# 通用低深度双幺正设计的可编程光子电路

S. A. Fldzhyan,[1] M. Yu. Saygin,[2, 1, *] and S. S. Straupe[2, 1]

[1] *Faculty of Physics, M. V. Lomonosov Moscow State University, Leninskie Gory 1, Moscow, 119991, Russia*

[2] *Sber Quantum Technology Center, Kutuzovski prospect 32, Moscow, 121170, Russia*

大规模、可编程光子电路的发展对于执行通用矩阵-向量乘法至关重要，这对于经典和量子信息处理都是必不可少的。然而，这一目标受到了高损耗、硬件错误以及可编程性困难的阻碍。我们提出了一种增强的可编程光子电路架构，它最小化了电路深度并提供了分析可编程性，这些特性在之前的电路设计中尚未同时实现。我们的提案利用了一种以前被忽视的一般非幺正矩阵表示形式，即两个幺正矩阵之和。此外，类似于传统的基于奇异值分解（SVD）的电路，我们基于幺正和的架构继承了构成幺正电路的优势。总体而言，与现有方法相比，我们的提案为矩阵-向量乘法提供了显著改进的解决方案。



## 介绍

可编程光子电路越来越多地被用作经典信息处理的节能解决方案 [1–3] ，并在光学量子计算中发挥着至关重要的作用 [4–7] 。然而，作为模拟设备，光子电路面临着诸如误差、损耗和有限可编程性等挑战，这些都阻碍了它们的可扩展性，并使它们无法与传统的数字电子方法竞争。

支撑可编程多模式光子计算能力的核心数学运算是矩阵向量乘法（MVM）。该运算对于数据变换 [8–10]、权重调整以及光子神经网络中的学习过程 [1] 至关重要。在光子架构的最新进展中，主要集中在执行单位传输矩阵乘法的可编程干涉仪上。已经提出了几种用于可编程光子电路的设计，并被研究者广泛采用。值得注意的是，Reck 等人 [11] 和 Clements 等人 [12] 的设计尤为突出，因为它们得益于一种计算给定目标矩阵所需相移值的分析过程，假设电路无错误。此外，还提出了单位可编程电路中的误差校正方法 [13]，并且提出了一些新的电路设计，即使在高错误率的情况下也能保持通用性 [14–16]。

然而，许多信息处理任务需要涉及超出酉群范围的更广泛的矩阵的 MVM 操作。这种情况在经典光子神经网络 [3]、迭代求解器 [2] 和量子图问题求解器 [7] 中很常见。已经提出了两种主要方法来实现非酉矩阵的光子乘法操作。第一种是已建立的奇异值分解（SVD）方法，它使用一系列可编程多模酉电路、模式内幅度调制和另一个酉电路 [17]。这种方法的优势在于其直接的可编程性，这一特性继承自分析上可编程的酉电路。

第二种方法涉及将目标非幺正矩阵嵌入到更大的幺正矩阵中 [16, 18, 19] 。这些方法通常需要的电路深度比基于 SVD 的电路短。例如，唐等人 [18] 提出将目标非幺正矩阵嵌入到深度仅为相应基于 SVD 的电路一半的幺正电路中。最近，我们提出了一种更加紧凑且与平面集成光子学兼容的架构 [19] 。然而，这两种方法都牺牲了分析编程的便利性，由于缺乏有效的编程算法和现场训练方法，使得它们的实际使用变得具有挑战性。这反过来可能会破坏光电 MVM 的所有计算优势。

在这项工作中，我们提出了一种新的可编程光子电路架构，该架构显著改进了基于 SVD 的电路 [17] 和我们最近提出的低深度架构 [19]。所提架构基于将目标非幺正矩阵表示为两个幺正矩阵之和。我们的方法结合了基于 SVD 的方法和嵌入到幺正中的方法的优点。具体而言，从我们架构中导出的可编程电路既具有低深度又可以进行解析编程。

* saygin@physics.msu.ru

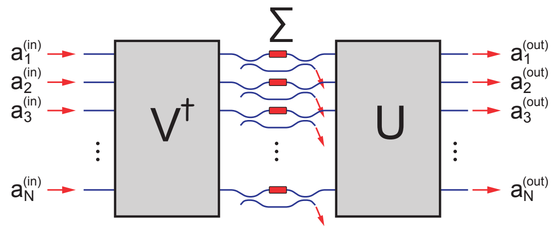


Figure 1. 基于 SVD 的通用矩阵向量乘法操作可编程电路。

## I. 低深度电路

我们关注通用矩阵向量乘法运算，在此过程中，场振幅输入向量 $\boldsymbol{a}^{(in)} = (a_1^{(in)}, \dots, a_N^{(in)})^{\mathrm{T}}$ 与目标 $M \times N$ 复数矩阵 $W$ 相乘：

$$\boldsymbol{a}^{(out)} = W\boldsymbol{a}^{(in)}, \tag{1}$$

在传播通过编程的多端口线性光学电路时，以在其输出处获得结果 $\boldsymbol{a}^{(out)} = (a_1^{(out)}, \dots, a_M^{(out)})^{\mathrm{T}}$。在下文中，我们假设方阵 $W$ 具有 $N = M$，这不会削弱我们成果的一般性。通过光学执行 MVM 的一般任务归结为构建一个可编程多端口电路，该电路可以实现一类感兴趣的传输矩阵 $W$。

在这项工作中，我们感兴趣的是通用 MVM——MVM 的最一般形式，在这种形式中，电路需要实现任意复数矩阵 $W$，该矩阵可以由 $2N^2$ 个实参数进行参数化。截至今天，为了达到这一目标，最常利用的方法是使用基于 SVD 的电路，如图 1 所示。SVD 将任意非幺正复数矩阵 $W$ 表示为三个矩阵的乘积：

$$W = U\Sigma V^\dagger, \tag{2}$$

每个矩阵都由一个不同的光子单元实现，如图 1 所示。在 (2) 中，$U$ 和 $V$ 是酉矩阵，并且 $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_N)$ 是由矩阵 $W$ 的实奇异值 $\sigma_j$ 组成的对角矩阵。严格来说，对于任意矩阵 $W$，奇异值 $\sigma_j$ 是无界的，然而，SVD 的线性光学实现施加了约束条件 $|\sigma_j| \le 1$。幸运的是，这并没有减少线性光学实现的应用范围，因为目标矩阵 $W$ 可以被重新缩放以满足该约束。

### 1. 提案

我们在本文中提出的电路基于一个简单的观察，即乘积 (2) 可以重写为两个幺正矩阵的和。为了说明这一点，我们将对角矩阵 $\Sigma$ 重写如下：

$$\Sigma = \frac{1}{2}(D + D^*), \tag{3}$$

其中 $D = \mathrm{diag}(e^{i\psi_1}, \dots, e^{i\psi_N})$ 是对角幺正矩阵，具有 $e^{i\psi_j} = \sigma_j + i\sqrt{1-\sigma_j^2}(j = \overline{1, N})$。然后，将 (3) 代入 (2)，我们得到表达式：

$$W = \frac{1}{2}(UDV^\dagger + UD^*V^\dagger). \tag{4}$$

执行根据 (4) 的 MVM 的光学方案如图 2 所示。该方案的主要部分由两个 $N$ 模式的幺正电路组成，即 $U^{(1)} = UDV^\dagger$ 和 $U^{(2)} = UD^*V^\dagger$，并行工作而不是顺序工作，与图 1 中基于 SVD 的电路不同。MVM 乘法首先将要乘的场向量 $\boldsymbol{a}^{(in)}$ 在一个由 $N$ 个平衡分束器 BS 上进行分割，使得其每个成分均匀分布在幺正电路中。不失一般性，我们取分束器传输矩阵为

$$U_{\mathrm{BS}} = \begin{pmatrix} \sqrt{R} & i\sqrt{1-R} \\ i\sqrt{1-R} & \sqrt{R} \end{pmatrix}, \tag{5}$$

其中 $R$ 是理想情况下应保持平衡的 BS 的反射率（即 $R = R_0 = 1/2$）。在下面的内容中，我们写 $R = \cos^2(\pi/4 + \alpha)$，其中 $\alpha$ 量化了 BS 与平衡操作之间的偏差。经过酉电路的乘法后，振幅在另一组平衡耦合器 BS 中合并以产生 $2N$ 个振幅 $a_1^{(out)}, \dots, a_{2N}^{(out)}$。然后，MVM 结果 (1) 由 $N$ 模式 $a_1^{(out)}, \dots, a_N^{(out)}$ 携带。

利用两个并行可编程幺正变换的电路具有比基于 SVD 的方法深度低两倍的优点，这导致了相应更低的损耗。同时，我们的电路架构保留了根据给定的目标矩阵轻松重新编程的优势，而无需依赖其他低深度电路固有的计算密集型数值优化 [18, 19]。

### 2. 对硬件错误的容错能力

考虑到各种通用幺正操作的容错设计，例如电路 [14–16]，只要纠正连接两个幺正操作的

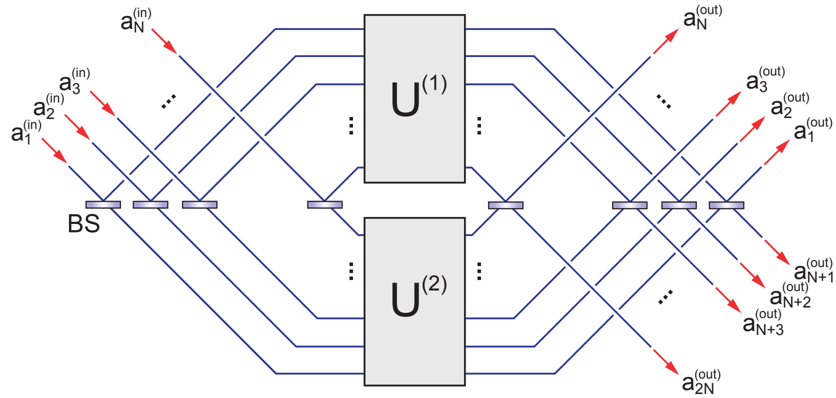


Figure 2. 实现两个酉矩阵 $U^{(1)}$ 和 $U^{(2)}$ 之和的可编程电路，提出用于与 $N\times N$ 非酉矩阵进行乘法运算。此处，在酉电路前后的 BSs 是平衡的。

BSs 中的错误，正在研究的非幺正电路也可以实现容错。这可以通过用可编程马赫-曾德尔干涉仪替换每个 $2N$BS 来实现，相对于 $N$ 引入了线性开销，与幺正操作中可编程元件的二次开销相比相对较小。然而，在我们的研究中，我们关注硬件错误的影响，假设这些 BSs 保持静态，但转移矩阵 $U^{(1)}$ 和 $U^{(2)}$ 不受错误影响。我们不假定特定的容错光子架构用于幺正电路；相反，我们将它们视为参数化矩阵。具体来说，我们通过非幺正矩阵 $K=\frac{U^{(1)}+U^{(2)}}{2}$ 的 $2N^2$ 个参数来参量化幺正算子 $U^{(1)}$ 和 $U^{(2)}$，这些参数的奇异值不超过 1。需要注意的是，如果没有硬件错误，则 $K=W$，然而，在一般存在误差的情况下则是 $K\neq W$。

因此，唯一有问题的元素是 BSs(5)，其值 $\alpha$' 表示相应的错误。考虑了两种误差模型的极限情况：i) 相关误差，所有 BS 具有相同的误差 $\alpha$，和 ii) 独立随机误差，每个 BS 的误差 $\alpha_j$ 遵循正态分布 $\sim\exp(-\alpha_j^2/2\sigma^2)$，其中 $\sigma$ 表示误差中的随机性程度。

为了评估实际电路在 BS 误差下实现 MVM 的准确性，我们将目标矩阵 $W^{(0)}$ 与电路实现的相应传输矩阵 $W$ 进行比较。为此，我们使用均方根误差（RMSE）：

$$\mathrm{RMSE}(W^{(0)},W)=\frac{1}{N}\sqrt{\sum_{i,j=1}^{N}|W_{ij}^{(0)}-W_{ij}|^2}, \tag{6}$$

其中 $W_{ij}$ 和 $W_{ij}^{(0)}$ 是矩阵的元素。此外，我们使用一种更为宽松的矩阵接近度量标准，允许 $W^{(0)}$ 和 $W$ 在全局缩放因子 $s$ 内重合。相应的接近度量是 $\mathrm{RMSE_s}(W^{(0)},W)=\mathrm{RMSE}(sW^{(0)},W)/|s|$ [19]。较小的 $|s|$ 值与较大的损失相关，因此它们在 $\mathrm{RMSE_s}$ 的分母中通过 $|s|$ 受到惩罚。为了确定实现目标非幺正矩阵 $W^{(0)}$ 并使其 RMSE/$\mathrm{RMSE_s}$ 尽可能低的幺正电路参数，我们采用了 L-BFGS 优化算法 [20]。在优化 $\mathrm{RMSE_s}$ 时，我们使用了之前工作中的一种方法 [19] 来找到使矩阵向量乘法误差最小化的同时确保其绝对值最大化的缩放因子 $s$。

我们使用了奇异值分解 (2) 来生成目标矩阵 $W^{(0)}$，其中酉复数矩阵 $U$ 和 $V$ 是通过基于 Ginibre 集合的随机矩阵的 QR 分解的方法从 Haar 随机分布中抽取的 [21]。对角矩阵 $\Sigma$ 的值由均匀分布范围 $[0,1]$ 内独立生成的值填充，然后进行重新缩放，使得 $\sigma_j$ 的最大值等于 1。

图 3 显示了模拟的结果。在图 3a 和 3b 中，分别绘制了 RMSE 和 $\mathrm{RMSE_s}$ 分布作为电路尺寸为 $N=10$ 和 $N=20$ 的功率反射率的函数。如图所示，即使 $R$ 偏离其平衡值 1/2 仅有很小的偏差也会降低矩阵运算的质量。然而，当允许转移矩阵通过全局因子 $s$ 进行缩放时，相干误差可以通过幺正电路完美纠正，并通过 $\mathrm{RMSE_s}$ 测量 MVM 的质量。请注意，用于校正相干误差所需的缩放因子的绝对值会随着误差值的增加单调减少。

图 3c 和图 3d 表明了 MVM 质量的不同行

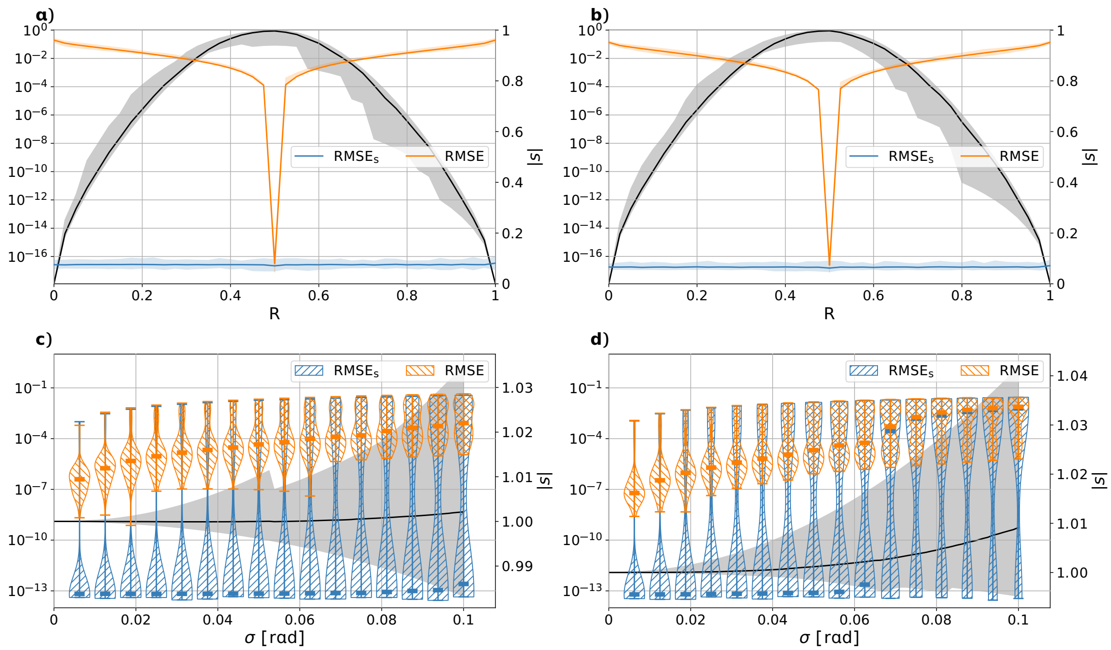


Figure 3. 分束器误差对光子矩阵向量乘操作质量的影响。RMSE 和 RMSE$_s$ 的分布如下所示：i) 作为 BS 反射率 $R$ 函数的相干误差在 $N = 10$(a) 和 $N = 20$(b)，以及 ii) 作为分布宽度 $\sigma$ 函数的独立采样随机误差在 $N = 10$(c) 和 $N = 20$(d)。每个 RMSE/RMSE$_s$ 分布都是基于 100 随机抽样的目标矩阵。对应于 RMSE$_s$ 的 $|s|$ 的分布也在图中绘制。对于独立采样随机误差对应的分布，最小值、最大值和中位数条形被绘制。

为，这种行为通过未缩放和缩放的度量进行了量化。然而，用 RMSE$_s$ 量化的缩放矩阵的实现要比用 RMSE 量化的未缩放矩阵好得多。需要注意的是，这一校正是通过与 1 相差很小值的缩放因子获得的。

总结来说，我们提出了一种可编程光子电路的架构，能够执行 MVM 操作，这一构想源于非幺正矩阵分解为两个幺正矩阵之和的方法——该方法之前在量子计算理论 [22] 中被使用过，并且据我们所知，在光子 MVM 的背景下从未应用过。这种架构的优势在于其电路深度仅为基于 SVD 干涉仪电路的一半，而后者是当今利用可编程光子学执行 MVM 的标准方法。此外，与最近研究的低深度电路 [18, 19] 相比，使用我们提出的可编程电路是一个更高效的选择，因为我们的电路可以进行解析编程同时保持相同的电路深度。我们认为我们的结果将对光子信息处理的各种应用具有价值。

## II. 致谢

S.A.F. 感谢俄罗斯理论物理与数学推进基金会（BASIS）(项目编号：23-2-10-15-1)。

[1] T. F. de Lima, H.-T. Peng, A. N. Tait, M. A. Nahmias, H. B. Miller, B. J. Shastri, and P. R. Prucnal, Machine learning with neuromorphic photonics, Journal of Lightwave Technology **37**, 1515 (2019).

[2] M. Chen, Q. Cheng, M. Ayata, M. Holm, and R. Penty, Iterative photonic processor for fast complex-valued matrix inversion, Photon. Res. **10**, 2488 (2022).

[3] L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, Photonic neural networks: A survey, IEEE Access **7**, 175827 (2019).

[4] J. Carolan, C. Harrold, C. Sparrow, *et al.*, Universal linear optics, Science **349**, 711 (2015).

[5] K. Alexander, A. Benyamini, D. Black, *et al.*, A manufacturable platform for photonic quantum computing, Nature 10.1038/s41586-025-08820-7 (2025).

[6] H. Aghaee Rad, T. Ainsworth, R. N. Alexander, *et al.*, Scaling and networking a modular photonic quantum computer, Nature **638**, 912 (2025).

[7] R. Mezher, A. F. Carvalho, and S. Mansfield, Solving graph problems with single photons and linear optics, Phys. Rev. A **108**, 032405 (2023).

[8] A. Melloni, A. Martinez, G. Cavicchioli, S. Seyedinnavadeh, F. Zanetto, D. Miller, and F. Morichetti, Programmable photonics for free space optics communications and computing, in *2023 IEEE Photonics Conference (IPC)* (2023) pp. 1–2.

[9] D. Pérez, I. Gasulla, P. D. Mahapatra, and J. Capmany, Principles, fundamentals, and applications of programmable integrated photonics, Adv. Opt. Photon. **12**, 709 (2020).

[10] D. Pérez-López, A. Gutierrez, D. Sánchez, A. López-Hernández, M. Gutierrez, E. Sánchez-Gomáriz, J. Fernández, A. Cruz, A. Quirós, Z. Xie, J. Benitez, N. Bekesi, A. Santomé, D. Pérez-Galacho, P. DasMahapatra, A. Macho, and J. Capmany, General-purpose programmable photonic processor for advanced radiofrequency applications, Nature Communications **15**, 1563 (2024).

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

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

[13] S. Bandyopadhyay, R. Hamerly, and D. Englund, Hardware error correction for programmable photonics, Optica **8**, 1247 (2021).

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

[15] M. Y. Saygin, I. V. Kondratyev, I. V. Dyakonov, S. A. Mironov, S. S. Straupe, and S. P. Kulik, Robust architecture for programmable universal unitaries, Phys. Rev. Lett. **124**, 010501 (2020).

[16] R. Hamerly, S. Bandyopadhyay, and D. Englund, Asymptotically fault-tolerant programmable photonics, Nature Communications **13**, 6831 (2022).

[17] L. N. Trefethen and D. Bau, *Numerical Linear Algebra* (SIAM, 1997).

[18] R. Tang, R. Tanomura, T. Tanemura, and Y. Nakano, Lower-depth programmable linear optical processors, Phys. Rev. Appl. **21**, 014054 (2024).

[19] S. A. Fldzhyan, M. Y. Saygin, and S. S. Straupe, Low-depth, compact, and error-tolerant photonic matrix-vector multiplication beyond the unitary group, Opt. Express **32**, 46239 (2024).

[20] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods **17**, 261 (2020).

[21] F. Mezzadri, How to generate random matrices from the classical compact groups (2007), arXiv:math-ph/0609050 [math-ph].

[22] N. Suri, J. Barreto, S. Hadfield, N. Wiebe, F. Wudarski, and J. Marshall, Two-Unitary Decomposition Algorithm and Open Quantum System Simulation, Quantum **7**, 1002 (2023).