Research Article

Jasvith Raj Basani*, Sri Krishna Vadlamani, Saumil Bandyopadhyay, Dirk R. Englund and Ryan Hamerly

A self-similar sine–cosine fractal architecture for multiport interferometers

Abstract: Multiport interferometers based on integrated beamsplitter meshes have recently captured interest as a platform for many emerging technologies. In this paper, we present a novel architecture for multiport interferometers based on the sine–cosine fractal decomposition of a unitary matrix. Our architecture is unique in that it is self-similar, enabling the construction of modular multi-chiplet devices. Due to this modularity, our design enjoys improved resilience to hardware imperfections as compared to conventional multiport interferometers. Additionally, the structure of our circuit enables systematic truncation, which is key in reducing the hardware footprint of the chip as well as compute time in training optical neural networks, while maintaining full connectivity. Numerical simulations show that truncation of these meshes gives robust performance even under large fabrication errors. This design is a step forward in the construction of large-scale programmable photonics, removing a major hurdle in scaling up to practical machine learning and quantum computing applications.

Keywords: integrated photonics; optical information processing; optical interferometry.

---

*Corresponding author: Jasvith Raj Basani, Department of Electrical and Computer Engineering, Institute for Research in Electronics and Applied Physics, and Joint Quantum Institute, University of Maryland, College Park 20742, USA, E-mai: jasvith@umd.edu. https://orcid.org/0000-0001-8453-8104
Sri Krishna Vadlamani, **Saumil Bandyopadhyay and Dirk R. Englund**, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. https://orcid.org/0000-0002-7761-337X (S. Bandyopadhyay)
Ryan Hamerly, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; and PHI Laboratories, NTT Research Inc., 940 Stewart Drive, Sunnyvale, CA 94085, USA. https://orcid.org/0000-0003-4976-2236

1 Introduction

Photonic integrated circuits (PICs) have recently captured interest as a promising time- and energy-efficient platform for classical and quantum optical information processing. They have been used to accelerate tasks in signal processing [1–5], machine learning [6, 7], optimization [8], and quantum simulation [9–13]. Scaling these systems up in order to tackle real-world problems requires careful attention to issues such as the effect of analog component imperfections on performance, and the scaling of chip area with system size.

For instance, it has been shown that the test accuracy of optical neural networks (ONNs) based on Mach–Zehnder Interferometer (MZI) meshes [6] drops rapidly as soon as the constituent beamsplitters deviate from 50 to 50 splitting ratio by a couple of percent [14]. A variety of error-correction techniques have been proposed for the MZI-based platform—global optimization [15–20], local correction [21, 22], self-configuration [23–26], and hardware augmentation [27–29]. The behavior of many of these techniques can be better understood by considering an important insight derived in Ref. [21]—MZIs with imperfect beamsplitters implement only a subset of all the 2×2 unitary matrices that a perfect MZI can implement. This fact explains the observed imperfection-induced reduction in ONN performance since circuits composed of imperfect MZIs implement fewer functions than those with perfect MZIs [6, 14].

We show in this paper that the extent of reduction of the expressivity of a faulty MZI mesh depends strongly on its geometry and that a careful choice of mesh geometry can significantly soften the negative impact of hardware errors. We do so by introducing a novel self-similar MZI-mesh architecture based on the recursive sine–cosine unitary decomposition of Polcari [30] and demonstrating that it is more robust to MZI errors than the conventional Reck (triangular) [31] and Clements (rectangular) [32] mesh geometries. The recursive sine–cosine decomposition [30] is a generalization of the standard FFT decomposition

of Fourier transform matrices [33] to arbitrary unitary matrices. We shall refer to MZI meshes constructed using this decomposition as sine–cosine fractal (SCF) meshes. Like the FFT mesh [34, 35], the SCF mesh has a recursive, self-similar structure; the FFT mesh can in fact be obtained from the SCF mesh by the mere pruning (omission) of certain columns of MZIs.

While SCF meshes have greater error robustness than other architectures,they can also be systematically shrunk in size for use in machine learning applications. In analogy with pruning in conventional neural networks [36, 37], we introduce a systematic mesh pruning scheme that interpolates between the simple FFT and the full sine–cosine fractal, and numerically demonstrate that ONNs composed of pruned meshes still achieve excellent performance at benchmark learning tasks.

The paper is organized as follows: Section 2 introduces and discusses the sine–cosine fractal architecture. Section 3 contains analytical and numerical results on the expressivity of SCF meshes in the presence of beamsplitter errors. Section 4 reports the performance of ONNs constructed from both complete and pruned SCF meshes, and Section 5 concludes the paper with a further discussion of the scope and impact of our work.

2 The sine–cosine fractal architecture

In order to construct a photonic circuit that implements a given unitary matrix U, one first decomposes U into a product of 2×2 unitary matrices and diagonal phase shifts. MZIs are used to implement the 2×2 unitary matrices in the hardware; the transfer function of an MZI with two phase-shifters θ, ϕ (Figure 1(a)) is given by:

$$
\begin{aligned}
T(\theta, \phi) &= \frac{1}{2} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{bmatrix} \\
&= ie^{i\theta/2} \begin{bmatrix} e^{i\phi} \sin(\theta/2) & \cos(\theta/2) \\ e^{i\phi} \cos(\theta/2) & -\sin(\theta/2) \end{bmatrix}
\end{aligned} \tag{1}
$$

The arrangement of MZIs in the circuit, that is, the geometry of the MZI mesh, determines the order of appearance of the corresponding 2×2 unitaries in the decomposition. Figure 1(a) depicts a mesh that implements an 8×8 matrix via the Clements decomposition [32]. The Reck [31] and balanced binary tree [38] are other important decompositions that have been used to construct unitary meshes.

This paper proposes a new mesh architecture based on the sine–cosine decomposition, a block diagonalization
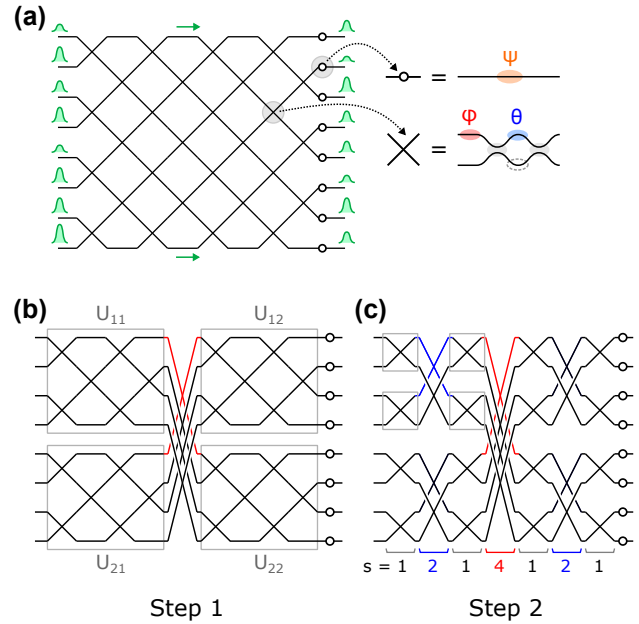

Figure 1: Illustration of optical mesh designs. (a) 8×8 Clements mesh. (b) First step of the block decomposition of the mesh. This results in 4×4 quadrants. (c) Further decomposition of the quadrants, which results in the SCF mesh.

of unitary matrices, which on an $N \times N$ matrix U consists of partitioning U into four $N/2 \times N/2$ blocks and performing the singular value decomposition (SVD) on each block. The unitarity of U imposes special constraints that force the blocks to share singular vectors. Polcari [30] shows that the block-wise SVD of U yields:

$$
U = \begin{bmatrix} U_{12} & 0 \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} U_{11} & 0 \\ 0 & U_{21} \end{bmatrix} \tag{2}
$$

where $U_{11}, U_{12}, U_{21}, U_{22}$ are unitary matrices and $D_{11}, D_{12}, D_{21}, D_{22}$ are diagonal matrices encoding the singular vectors and values, respectively. Unitarity constrains the D_{ij} to take the following form:

$$
\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = ie^{i\Theta/2} \begin{bmatrix} e^{i\Phi} \sin(\Theta/2) & \cos(\Theta/2) \\ e^{i\Phi} \cos(\Theta/2) & -\sin(\Theta/2) \end{bmatrix} \tag{3}
$$

with (Θ, Φ) representing diagonal matrices that encode phase shifts. Figure 1(b) depicts Eq. (2) graphically for the 8×8 case—the 4×4 unitary matrices U_{ij} are implemented by Clements meshes while the diagonal matrices D_{ij} are implemented by MZIs in the center that couple the four unitary blocks. One can actually go further and perform the block-wise SVD again on each of the U_{ij} sub-blocks to obtain the mesh of Figure 1(c); the 2×2 unitaries obtained from the 4×4 unitaries are now directly implemented by

MZIs. Because of its self-similar structure, we denote this geometry the sine–cosine fractal (SCF) mesh. In the general case, an SCF mesh can be constructed from any radix-2 ($N = 2^n$) matrix: one recursively performs blockwise SVDs on each unitary matrix of size greater than 2 until the full decomposition consists only of 2×2 matrices that connect different modes.

Like the `Clements` and other conventional architectures, the SCF mesh has a depth that scales as $O(N)$, has N^2 degrees of freedom, and is universal, i.e. it can represent the entire unitary group. However, the SCF mesh also possesses tunable crossings that couple non-neighboring waveguides (Figure 1(c)), i.e. crossings of *stride s* > 1, interleaved between conventional crossings with $s = 1$. This is in contrast to conventional architectures, where all MZIs have unit stride.

3 Error correction and matrix fidelity

The long-stride crossings give the SCF mesh greater robustness to hardware imperfections. To see how, consider the problem of realizing a target at high fidelity on imperfect hardware. As mentioned previously, MZI meshes with perfect beamsplitters can implement any unitary matrix; the introduction of faults, however, reduces the expressivity of the mesh and consequently the fraction of matrices that are

implementable drops below unity. In this section, we show that sine–cosine fractal meshes can perfectly implement a greater fraction of random matrices than Clements meshes can for the same beamsplitter error level.

3.1 Distribution of mesh phase-shifts for Haar-random matrices

A specific setting of phase shifts (θ, ϕ) is required to make a mesh implement a given target unitary matrix. Drawing the target matrix from the Haar (uniform) distribution [39] induces a distribution $P(\theta)$ over the phase-shifts. Russell et al. [40] show that the phase-shift distribution of the nth MZI (according to any indexing scheme) in either the `Reck` or `Clements` meshes is given by $P_n(\theta) = k_n \sin(\theta/2) \cos(\theta/2)^{2k_n - 1}$, where $k_n \in \{1, \ldots, N-1\}$, called the *rank* of the nth MZI, is a function of the physical position of the MZI within the mesh. There are $(N - k)$ MZIs of rank k [40]. For larger meshes, the rank of an average MZI increases, and this results in $P(\theta)$, which is an average of the $P_n(\theta)$ over all the MZIs, clustering around the cross state $\theta = 0$ (top row of Figure 2). The phase shift distribution for both the `Clements` and `Reck` meshes are identical, with $P(\theta)$ clustering near $\theta = 0$ [40].

However, unlike the `Reck` and `Clements` meshes, the SCF mesh is configured from a top–down block decomposition of the matrix. For a given matrix U, sampled over the Haar measure, the singular-vector matrices $U_{ij}, (i, j) \in \{1, 2\}$ in Eq. (2) are also Haar-random and independent of
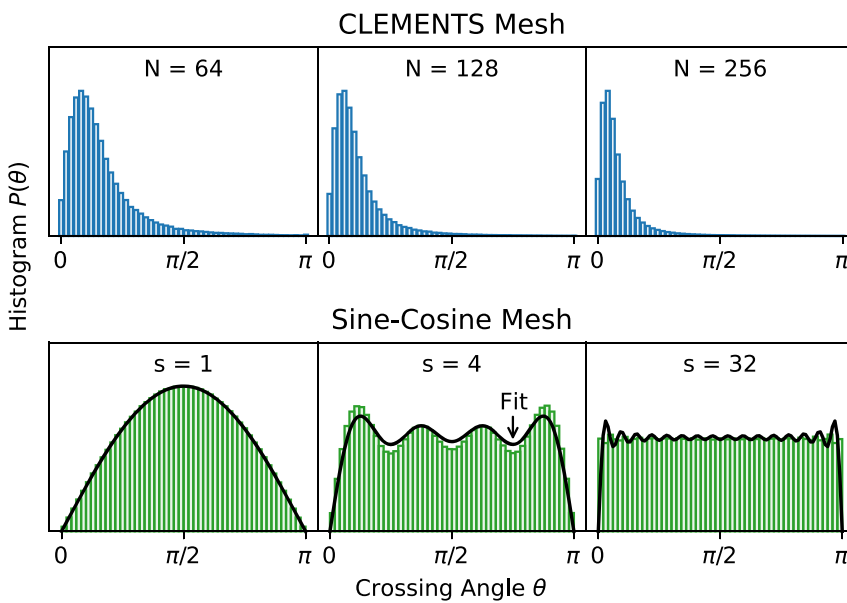

Figure 2: Distribution of crossing angles for the `Clements`, `Reck` and SCF meshes as a function of mesh size and stride, respectively. The crossing angles of the `Clements` mesh and the `Reck` mesh increasingly clusters near $\theta = 0$ for larger meshes. Crossing angles for the SCF mesh become increasingly uniformly distributed as the stride increases.

each other. As a result, the distribution $P_n(\theta)$ depends only on the stride s_n of the nth MZI, not on its location in the mesh. The bottom row of Figure 2 shows the distribution of angles for MZIs of stride $s = 1, 4, 32$ for mesh size $N = 64$. The majority of MZIs have unit stride and $P_n(\theta) \propto \sin(\theta)$. As the stride increases, this distribution begins to resemble a uniform distribution. We find that $P_n(\theta)$ for stride s_n can be fit well by the normalized finite Fourier series of a constant given by:

$$P_n(\theta) \approx \frac{1}{2 \sum\limits_{q=1}^{s_n} (2q-1)^{-2}} \sum_{q=1}^{s_n} \frac{\sin\big((2q-1)\theta\big)}{2q-1} \tag{4}$$

3.2 Error in implementing Haar-random matrices in the presence of MZI errors

The deviations from 50:50 of the two constituent beamsplitters of fabricated MZIs are captured by the phase angles (α, β). These splitter errors perturb the MZI transfer matrix as follows:

$$T'(\theta', \phi', \alpha, \beta) = \begin{bmatrix} \cos\left(\frac{\pi}{4} + \beta\right) & i\sin\left(\frac{\pi}{4} + \beta\right) \\ i\sin\left(\frac{\pi}{4} + \beta\right) & \cos\left(\frac{\pi}{4} + \beta\right) \end{bmatrix} \begin{bmatrix} e^{i\theta'} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} \cos\left(\frac{\pi}{4} + \alpha\right) & i\sin\left(\frac{\pi}{4} + \alpha\right) \\ i\sin\left(\frac{\pi}{4} + \alpha\right) & \cos\left(\frac{\pi}{4} + \alpha\right) \end{bmatrix} \begin{bmatrix} e^{i\phi'} & 0 \\ 0 & 1 \end{bmatrix} \tag{5}$$

Bandyopadhyay et al. [21] show that it is always possible to choose phase-shifts θ', ϕ' for a faulty MZI with errors (α, β) such that it implements the transfer matrix of an ideal MZI $T(\theta, \phi)$ as long as:

$$\underbrace{2|\alpha + \beta|}_{\theta_{\min}} \leq \theta \leq \underbrace{\pi - 2|\alpha - \beta|}_{\theta_{\max}} \tag{6}$$

If θ is outside this range, the faulty MZI cannot exactly emulate the ideal MZI and the faulty mesh transfer function deviates from that of the ideal mesh which implements the target matrix. Alternate approaches to correct for these errors have been proposed in refs. [15, 27, 29, 41].

For a given target unitary U, we quantify the deviation of the faulty mesh using the Frobenius norm $\mathcal{E} = \|\Delta U\|/\sqrt{N}$ which computes the average relative error per matrix element. This quantity is then averaged over both the choice of target unitary (from the Haar distribution), and the distribution of MZI splitter errors, which are assumed to be independent Gaussians ($\alpha, \beta \sim \mathcal{N}(0, \sigma)$). In the case

of correlated errors, the effect of most correlations vanishes over the Haar measure as proven in Ref. [24].

In the absence of error-correction techniques, all the MZIs of the ideal mesh are implemented incorrectly by the corresponding MZIs in the faulty mesh and the error Frobenius norm $\mathcal{E}_0$ (for both `Clements` and SCF meshes) has been shown [24] to scale linearly with the MZI error σ as $\mathcal{E}_0 = \sqrt{2N}\sigma$. When one uses the error-correction techniques of Refs. [21, 24, 25], only the MZIs in the ideal mesh that do not satisfy Eq. (6) are implemented incorrectly by the corresponding MZIs in the faulty mesh—only those MZIs contribute to the error Frobenius norm. Using n to denote the MZI location in the mesh as before, the average "corrected" error, $\mathcal{E}_c$ (assuming uncorrelated errors, see [27, Supp. Sec. 1]), is computed by integrating over the probability that each MZI does not satisfy Eq. (6):

$$(\mathcal{E}_c)^2 = \frac{1}{2N} \sum_n \left[\int_0^{\theta_{\min}} P_n(\theta)(\theta - \theta_{\min})^2 \, d\theta \right.$$

$$\left. + \int_{\theta_{\max}}^{\pi} P_n(\theta)(\theta - \theta_{\max})^2 \, d\theta \right] \tag{7}$$

Eq. (7) indicates that the effectiveness of error-correction, measured by $\mathcal{E}_c$, is strongly dependent on the distribution $P_n(\theta)$. For the `Clements` mesh, Ref. [24] proves that $\mathcal{E}_c^{(\text{clem})} = \sqrt{\frac{2}{3}}N\sigma^2$, which is a quadratic improvement over $\mathcal{E}_0$.

Since the integral is over angles close to either $\theta = 0$ or $\theta = \pi$, the two terms in Eq. (7) are estimated for the sine–cosine fractal mesh by Taylor-expanding Eq. (4) to first order about $\theta = 0, \pi$, respectively. The result is:

$$\mathcal{E}_c^{(\text{scf})} = \frac{4}{\pi} \sqrt{N \log_2(N)} \, \sigma^2. \tag{8}$$

The ratio of corrected errors for both meshes is:

$$\frac{\mathcal{E}_c^{(\text{clem})}}{\mathcal{E}_c^{(\text{scf})}} = \sqrt{\frac{\pi^2}{24} \frac{N}{\log_2(N)}} \tag{9}$$

which is greater than 1 for all but very small N.

We performed numerical experiments on meshes up to size $N = 1024$ to validate the above expressions (Figure 3). The observed uncorrected error $\mathcal{E}_0$ of both meshes (Figure 3(a)) scales as σ while error-correction on both meshes improves this scaling to σ^2. For meshes of size 256×256, error correction shows over an order of magnitude improvement in matrix error Frobenius norm over the uncorrected case when $\sigma \leq 0.01$, with the Sine–Cosine Fractal mesh performing better than the `Clements` mesh. Figure 3(b) illustrates the distribution of post-correction
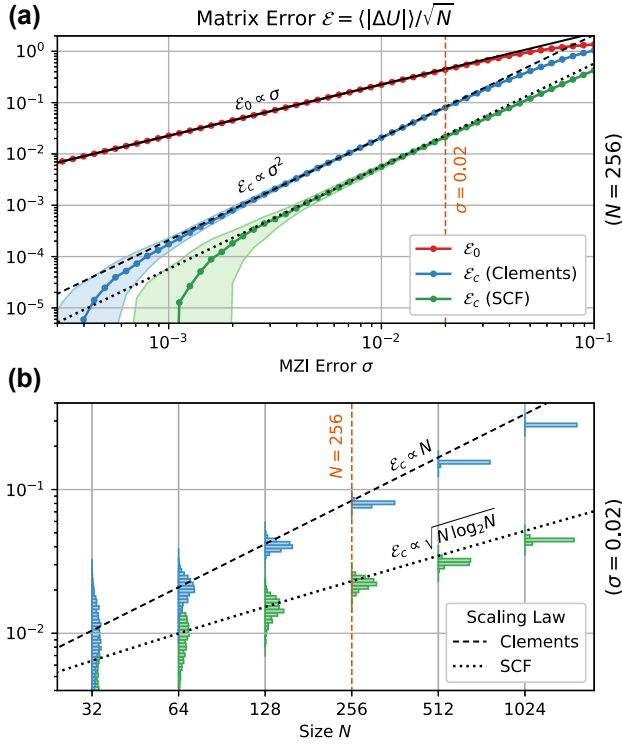
(a)

(b)

Figure 3: Comparision of matrix error scaling for the SCF and Clements meshes. (a) Scaling of matrix error with process variation σ. (b) Scaling of matrix error with mesh size, showing the advantage of the SCF mesh over the Clements mesh for larger mesh sizes.

error Frobenius norm as a function of mesh size N for a fixed $\sigma = 0.02$ (which is a typical value for directional couplers under wafer-scale process variation). While the factor $\sqrt{\frac{N}{\log_2(N)}}$ is modest for small meshes ($N < 64$), it clearly opens up a significant accuracy gap in Figure 3(b) in the large scale regime ($N > 1024$).

3.3 Fraction of Haar-random matrices those are exactly realizable in the presence of MZI errors

The fraction of the unitary group $U(N)$ that can be realized by an imperfect mesh is equal to the probability that, under the Haar measure [40], all target splitting angles θ are realizable. For convenience, we derive a quantity called the *coverage*, $\mathrm{cov}(N)$ [24], from this probability:

$$\mathrm{cov}(N) = \prod_n \left(1 - P_n(\theta < \theta_{\min}) - P_n(\theta > \theta_{\max}) \right)$$

$$\approx \exp\left(-\sum_n \left[P_n(\theta < \theta_{\min}) + P_n(\theta > \theta_{\max}) \right] \right) \quad (10)$$

For a `Clements` mesh, Ref. [24] proves that $\mathrm{cov}^{(\mathrm{clem})}(N) = \exp\left(-\frac{N^3 \sigma^2}{3} \right)$. Like the matrix error in the previous subsection, the coverage of an SCF mesh is computed using Taylor series expansions. The result is:

$$\mathrm{cov}^{(\mathrm{scf})}(N) = \exp\left(-\frac{8N^2 \log_2(N)}{\pi^2} \sigma^2 \right) \quad (11)$$

which is greater than $\mathrm{cov}^{(\mathrm{clem})}(N)$ for the same σ for all but very small N.

This coverage, however, expresses the ability of an imperfect mesh to realize random matrices under the Haar measure. It is often the case that such meshes will be used to realize matrices that are very commonly used in quantum computation, signal processing or machine learning. In Figure 4(a), we compare the ability of the SCF and `Clements` meshes to realize Hadamard matrices and discrete-Fourier transform (DFT) matrices of size $N = 128$. These examples confirm that many common matrices possess structure that does not follow the statistics of Haar random unitaries. For
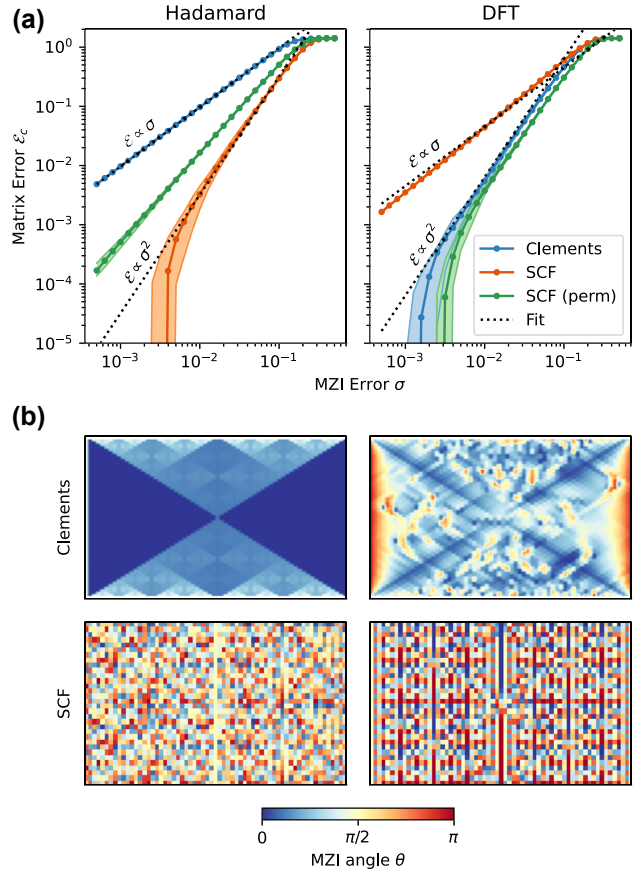
(a)

(b)

Figure 4: Realization of Hadamard and DFT matrices with Clements and SCF meshes. (a) Post-correction error $\mathcal{E}_c$ for matrix size $N = 128$. (b) MZI splitter angles θ for meshes of size $N = 64$ realizing Hadamard and DFT matrices.

the Hadamard matrix, the Clements decomposition yields a mesh where 50% of all MZIs are programmed to the crossing state $\theta = 0$ (Figure 4(b)); as a result, the matrix fidelity is much worse than for a typical Haar-random instance, and scales as $\mathcal{E}_c \propto \sigma$ as opposed to σ^2. The SCF mesh, by contrast, has a more uniform distribution of splitter angles, recovering $\mathcal{E}_c \propto \sigma^2$ scaling with overall matrix errors about an order of magnitude smaller.

For the DFT matrix, the situation is reversed. Here, the Clements mesh shows a more uniform distribution of angles, while the SCF mesh has a large number of MZIs with extreme splitting ratios $\theta \in \{0, \pi\}$. Therefore, the SCF mesh achieves significantly worse matrix fidelity when straightforwardly implementing the DFT. The reason is connected with the highly non-uniform distribution of block-singular values (i.e. D_{ij} in Eq. (3)) in the standard DFT. We can circumvent this problem using the technique of port allocation [22]. A random permutation of the input and output ports yields a more even distribution of singular values, which reduces the number of MZIs with $\theta \in \{0, \pi\}$. The resulting matrix fidelity is comparable to that of the Clements mesh (green curve in Figure 4(a), right).

4 Use in optical neural networks

In this section, we study the performance of an optical neural network (ONN) built from sine–cosine fractal meshes. We also propose and evaluate a pruning scheme for these meshes that allows areal footprint reduction while maintaining test performance. The neural network configuration is similar to those studied in Refs. [6, 14, 26]—each

neural net layer is implemented by an SCF mesh connected to an electro-optic nonlinearity [42] (Figure 6(a)). All our networks had two layers. Our simulations used the meshes [43] package, and results are presented for the MNIST [44] image classification task. The preprocessing of the images involved low-pass filtering and was identical to the procedure adopted in Refs. [21, 26]. The standard cross entropy loss and the Adam optimizer were used for training.

ONNs trained with SCF meshes achieved accuracies that matched the Clements mesh [21, 24] of ∼95%–96% for small meshes ($N = 64$) and ∼97% for larger meshes ($N = 256$). Next, we simulated the effect of MZI errors on the trained SCF mesh neural net; Figure 6(b) shows the median classification accuracy of 10 independently trained networks as a function of splitter errors. SCF networks of size $N = 64$ yielded ∼95% test accuracy while those of size $N = 256$ reached ∼97%. The presence of MZI errors rapidly degrades the performance of the network, with accuracy dropping to below 90% with splitter variation as low as ∼2%. The use of hardware error correction, however, extends this cutoff to greater than 12% even for bigger meshes, which is well above present-day process error [45] and larger than the corresponding cutoff for Clements meshes (which is 6% [21]).

4.1 Weight pruning

The number of columns of MZIs with stride s is $\frac{N}{2s}$ in the Sine–Cosine Fractal mesh while the standard FFT mesh contains a single column of each stride. We introduce a pruning scheme that interpolates between these extremes
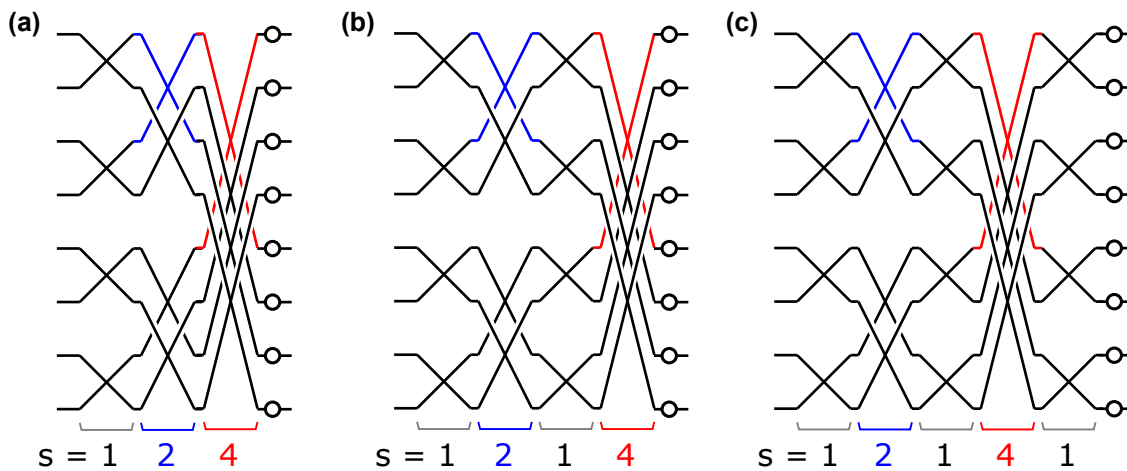

Figure 5: The SCF mesh at different stages of pruning constructed by interleaving strides from the largest successively to the smallest, from left to right with the number of columns limited by $\left(\frac{N}{2s}\right)^{D-1}$: (a) A maximally pruned mesh, which is the standard FFT mesh, $D = 1.0$. (b) $D = 1.5$ (c) $D = 1.75$.

by introducing a *fractal dimension* $D \in [1, 2]$—in a partially pruned mesh, the number of columns of stride s is $\left(\frac{N}{2s}\right)^{D-1}$. Setting $D = 1$ and $D = 2$ yields the FFT and SCF meshes, respectively. Controlling D allows us to tune the number of degrees of freedom and reduce the areal footprint of the device while ensuring full connectivity. Figure 5 illustrates partially pruned 8×8 SCF meshes for different values of D. The depth of the pruned network (approximated to leading order in N) scales as $\frac{N^{D-1}}{2^{D-1}-1}$ while the number of MZIs is approximately $\frac{N^D}{2(2^D-1)}$. To construct a pruned mesh, the column with the largest stride is chosen initially and is interleaved with columns of the next smaller stride. Each of these columns is then interleaved with the following smaller stride from left to right until $s = 1$ is reached. When the maximum number of columns is reached for a given stride, the interleaving is terminated and the process is continued with a column of the next smaller stride. This ensures that no two consecutive columns have the same stride since any such columns would collapse into a single column.

Figure 6(c) illustrates the results of training networks with pruned meshes of ideal MZIs for different D. Increasing D (decreasing the amount of pruning, or increasing the size of the network) increases the classification accuracy as one would expect. Interestingly, a maximally pruned (that is, a standard FFT) 64×64 2-layer ONN already achieves 95% accuracy, which is commensurate with the performance of present-day DNN accelerators [46].

4.2 Lower bounds on SCF mesh performance

We performed error-aware "maximally faulty mesh" training [47] to obtain non-trivial (empirical) lower bounds on the performance of SCF mesh ONNs with MZI errors as large as 30%. For each error level σ, we train an SCF mesh ONN composed of MZIs that all have $\alpha = 2\sigma$ and $\beta = 0$. Ref. [47] shows that the resultant trained matrices of this maximally faulty network can be exactly transferred to any other SCF mesh ONN while MZIs have errors $-\sigma \leq \alpha, \beta \leq \sigma$. This implies that the test accuracy of the maximally faulty SCF mesh ONN is a lower bound on the performance of any SCF mesh ONN whose MZI errors are bounded by σ. Maximally faulty 2-layer SCF mesh ONNs of sizes 64 and 256 were trained for several values of MZI error level σ and fractal dimension D; the resultant test accuracies are depicted in Figure 6(d) and (e). As expected, the 256 sized ONNs perform better than the 64 sized ones. However, it is particularly striking in the size 256 case that one can prune the network up to $D \sim 1.1$ and still lose only 1% in
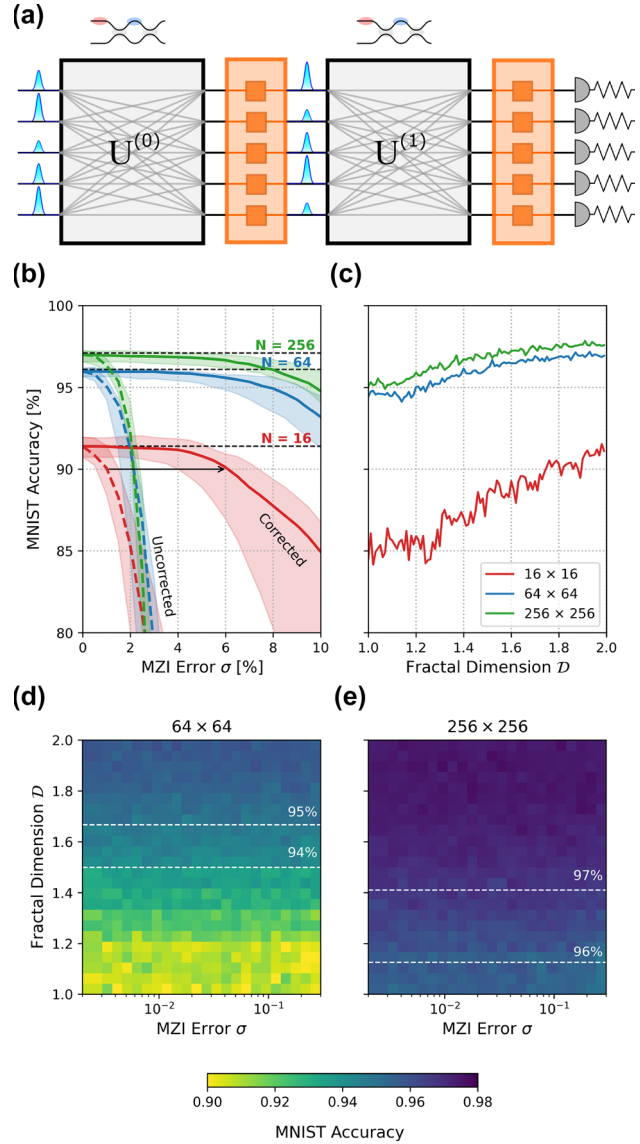

Figure 6: Performance of optical neural networks (ONNs) constructed using SCF meshes. (a) 2-Layer deep neural network architecture for MNIST classification where the blocks $U^{(0,1)}$ represent unitary transforms by the sine–cosine fractal mesh. (b) Simulated classification accuracy as a function of mesh size ($N = 64$ and 256) and MZI error σ. (c) Classification accuracy as a function of degree of pruning $D \in [1, 2]$ and mesh size ($N = 16$ and 64 and 256). (d) and (e) Classification accuracy as a function of the degree of pruning D and MZI error, trained on maximally faulty mesh of size $N = 64$ and 256.

classification accuracy. The size 64 case allows less freedom in pruning because it had a lower number of parameters to begin with. The performance also seems to be a nearly constant function of the MZI error. This investigation indicates that one can aggressively prune very faulty ONNs but still achieve excellent performance.

5 Conclusions

This work presented a novel architecture for multiport interferometers based on the sine–cosine decomposition of unitary matrices. The proposed scheme is self-similar, and therefore modular. As a result, this architecture is ideal to construct multi-chiplet modules for large-scale devices that are typically limited by device yield. We showed that SCF meshes show improved scaling under error correction when compared to conventional multiport interferometers. Finally, this design allows for systematic re-wiring of MZI layers, which is an efficient way of reducing the areal footprint of the mesh while maintaining full connectivity.

The proposed design has multiple advantages over the traditional architectures for multiport interferometers. Due to the uniform distribution of coupling angles (which is in contrast to conventional mesh architectures where crossing angles are clustered around the cross state), error correction techniques are far more effective in the case of SCF meshes. Truncated meshes used in ONNs can be trained to perform on par with the performances of present-day DNN accelerators. While the benefits of modularity and stronger robustness to error are small for present-day mesh size, the scaling gap of $O(N)$ versus $O(\sqrt{N \log_2(N)})$ significantly impacts the performance for larger meshes. This reduced scaling with N implies that SCF meshes are more sensitive to improvements in foundry process. Reducing σ by 2× corresponds to a 4× increase in the maximum mesh size for RECK and CLEMENTS owing to the $O(N\sigma^2)$ scaling; for the butterfly fractal, the corresponding increase would be 16×.

A potential drawback of this scheme is the presence of multiple large stride crossings with non-zero crosstalk. The loss/crosstalk introduced by these crossings can be minimized by the following methods that are enabled by the modularity of the SCF architecture:

- *Incomplete Decomposition*: To minimize the number of crossings, the block decomposition of the mesh can be terminated such that the smallest block size $N_{\text{blk}} > 2$. Each block will then take the form of a standard `Clements` geometry with no intra-block crossings.
- *Out-of-plane Crossings*: Inter-block crossings ($s > 1$ among arbitrary block sizes) can be implemented using waveguide escalators and out-of-plane crossings. Crossings of different stride can be fabricated into multiple layers of the photonic circuit and can be connected with inter-planar couplers. The structure of the SCF mesh ensures that each spatial mode would travel through the same number of inter-planar couplers, which results in losses that are uniform. This would only impact the power scaling of the output vector, and

not the matrix realized by the mesh. Recent work on the construction of these couplers has shown losses as low as 0.05 ± 0.02 dB or approximately 98.7% per coupler [48]. Scaling up the input power by a factor of $1/0.987^{(N-2)}$ (up to heating limitations posed by the hardware) for a mesh of size N would allow us to account for losses due to the couplers. Out-of-plane crossings have also been shown to have much lower crosstalk than in-plane crossings [49, 50]. Crosstalk on the order of -50 dB for input powers of 1 mW has been achieved in non-planar topologies, and can be fabricated scalably for such an integrated photonic mesh [51, 52]. Since the crosstalk due to induced fields would decrease exponentially with increasing interplanar distance, it would be easily possible to further reduce the crosstalk. In the case of a multi-chip module, each chiplet will be connected by a set of waveguide crossings with a large stride. These crossings can be fabricated using present-day lithography and laser-writing techniques in either polymer [53, 54] or glass [55, 56], and could utilize "hockey-stick" escalator couplers to reduce the alignment tolerances for each chiplet [57].

This suggests that there exists an optimal block or chiplet size N_{blk} for multi-chip modules, which will have to be determined by the trade-off between intra-block errors (that favors small N_{blk}) and inter-block losses (that favors large N_{blk}).

For practical use as energy efficient deep learning accelerators, nanophotonic circuits will need to be scaled to reach the sub-fJ/MAC energy target. Present-day foundry processes face immense difficulties in scaling conventional multiport interferometers to the large size required to achieve this energy efficiency target. Furthermore, these meshes suffer unacceptably high errors $\mathcal{E}_c^{(\text{clem})} \gtrsim 0.2$ even in the presence of error correction [21, 24, 25]. This brings us to the regime below the recommended 4–8 bits of precision that is usually targeted for DNN accelerators [58–60]. The modularity of the SCF mesh, and its improved tolerance to error, open up the regime of large-scale programmable photonics, making the SCF mesh frontrunning candidate architecture for future systems.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

[1] A. Annoni, E. Guglielmi, M. Carminati, et al., "Unscrambling light—automatically undoing strong mixing between modes," *Light Sci. Appl.*, vol. 6, p. e17110, 2017.

[2] A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a 4 × 4-port universal linear circuit," *Optica*, vol. 3, p. 1348, 2016.

[3] M. Milanizadeh, P. Borga, F. Morichetti, D. Miller, and A. Melloni, "Manipulating free-space optical beams with a silicon photonic mesh," in *2019 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, IEEE, 2019, pp. 1–2.

[4] L. Zhuang, C. G. Roeloffzen, M. Hoekman, K.-J. Boller, and A. J. Lowery, "Programmable photonic signal processor chip for radiofrequency applications," *Optica*, vol. 2, p. 854, 2015.

[5] J. Notaros, J. Mower, M. Heuck, et al., "Programmable dispersion on a photonic integrated circuit for classical and quantum applications," *Opt. Express*, vol. 25, p. 21275, 2017.

[6] Y. Shen, N. C. Harris, S. Skirlo, et al., "Deep learning with coherent nanophotonic circuits," *Nat. Photonics*, vol. 11, p. 441, 2017.

[7] J. R. Basani, M. Heuck, D. R. Englund, and S. Krastanov, "All-photonic artificial neural network processor via non-linear optics," arXiv preprint arXiv:2205.08608, 2022.

[8] M. Prabhu, C. Roques-Carmes, Y. Shen, et al., "Accelerating recurrent Ising machines in photonic integrated circuits," *Optica*, vol. 7, p. 551, 2020.

[9] N. C. Harris, G. R. Steinbrecher, M. Prabhu, et al., "Quantum transport simulations in a programmable nanophotonic processor," *Nat. Photonics*, vol. 11, p. 447, 2017.

[10] J. Wang, S. Paesani, Y. Ding, et al., "Multidimensional quantum entanglement with large-scale integrated optics," *Science*, vol. 360, p. 285, 2018.

[11] X. Qiang, X. Zhou, J. Wang, et al., "Large-scale silicon quantum photonics implementing arbitrary two-qubit processing," *Nat. Photonics*, vol. 12, p. 534, 2018.

[12] C. Sparrow, E. Martín-López, N. Maraviglia, et al., "Simulating the vibrational quantum dynamics of molecules using photonics," *Nature*, vol. 557, p. 660, 2018.

[13] J. Carolan, C. Harrold, C. Sparrow, et al., "Universal linear optics," *Science*, vol. 349, p. 711, 2015.

[14] M. Y.-S. Fang, S. Manipatruni, C. Wierzynski, A. Khosrowshahi, and M. R. DeWeese, "Design of optical neural networks with component imprecisions," *Opt. Express*, vol. 27, p. 14009, 2019.

[15] R. Burgwal, W. R. Clements, D. H. Smith, et al., "Using an imperfect photonic network to implement random unitaries," *Opt. Express*, vol. 25, p. 28236, 2017.

[16] J. Mower, N. C. Harris, G. R. Steinbrecher, Y. Lahini, and D. Englund, "High-fidelity quantum state evolution in imperfect photonic integrated circuits," *Phys. Rev. A*, vol. 92, p. 032322, 2015.

[17] D. P. López, "Programmable integrated silicon photonics waveguide meshes: optimized designs and control algorithms," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, p. 1, 2019.

[18] A. López, D. Pérez, P. DasMahapatra, and J. Capmany, "Auto-routing algorithm for field-programmable photonic gate arrays," *Opt. Express*, vol. 28, p. 737, 2020.

[19] D. Pérez-López, A. López, P. DasMahapatra, and J. Capmany, "Multipurpose self-configuration of programmable photonic circuits," *Nat. Commun.*, vol. 11, p. 1, 2020.

[20] S. Pai, B. Bartlett, O. Solgaard, and D. A. Miller, "Matrix optimization on universal unitary photonic devices," *Phys. Rev. Appl.*, vol. 11, p. 064044, 2019.

[21] S. Bandyopadhyay, R. Hamerly, and D. Englund, "Hardware error correction for programmable photonics," *Optica*, vol. 8, p. 1247, 2021.

[22] S. P. Kumar, L. Neuhaus, L. G. Helt, et al., "Mitigating linear optics imperfections via port allocation and compilation," arXiv preprint arXiv:2103.03183, 2021.

[23] D. A. Miller, "Setting up meshes of interferometers—reversed local light interference method," *Opt. Express*, vol. 25, p. 29233, 2017.

[24] R. Hamerly, S. Bandyopadhyay, and D. Englund, "Stability of self-configuring large multiport interferometers," *Phys. Rev. Appl.*, vol. 18, p. 024018, 2022.

[25] R. Hamerly, S. Bandyopadhyay, and D. Englund, "Accurate self-configuration of rectangular multiport interferometers," *Phys. Rev. Appl.*, vol. 18, p. 024019, 2022.

[26] S. Pai, I. A. Williamson, T. W. Hughes, et al., "Parallel programming of an arbitrary feedforward photonic network," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, p. 1, 2020.

[27] R. Hamerly, S. Bandyopadhyay, and D. Englund, "Infinitely scalable multiport interferometers," arXiv preprint arXiv:2109.05367, 2021.

[28] K. Suzuki, G. Cong, K. Tanizawa, et al., "Ultra-high-extinction-ratio 2 × 2 silicon optical switch with variable splitter," *Opt. Express*, vol. 23, p. 9086, 2015.

[29] D. A. Miller, "Perfect optics with imperfect components," *Optica*, vol. 2, p. 747, 2015.

[30] J. Polcari, "Generalizing the butterfly structure of the FFT," in *Advanced Research in Naval Engineering*, Cham, Switzerland, Springer, 2018, pp. 35–52.

[31] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, p. 58, 1994.

[32] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," *Optica*, vol. 3, p. 1460, 2016.

[33] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, p. 297, 1965.

[34] F. Flamini, N. Spagnolo, N. Viggianiello, A. Crespi, R. Osellame, and F. Sciarrino, "Benchmarking integrated linear-optical architectures for quantum information processing," *Sci. Rep.*, vol. 7, p. 1, 2017.

[35] J. Gu, Z. Zhao, C. Feng, M. Liu, R. T. Chen, and D. Z. Pan, "Towards area-efficient optical neural networks: an FFT-based architecture," in *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, 2020, pp. 476–481.

[36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv:1207.0580, 2012.

[37] H. Tanaka, D. Kunin, D. L. Yamins, and S. Ganguli, "Pruning neural networks without any data by iteratively conserving synaptic flow," *Adv. Neural Inf. Process. Syst.*, vol. 33, p. 6377, 2020.

[38] H.-Y. Lee and I.-C. Park, "Balanced binary-tree decomposition for area-efficient pipelined FFT processing," *IEEE Trans. Circ. Syst.*, vol. 54, p. 889, 2007.

[39] W.-K. Tung, *Group Theory in Physics: An Introduction to Symmetry Principles, Group Representations, and Special Functions in Classical and Quantum Physics*, Shanghai, China, World Scientific, 2003.

[40] N. J. Russell, L. Chakhmakhchyan, J. L. O'Brien, and A. Laing, "Direct dialling of Haar random unitary matrices," *New J. Phys.*, vol. 19, p. 033007, 2017.

[41] S. A. Fldzhyan, M. Y. Saygin, and S. P. Kulik, "Optimal design of error-tolerant reprogrammable multiport interferometers," *Opt. Lett.*, vol. 45, p. 2632, 2020.

[42] I. A. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, p. 1, 2019.

[43] R. Hamerly, *Meshes: Tools for Modeling Photonic Beamsplitter Mesh Networks*, 2021. Available at: https://github.com/QPG-MIT/meshes.

[44] Y. LeCun, *The MNIST Database of Handwritten Digits*, 1998. Available at: http://yann.lecun.com/exdb/mnist/.

[45] J. C. Mikkelsen, W. D. Sacher, and J. K. Poon, "Dimensional variation tolerant silicon-on-insulator directional couplers," *Opt. Express*, vol. 22, p. 3145, 2014.

[46] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," *Phys. Rev. X*, vol. 9, p. 021032, 2019.

[47] S. K. Vadlamani, D. Englund, and R. Hamerly, "Transferable learning on analog hardware," arXiv preprint arXiv:2210.06632, 2022.

[48] J. Chiles, S. Buckley, N. Nader, S. W. Nam, R. P. Mirin, and J. M. Shainline, "Multi-planar amorphous silicon photonics with compact interplanar couplers, cross talk mitigation, and low crossing loss," *APL Photonics*, vol. 2, p. 116101, 2017.

[49] A. M. Jones, C. T. DeRose, A. L. Lentine, D. C. Trotter, A. L. Starbuck, and R. A. Norwood, "Ultra-low crosstalk, CMOS compatible waveguide crossings for densely integrated photonic interconnection networks," *Opt. Express*, vol. 21, p. 12002, 2013.

[50] W. D. Sacher, J. C. Mikkelsen, Y. Huang, et al., "Monolithically integrated multilayer silicon nitride-on-silicon waveguide platforms for 3-D photonic circuits and devices," *Proc. IEEE*, vol. 106, p. 2232, 2018.

[51] A. Nesic, M. Blaicher, T. Hoose, et al., "Photonic-integrated circuits with non-planar topologies realized by 3D-printed waveguide overpasses," *Opt. Express*, vol. 27, p. 17402, 2019.

[52] M. Johnson, M. G. Thompson, and D. Sahin, "Low-loss, low-crosstalk waveguide crossing for scalable integrated silicon photonics applications," *Opt. Express*, vol. 28, p. 12498, 2020.

[53] N. Lindenmann, G. Balthasar, D. Hillerkuss, et al., "Photonic wire bonding: a novel concept for chip-scale interconnects," *Opt. Express*, vol. 20, p. 17667, 2012.

[54] M. R. Billah, M. Blaicher, T. Hoose, et al., "Hybrid integration of silicon photonics circuits and InP lasers by photonic wire bonding," *Optica*, vol. 5, p. 876, 2018.

[55] P. L. Mennea, W. R. Clements, D. H. Smith, et al., "Modular linear optical circuits," *Optica*, vol. 5, p. 1087, 2018.

[56] A. Szameit and S. Nolte, "Discrete optics in femtosecond-laser-written photonic structures," *J. Phys. B: At., Mol. Opt. Phys.*, vol. 43, p. 163001, 2010.

[57] S. Bandyopadhyay and D. Englund, "Alignment-free photonic interconnects," arXiv preprint arXiv:2110.12851, 2021.

[58] S. Friedmann, N. Frémaux, J. Schemmel, W. Gerstner, and K. Meier, "Reward-based learning under hardware constraints—using a RISC processor embedded in a neuromorphic substrate," *Front. Neurosci.*, vol. 7, p. 160, 2013.

[59] F. Akopyan, J. Sawada, A. Cassidy, et al., "Truenorth: design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip," *IEEE Trans. Comput. Aided Des. Integrated Circ. Syst.*, vol. 34, p. 1537, 2015.

[60] N. P. Jouppi, C. Young, N. Patil, et al., "In-datacenter performance analysis of a tensor processing unit," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 2017, pp. 1—12.