilocks region. Due to the existence of false-negative results, we deem the density criterion as only a sufficient condition. Likewise, a similar analysis can be carried out for the testing matrices F_{B_j} and F_{C_j}. To complement the analysis, and to better visualize and assess the Goldilocks region, we put forward an alternative representation in Fig. 3c, where we depict both $N\widetilde{\mu}$ and $N\widetilde{\sigma}$ separately. Here, the Goldilocks principle is established if both quantities lie below their respective thresholds $N\widetilde{\mu}$ and $N\widetilde{\sigma}$. In this form, it is no longer required to draw the universality region, and one can assess the Goldilocks principle in a simple plot.

As previously discussed, the the computational time required to optimize (1) and test the corresponding g universality for a given choice of F increases with the total number of ports N. In turn, the identification of the Goldilocks region and the associated vector $\vec{R}$ for a specific matrix F enables a quick classification for preselecting the feasibility of the matrix. Furthermore, the Goldilocks region spanned in the $\vec{R}$-space is finite and non-null for $N \to \infty$, making it a suitable measure for architectures with an arbitrary number of ports.

Photonic platform and feasible realizations

So far, the universality of the proposed architecture has been numerically established using different choices of the intervening matrix F as the passive mixing layer possessing the required density criterion. In the following, we will discuss potential candidates for creating F matrices using photonic systems, with a specific emphasis on photonic lattices and meshes of directional couplers. Such systems can be readily implemented with silicon photonics. One can build such structures, e.g., using buried silicon waveguides at the telecommunication wavelength of 1550nm. The waveguide system comprises a silicon (Si) core surrounded by a silica (SiO2) cladding and substrate with refractive indices $n_{Si} = 3.47$ and $n_{SiO2} = 1.47$, respectively. The core dimensions are 500nm in width and 220nm in height. Such a geometry renders the fundamental quasi-TE01 mode with an effective mode index of $n_{mode}^{(eff)} = 2.4456$, which is the operational mode used for the unitary devices discussed in the following. This applies to passive F matrix solutions based on both waveguide arrays and directional coupler meshes. For the active layers, phase shifters based on thermo-optical effects can be considered. This might include solutions based on metal heaters[55,56], which are widely implemented by open-access foundries and occupy an approximate area of $370\,\mu m \times 30\,\mu m$. Alternatively, one can consider ultra-compact phase shifters based on phase-change materials (PCMs)[7], achieving phase shifts of approximately $\pi/11$ radians per one-micron length[57]. Although the

latter is not as broadly implemented as metal-heater solutions in foundries, it lights up the path for the future of dense programmable photonic chips.

Waveguide lattices

Waveguide arrays can be modeled with high precision using coupled-mode theory[58]. The latter takes into account the coupling of evanescent waves from one waveguide interacting with its nearest neighbor while neglecting farther neighbors due to their weak coupling. In this form, the effective Hamiltonian describing an array of N waveguides is characterized by a tridiagonal and symmetric matrix Hamiltonian $\mathbb{H}$ of dimension N. The wave evolution through the lattice is ruled by the dynamical law $i\frac{d}{dz}\vec{u}(z) = \mathbb{H} \cdot \vec{u}(z)$, where $\vec{u}(z) \in \mathbb{C}^N$ is the mode field amplitude at each waveguide at the propagation distance z. Since $\mathbb{H} \neq \mathbb{H}(z)$, the wave evolution is determined through the unitary evolution operator $\mathbb{F}(z) = e^{-iz\mathbb{H}}$ as $\vec{u}(z) = \mathbb{F}(z)\vec{u}(z=0)$.

We can thus implement waveguide arrays in the universal architecture, provided they fulfill the desired universality. To this end, we can test the behavior of a given lattice evolution operator for specific lengths using the Goldilocks principle. Particularly, we consider the photonic J_x lattice[46], the homogeneous lattice[59], and the disordered homogeneous lattice[60] as the physical waveguide arrays under consideration described by the respective Hamiltonians $\mathbb{H}^{(J_x)}$, $\mathbb{H}^{(h)}$, and $\mathbb{H}^{(h,d)}$. The matrix elements of the latter are explicitly given by

$$\begin{aligned}
\mathbb{H}^{(J_x)}_{p,q} &:= \kappa(p)\delta_{p+1,q} + \kappa(p-1)\delta_{p-1,q}, \\
\mathbb{H}^{(h)}_{p,q} &:= \kappa_0\delta_{p+1,q} + \kappa_0\delta_{p-1,q}, \\
\mathbb{H}^{(h,d)}_{p,q} &:= (\kappa_0 + \Delta\kappa_p)\delta_{p+1,q} + (\kappa_0 + \Delta\kappa_{p-1})\delta_{p-1,q},
\end{aligned}$$

(4)

with $p, q \in \{1, \ldots, N\}$. Here, $\kappa(p) = \frac{\kappa_0}{2}\sqrt{(N-p)p}$ stands for the coupling parameter between nearest waveguide neighbors in the J_x lattice, where κ_0 is a design scaling factor, and $\Delta\kappa_p \in N(\mu, \sigma)$ are random numbers taken from the *normal distribution* $N(\mu, \sigma)$ characterizing the disorder effects. The coupled waveguide implementation for each Hamiltonian is depicted in Fig. 4.

The corresponding unitary evolution operators are simply given by $\mathbb{F}^{(J_x)}(z) = e^{iz\mathbb{H}^{(J_x)}}$ and $\mathbb{F}^{(h)}(z) = e^{iz\mathbb{H}^{(h)}}$. Although both are functions of the lattice length z, the J_x lattice (Fig. 4a) has equidistant eigenvalues that lead to a periodic unitary evolution operator $\mathbb{F}^{(J_x)}(z)$ in z, so that we can simply focus on the interval $z \in [0, 2\pi)$. We first estimate the lengths that induce universality in our architecture, which is depicted in Fig. 4b. In the latter, we mark the particular lengths $z_j^{(m)}$ and $z_j^{(M)}$ that denote the local minima and maxima of the standard deviation $N\widetilde{\sigma}$, the exact values of which have been determined numerically and presented in Table 1. In turn, the black-thick line in Fig. 4b denotes the lengths where both $N\widetilde{\mu}$ and $N\widetilde{\sigma}$ are below the universality threshold; i.e., the lengths where universality is expected. Without any prior performance test, one can see that the lengths
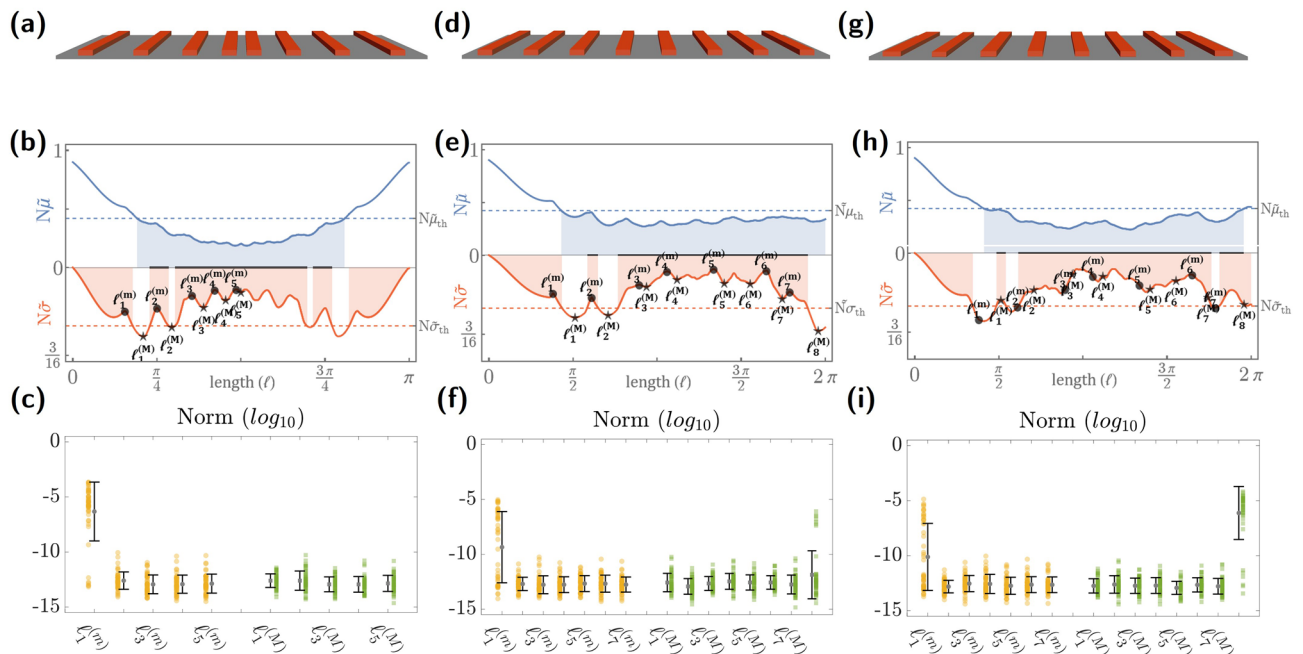

Figure 4. Photonic platform and lattice universality. Sketch for the waveguide array associated with the J_x lattice (**a**), homogeneous lattice (**d**), and homogenous lattice with disorder effects (**e**). Density criterion as a function of the lattice length ℓ for $N = 10$ considering the J_x (**b**), homogeneous (**e**), and disordered (**h**) lattices. The corresponding numerical performance test for $N = 10$ at the reference lengths $\ell_j^{(m)}$ and $\ell_j^{(M)}$ for the J_x (**c**), homogeneous (**f**), and disordered (**i**) lattices.

z	J_x	Homogeneous	z	J_x	Homogeneous
$\ell_1^{(m)}$	0.48781	1.1968	$\ell_1^{(M)}$	0.6579	1.6099
$\ell_2^{(m)}$	0.78735	1.9211	$\ell_2^{(M)}$	0.9256	2.2285
$\ell_3^{(m)}$	1.11393	2.8088	$\ell_3^{(M)}$	1.2227	2.9441
$\ell_4^{(m)}$	1.32647	3.3290	$\ell_4^{(M)}$	1.4293	3.5156
$\ell_5^{(m)}$	1.52812	4.1988	$\ell_5^{(M)}$	$\pi/2$	4.3984
$\ell_6^{(m)}$		5.1835	$\ell_6^{(M)}$		4.8832
$\ell_7^{(m)}$		5.6182	$\ell_7^{(M)}$		5.4811
			$\ell_8^{(M)}$		6.1544

Table 1. Lattice lengths at the local minima $\ell_j^{(m)}$ and local maxima $\ell_j^{(M)}$ of the density criterion $N\widetilde{\sigma}$ for the J_x and homogeneous lattice.

$z_1^{(m)}$, $z_1^{(M)}$, and $z_2^{(M)}$ may fail in obtaining the desired universality. Recall that our estimation criterion is only a sufficient condition and may rule out positive cases. Nevertheless, all the other marked points can be considered candidates for the matrix F in our architecture, as no false positive cases will be included. This is indeed verified in the performance test portrayed in Fig. 4c, where, for each point, we have used fifty randomly generated unitary matrices as targets. The latter confirms our predictions, where the only bad-performing length is found at $z_1^{(M)}$, reinforcing the fact that only two positive cases were discarded, but no false positives were included. In this form, we can confidently conclude that a universal architecture can be built using J_x lattices with lengths as small as $z = \pi/4$ for $N = 10$.

We alternatively consider the homogeneous lattice, which contains waveguide arrays homogeneously distributed (Fig. 4d). The eigenvalues accordingly distributed as $\hat{\lambda}_n^{(h)} = 2\kappa_0 \cos(\frac{n\pi}{N+1})$, and no periodic behavior is expected. We thus focus on the interval $\ell \in [0, 4\pi]$ for this particular lattice. The density criterion shown in Fig. 4e reveals that lattice lengths in the interval $z/\kappa_0 \in (2.4098, 5.9595)$ are suitable for our universal architecture. Particularly, note that the interval $z/\kappa_0 \in [\ell_4^{(m)}, \ell_6^{(m)}]$ contains lengths so that $N\widetilde{\mu}$ and $N\widetilde{\sigma}$ remain mostly constant with minor variations. Thus, the performance test in this interval is expected to perform well. In the Goldilocks region, there is a local minimum $\ell_2^{(m)}$ isolated and associated with a shorter lattice length. This reference point may be useful for reducing the size of the universal structure. Figure 4f displays the corresponding performance test, which supports our previous statements. As expected, the performance for lattices with length $\ell_1^{(m)}$ and $\ell_8^{(M)}$ is particularly poor. However, the length $\ell_1^{(M)}$ has a generally good performance, with only two test targets displaying slightly higher errors than the other well-performing cases.

We additionally take into account the effects of disorder on the homogeneous lattice, which may be caused by impurities or imperfections during the manufacturing process, resulting in waveguides not being in their ideal positions or displaying deviations in their sizes (see Fig. 4g). The defects here considered are such that the nearest-neighbor interactions deviate from the ideal homogeneous lattice by a factor of twenty percent; i.e., the disorder couplings in (4) take values from the normal distribution as $\Delta\kappa_p \in N(\mu = 0, \sigma = 0.2\kappa_0)$. Although the lattice structure is modified in the latter disorder, the estimation of density does not differ significantly from the ideal case, as shown in Fig. 4h–i.

Directional coupler mesh
Alternatively, the matrix F can be optically realized through proper mesh arrays of directional couplers. Particularly, we consider a construction based on two-port passive elements, which act as a power divider (3-dB directional coupler) equivalent, up to a global phase, to the unitary matrix $U(2) \ni \mathscr{T}_0 = \frac{1}{\sqrt{2}}(\sigma_0 - i\sigma_1)$, with σ_j the conventional Pauli matrices for $j \in \{1, 2, 3\}$ and σ_0 the 2×2 identity matrix. The latter can be used as a building block to construct other $U(2)$ matrices[61] as well as higher dimensional unitary matrices $U(N)$ through appropriate Kronecker products[22]. The silicon photonics platform discussed above allows the implementation of each 3-dB directional coupler through a coupling length and waveguide separation of 29.35 μm and 630 nm, respectively. In this section, we consider the symmetric 10-port array portrayed in Fig. 5a, composed of power dividers $\mathscr{T}_0$ interconnected through different layers $\mathscr{L}_1$ and $\mathscr{L}_2$. Here, each layer is described by the $U(10)$ matrices $\mathscr{L}_1 = \mathbb{I}_5 \otimes \mathscr{T}_0$ and $\mathscr{L}_2 = \mathbb{I}_1 \oplus (\mathbb{I}_4 \otimes \mathscr{T}_0) \oplus \mathbb{I}_1$, with $\otimes$ and $\oplus$ the Kronecker product and Kronecker sum (Such operations are also known as direct product and direct sum.), respectively, and $\mathbb{I}_n$ the $n \times n$ identity matrix. The p-layered unitary matrix describing the power divider array is thus given by

$$\mathbb{F}_p = \underbrace{\mathscr{L}_{\widetilde{p}} \dots \mathscr{L}_1 \mathscr{L}_2 \mathscr{L}_1}_{p\text{-times}}, \quad \widetilde{p} = \begin{cases} 1, & p \in \{1, 3, 5, 7, 9\} \\ 2, & p \in \{2, 4, 6, 8, 10\} \end{cases}, \tag{5}$$

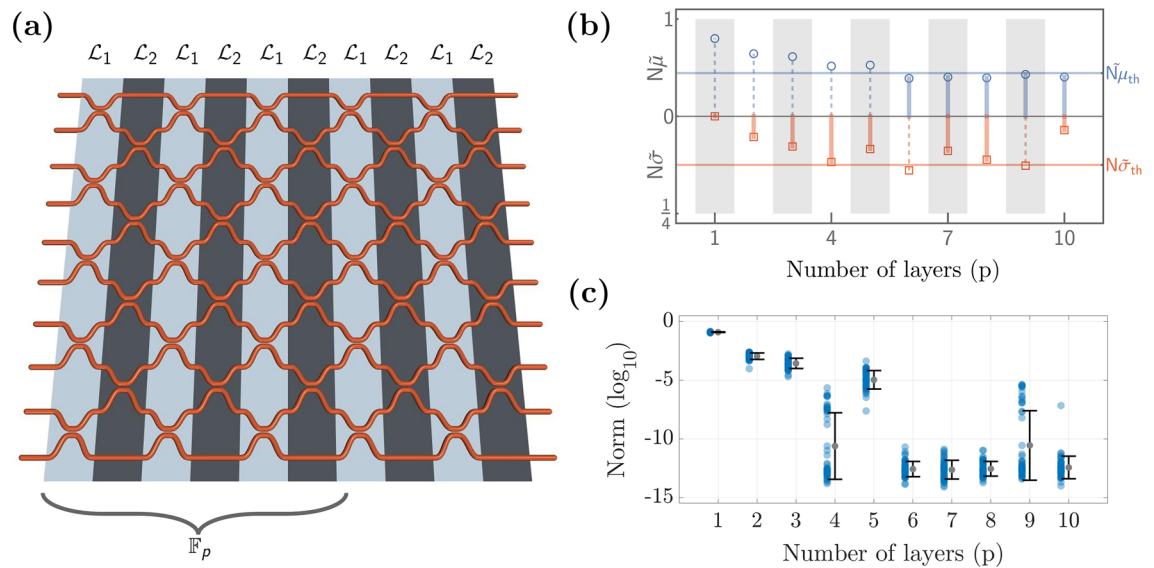where we have truncated the maximum number of layers to ten.

Figure 5. Geometric array for the passive matrix F using power dividers (3-dB directional coupler). (**a**) Power divider array composed of p layers as defined in (5). Light-shaded and dark-shaded layers denote $\mathscr{L}_1 = \mathbb{I}_5 \otimes \mathscr{T}_0$ and $\mathscr{L}_2 = \mathbb{I}_1 \oplus (\mathbb{I}_4 \otimes \mathscr{T}_0) \oplus \mathbb{I}_1$, respectively. Density criterion (**b**) and error norm ($\log_{10} L$) (**c**) of the mesh architecture in (**a**) as a function of the number of layers p.

It is not mandatory to truncate these layers, and the procedure can involve additional layers if needed. However, for practical physical implementations and to reduce the device footprint, we aim for devices with a minimal number of layers. To this end, we estimate the Goldilocks region in Fig. 5b as a function of the number of layers p, revealing that $p = 7, 8, 10$, are indeed good candidates. This is further corroborated through the LMA optimization results shown in Fig. 5c, which indeed shows that $p = 7, 8, 10$ layers render a well-performing F layer. Notably, the latter also indicates that $p = 6$ is also a valid choice. The case $p = 6$ was originally deemed inadequate from the Goldilocks principle, but Fig. 5b shows that $N\widetilde{\sigma}$ is in the vicinity of the threshold, which, as discussed above, usually renders false negative outcomes. Despite the latter, no false positives were detected during the analysis; that is, the Goldilocks principle did not show good-performing cases that contained high error norms L. This is strictly necessary to avoid faulty designs in the final architecture.

Conclusions

We have introduced the design for a lossless universal photonic architecture based on a layered scheme of interlaced active phase shifter layers and passive random matrices. Numerical results obtained from the LMA optimization revealed that generating Haar random matrices F leads, in a vast majority of cases, to the desired universal architecture. It is observed that well-behaved matrices F show a phase transition on the error norm L of the reconstructed target at $M = N + 1$, with M the total number of phase shifter layers. In such a layer number, the error drops significantly to numerical noise values. While this is not proof that the factorization is exact, the error involved in the reconstruction process lies in the numerical error regime, and it is thus low enough to ensure that any unitary matrix is reconstructed with the desired accuracy.

Despite the accuracy of the LMA optimization, the computational time required for testing the universality of the random matrices F scales with the total number of ports, which becomes impractical for particularly large architectures. Numerical evidence shows that denser matrices perform better than sparse ones, usually involving relatively large errors. Therefore, a density criterion has been devised and introduced to classify the candidates for the matrix F used in the architecture. This criterion is built on preliminary knowledge of bad-performing matrices, such as diagonal and block diagonal matrices, which are analytically known to fail but serve as reference points to look for good-performing matrices, such as the DFT case. In this form, instead of performing a long optimization routine on the candidate for F, we simply analyze the standard deviation of the modulus of its columns or rows, which provides information about its density. This allows defining a mapping $\vec{R} : U(N) \to \mathbb{R}^2$, which renders a vector that estimates whether F is suitable for the architecture. We thus possess a tool to preselect matrices F beforehand, making the design process more practical than generating and testing several random matrices.

Our tests using randomly generated unitary matrices showed that matrices within the threshold marked by the density criterion led to the required universality. Thus, universality is not limited to a specific realization of F; as shown in the results, infinitely many unitary matrices can meet our requirements. This paves the way for more efficient construction and optimization of compact devices that are simultaneously resilient to random defects. Particularly, the photonic Jx lattice was found suitable for this task at lengths different than the previously reported critical value $\ell = \pi/2$[37]. This defines intervals in the lattice length for which the architecture is universal, leading to more flexibility in the manufacturing process so that one can allow for deviations in lattice length.

This fact is further supported in the context of homogeneous lattices, which are also suitable for our architecture and robust against disorder effects due to waveguide impurities or mismatching sizes. The latter was tested by introducing deviations of up to 20% into the homogeneous lattice, from which the density estimation showed no significant difference in the universality performance for the lattice lengths considered. Further constructions for the F matrices are indeed allowed, and an alternative construction based on a layered array of power dividers was shown to be efficient for our purposes, the analysis of which allowed us to determine the optimal number of passive elements required for the architecture.

The universality of the interlaced architecture in (1) can be further assessed using the density criterion for optical implementations of the passive matrix F beyond the waveguide arrays and directional coupler mesh discussed in the manuscript, as long as F is described by a unitary matrix. For instance, the DFT can be implemented using MMI couplers using the self-image property[48]. The MMI construction of the DFT has been used to develop an interlaced unitary akin to (1), but in such a construction, it was analytically proved that $6N + 1$ phase layers are required using a DFT as the intervening layer[12]. Our results in Fig. 2b provide numerical evidence that such a construction can be realized with only $N + 1$ phase layers with the prescribed accuracy. Furthermore, a microwave implementation of the DFrFT has been theoretically and experimentally validated[62], offering an alternative approach for the interlaced architecture. The latter helps facilitate the design of the passive layer to reduce the overall architecture size, account for potential manufacturing errors, and diminish the device footprint. This is particularly handy when deploying more complex optical circuits.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References
1. Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: Putting a new twist on light. *Nature* **386**, 143–149. https://doi.org/10.1038/386143a0 (1997).
2. Yan, R., Gargas, D. & Yang, P. Nanowire photonics. *Nat. Photon.* **3**, 569–576. https://doi.org/10.1038/nphoton.2009.184 (2009).
3. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446. https://doi.org/10.1038/nphoton.2017.93 (2017).
4. Zelaya, K. & Miri, M.-A. Integrated photonic fractional convolution accelerator (2023). ArXiv:2307.10976 [physics.optics]
5. Liao, K. *et al.* Matrix eigenvalue solver based on reconfigurable photonic neural network. *Nanophotonics* **11**, 4089–4099. https://doi.org/10.1515/nanoph-2022-0109 (2022).
6. Mao, J.-Y., Zhou, L., Zhu, X., Zhou, Y. & Han, S.-T. Photonic memristor for future computing: A perspective. *Adv. Opt. Mater.* **7**, 1900766. https://doi.org/10.1002/adom.201900766 (2019).
7. Youngblood, N., Ríos Ocampo, C. A., Pernice, W. H. & Bhaskaran, H. Integrated optical memristors. *Nat. Photon.* https://doi.org/10.1038/s41566-023-01217-w *(2023)*.
8. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506. https://doi.org/10.1038/s41586-022-04714-0 (2022).
9. Liao, K., Dai, T., Yan, Q., Hu, X. & Gong, Q. Integrated photonic neural networks: Opportunities and challenges. *ACS Photon.* https://doi.org/10.1021/acsphotonics.2c01516 *(2023)*.
10. Saygin, M. Y. *et al.* Robust architecture for programmable universal unitaries. *Phys. Rev. Lett.* **124**, 010501. https://doi.org/10.1103/PhysRevLett.124.010501 (2020).
11. Zhou, H. *et al.* Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci. Appl.* **11**, 30. https://doi.org/10.1038/s41377-022-00717-8 (2022).
12. Pastor, V. L., Lundeen, J. & Marquardt, F. Arbitrary optical wave evolution with Fourier transforms and phase masks. *Opt. Express* **29**, 38441–38450. https://doi.org/10.1364/OE.432787 (2021).
13. Madsen, L. S. *et al.* Quantum computational advantage with a programmable photonic processor. *Nature* **606**, 75–81. https://doi.org/10.1038/s41586-022-04725-x (2022).
14. Harris, N. C. *et al.* Quantum transport simulations in a programmable nanophotonic processor. *Nat. Photon.* **11**, 447–452. https://doi.org/10.1038/nphoton.2017.95 (2017).
15. Slussarenko, S. & Pryde, G. J. Photonic quantum information processing: A concise review. *Appl. Phys. Rev.* https://doi.org/10.1063/1.5115814 *(2019)*.
16. Notaros, J. *et al.* Programmable dispersion on a photonic integrated circuit for classical and quantum applications. *Opt. Express* **25**, 21275–21285. https://doi.org/10.1364/OE.25.021275 (2017).
17. Xu, X. *et al.* Neuromorphic computing based on wavelength-division multiplexing. *IEEE J. Sel. Top. Quantum Electron.* **29**, 1–12. https://doi.org/10.1109/JSTQE.2022.3203159 (2023).
18. Zhu, H. *et al.* Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044. https://doi.org/10.1038/s41467-022-28702-0 (2022).
19. Li, X.-K. *et al.* High-efficiency reinforcement learning with hybrid architecture photonic integrated circuit. *Nat. Commun.* **15**, 1044. https://doi.org/10.1038/s41467-024-45305-z (2024).
20. Tang, R. *et al.* Two-layer integrated photonic architectures with multiport photodetectors for high-fidelity and energy-efficient matrix multiplications. *Opt. Express* **30**, 33940–33954. https://doi.org/10.1364/OE.457258 (2022).
21. Xu, S. *et al.* Parallel optical coherent dot-product architecture for large-scale matrix multiplication with compatibility for diverse phase shifters. *Opt. Express* **30**, 42057–42068. https://doi.org/10.1364/OE.471519 (2022).
22. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58. https://doi.org/10.1103/PhysRevLett.73.58 (1994).
23. Miller, D. A. All linear optical devices are mode converters. *Opt. Express* **20**, 23985–23993. https://doi.org/10.1364/OE.20.023985 (2012).
24. Miller, D. A. Self-configuring universal linear optical component. *Photon. Res.* **1**, 1–15. https://doi.org/10.1364/PRJ.1.000001 (2013).
25. Miller, D. A. Self-aligning universal beam coupler. *Opt. Express* **21**, 6360–6370. https://doi.org/10.1364/OE.21.006360 (2013).
26. 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. https://doi.org/10.1364/OPTICA.3.001460 (2016).

27. Burgwal, R. *et al.* Using an imperfect photonic network to implement random unitaries. *Opt. Express* **25**, 28236–28245. https://doi.org/10.1364/OE.25.028236 (2017).
28. Shokraneh, F., Geoffroy-Gagnon, S. & Liboiron-Ladouceur, O. The diamond mesh, a phase-error-and loss-tolerant field-programmable mzi-based optical processor for optical neural networks. *Opt. Express* **28**, 23495–23508. https://doi.org/10.1364/OE.395441 (2020).
29. Mojaver, K. H. R., Zhao, B., Leung, E., Safaee, S. M. R. & Liboiron-Ladouceur, O. Addressing the programming challenges of practical interferometric mesh based optical processors. *Opt. Express* **31**, 23851–23866. https://doi.org/10.1364/OE.489493 (2023).
30. On, M. B. et al. Programmable integrated photonics for topological hamiltonians (2023). Arxiv:2307.05003 [physics.optics]
31. Wang, M. *et al.* Topologically protected entangled photonic states. *J. Nanophoton.* **8**, 1327–1335. https://doi.org/10.1515/nanoph-2019-0058 (2019).
32. Labroille, G. *et al.* Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion. *Opt. Express* **22**, 15599–15607. https://doi.org/10.1364/OE.22.015599 (2014).
33. Morizur, J.-F. *et al.* Programmable unitary spatial mode manipulation. *JOSA A* **27**, 2524–2531. https://doi.org/10.1364/JOSAA.27.002524 (2010).
34. Kulce, O., Mengu, D., Rivenson, Y. & Ozcan, A. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. *Light Sci. Appl.* **10**, 196. https://doi.org/10.1038/s41377-021-00623-5 (2021).
35. Tanomura, R., Tang, R., Ghosh, S., Tanemura, T. & Nakano, Y. Robust integrated optical unitary converter using multiport directional couplers. *J. Lightw. Technol.* **38**, 60–66. https://doi.org/10.1109/JLT.2019.2943116 (2020).
36. Tanomura, R. *et al.* Scalable and robust photonic integrated unitary converter based on multiplane light conversion. *Phys. Rev. Appl.* **17**, 024071. https://doi.org/10.1103/PhysRevApplied.17.024071 (2022).
37. Markowitz, M. & Miri, M.-A. Universal unitary photonic circuits by interlacing discrete fractional fourier transform and phase modulation (2023). ArXiv:2307.07101 [physics.optics]
38. Ribeiro, A., Ruocco, A., Vanacker, L. & Bogaerts, W. Demonstration of a 4 × 4-port universal linear circuit. *Optica* **3**, 1348–1357. https://doi.org/10.1364/OPTICA.3.001348 (2016).
39. Taballione, C. *et al.* 8 × 8 reconfigurable quantum photonic processor based on silicon nitride waveguides. *Opt. Express* **27**, 26842–26857. https://doi.org/10.1364/OE.27.026842 (2019).
40. Tang, R., Tanomura, R., Tanemura, T. & Nakano, Y. Ten-port unitary optical processor on a silicon photonic chip. *ACS Photon.* **8**, 2074–2080. https://doi.org/10.1021/acsphotonics.1c00419 (2021).
41. Taballione, C. *et al.* A universal fully reconfigurable 12-mode quantum photonic processor. *Mater. Quant. Technol.* **1**, 035002. https://doi.org/10.1088/2633-4356/ac168c (2021).
42. Tanomura, R., Tanomura, T. & Nakano, Y. Multi-wavelength dual-polarization optical unitary processor using integrated multiplane light converter. *Jpn. J. Appl. Phys.* **62**, 1029. https://doi.org/10.35848/1347-4065/acab70 (2023).
43. Skryabin, N., Dyakonov, I., Saygin, M. Y. & Kulik, S. Waveguide-lattice-based architecture for multichannel optical transformations. *Opt. Express* **29**, 26058–26067. https://doi.org/10.1364/OE.426738 (2021).
44. Markowitz, M., Zelaya, K. & Miri, M.-A. Auto-calibrating universal programmable photonic circuits: Hardware error-correction and defect resilience. *Opt. Express* **31**, 37673–37682. https://doi.org/10.1364/OE.502226 (2023).
45. Markowitz, M., Zelaya, K. & Miri, M.-A. Learning arbitrary complex matrices by interlacing amplitude and phase masks with fixed unitary operations (2023). ArXiv:2312.05648 [physics.optics]
46. Weimann, S. *et al.* Implementation of quantum and classical discrete fractional Fourier transforms. *Nat. Commun.* **7**, 11027. https://doi.org/10.1038/ncomms11027 (2016).
47. Cooney, K. & Peters, F. H. Analysis of multimode interferometers. *Opt. Express* **24**, 22481–22515. https://doi.org/10.1364/OE.24.022481 (2016).
48. Bachmann, M., Besse, P. A. & Melchior, H. General self-imaging properties in n × n multimode interference couplers including phase relations. *Appl. Opt.* **33**, 3905–3911. https://doi.org/10.1364/AO.33.003905 (1994).
49. Taguchi, Y., Wang, Y., Tanomura, R., Tanemura, T. & Ozeki, Y. Iterative configuration of programmable unitary converter based on few-layer redundant multiplane light conversion. *Phys. Rev. Appl.* **19**, 054002. https://doi.org/10.1103/PhysRevApplied.19.054002 (2023).
50. Mezzadri, F. How to generate random matrices from the classical compact groups (2007). ArXiv:math-ph/0609050
51. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944).
52. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11**, 431–441 (1963).
53. Tang, R., Tanemura, T. & Nakano, Y. Integrated reconfigurable unitary optical mode converter using mmi couplers. *IEEE Photon. Technol. Lett.* **29**, 971–974. https://doi.org/10.1109/LPT.2017.2700619 (2017).
54. Livan, G., Novaes, M. & Vivo, P. *Introduction to Random Matrices: Theory and Practice* (Springer, 2018).
55. Harris, N. C. *et al.* Efficient, compact and low loss thermo-optic phase shifter in silicon. *Opt. Express* **22**, 10487–10493. https://doi.org/10.1364/OE.22.010487 (2014).
56. Liu, S. *et al.* Thermo-optic phase shifters based on silicon-on-insulator platform: State-of-the-art and a review. *Front. Optoelectron.* **15**, 9. https://doi.org/10.1007/s12200-022-00012-9 (2022).
57. Ríos, C. *et al.* Ultra-compact nonvolatile phase shifter based on electrically reprogrammable transparent phase change materials. *PhotoniX* **3**, 26. https://doi.org/10.1186/s43074-022-00070-4 (2022).
58. Huang, W.-P. Coupled-mode theory for optical waveguides: An overview. *JOSA A* **11**, 963–983. https://doi.org/10.1364/JOSAA.11.000963 (1994).
59. Christodoulides, D. N., Lederer, F. & Silberberg, Y. Discretizing light behaviour in linear and nonlinear waveguide lattices. *Nature* **424**, 817–823. https://doi.org/10.1038/nature01936 (2003).
60. Miri, M.-A. Integrated random projection and dimensionality reduction by propagating light in photonic lattices. *Opt. Lett.* **46**, 4936–4939. https://doi.org/10.1364/OL.433101 (2021).
61. de Guise, H., Di Matteo, O. & Sánchez-Soto, L. L. Simple factorization of unitary transformations. *Phys. Rev. A* **97**, 022328. https://doi.org/10.1103/PhysRevA.97.022328 (2018).
62. Keshavarz, R., Shariati, N. & Miri, M.-A. Real-time discrete fractional Fourier transform using metamaterial coupled lines network. *IEEE Trans. Microw. Theory Tech.* **71**, 3414–3423. https://doi.org/10.1109/TMTT.2023.3278929 (2023).

Acknowledgements

Author contributions

All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interest

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.