Hybrid spatio-temporal architectures for universal linear optics

Daiqin Su,1, * Ish Dhand,1, 2, * Lukas G. Helt,1 Zachary Vernon,1 and Kamil Braďler1

1 Xanadu, 372 Richmond Street West, Toronto, Ontario M5V 1X6, Canada
2 Institut für Theoretische Physik and Center for Integrated Quantum Science and Technology (IQST), Albert-Einstein-Allee 11, Universität Ulm, 89069 Ulm, Germany

(Dated: December 20, 2018)

We present a hybrid linear-optical architecture that simultaneously exploits spatial and temporal degrees of freedom of light to effect arbitrary discrete unitary transformations. Our architecture combines the benefits of spatial implementations of linear optics, namely low loss and parallel operation, with those of temporal implementations, namely modest resource requirements and access to transformations of potentially unbounded size. We arrive at our architecture by devising decompositions of large discrete unitary transformations into smaller ones, decompositions we expect to have broad utility beyond spatio-temporal linear optics. We show that the hybrid architecture promises important advantages over spatial-only and temporal-only architectures.

Introduction. — Scalable linear optical interferometers are key to unlocking important quantum technologies such as quantum computation [1, 2], quantum metrology [3], quantum simulations (including vibronic spectroscopy [4]), boson sampling [5], and Gaussian boson sampling [6]. Obtaining a quantum advantage in these applications requires overcoming the challenging problem of scaling linear optical interferometers to a large number of photons and modes. For instance, recent classical algorithms have raised the bar for a quantum advantage in boson sampling to requiring more than 30–50 indistinguishable photons in thousands of modes [7, 8].

Implementing a programmable linear optical interferometer on thousands of modes is infeasible in bulk optics because of stability requirements. An alternative is to integrate the components onto a monolithic photonic chip, which alleviates the stability issues [9–11]. However, scaling these chips to include a large number of modes is challenging as the dimensions of these chips are limited, and the optical elements and their corresponding classical control elements have a certain minimal area footprint. Indeed, whereas achieving a verifiable quantum advantage via boson sampling in integrated chips could require $10^6$–$10^7$ optical elements [7, 8], current implementations are limited to tens of modes or hundreds of elements [12].

Exploiting the potentially unbounded temporal degree of freedom provides an alternative advantage to using only the spatial modes of light on an integrated chip. Existing architectures allow for implementing arbitrary discrete unitary transformations on the temporal modes of light in a single spatial mode [13–15]. Fully connected implementations of up to eight modes [16] and partially connected implementations of up to a million modes [17] have been demonstrated. Although scaling to an arbitrarily large number of modes using only a limited number of optical elements is possible in principle using this architecture, two primary challenges remain. First, this architecture requires using optical delay lines with lengths on the order of tens of kilometers [18], which can induce significant loss and other imperfections as modes are switched in, propagated through, and switched out. Secondly, the overall rate of performing a single run of the experiment decreases as the number of implemented modes is increased.

To date, only the two extremes of fully spatial or fully temporal interferometers have been explored, each with their individual advantages and shortcomings. In this paper, we present a hybrid architecture that exploits both the spatial and temporal degrees of freedom of light to effect arbitrary discrete unitary transformations. This hybrid architecture allows the exploration of the middle ground between spatial and temporal architectures and optimal trade-offs for specific implementations, enabling the construction of large interferometers with minimal experimental challenge. Our architecture achieves this by combining the advantages of spatial implementations, i.e., lower switching and transmission losses and parallel operation, with those of temporal implementations, i.e., reduced number of optical elements and the possibility of implementing arbitrarily large transformations.

In devising the architecture, we present two decompositions of SU($N$) transformations into products of U($M$) transformations. Our hybrid architecture uses these decompositions to effect arbitrary $N \times N$ discrete unitary transformations on the temporal modes of light in $M$ spatial modes. These spatial modes are acted upon by the obtained U($M$) transformations at different times, with optical delay lines connecting together different temporal modes within these spatial modes. Before detailing the decompositions and the resulting architectures, we first introduce some architecture independent definitions.

Definitions. — An $N$-mode linear optical interferometer is characterized by a unitary operator $U$ that transforms $N$ bosonic annihilation operators $a_1, a_2, \ldots, a_N$ according to $a_i \to a'_i = U^\dagger a_i U = \sum_{j=1}^N U_{ij} a_j$, where $U = U_{ij}$ is an $N \times N$ unitary matrix. If the global phase of the emitted light is inconsequential, we can assume that the matrix $U$ is in SU($N$), the group of special unitary matrices. Existing architectures for implementing a given SU($N$) matrix $U$ via linear optics rely on first systematically decomposing $U$ into U(2) matrices, which are then iden-
tified as beamsplitter and phase transformations acting on different modes of light [19–21].

**Elimination-based decomposition.**— Our first procedure decomposes a given SU($N$) matrix into two types of elementary matrices: standard U($M$) matrices and specialized “residual” ($2M - 3 \times 2M - 3$) unitary matrices as illustrated in Fig. 1a. The elimination decomposition is obtained by eliminating elements from the given unitary matrix in an order that is motivated by decomposing into U($M$) matrices. For the purpose of elimination, we define an $N \times N$ unitary matrix $T_{mn}(\theta, \phi)$ with $n > m$ following earlier work [19, 20]. $T_{mn}(\theta, \phi)$ is obtained from the $N \times N$ identity matrix $I_N$ by changing the entries at the intersection of the $m$-th and $n$-th rows and columns to

$$
\begin{pmatrix}
\cos \theta e^{-i \phi} & -\sin \theta e^{-i \phi} \\
\sin \theta e^{-i \phi} & \cos \theta
\end{pmatrix},
$$

and leaving the other entries unchanged. These matrices, with suitably chosen values of $\theta$ and $\phi$, can be multiplied into matrices from the right in order to obtain new matrices in which specific entries are zero [19, 20]. Physically, each $T_{mn}(\theta, \phi)$ can be realized using a beamsplitter and phase shifters parameterized by $\theta, \phi$ acting on modes labeled $m$ and $n$. Henceforth, we drop the arguments ($\theta, \phi$) for simplicity.

We illustrate our procedure with the concrete example of decomposing an SU(7) matrix into U(3) matrices. A general SU(7) matrix is represented by

$$
\begin{pmatrix}
\ast & C^8_{(1,2)} & C^7_{(2,3)} & B^5_{(3,4)} & B^4_{(4,5)} & A^2_{(5,6)} & A^1_{(6,7)} \\
\ast & C^9_{(2,3)} & E^{11}_{(2,4)} & B^7_{(4,5)} & \bar{V}_{10}^{(6,4)} & A^2_{(6,7)} \\
\ast & G^{13}_{(3,4)} & G^{15}_{(4,5)} & F^{13}_{(5,6)} & F^{12}_{(6,7)} \\
\ast & G^{14}_{(5,7)} & \bar{H}^{18}_{(5,6)} & F^{14}_{(6,7)} \\
\ast & \ast & \ast & \bar{T}^{14}_{(6,7)} & \bar{T}^{14}_{(6,7)} & \ast \\
\ast & \ast & \ast & \ast & \ast & \ast
\end{pmatrix},
$$

where the bottom off-diagonal part is not explicitly shown for simplicity and the elements are complex-valued in general. The matrix elements with subscripts ($m, n$) above the diagonal are mulled systematically in the order of their superscripts using $T_{mn}$ matrices (see supplementary material (SM) Section A). For SU(7), this leads to the decomposition

$$
U (T_{67} T_{56} T_{67})^{-1} (T_{45} T_{34} T_{45})^{-1} (T_{23} T_{12} T_{23})^{-1} 
T_{46}^{-1} (T_{24}^{-1} (T_{67} T_{56} T_{67})^{-1} (T_{45} T_{34} T_{45})^{-1} 
T_{46}^{-1} (T_{27} T_{36} T_{27})^{-1} = D,
$$

which gives $U$ in terms of $T_{mn}$ and diagonal matrix $D = \text{diag}(e^{i \delta_1}, e^{i \delta_2}, \ldots, e^{i \delta_7})$. Here, the factors in the brackets are combinations that can be grouped together into U(3) matrices acting on three adjacent rows and leaving the other rows unchanged. Thus, the SU(7) matrix can be decomposed into six U(3) matrices denoted $\tilde{V}$, three residual matrices $\tilde{W}$, and seven phases corresponding to the diagonal matrix $D$. As $U$ is a special unitary matrix, one of these phases can be set equal to unity so only six additional phase shifters are required. In general, assuming $N = k(M - 1) + 1$ for positive integer $k$, any given $U \in \text{SU}(N)$ can be decomposed into $k(k + 1)$ universal $\tilde{V} \in \text{U}(M)$ matrices and $k(k - 1)$ residual matrices $\tilde{W}$.

**Cosine-sine-based decomposition.**— Our second decomposition is based on the cosine-sine decomposition (CSD), which factorizes any arbitrary $(m + n) \times (m + n)$ unitary matrix $U_{m+n}$ into unitary matrices $\tilde{L}_{m+n}$, $\tilde{S}_{2m}$, and $\tilde{R}_{m+n}$ according to [22–24]

$$
U_{m+n} = \tilde{L}_{m+n} \left( \tilde{S}_{2m} \oplus \mathbb{I}_{n-m} \right) \tilde{R}_{m+n},
$$

$$
\tilde{L}_{m+n} = \begin{pmatrix}
L_m & 0 \\
0 & P_n
\end{pmatrix}, \quad \tilde{R}_{m+n} = \begin{pmatrix}
R_m & 0 \\
0 & R_n
\end{pmatrix},
$$

where $\tilde{S}_{2m}$ is a cosine-sine (CS) matrix of the form

$$
\begin{pmatrix}
\text{diag}(\cos \theta_1, \ldots, \cos \theta_m) & \text{diag}(\sin \theta_1, \ldots, \sin \theta_m) \\
-\text{diag}(\sin \theta_1, \ldots, \sin \theta_m) & \text{diag}(\cos \theta_1, \ldots, \cos \theta_m)
\end{pmatrix}.
$$

Note that the matrix subscripts give the dimensions of the matrices. The matrices $L_m, L'_m, R_m, R'_n$ and the angles $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ can be determined using stable numerical methods [25]. The CSD can be applied repeatedly to decompose an $N \times N$ unitary into smaller $M \times M$ universal unitary matrices and specialized (i.e., non universal) $2M \times 2M$ CS matrices, which are collectively referred to as elementary matrices [25]. The decomposition into elementary matrices, depicted in Fig. 1b, proceeds as follows.

Taking $N = \ell M$ for integer $\ell$, the decomposition is an iterative process comprising $\ell - 1$ iterations. In the first iteration, the full $N \times N$ matrix is decomposed via $\ell - 1$ applications of the CSD into a single $(N - \ell)$-dimensional unitary matrix along with a “layer” of elementary matrices comprising $2\ell - 1 M \times M$ unitary matrices $(U^{(j)}_i, V^{(j)}_i)$ in Fig. 1b) and $\ell - 1$ CS matrices $(\tilde{S}^{(j)}_i$ in Fig. 1b). The first layer is depicted in Fig. 1b by the boxes with subscript 1. In general, the $(i + 1)$-th iteration uses the CSD to decompose the $(N - i\ell)$-dimensional unitary matrix into a layer of elementary matrices and a smaller $(N - (i + 1) \ell)$-dimensional unitary matrix, which is then decomposed in subsequent iterations. Eventually, the full unitary is decomposed into $\ell$ layers of elementary matrices, of which the last layer comprises a single $M \times M$ universal unitary matrix.

**Hybrid spatial-temporal architectures.**— Based on these two decompositions, we present corresponding architectures for implementing an arbitrary SU($N$) matrix on the combined temporal and spatial modes of light. First, we consider the elimination-based decomposition, which returns $k = (N - 1)/(M - 1)$ “layers” of $M$-mode universal
matrices $\tilde{V}$ and $k - 1$ layers of $2M - 3$ residual unitary matrices $\tilde{W}$. Different layers are labeled by different subscripts in Fig. 1a.

The key insight behind our hybrid architecture is that the action of one layer of interferometers on spatial modes of light can be replaced by the action of a single tunable interferometer with suitable delay lines on the spatial and temporal modes of light.

To implement the hybrid architecture one needs to shift the bottom $M - 1$ rows of the $V$ and $W$ matrices to the top, and we denote the resultant matrices $V$ and $W$ (see SM Section A). Without embedding these swaps in the $V$ and $W$ matrices, $(M - 1)^2$ additional swap gates would be needed for our hybrid spatial-temporal architecture. The action of tunable interferometer $V$ in Fig. 2a on $M$ spatial modes of light is equivalent to the action of the $V_j$ for $j = 1, 2, \ldots, k$ in Fig. 1a. Here, $V$ is a single tunable $M$-mode universal spatial interferometer that has $M - 1$ free input and output ports and one output port connected to one input port with a delay line of length equal to the separation $\tau$ between subsequent temporal modes.

Implementing $V$ requires $M(M - 1)/2$ beamsplitters [19]. Similarly, the residual $W$ matrices are implemented using a specialized $(2M - 3)$-dimensional spatial-mode interferometer and $M - 2$ delay lines. As detailed in SM Section A, realizing these requires $(M - 1)(M - 2)/2$ beamsplitters. Note that, altogether, this is fewer elements than the $N(N - 1)/2$ beamsplitters required in a purely spatial scheme.

The sequence of operations for effecting a single layer of $V$ matrices is as follows. Initially, a single temporal mode (pulse) impinges on $V$ at the first port and the $V$ interferometer is set such that the pulse moves into the delay line. When this pulse is guided to the $M$-th input of

![FIG. 1. Depiction of the two decompositions](image)

(a)

(b)

![FIG. 2.](image)
$V$, another $M - 1$ pulses impinge simultaneously via the first $M - 1$ inputs of $V$. Then the first $M$-mode unitary $V^{(1)}_1$ is implemented. After this action, the first $M - 1$ output pulses from $V$ move on to the next layer while the last output pulse moves into the delay loop and will couple with another $M - 1$ pulses that arrive after an interval $\tau$ on unitary $V^{(2)}_1$. This process continues until all $V^{(j)}_1$ unitary cells in the first layer are implemented. A similar sequence of operations effects the $W_1$ layer.

The full $N \times N$ unitary matrix is a composition of the action of $k$ layers of $V$ matrices and $k - 1$ layers of $W$ matrices. Multiple layers of unitary matrices can be realized either by reusing these two interferometers using a dual-loop architecture along the lines of Refs. [13, 14] or by chaining together two pairs of such interferometers in series [26]. In the dual-loop architecture, only a single block (as in Fig. 2a) of two interferometers is required. Also, a total of $M - 1$ optical delay lines are used to feed the light emitted from the block back into the input of the block. These delay lines, which implement time delays $\geq \kappa \tau$, are attached to the output ports via switches that can guide some of the pulses into the delay lines while other pulses are transmitted onwards. Each action of the two interferometers (Fig. 2a) effects a single layer. Thus, implementing a total of $k$ layers requires $k$ passes of the pulses through the $V$ and $W$ interferometers and $k - 1$ passes through the delay lines.

The second, CS-based, architecture effects SU($N$) transformations on $M$ spatial and $\ell = N/M$ temporal modes of light. The scheme employs tunable universal interferometers each acting on $M$ spatial modes and non-universal $2M$-mode interferometers, each requiring only $M$ beamsplitters. A single layer of unitary blocks can be implemented on spatial and temporal modes using three optical elements (Fig. 2b). The matrices $U$ and $V$ are universal tunable interferometers. Finally, we replace $\Theta$ parameterizing each of the CS matrices obtained from (5) by their respective complements $\Theta' = \{\pi/2 - \theta_1, \pi/2 - \theta_2, \ldots, \pi/2 - \theta_M\}$, thus replacing the obtained $S(\Theta)$ matrices by $S(\Theta')$. This replacement eliminates the need for $\ell^2$ swap gates that would otherwise be required in the CS-based hybrid spatial-temporal architecture. The $S$ matrix can be realized using $M$ beamsplitters.

The sequence of operations for this architecture is as follows. The first $M$ pulses, one in each of the $M$ spatial modes arrive simultaneously at $U$. On the first set of pulses, $U$ implements an identity transformation, letting these pulses pass unchanged. Also only for the first set of pulses, $S$ redirects the pulses into the $M$ delay lines by tuning all the beamsplitters to unit transmissivity $T = \sin \pi/2 = 1$. Because of the delay lines, the next $M$ pulses arrive at $U$ at the same time that these cycling pulses arrive at $V$. Now these two interferometers enact the first two unitary transformations of the decomposition procedure, i.e., the blocks $U_1^{(1)}$ and $V_1^{(1)}$ in Fig. 1b.

Together, these $2M$ pulses are acted upon by $S$. Note that $M$ of these $2M$ pulses leave the interferometer and $M$ pulses enter the delay lines to arrive at the interferometers synchronously with the next set of $M$ pulses. In the next round, the two universal and one non-universal interferometers are tuned to their next values, i.e., those corresponding to the next superscript in the first layer of Fig. 1b. This process is repeated $\ell - 1$ times, until the complete first layer is implemented. As in the case of the elimination-based decomposition, the full unitary is implemented by concatenating multiple layers. This is performed by chaining together a sequence of $\ell$ spatial interferometers one after the other or using an appropriate dual-loop architecture [13, 14].

Furthermore, the number of optical elements required to implement a single layer can be reduced. In particular, two universal interferometers can be implemented by a single interferometer if $2M$ additional switches are available. In this implementation (Fig. 2c), a single tunable interferometer plays the role of both $U_V$ and $V_M$ by switching between these two operating states after time $\tau/2$. Two sets of $M$ optical lines implementing time delays of $\tau/2$ are used: one from the universal interferometer to the CS interferometer and another from the CS to the universal interferometer.

Discussion.— To summarize, we have presented two architectures for the hybrid spatio-temporal implementation of a linear optical interferometer based on decompositions of SU($N$) into products of U($M$). These decomposition schemes also open the possibility of other hybrid architectures such as those involving frequency [27, 28] or orbital angular momenta [29] of light in addition to the spatial or temporal degrees of freedom.

Our hybrid spatio-temporal architecture fills the space between the two extremes of fully spatial and fully temporal architectures. It maintains two advantages of temporal architectures, namely a potentially unlimited number of...
realizable modes and a small number of required optical elements. Our two decompositions lead to architectures that have $O(N^2/M^2)$ fewer optical elements than fully spatial realizations. This reduction in elements comes at the experimental cost of having to stabilize $O(M)$ delay lines and also with a concomitant $O(N/M)$ increase in the time required to implement an $SU(N)$ transformation.

Furthermore, the hybrid architecture also allows for two advantages of fully spatial architectures which are not present in fully temporal architectures. First, the parallel operation on $M$ modes of light leads to a factor $O(M)$ speedup over fully temporal architectures. Secondly, each of the pulses needs to cycle in the outer loop $O(M)$ fewer times as compared to the fully temporal architecture. As a result, for large $M$, the hybrid architecture avoids losses associated with repeatedly cycling in the delay lines. Fig. 3 presents a comparison of the effect of this loss in our hybrid architecture with that in the temporal-only architecture. For the large number $N$ of modes required for demonstrating quantum advantage in boson sampling [7, 8], the hybrid architecture promises many orders of magnitude improvement over an all-temporal architecture assuming realistic values of losses (see SM Section B for details). Thus, based on experimental capabilities and requirements, our architecture enables optimized implementations of $SU(N)$ unitary transformations.

Acknowledgments: The authors thank Christian Weedbrook, Casey Myers and especially Haoyu Qi for discussions and insightful comments. ID acknowledges support from the Alexander von Humboldt Foundation via the Humboldt Research Fellowship for Postdoctoral Researchers.

* These two authors contributed equally

[1] E. Knill, R. Laflamme, and G. J. Milburn, Nature 409, 46 (2001).
[2] T. Rudolph, APL Photonics 2, 030901 (2017).
[3] K. R. Motes, J. P. Olson, E. J. Rabeaux, J. P. Dowling, S. J. Olson, and P. P. Rohde, Phys. Rev. Lett. 114, 170802 (2015).
[4] J. Huh, G. G. Guerreschi, B. Peropadre, J. R. McClean, and A. Aspuru-Guzik, Nat. Photonics 9, 615 (2015).
[5] S. Aaronson and A. Arkhipov, Theory Comput. 9, 143 (2013).
[6] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Phys. Rev. Lett. 119, 170501 (2017).
[7] A. Neville, C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, Nat. Phys. 13, 1153 (2017).
[8] P. Clifford and R. Clifford, in Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (Society for Industrial and Applied Mathematics, 2018) pp. 146–155.
[9] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguna, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O’Brien, and A. Laing, Science 349, 711 (2015).
[10] N. C. Harris, D. Bunandar, M. Pant, G. R. Steinbrecher, J. Mower, M. Prabhu, T. Baehr-Jones, M. Hochberg, and D. Englund, Nanophotonics 5, 456 (2016).
[11] N. C. Harris, G. R. Steinbrecher, M. Prabhu, Y. Lahini, J. Mower, D. Bunandar, C. Chen, F. N. C. Wong, T. Baehr-Jones, M. Hochberg, S. Lloyd, and D. Englund, Nat. Photonics 11, 447 (2017).
[12] See Table 3 in F. Flamini, N. Spagnolo, and F. Sciarrino, Rep. Prog. Phys. 82, 016001 (2019).
[13] K. R. Motes, A. Gilchrist, J. P. Dowling, and P. P. Rohde, Phys. Rev. Lett. 113, 120501 (2014).
[14] K. R. Motes, J. P. Dowling, A. Gilchrist, and P. P. Rohde, Phys. Rev. A 92, 052319 (2015).
[15] S. Takeda and A. Furusawa, Phys. Rev. Lett. 119, 120504 (2017).
[16] Y. He, X. Ding, Z.-E. Su, H.-L. Huang, J. Qin, C. Wang, S. Unslieber, C. Chen, H. Wang, Y.-M. He, X.-L. Wang, W.-J. Zhang, S.-J. Chen, C. Schneider, M. Kamp, L.-X. You, Z. Wang, S. Höfling, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. 118, 190501 (2017).
[17] J. ichi Yoshikawa, S. Yokoyama, T. Kaji, C. Sornphiphatphong, Y. Shiozawa, K. Makino, and A. Furusawa, APL Photonics 1, 060801 (2016).
[18] For 2500 modes required for verifiable quantum advantage with 50-photon boson sampling assuming current optical technology.
[19] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. 73, 58 (1994).
[20] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walsley, Optica 3, 1460 (2016).
[21] H. de Guise, O. Di Matteo, and L. L. Sánchez-Soto, Phys. Rev. A 97, 023238 (2018).
[22] G. W. Stewart, SIAM Rev. 19, 634 (1977).
[23] G. W. Stewart, Numer. Math. 40, 297 (1982).
[24] B. D. Sutton, Numer. Algorithms 50, 33 (2009).
[25] I. Dhand and S. K. Goyal, Phys. Rev. A 92, 043813 (2015).
[26] H. Qi, L. G. Helt, D. Su, Z. Vernon, and K. Brádler, arXiv preprint arXiv:1812.07015 (2018).
[27] H.-H. Lu, J. M. Lukens, N. A. Peters, O. D. Odele, D. E. Leaird, A. M. Weiner, and P. Lougovski, Phys. Rev. Lett. 120, 030502 (2018).
[28] C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, B. Fischer, S. Loranger, R. Kashyap, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, W. J. Munro, J. Azaña, M. Kues, and R. Morandotti, Nat. Phys. 14 (2018), 10.1038/s41567-018-0347-x.
[29] J. C. García-Escartín and P. Chamorro-Posada, J. Opt. 13, 064022 (2011).
A: Elimination-based decomposition details

In this section, we provide details for the elimination-based decomposition. The procedure to decompose $U$ into SU(2) matrices is to multiply $U$ by a series of $T_{mn}$ in an appropriate order such that $U$ is diagonalized to become a matrix $D$ with complex elements of unit modulus on the diagonal [19, 20]. We follow the same procedure as in Refs. [19, 20] but combine an appropriate number of beamsplitters to form universal U($M$) transformations, i.e., each U($M$) requires $M(M - 1)/2$ beamsplitters. To clarify the procedure, we consider the concrete example of decomposing a SU(7) matrix into U(3) and U(2) matrices (plus some phase shifters).

A general SU(7) matrix is represented by Fig. 4, where the bottom off-diagonal part is not explicitly shown for simplicity. In the first step, we multiply $U$ by $(T_{67}T_{56}T_{67})^{-1}$ to get $U^{(1)}$ for which the three entries represented by the “heart” in Fig. 4 are zero. The order of elimination is: $U_{17} \to U_{16} \to U_{27}$, and generally the parameters of the above two $T_{67}$ are not the same.

In the second step, we multiply $U^{(1)}$ by $(T_{23}T_{12}T_{23})^{-1}$ to get $U^{(2)}$ for which the three entries represented by the “sun” are zero. The order of elimination is similar: $U_{15} \to U_{14} \to U_{25}$. In the third step, we multiply $U^{(2)}$ by $T_{46}^{-1}$ to get $U^{(3)}$ for which the entry $U_{26}$ represented by a “diamond” is set to be zero. In the fourth step, we multiply $U^{(3)}$ by $(T_{23}T_{12}T_{23})^{-1}$ to get $U^{(4)}$ for which the three entries represented by the “moon” are zero. The order of elimination is: $U_{13} \to U_{12} \to U_{23}$. In the fifth step, we multiply $U^{(4)}$ by $T_{24}^{-1}$ to get $U^{(5)}$ for which the entry $U_{24}$ represented by a “pentagon” is set to be zero. At this point all entries in the first two rows of the top off-diagonal part are zero. We continue the above procedure to set all entries in the top off-diagonal part are zero. The whole process can be described by Eq. (3) in the main text and the SU(7) unitary matrix can be written as

$$U = D(T_{67}T_{56}T_{67})U_{40}(T_{45}T_{34}T_{45})(T_{67}T_{56}T_{67})U_{24} \times (T_{23}T_{12}T_{23})U_{46}(T_{45}T_{34}T_{45})(T_{67}T_{56}T_{67}),$$

where $D = \text{diag}(e^{i\delta_1}, e^{i\delta_2}, \ldots, e^{i\delta_7})$. Therefore, the SU(7) matrix can be decomposed into 6 U(3) matrices as well as 3 U(2) matrices, and 6 phase shifts which are easy to implement. The decomposition of Eq. (6) can be represented by the circuit in Fig. 5 where the phase shifts are not explicitly shown. The three beam splitters in each bracket in Eq. (6) implement a three-mode unitary (a tritter) and are each represented by a V-box in Fig. 5. The remaining single beam splitters $T_{46}$ and $T_{24}$ are represented by $W$-boxes. Note that $T_{46}$ does not represent a coupling between adjacent modes. In the physical implementation, one has to swap modes 4 and 5 such that mode 4 can interact with mode 6, and swap them back to the original order after the interaction. One has to do similar swaps when implementing the single beamsplitter $T_{24}$. Physically, the swap is a special beamsplitter. We can combine two swaps and a single beam splitter to form a three-mode unitary, which we will call “residual” unitary and is included in a brown box in Fig. 5. However, note that the three-mode residual unitary is not universal because only one beamsplitter is tunable.

Our next step is to eliminate the 6 nonzero entries $U_{2,n-4}^{(1)}$, $U_{3,n-3}^{(1)}$, $U_{4,n-2}^{(2)}$, $U_{4,n-1}^{(1)}$, $U_{4,n+1}^{(1)}$, and $U_{4,n+2}^{(1)}$, where the superscripts represent the order of elimination. This can be done by sequentially applying $T_{n-4,n}^{-1}$, $T_{n,n+1}^{-1}$, $T_{n-3,n}^{-1}$, $T_{n-2,n}^{-1}$, $T_{n-1,n}^{-1}$, and $T_{n-2,n}^{-1}$. These 6 beamsplitter transformations are combined together to form the specialized non-universal unitary $W$. Note that $T_{n-4,n}^{-1}$, $T_{n-3,n}^{-1}$, and $T_{n-2,n}^{-1}$ involve nonadjacent interactions, which means swaps are required to bring nonadjacent modes adjacent in the physical implementation. To physically implement

FIG. 4. Decomposition of a SU(7) matrix. Each off-diagonal entry has to be null by a $2 \times 2$ beam splitter unitary. The beam splitters that nullify the three nearby entries represented by the same shape are collected to form a U(3) unitary matrix. The three entries represented by “diamond”, “pentagon” and “hexagon” are nullled by three beam splitters independently.

For a larger positive integer $M$, the procedure to get U($M$) is straightforward but the determination of the residual unitary $W$ requires more explanation. We start from a concrete example of decomposing a U($N$) unitary matrix into U($5$) unitary matrices. After applying U($5$) unitaries, a submatrix of a SU($N$) matrix can be written as

$$
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & U_{4,n-4}^{(1)} & 0 & 0 & U_{4,n-1}^{(2)} \\
U_{4,n-4}^{(2)} & 0 & U_{4,n-2}^{(3)} & 0 & 0 \\
U_{4,n-3}^{(3)} & U_{4,n-4}^{(1)} & 0 & U_{4,n+1}^{(1)} & 0 \\
U_{4,n-1}^{(4)} & U_{4,n-2}^{(3)} & U_{4,n+1}^{(1)} & U_{4,n+2}^{(1)} & 0
\end{pmatrix}.
$$

(7)
In this section, we make a comparison of photon loss between the fully temporal architecture and hybrid architecture. To characterize the loss of an architecture, we define the overall transmission as the transmission coefficient when the architecture implements an identity unitary transformation. For the temporal architecture, the main sources of photon loss are the propagation loss in the inner loop and outer loop, loss in the beam splitter, and the switching and coupling loss to the outer loop. We use the transmission coefficients \( \eta_i, \eta_o, \eta_{BS} \) and \( \eta_{sc} \) to represent those losses in the temporal architecture. Among these losses, the switching and coupling loss is dominant.

FIG. 6. A residual unitary \( \tilde{W} \) for \( M = 5 \). It consists of 6 tunable beamsplitters (orange B-boxes) and 12 swaps.

B: Loss analysis

In this section, we make a comparison of photon loss between the fully temporal architecture and hybrid architecture. To characterize the loss of an architecture, we define the overall transmission as the transmission coefficient when the architecture implements an identity unitary transformation. For the temporal architecture, the main sources of photon loss are the propagation loss in the inner loop and outer loop, loss in the beam splitter, and the switching and coupling loss to the outer loop. We use the transmission coefficients \( \eta_i, \eta_o, \eta_{BS} \) and \( \eta_{sc} \) to represent those losses in the temporal architecture. Among these losses, the switching and coupling loss is dominant.

To implement a layer of beam splitters, each pulse has to pass through the inner loop and the outer loop once, and the beamsplitter and the switch twice. Therefore, the overall transmission coefficient when implementing an \( N \)-mode interferometer is

\[
\eta_{t} = \left( \eta_i \eta_o \eta_{BS} \eta_{sc} \right)^{N-1} \eta_{o}^{-1},
\]

where we have assumed that each mode has identical loss.

For the spatial-temporal hybrid architecture, the main sources of photon loss are similar. If the block unitary is integrated on chip, then the dominant loss would come from the coupling in and out of the chip. We use the transmission coefficients \( \tilde{\eta}_i, \tilde{\eta}_o, \tilde{\eta}_{BS} \) and \( \tilde{\eta}_c \) to represent, respectively, each of these losses in the hybrid architecture. To implement a layer of block unitaries, each pulse has to pass through the inner loop and the outer loop once, and the beamsplitter and the switch twice. Therefore, the overall transmission coefficient when implementing an \( N \)-mode interferometer is

\[
\tilde{\eta}_t = \left( \tilde{\eta}_i \tilde{\eta}_o \tilde{\eta}_{BS} \tilde{\eta}_{c} \right)^{N-1} \tilde{\eta}_{o}^{-1} \tilde{\eta}_{c},
\]

where \( k = (N-1)/(M-1) \) is the number of layers of block unitaries.

To make a comparison, we have to specify numeric values of the transmission coefficients, values that depend crucially on the platforms the architectures are implemented on. In general, we can assume \( \tilde{\eta}_i \approx \eta_i \) and \( \tilde{\eta}_{BS} \approx \eta_{BS} \). Since the outer loop of the temporal architecture is \( k \) times longer than that of the hybrid architecture, we have \( \eta_o = \tilde{\eta}_{o}^k \). The loss in the coupling in and out of the chip is usually higher than the loss in the coupling in and out of the outer loop (likely implemented in fiber), namely \( \eta_c < \eta_{sc} \). However, it is still possible that the overall loss of the hybrid architecture is lower than the
FIG. 7. A residual unitary $W$ for $M = 5$. From $\tilde{W}$ to $W$, another 12 swaps are needed. However 6 swaps (red circles) cancel with each other and one swap (blue rectangle) can be absorbed into a tunable beam splitter.

temporal architecture. From the above assumptions, the ratio between $\eta_t$ and $\tilde{\eta}_t$ is

$$\frac{\eta_t}{\tilde{\eta}_t} \approx \eta_i^{N-k} \eta_o^{N-3} \eta_{sc}^{2(N-1)} \eta_c^{-2(k-1)} \approx \eta_{sc}^{2(N-1)} \eta_c^{-2(k-1)},$$

(10)

where in the last step we assume that $\eta$ and $\eta_o$ are very close to one. From Eq. (10) it can be estimated that $\eta_t/\tilde{\eta}_t \leq 1$ when $\eta_c \geq \eta_{sc}^{3/2}$. This means the overall loss of the hybrid architecture can be lower than that of the temporal architecture when $\eta_c$ is sufficiently large. As a concrete example, we compare $\eta_t$ and $\tilde{\eta}_t$ by choosing representative values: $\eta_i = \eta_o = 0.9999$, $\eta_{sc} = 0.95$, $\eta_{BS} = 0.96$ and $\eta_c = 0.5$, as shown in Fig. 3 of the main text.