arXiv:2301.13658v2 [cs.ET] 2 May 2023

基于少层冗余多平面光转换的可编程酉变换器迭代配置

Yoshitaka Taguchi,[1,*] Yunzhuo Wang,[2] Ryota Tanomura,[1] Takuo Tanemura,[1] 和 Yasuyuki Oze ki[1]

[1] *Department of Electrical Engineering and Information Systems, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan* [2] *Preferred Networks Inc. Otemachi Bldg., 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-0004 Japan*（日期：2023年5月4日）

可编程酉光子器件正逐渐成为实现量子信息处理、机器学习和光通信中酉变换的有前景工具。这些器件通常采用由马赫-曾德尔干涉仪（MZI）构成的矩形网格，该网格具有清晰的数学结构，且可通过确定性方法进行配置。然而，这种网格架构对制造误差较为敏感，其校正技术仍在研究之中。相比之下，多平面光转换（MPLC）架构对制造误差具有更强的鲁棒性，但由于其数学结构复杂，尚未开发出配置该变换器的确定性方法。在本研究中，我们对矩阵距离进行数学回顾并提出一种新的范数，在此基础上提出了一种快速的 MPLC 迭代配置方法。数值模拟结果表明，增加少量冗余层可显著提高 MPLC 架构的收敛性，使其成为一种实用且有吸引力的方案。我们还在模拟中考察了有限分辨率和移相器串扰的影响。此外，我们提出了一种适用于仅使用强度检测的应用的相位不敏感距离。与以往方法相比，我们的方法准确度提高了多个数量级，速度提升了 20 倍。

I. 引言

在集成光子平台上实现的可编程酉变换正日益成为一种强大的工具，可用于多种应用，包括量子光子学 [1–7], 机器学习 [8–14], 和光通信 [15–19]。准确实现给定的酉变换至关重要，因为计算结果的保真度和光通信中的误差可能会受到所实现变换精度的显著影响。合成酉变换的一种常见方法是使用由马赫–曾德尔干涉仪（MZI）组成的网格，即Clements架构 [20], ，该架构包含移相器和分束器（BS）。这一架构颇具吸引力，因为其数学结构具有可分解性，从而可以根据给定的酉变换明确确定每个MZI所需的相移。然而，诸如分束器分束比偏差等物理实现伪影，可能导致合成变换出现误差。随着光学模式数量的增加，这些误差可能变得十分显著 [21]。为减少或消除这一误差，人们提出了多种设计方案，旨在实现一种精确、可定制且能够容忍制造误差的酉变换，以应用于可扩展且可靠的场景。

为应对Clements架构中的实现伪影这一挑战，已有研究提出了几种方法。其中一种方法是局部误差校正，它涉及修正每个MZI，可应用于任何基于MZI的架构，但需要预先了解无源和有源组件 [22]。另一种方法是利用芯片上的功率监测器测量组件，这样可以对每个MZI进行校准，但也会增加芯片尺寸和布线复杂度 [23, 24]。自配置和三MZI方法利用额外的分束器实现部分完美的线性操作，并采用反馈回路，仅使用输出信号调节每个相移 [25, 26]。尽管该方法能够实现无限扩展，但也会增加电路尺寸，并可能存在稳定性问题 [27]。值得注意的是，这些方法主要考虑电路中无源分束器的实现伪影，而没有充分考虑移相器的实现伪影，例如串扰。

另一种架构采用移相器阵列与酉变换的串联连接，以实现一种高度稳健、能够抵抗制造误差的酉矩阵通用综合。该架构也称为多平面光转换（MPLC）架构 [28–31], ，之所以特别稳健，是因为每个酉变换都可以从广泛的可能酉变换中进行选择 [32–35]。酉变换几乎可以采用任何广为人知的N模混合器，从而显著提高灵活性和对制造误差的容差。然而，在该架构中配置移相器颇具挑战性，由于其复杂的数学结构，目前尚无明确的配置方法。该架构的优化必须应对其高维参数空间中存在的大量局部极小值[34]。因此，先前的研究一直依赖启发式全局搜索方法，例如盆地跳跃和模拟

* ytaguchi@ginjo.t.u-tokyo.ac.jp

退火，来配置移相器 [34–36]。然而，这些方法耗时较长，并且随着参数空间维度的增加，搜索时间呈指数级增长。为解决这一问题，有研究提出了一种基于机器学习的配置算法 [37]。尽管该算法可能提供一种解决方案，但它需要对结构进行准确的初始估计；如果初始估计包含误差，则可能导致矩阵保真度下降。

在本研究中，我们提出了一种新型、快速且可迭代配置的 MPLC架构，无需先验知识，仅依赖输出信号。该方法通过在现有 MPLC架构中添加少量冗余层，并采用基于梯度近似的基于导数的优化来实现。与 Clements架构中采用的类似方法相比，后者添加了大量冗余层；这种额外的层冗余显著提升了 MPLC架构的优化性能。与不含冗余的 Clements架构数值优化相比 [38], ，对于 $N = 128$ 种变换模式，我们提出的方法以少 1/20 的迭代次数，实现了高 5 个数量级的准确度 [21, 38], 。此外，与先前使用启发式算法优化 MPLC架构的研究相比 [39]，我们提出的方法实现了高 5 个数量级的准确度，配置速度也快了 23 倍。

本文结构如下。在讨论主要结果之前，我们首先在第二节讨论酉矩阵优化的重要一般性质，并引入一种新的距离度量。我们介绍的一个关键性质是，酉矩阵优化基本上不存在局部极小值。此外，我们提出了一种新的距离度量，即弗罗贝尼乌斯范数的一种与相位无关的变体，该度量对于输出模式上的相移具有不变性。此前，标准弗罗贝尼乌斯范数甚至也曾被用于与相位无关的应用中。在第三节中，我们研究采用少量冗余参数化层的 MPLC架构的优化性质。尽管酉矩阵的参数化可能导致优化陷入局部极小值，但我们通过数值模拟证明，添加少量冗余层可以有效避免这一问题。结果表明，该架构可以利用梯度下降算法等广为人知的局部最小化算法进行高效优化，而Clements架构则无法做到这一点。我们还研究了收敛的统计性质。在第四节中，我们考察一些实际场景，例如系统的梯度不可用、仅使用强度检测，以及移相器之间存在串扰的情况。我们评估了梯度近似和串扰对所提方法的影响，并表明该方法仍能保持良好性能，尽管优化后所达到的矩阵准确度会有所降低，或者达到收敛所需的迭代次数会增加。与相位无关的距离度量表现出相似的优化性质。在第五节中，我们总结全文。

II. 使用弗罗贝尼乌斯范数的矩阵距离

本节介绍弗罗贝尼乌斯范数的一些一般数学性质，并提出一种新的距离。我们首先定义酉群 $U(N)$ 上函数的单峰性概念，并证明使用弗罗贝尼乌斯范数的矩阵距离表现出这种单峰性。随后，我们明确范数的范围和期望值。此外，我们还为仅使用强度检测的应用引入了与相位无关的矩阵距离。

A. $U（N）$上的单峰性

这里介绍 $U(N)$ 上函数的单峰性概念。单峰性通常用于概率分布 [40] 的定义。对于多变量函数 $f: \mathbb{R}^N \to \mathbb{R}$，单峰性通过水平集 $L(f, \alpha) = \{\boldsymbol{x} | f(\boldsymbol{x}) \leq \alpha, \boldsymbol{x} \in \mathbb{R}^N\}$ 以及 $L(f, \alpha)$ [41, 42] 的凸性来定义。本文通过考察 $L(f, \alpha)$ 的路径连通性，将这一定义扩展到 $U(N)$ 上的函数，因为 $U(N)$ 不是凸集。

定义若函数 $f: U(N) \to \mathbb{R}$ 称为单峰函数，则对于任意 $\alpha \in \mathbb{R}$，水平集 $L(f, \alpha) = \{X | f(X) \leq \alpha, X \in U(N)\}$ 都是路径连通的。

换句话说，定义在 $U(N)$ 上的单峰函数的任何局部最小值也是全局最小值。

B. 弗罗贝尼乌斯范数的单峰性

我们证明，使用弗罗贝尼乌斯范数的酉矩阵距离是单峰的。两个酉矩阵 $d(X, U)$ 之间的距离定义为 $\|X - U\|_F$，其中 $\|A\|_F = \sqrt{\mathrm{Tr}\,[A^\dagger A]}$ 是弗罗贝尼乌斯范数。值得注意的是，使用均方误差（MSE）的矩阵距离 $\sum_{i,j} |X_{ij} - U_{ij}|^2$ 等价于 $d(X, U)^2$。给定一个酉矩阵 $U \in U(N)$，我们证明函数$f_U: U(N) \to \mathbb{R}$，定义为 $f_U(X) = d(X, U)$，是单峰的。首先，根据弗罗贝尼乌斯范数的定义，$f_U(X)^2$ 可简化为

$$\begin{aligned} f_U(X)^2 &= \mathrm{Tr}\left[(X-U)^\dagger (X-U)\right] \\ &= 2N - 2\,\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]. \end{aligned} \tag{1}$$

我们将 $U^\dagger X$ 的特征值写为 $\lambda_k (1 \leq k \leq N)$。由于 U 和 X 都是酉矩阵，所有特征值 λ_k 都满足 $|\lambda_k| = 1$。因此，特征值可以写为 $\lambda_k = e^{i\theta_k}$，其中 $-\pi \leq \theta_k \leq \pi$。利用这些特征值，公式 1 可进一步简化为

$$f_U(X)^2 = 2N - 2\sum_{k=1}^{N} \cos\theta_k. \tag{2}$$

公式 2 表明， $f_U(X)^2$ 是单峰的，因为 cos θ_k 在区间 $-\pi \le \theta \le \pi$ 上具有单峰性，且它们的和也具有单峰性。附录A中给出了这一单峰性的代数证明。

C. 范围与归一化

我们根据公式 2 推导出 $f_U(X)^2$ 的取值范围，并提出一种适用于距离的恰当归一化方法。当且仅当对所有 k 均有 $\theta_k = \pm\pi$ 时， $f_U(X)^2$ 取得最大值 $4N$；当且仅当对所有 k 均有 $\theta_k = 0$ 时，其最小值为 0。在以往的研究中， $f_U(X)^2$ 已通过 N [25]、$2N$ [38] 或 N^2 [35] 进行归一化。本文提出采用 $4N$ 进行归一化，从而得到 $0 \le f_U(X)^2/4N \le 1$。这是对该范数的一种良好归一化，其取值范围为 0 到 1，且与 N 无关。

D. 期望值

为计算期望值 $\mathbb{E}\big[f_U(X)^2/4N\big]$，我们考虑 θ_k 的分布。若 X 是从 Haar测度中采样得到的，则由于 Haar 测度的不变性， $U^\dagger X$ 也服从 Haar 随机分布。因此， $U^\dagger X$ 的特征值在单位圆 $|c| = 1$ 上均匀分布，并且有 $\theta_k \sim U(-\pi,\ \pi)$。因为

$$\mathbb{E}[\cos\theta_k] = \int_{-\pi}^{\pi} \frac{1}{2\pi}\cos\theta\,\mathrm{d}\theta = 0, \tag{3}$$

式 2 中第二项的期望值为 0。现在我们得出结论： $\mathbb{E}\big[f_U(X)^2/4N\big] = 2N/4N = 0.5$。在第 III 节收敛图的初始值中，可以通过数值方法观察到这一事实。

E. 相位不敏感距离

这里，我们使用弗罗贝尼乌斯范数介绍一种与相位无关的矩阵距离变体。该变体适用于仅检测输出模式强度的应用，因为距离不应受到酉转换器输出相位的影响。受益于这种与相位无关的距离的应用包括机器学习和量子光子学，在这些应用中，光电二极管或光子数计数器被置于输出端口处。图 1 展示了这样一种情形：一个复向量（s_1，s_2，……，s_n）$^\top$ 被输入两个酉变换器件。这些器件的传输矩阵分别表示为 P 和 Q，其复输出以极坐标形式表示为 $te^{i\theta}$。这两个器件的输出向量之间唯一的差异在于其相位，即$\theta_i \neq \theta_i'$。在仅检测输出模式强度的应用中，这两个矩阵 P 和 Q 被视为相同，因此引入一种适用于此目的的矩阵距离。在下文中，矩阵 U 表示给定的目标酉矩阵，而矩阵 X 表示酉转换器件所实现的实际转换。

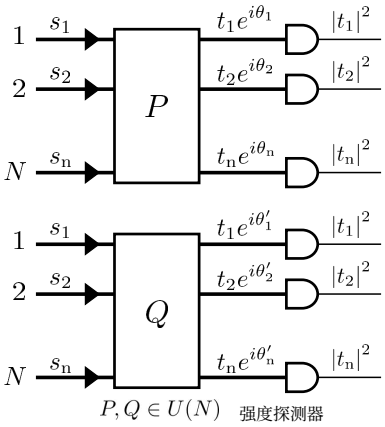


图 1：一个复向量（s_1，s_2，……，s_n）$^\top$ 被输入两个酉转换器件，其传输矩阵分别表示为 P 和 Q。使用相位不敏感距离进行评估时，这些器件的输出完全相同，唯一区别在于输出端口处的相位自由度。

为研究酉变换器输出相位的影响，我们将酉矩阵 $U^\dagger$ 和X 表示为

$$U^\dagger = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}$$
$$X = \begin{bmatrix}\mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top\end{bmatrix}, \tag{4}$$

其中，$\mathbf{u}_i$是 $U^\dagger$的列向量，而 $\mathbf{x}_i^\top$是 X的行向量。由于 $U^\dagger$和 X是酉矩阵，$\mathbf{u}_i$ 和 $\mathbf{x}_i$ 的范数均等于 1。在此背景下，酉变换器的输出相位对应于$\mathbf{x}_i$的相位。原始的弗罗贝尼乌斯范数会随$\mathbf{x}_i$的全局相位变化而改变，其定义为将$\mathbf{x}_i$ 替换为 $e^{i\theta_i}\mathbf{x}_i$。

我们分析弗罗贝尼乌斯范数对每个$\mathbf{x}_i$的全局相位的依赖关系。弗罗贝尼乌斯范数利用列向量$\mathbf{u}_i$ 展开为

$$\begin{aligned}&\|X-U\|_F^2 = \left\|XU^\dagger - I\right\|_F^2 \\ &= \|X\mathbf{u}_1 - \mathbf{e}_1\|^2 + \|X\mathbf{u}_2 - \mathbf{e}_2\|^2 + \cdots + \|X\mathbf{u}_n - \mathbf{e}_n\|^2,\end{aligned} \tag{5}$$

其中 $\mathbf{e}_i$ 是一个单位列向量，其第 i 个元素为 1，其余元素为 0。展开项 $\|X\mathbf{u}_1 - \mathbf{e}_1\|^2$，得到

$$\begin{aligned}&\|X\mathbf{u}_1 - \mathbf{e}_1\|^2 \\ &= |\mathbf{x}_1^\top\mathbf{u}_1 - 1|^2 + |\mathbf{x}_2^\top\mathbf{u}_2|^2 + \cdots + |\mathbf{x}_n^\top\mathbf{u}_n|^2.\end{aligned} \tag{6}$$

除第一项 $|\mathbf{x}_1^\top \mathbf{u}_1 - 1|^2$ 外，其余各项$|\mathbf{x}_i^\top \mathbf{u}_i|^2 (i \geq 2)$ 在 $\mathbf{x}_i$ 的全局相位变化下保持不变。只有第一项取决于 $\mathbf{x}_1$ 的全局相位。因此，项 $\|X\mathbf{u}_1 - \mathbf{e}_1\|^2$ 在行向量 $\mathbf{x}_2, \mathbf{x}3, \cdots \mathbf{x}_n$ 的全局相位变化下保持不变。类似地，$\|X\mathbf{u}_i - \mathbf{e}_i\|^2$ 与 $\mathbf{x}^j$ 的全局相位无关，其中 $j \neq i$。当 $\mathbf{x}_i$ 的全局相位发生变化时，仅当 $\mathbf{x}_i^\top \mathbf{u}_i$ 为正实数时，项 $|\mathbf{x}_i^\top \mathbf{u}_i - 1|^2$ 才取最小值，因为

$$|\mathbf{x}_i^\top \mathbf{u}_i - 1|^2 = |\mathbf{x}_i^\top \mathbf{u}_i|^2 - 2\,\mathrm{Re}\left[\mathbf{x}_i^\top \mathbf{u}_i\right] + 1 \tag{7}$$

并且只有第二项 $2\,\mathrm{Re}\left[\mathbf{x}_i^\top \mathbf{u}_i\right]$ 依赖于$\mathbf{x}_i$的全局相位。这启发我们考虑在弗罗贝尼乌斯范数 $\|X - U\|_F^2$ 中，将所有 $|\mathbf{x}_i^\top \mathbf{u}_i - 1|^2$ 替换为 $\left(|\mathbf{x}_i^\top \mathbf{u}_i| - 1\right)^2$。如公式 6 和公式 7 所述，由于相位依赖性质，在进行此替换后，弗罗贝尼乌斯范数的最小值与原弗罗贝尼乌斯范数的最小值相同。

基于上述考虑，我们提出一种相位不敏感矩阵距离。给定酉矩阵 U，我们定义距离函数 $h_U\colon\ U(N) \to \mathbb{R}$，仅使用形式为 $|\mathbf{x}_i^\top \mathbf{u}_j|$的项；这些项可通过对酉变换器输出端口进行强度测量获得。我们定义正实数 $a_{ij} = |\mathbf{x}_i^\top \mathbf{u}_j|^2 \geq 0$。所有 a_{ij} 均可通过酉变换器将列向量$\mathbf{u}_i$ 与矩阵 X 相乘获得，因为

$$X\mathbf{u}_j = \begin{bmatrix} \mathbf{x}_1^\top \mathbf{u}_j \\ \mathbf{x}_2^\top \mathbf{u}_j \\ \vdots \\ \mathbf{x}_n^\top \mathbf{u}_j \end{bmatrix} \tag{8}$$

以及 $|\mathbf{x}_i^\top \mathbf{u}_j|^2$ 可以通过对$\mathbf{x}_i^\top \mathbf{u}_j$进行强度测量获得。将式 6 代入式 5 展开所有项，并应用式 7 相关讨论中的替换，我们将 $h_U(X)$ 定义为

$$\begin{aligned} &h_U(X) = \\ &(\sqrt{a_{11}} - 1)^2 + a_{12} + \cdots + a_{1n} \\ &+ a_{21} + (\sqrt{a_{22}} - 1)^2 + \cdots + a_{2n} \\ &\vdots \quad\quad \vdots \quad\quad + (\sqrt{a_{ii}} - 1)^2 \cdots \\ &+ a_{n1} + a_{n2} + \cdots + (\sqrt{a_{nn}} - 1)^2 \\ &= \sum_{ij} \left(\delta_{ij} - \left|\left[XU^\dagger\right]_{ij}\right|\right)^2. \end{aligned} \tag{9}$$

该函数与$\mathbf{x}_i$的全局相位无关，也就是说，与单位变换器各输出的相位无关；其最小值与原始弗罗贝尼乌斯范数距离函数 $f_U(X)$ 的最小值相同。

函数 $h_U(X)$ 也是单峰函数。这可以通过反证法证明。假设 $h_U(X)$ 不是单峰函数，并且存在多个不连通的局部最小值。那么，对于所有满足 $h_U(X')$ 是局部最小值的矩阵 X'，必然至少存在一组相位（θ_1，θ_2，$\cdots\theta_n$），用于定义一个矩阵

$$Y = \begin{bmatrix} e^{i\theta_1}\mathbf{x}_1'^\top \\ e^{i\theta_2}\mathbf{x}_2'^\top \\ \vdots \\ e^{i\theta_n}\mathbf{x}_n'^\top \end{bmatrix} \tag{10}$$

使得 $f_U(Y)$ 也是一个局部最小值，其中$\mathbf{x}_i'^\top$是 X'的行向量。根据 $h_U(X)$ 的定义，可知这样的矩阵Y 存在。然而，这意味着 $f_U(X)$ 也有多个彼此不连通的局部最小值，这与前提矛盾。因此，$h_U(X)$ 必须是单峰的。

III. 少层冗余参数化

上一节讨论的由单峰性保证的全局优化性质，是在不对所优化矩阵作任何假设的情况下推导出的；然而，该器件合成的矩阵由物理参数参数化，而在此参数空间中，单峰性可能并不总是成立。我们将研究这一问题，并提出一种缓解该困难的方法。令 $l_U(X) \geq 0$ 表示一个从期望酉矩阵 U出发的距离函数，令 $X(\mathbf{p})$ 表示由物理转换器实现的酉矩阵，该转换器由实参数向量 $\mathbf{p}$ 参数化。$\mathbf{p}$ 中的每个元素都对应于实际器件中的相移量。如果 $l_U(X)$ 具有单峰性，则其梯度仅在 $X = \pm U$时才变为零向量。然而，在优化实际的酉转换器件时，我们需要考虑对 $l_U(X(\mathbf{p}))$ 进行标量优化。如果 $X(\mathbf{p})$ 的雅可比矩阵在任意 $\mathbf{p}$ 处均为满秩，这意味着对于任意无穷小的矩阵变化ΔX，都存在无穷小的参数变化 $\Delta\mathbf{p}$，那么由于前述 $l_U(X)$ 的单峰性质，函数 $l_U(X(\mathbf{p}))$ 也具有唯一的最小值。通过增加酉转换器件中的层数，参数空间中的自由度随之增加，这可能使 $X(\mathbf{p})$ 的雅可比矩阵更有可能为满秩。在本节中，我们将证明，在满足最小要求的基础上，将酉转换器件的层数增加少数几层，能够显著改善使用基于梯度的优化算法对 MPLC 架构进行的优化。

A. 器件定义与冗余

我们给出了酉变换器和少层冗余参数化的数学定义。图2a展示了基于MZI的酉变换器架构

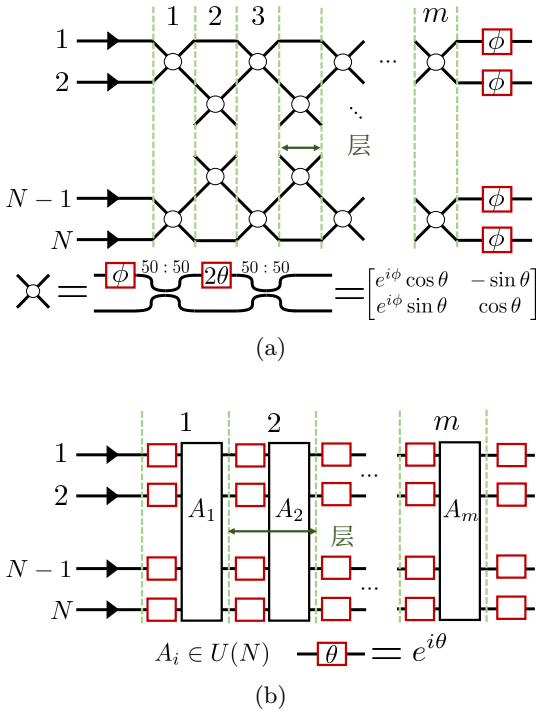


图 2： $N \times m$ Clements架构（a）和 $N \times m$ MPLC架构（b）的示意图。左侧端口为输入端口，右侧端口为输出端口。每种架构中的层数由 m指定。在Clements架构中，每层包含$N/2$ 个或（$N-1$）$/2$ 个MZI节点，图中以白色圆圈表示。每个MZI节点由两个移相器组成，分别用变量 ϕ 和 θ表示。在基于MPLC的幺正转换器（b）中，该架构由一个 N端口固定幺正转换器（以 A表示）以及其后的 N个单模移相器组成。

，通常称为Clements架构 [20]。MZI由两个50:50分束器和两个移相器组成，可以实现任意的 $U(2)$变换。本文不考虑MZI的任何缺陷。图2b展示了MPLC架构的结构。每层由一个 N端口固定幺正转换器 A_i 和一组 N 单模移相器组成。经过 m 层后，以类似于Clements架构的方式放置另一组移相器。该器件的整体变换记为 X，表示为

$$X = L_{m+1} A_m L_m \cdots A_2 L_2 A_1 L_1, \tag{11}$$

其中， A_i 是一个 N端口酉变换器的传输矩阵， L_i 表示为

$$L_i = \begin{bmatrix} e^{i\theta_{i1}} & & & \\ & e^{i\theta_{i2}} & & \\ & & \ddots & \\ & & & e^{i\theta_{in}} \end{bmatrix}. \tag{12}$$

对于任意 $i \neq j$，矩阵 A_i 和 A_j 都不相同。该矩阵的自由度总数为（$m+1$）（$N-1$）$+1$。这是因为每个移相器阵列有 $N-1$ 个自由度，这是由于全局相位损失了一个自由度；此外，整个器件还多出一个自由度，即全局相位。 N端口固定幺正转换器 A_i 可以通过多端口定向耦合器 [33],多模干涉耦合器 [30], 或其他多端口幺正变换器件实现。应谨慎选择器件，以确保整体变换 X 具有通用性。器件的混合熵可用作衡量通用性的指标 [43, 44]。要实现任意 U（N）变换，自由度总数必须大于 N^2 [45]。对于Clements架构，层数 m 必须满足 $m \geq N$[20]。同样，MPLC架构的层数 m 也必须满足 $m \geq N$，这可由（$m+1$）（$N-1$）$+1 \geq N^2$推出。在此背景下，少层冗余参数化架构定义为两种架构的层数均满足$m = N+1$， $N+2$。

B. 优化问题设置与算法

我们将矩阵优化问题表述如下。我们将实参数变量表示为向量 $\mathbf{p}$。参数的数量取决于 m 和 N。我们将两个矩阵之间的归一化代价函数 $\mathcal{L}$ 定义为

$$\mathcal{L}(\mathbf{p}) = \frac{1}{4N} \| X(\mathbf{p}) - U \|_F^2, \tag{13}$$

其中， $X(\mathbf{p})$是由参数向量$\mathbf{p}$在物理上实现的酉矩阵， U 是待实现的目标矩阵， $\|\cdot\|_F$ 是弗罗贝尼乌斯范数。代价函数 $\mathcal{L}$ 按第二节 C 所述除以 $4N$，且始终满足 $0 \leq \mathcal{L} \leq 1$。优化开始时，参数从 0 到 2π的均匀分布中初始化；目标酉矩阵 U 则使用 SciPy [46]的 stats 模块从 Haar 测度中采样。对于 MPLC 架构，矩阵 A_i 也从 Haar 测度中采样，用于 $1 \leq i \leq m$。初始化参数和矩阵后，使用 SciPy [46]的 optimize 模块中实现的 L-BFGS [47] 拟牛顿优化方法对代价函数 $\mathcal{L}$ 进行优化。L-BFGS 所需的代价函数 $\mathcal{L}$ 导数通过 JAX 框架[48]的自动微分计算。该方法从初始参数开始，

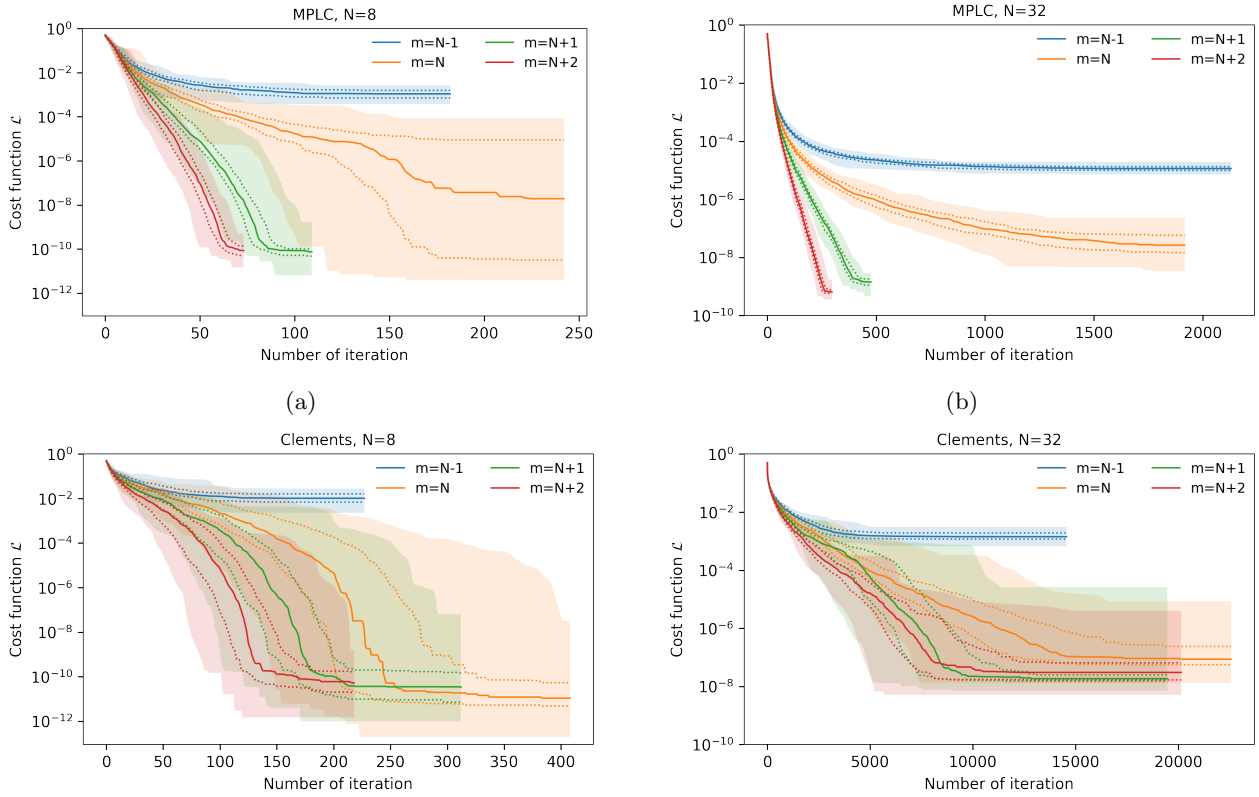


图3：MPLC架构和Clements架构的收敛图。纵轴显示代价函数的值 $\mathcal{L}$ ，该代价函数由公式13定义；横轴显示迭代次数。阴影区域表示最小值和最大值，实线表示中位数，点线表示64次优化试验中的第25%和第75%分位数。具有 (a) $N = 8$、(b) $N = 32$ 的MPLC架构，以及具有 (c) $N = 8$、(d) $N = 32$ 的Clements架构。

并在每一步对参数进行修改，直至收敛到局部最小值，其中 $d\mathcal{L}/d\mathbf{p} = \mathbf{0}$。在每一层中，我们使用$N$ 个实参数来表示 Clements 和 MPLC 架构中的 N 相移。通过改变初始参数，重复运行优化 64 次，以研究其统计行为。我们考察了 $N = 8$ 和$N = 32$ 的情况。

C. 结果

图3展示了层数变化时的收敛图。记录了64次优化试验的代价函数收敛图。阴影区域表示最小值与最大值的范围，虚线表示25%和75%分位数，实线表示各次试验的中位数。对于Clements架构和MPLC架构，参数化层设置不足都会导致大量误差。对于MPLC架构，$m = N$ 非冗余情形会导致较大的误差方差，尤其是在 $N = 8$时。这表明MPLC架构的参数空间中存在许多局部极小值，正如文献 $m = N$ 此前所报道的那样。

尽管 $N = 32$的非冗余设置 $m = N$ 的误差方差小于 $N = 8$，但 $N = 32$的误差仍然存在，这表明在该条件下同样存在不可避免的局部极小值。当增加层数并加入冗余自由度时，误差方差和误差都会变小，如 $m = N + 1, N + 2$的情况所示。相比之下，尽管Clements架构具有足够的自由度，但在所有条件下都会产生较大的误差方差。

对于端口数量较多的情况， $N = 128$，图4展示了与先前研究 [38] 中具有冗余的Clements架构的性能比较。与无冗余的情况相比，先前研究在经过20000次迭代后收敛于 $\mathcal{L} = 1.4 \times 10^{-2}$，而MPLC架构仅用其二十分之一的迭代次数，便取得了比具有冗余的Clements架构好6个数量级的结果。具有冗余层的MPLC架构 $m = N + 1$ 在收敛速度和准确度方面仍然表现更优，即使相比

与具有128个冗余层的Clements架构相比。

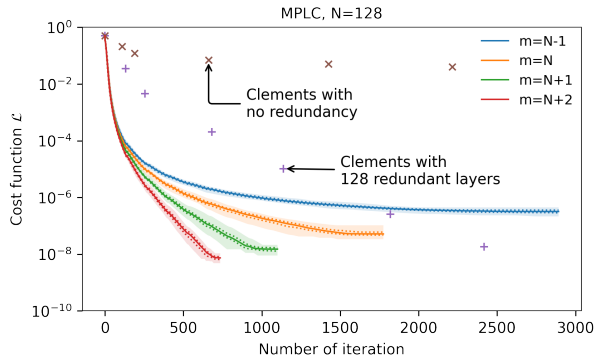
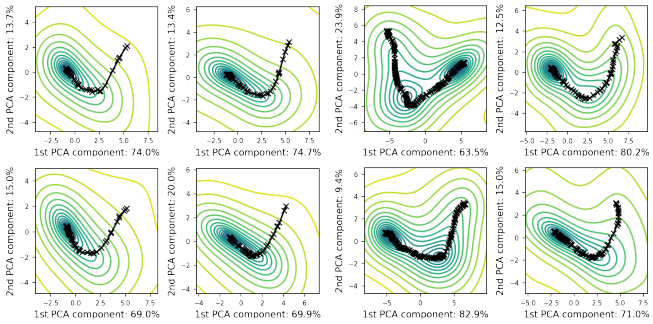


图4： $N = 128$ 的大规模情况与 [38]先前研究的比较。

我们采用 [49, 50] 中报告的方法以及[51]补充材料中的方法，在参数空间中可视化优化轨迹和损失函数。优化轨迹是优化过程在高维空间中形成的参数路径。我们记录了优化每一步的参数历史，并对这些记录应用主成分分析（主成分分析）。我们使用第一和第二个主成分，将高维路径投影到二维空间中。我们对 $N = 8$, $m = N + 1$进行了可视化。图5展示了投影后的轨迹，以及投影子空间中损失函数对数的等高线图。MPLC架构的等高线图类似于简单的椭圆单峰函数，而Clements架构的等高线图则更为复杂。这些架构的等高线图存在差异，或可解释图3所示的收敛图为何不同。

IV. 实际设置下的优化

我们讨论了将第三节提出的方法应用于真实器件优化时必须解决的三个挑战。首先，本节采用的优化方法需要目标函数的梯度。虽然可以通过物理方式获取梯度 [52], 但这需要额外的外部设备，使系统体积庞大且难以扩展。其次，该优化方法使用输出端的复振幅进行优化。在真实器件中读取复振幅需要在输出端配备相干探测器，这会使器件更加复杂。尽管某些应用需要相干探测，但许多基于光子学的光计算平台以及采用光子芯片的量子计算都使用强度检测。第三，真实器件中的移相器之间存在串扰，而优化方法并未考虑这一点。串扰在热光相移器中尤其成问题，



(a) (b)

图5：MPLC架构（a）和Clements架构（b）在 $N = 8$, $m = N + 1$下的优化轨迹的主成分分析投影，以及损失函数对数的等高线图。四幅图中的每一幅都展示了初始参数和目标矩阵随机变化时的优化轨迹。

尽管热光相移器因其较小的占地面积而颇具吸引力。

本节报告使用原始范数和相位不敏感范数、仅利用输出信号进行的无导数优化结果，并考察串扰的影响。我们首先介绍数学表述，然后给出数值结果。所采用的优化方法与第三节B部分相同。这些结果为设计无需额外组件的光学酉变换器铺平了道路，使该平台具有更好的可扩展性和多功能性。

A. 多元函数的梯度近似

我们采用数值梯度近似，仅使用函数值来近似解析梯度，并研究这种近似对优化行为的影响。多元函数 $\nabla f(x_1, x_2、\ldots\ldots、 x_n)$ 的梯度近似为

$$\nabla f(x_1, x_2, \ldots, x_n) \approx \begin{bmatrix} \frac{f(x_1+\Delta, x_2, \ldots, x_n) - f(x_1, \ldots, x_n)}{\Delta} \\ \frac{f(x_1, x_2+\Delta, \ldots, x_n) - f(x_1, \ldots, x_n)}{\Delta} \\ \vdots \\ \frac{f(x_1, x_2, \ldots, x_n+\Delta) - f(x_1, \ldots, x_n)}{\Delta} \end{bmatrix}. \tag{14}$$

其中， $\Delta \ll 1$ 表示有限差分。计算梯度近似所需的函数评估次数与参数数量相同。我们表明，使用这种梯度近似的基于导数的算法仍然能够有效地优化酉矩阵。

B. 强度检测器件的定义

我们评估了第 IIE 节介绍的相位不敏感距离的优化行为，该距离仅使用输出端口处的强度探测器。我们预计，相位不敏感距离在优化过程中的表现将与相位敏感距离相似，因为如第 IIE 节所示，它同样具有单峰性质。为检验这一点，我们移除了 MPLC架构中的最后一个移相器阵列，如图6所示。尽管这会从该架构中移除 N 自由度，但我们仍预计其优化行为将与标准相位敏感范数相似。

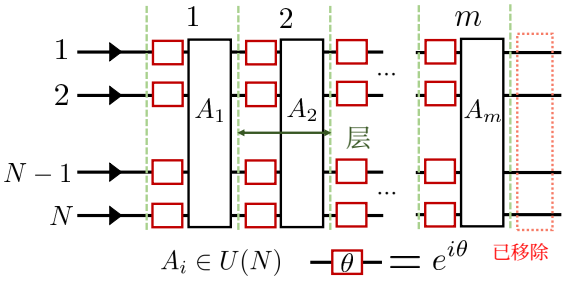


图 6：移除最后一个移相器阵列后的 MPLC架构示意图。

C. 串扰模型

我们通过考虑相邻移相器之间的相互作用来建立串扰模型。串扰表示为相移的线性组合，可表示为

$$\theta_i = \sum_j \alpha_{ij}\theta_j. \tag{15}$$

耦合模型和耦合系数 α_{ij} 如图7所示。以下模拟中的耦合表示为 $\theta_i' = 0.1\theta_{i-2} + 0.5\theta_{i-1} + \theta_i + 0.5\theta_{i+1} + 0.1\theta_{i+2}$。如果耦合参数$\mathbf{p}' = w(\mathbf{p})$是可逆的，则该耦合参数实现的酉矩阵 $X(\mathbf{p}')$将在原始 $X(\mathbf{p})$具有满秩雅可比矩阵的情况下，也具有满秩雅可比矩阵。我们使用梯度近似进行优化。

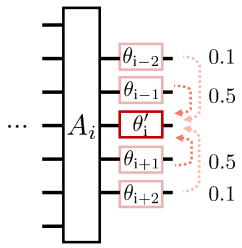


图7：MPLC架构的串扰模型。

D. 结果

图8展示了使用基于标准 Frobenius 范数的距离进行梯度近似时的收敛图。该近似使用 Δ 值 2^{-10}计算，对应于具有10位分辨率的相移。添加冗余层后，MPLC架构表现出受数值准确度限制的性能（对于 $m = N+1,\ N+2$），且误差的方差较小。由于采用梯度近似，每次优化迭代都需要评估距离，其次数与参数数量相同。例如，当 $N = 8$ 和 $m = N+1$时，每次优化都需要 $8\times(8+2) = 80$ 次距离评估。对于这种情况，使用MPLC架构时，优化在100次迭代后收敛，因此评估总次数约为8000次。

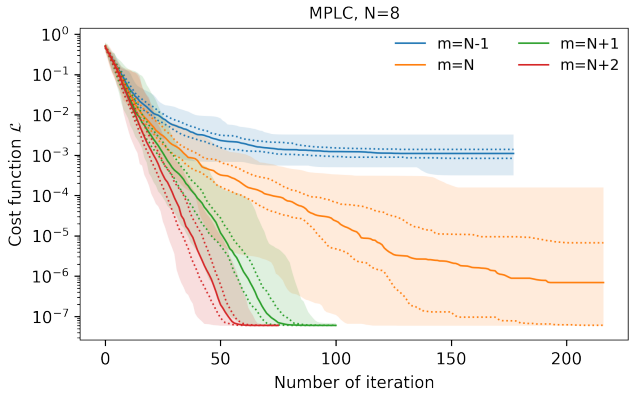


图8：采用梯度近似的MPLC架构的收敛图，其中 $\Delta = 2^{-10}$。64次优化试验以与图3相同的方式展示。

我们研究了图8中观察到的小误差方差对 Δ 的依赖性，该方差源于梯度近似。图9展示了在冗余层设置为 $m = N+1$ 时，每种梯度近似精度下的优化结果。随着有限差分 Δ 变小，最终误差也随之减小。如果梯度近似的准确度不足，即 Δ 不够小，优化结果的方差就非常小。例如，在 $\Delta = 2^{-6}$ 情况下，误差范围为 $[1.5\times10^{-5}, 1.8\times10^{-5}]$；在 $\Delta = 2^{-9}$ 情况下为 $[2.3\times10^{-7}, 2.5\times10^{-7}]$；在 $\Delta = 2^{-12}$ 情况下为 $[3.5\times10^{-9}, 4.5\times10^{-9}]$。如果梯度近似的准确度足够高（$\Delta \le 2^{-15}$），结果将具有不可忽略的方差，与图3a所示的结果相似。在 $\Delta = 2^{-15}$ 情况下，误差范围为 $[7.9\times10^{-10}, 6.8\times10^{-11}]$；在 $\Delta = 2^{-18}$ 情况下为 $[3.7\times10^{-10}, 4.6\times10^{-12}]$。这些结果为设计酉变换器系统的DAC分辨率提供了标准。

我们使用相位不敏感距离研究了优化性质。

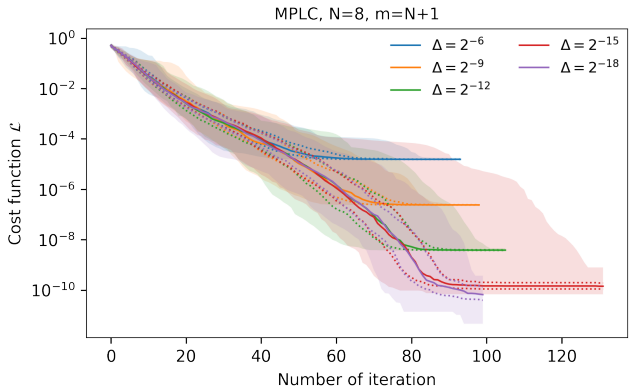


图9：改变MPLC架构梯度近似准确度时最终误差的比较。64次优化试验以与图3相同的方式显示。

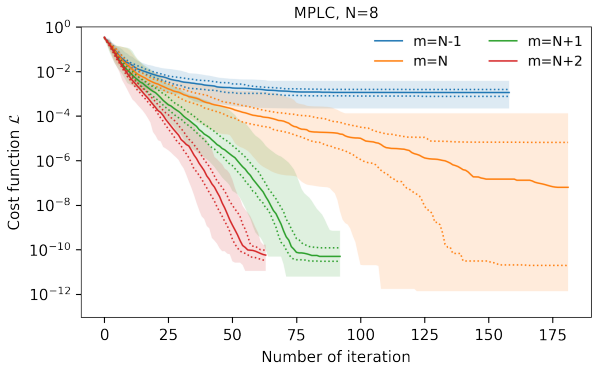


图10：使用相位不敏感距离并采用 $N = 8$的MPLC架构收敛图。64次优化试验以与图3相同的方式显示。

图10显示了使用解析梯度获得的优化结果。尽管自由度有所减少，但收敛曲线与图3所示的曲线相似。然而，在使用梯度近似并比较准确度依赖性时，如图11所示，相位不敏感距离与标准的基于弗罗贝尼乌斯范数的距离呈现出不同的结果。与图9中$\Delta = 2^{-6}$、2^{-9}、2^{-12} 所示的情况相反，所有优化结果都具有较大的方差。有限差分 Δ必须小于 2^{-18}，才能实现与使用解析梯度所获得的结果相当的优化结果。黑色虚线表示先前研究 [35]中通过模拟退火得到的优化结果。他们实现了 $f_{\mathrm{MSE}} = -50$ dB的误差，对应于 $\mathcal{L} = 2 \times 10^{-5}$。使用我们的方法，在有限差分为 $\Delta \leq 2^{-9}$时即可达到该误差；通过更准确的梯度近似，还可以进一步提升多个数量级。当准确度足够高时，我们的优化方法经过约100次迭代即可收敛。为了将我们的方法与先前一项使用强度检测进行MPLC架构优化的实验报告 [39],的速度进行比较，我们还针对 $N = 4$和 $m = N + 1$的情况进行了优化。该优化经过约45次迭代后收敛， $\Delta = 2^{-18}$。由于每次迭代需要$4\times(4+2) = 24$次评估，因此收敛所需的迭代总次数为1080次；与先前报告所需的（~ 25200 次评估）相比，速度提升了23倍。

使用相位不敏感距离和近似梯度时串扰的影响如图12所示。梯度近似使用$\Delta = 2^{-12}$计算。尽管串扰导致误差增大并增加了达到收敛所需的迭代次数，但通过增加几层额外冗余，可以减轻性能下降。这一结果表明，或许可以对器件进行端到端优化，

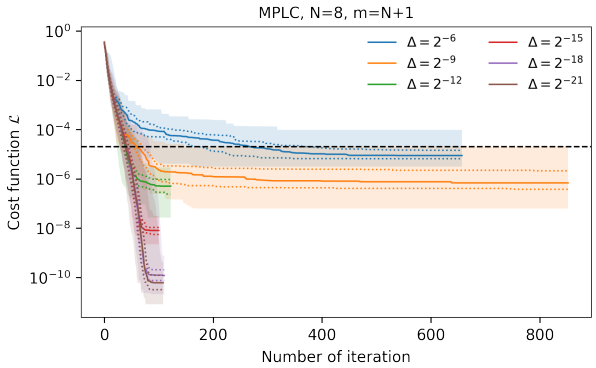


图11：使用相位不敏感距离的MPLC架构在改变梯度近似准确度时最终误差的比较。黑色虚线表示先前研究 [35], 中的优化结果，其中实现了 $\mathcal{L} = 2.0 \times 10^{-5}$。

包括矩阵优化和移相器校准。

V. 结论

我们提出了一种快速且可迭代配置的MPLC架构，用于实现精确且能够容忍制造误差的酉变换。数值结果表明，在MPLC架构中添加少量冗余层，可显著改善优化行为。我们还研究了串扰等伪影的影响，发现所提出的架构能够进行端到端优化。除提出新架构和优化方法外，我们还使用弗罗贝尼乌斯范数分析了酉矩阵之间的距离。我们引入了酉群上函数单峰性的概念，并证明使用弗罗贝尼乌斯范数计算的矩阵距离具有这种性

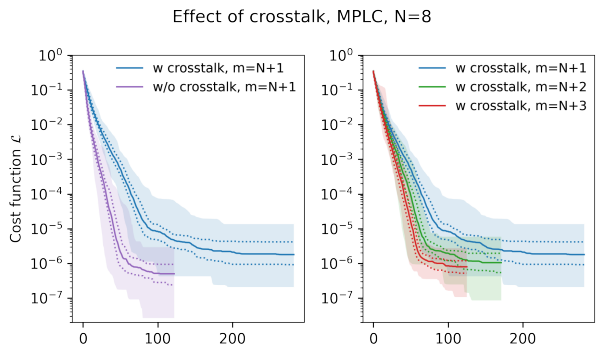


图12：比较使用近似梯度在存在串扰时的最终误差与使用 $\Delta = 2^{-12}$ 和相位无关距离时的最终误差，其中 $N = 8$。

质。我们还计算了矩阵距离的期望值和范围。此外，我们引入了与相位无关的范数，该范数适用于仅使用强度探测的应用。我们相信，这种方法将推动光学酉变换器的可扩展且稳健的实现，并拓展光子集成电路在各个领域中的应用。

致谢

我们谨向 Sho Yasui 致以谢意，感谢其富有成果的讨论。本研究获得日本 JST CREST项目编号JPMJCR1872 的支持。

附录A：弗罗贝尼乌斯范数的单峰性

这里，我们给出弗罗贝尼乌斯范数单峰性的代数证明。设 X，$U \in U(N)$。

定理 给定 U，若 $f_U(X) = \|X - U\|_F$ 在某个 X 处取得局部最小值，则该值为全局最小值。

证明。弗罗贝尼乌斯范数平方可表示为 $\|X-U\|_F^2 = 2N - 2\,\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]$；由于这是一个局部最小值，项 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]$ 是一个局部最大值。将 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]$ 视为流形 $U(N)$ 上的函数。我们可以通过考察该函数相对于 X 处切向量 X' 的方向导数来研究其临界点。切向量 X' 可表示为 $X' = ZX$，其中 Z 是反厄米矩阵，X 是酉群 $U(N)$ 上的任意矩阵。这是因为酉群 $U(N)$ 的李代数 $\mathfrak{u}(N)$ 由反厄米矩阵 [1] 构成。当 X 处于临界点时，对于任意 X'，均满足 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X'\right]\right] = 0$。代入 $X' = ZX$ 后，得到

$$\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X'\right]\right] = \mathrm{Re}\left[\mathrm{Tr}\left[ZXU^\dagger\right]\right] = 0. \tag{A1}$$

为进一步展开该方程，我们考虑两组特殊矩阵 Z^1 和 Z^2，其矩阵 $Z^1_{ij} \in Z^1$，$Z^2_{ij} \in Z^2$ 由 $1 \le i,\ j \le N$ 编索引，且满足 $i \neq j$。这些矩阵的第（k，l）个元素 $[\cdot]_{kl}$ 定义如下：

$$[Z^1_{ij}]_{kl} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} \tag{A2}$$

$$[Z^2_{ij}]_{kl} = i(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{A3}$$

例如，每个矩阵集合都包含以下矩阵：

$$Z^1_{12} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ -1 & 0 & & & \\ 0 & & & & \\ \vdots & & & & \vdots \\ 0 & & & \cdots & 0 \end{bmatrix}, Z^2_{23} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & i & 0 & & \\ 0 & i & 0 & & & \\ 0 & 0 & & & & \\ \vdots & & & & & \vdots \\ 0 & & & & \cdots & 0 \end{bmatrix}. \tag{A4}$$

将 Z^1_{ij}，Z^2_{ij}，$1 \le i,\ j \le N$ 代入式 A1 中的 Z 后，我们得到

$$\begin{cases} \mathrm{Re}\left[\left[XU^\dagger\right]_{ij}\right] - \mathrm{Re}\left[\left[XU^\dagger\right]_{ji}\right] = 0 \\ \mathrm{Im}\left[\left[XU^\dagger\right]_{ij}\right] + \mathrm{Im}\left[\left[XU^\dagger\right]_{ji}\right] = 0 \end{cases}, \tag{A5}$$

由此得到 $XU^\dagger = (XU^\dagger)^\dagger$。因此，

$$(XU^\dagger)^2 = I. \tag{A6}$$

酉矩阵 $XU^\dagger$ 可以使用可逆矩阵 V和对角矩阵 D 进行对角化，其中其对角元素 $d_i \in \mathbb{C}$ 满足 $|d_i| = 1$，因为 $XU^\dagger$ 是酉矩阵。我们可以将该对角化表示为$XU^\dagger = VDV^{-1}$。将 $XU^\dagger$ 代入式 A6，我们得到 $D = D^\dagger$。由于 $|d_i| = 1$，因此有 $d_i = \pm 1$。现在考虑原始的局部最大值项 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]$。我们可以将其重写为 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right] = \mathrm{Re}\left[\mathrm{Tr}\left[VDV^{-1}\right]\right] = \mathrm{Re}\left[\mathrm{Tr}\left[D\right]\right] = \mathrm{Re}\left[\sum_i d_i\right]$。当对所有 i均有 $d_i = +1$ 时，该值显然达到最大。如果对于某个 i有 $d_i = -1$，我们可以沿着复平面中的单位圆 $|c| = 1$ 将该值旋转至 $d_i = +1$，并仍然达到最大值。因此，如果 $\mathrm{Re}\left[\mathrm{Tr}\left[U^\dagger X\right]\right]$ 是局部最大值，那么它也是全局最大值。由此，我们现在得出结论：如果 $\|X - U\|_F \ge 0$ 是局部最小值，那么它也必定是全局最小值。证毕。

[1] $X' = ZX$的另一种证明。考虑恒等式 $XX^\dagger = I$。对等式两边求导，得到 $X'X^\dagger + X(X^\dagger)' = O$。然后可将其改写为 $X'X^\dagger = -X(X^\dagger)' = -(X'X^\dagger)^\dagger$。令 $Z = X'X^\dagger$。然后，我们有 $Z = -Z^\dagger$，这表明 Z 是一个斜厄米矩阵。由于 $Z = X'X^\dagger$，我们得出 $X' = ZX$。□

[1]J. Carolan、C. Harrold、C. Sparrow、E. Mart´ín- L´opez、N. J. Russell、J. W. Silverstone、P. J. Shadbolt、N. Matsuda、M. Oguma、M. Itoh、G. D. Marshall、M. G. Thompson、J. C. F. Matthews、T. Hashimoto、J. L. O'Brien 和 A. Laing，《通用线性光学》，《科学》**349**，711（2015年），https://www.science.org/doi/pdf/10.1126/science.aab364

[2] J. Wang、F. Sciarrino、A. Laing 和 M. G. Thompson，集成光子量子技术，《自然·光子学》**14**，273（2020）。[3] A. W. Elshaari、W. Pernice、K. Srinivasan、O. Benson 和 V. Zwiller，混合集成量子光子电路，《自然·光子学》**14**，285（2020）。[4] J. Carolan、M. Mohseni、J. P. Olson、M. Prabhu、C. Chen、D. Bunandar、M. Y. Niu、N. C. Harris、F. N. C. Wong、M. Hochberg、S. Lloyd 和 D. Englund，量子光子处理器上的变分量子反采样，《自然·物理学》**16**, 322 2020 1904 10463 （），[ ] arXiv：。quant-ph。[5] E. Pelucchi、G. Fagas、I. Aharonovich、D. Englund、E. Figueroa、Q. Gong、H. Hannes、J. Liu、C.-Y. Lu、N. Matsuda、J.-W. Pan、F. Schreck、F. Sciarrino、C. Silberhorn、J. Wang 和 K. D. J¨ons，集成光子学在量子技术中的潜力与全球展望，《自然综述：物理学》**4**，194（2022）。[6] Y. Chi、J. Huang、Z. Zhang、J. Mao、Z. Zhou、X. Chen、C. Zhai、J. Bao、T. Dai、H. Yuan、M. Zhang、D. Dai、B. Tang、Y. Yang、Z. Li、Y. Ding、L. K. Oxenløwe、M. G. Thompson、J. L. O'Brien、Y. Li、Q. Gong 和 J. Wang，可编程量子比特维数系统量子处理器，《自然通讯》**13**，1166（2022）。[7] L. S. Madsen、F. Laudenbach、M. F. Askarani、F. Rortais、T. Vincent、J. F. F. Bulmer、F. M. Miatto、L. Neuhaus、L. G. Helt、M. J. Collins、A. E. Lita、T. Gerrits、S. W. Nam、V. D. Vaidya、M. Menotti、I. Dhand、Z. Vernon、N. Quesada 和 J. Lavoie，利用可编程光子处理器实现量子计算优势，《自然》**606**，75（2022）。[8] Y. Shen、N. C. Harris、S. Skirlo、M. Prabhu、T. Baehr-Jones、M. Hochberg、X. Sun、S. Zhao、H. Larochelle、D. Englund 和 M. Solja˘ci´c，利用相干纳米光子电路进行深度学习，《自然·光子学》**11**，441（2017）。[9] M. Prabhu、C. Roques-Carmes、Y. Shen、N. Harris、L. Jing、J. Carolan、R. Hamerly、T. Baehr-Jones、M. Hochberg、V. ˇCeperi´c、J. D. Joannopoulos、D. R. Englund 和 M. Solja˘ci´c，在光子集成电路中加速循环伊辛机，《Optica》**7**，551（2020）。[10] H. Zhang、M. Gu、X. D. Jiang、J. Thompson、J. Cai、S. Paesani、R. Santagati、A. Laing、Y. Zhang、M. H. Yung、Y. Z. Shi、F. K. Muhammad、G. Q. Lo、X. S. Luo、B. Dong、D. L. Kwong、L. C. Kwek 和 A. Q. Liu，用于实现复值神经网络的光学神经芯片，《自然通讯》**12**，457（2021）。[11] S. Pai、Z. Sun、T. W. Hughes、T. Park、B. Bartlett、I. A. D. Williamson、M. Minkov、M. Milanizadeh、N. Abebe、F. Morichetti、A. Melloni、S. Fan、O. Solgaard 和 D. A. B. Miller，在纳米光子神经网络中通过实验实现用于深度学习的原位反向传播，arXiv 电子印本，arXiv：2205.08501（2022），arXiv：2205.08501 [cs.ET]。

2.

[12] F. Ashtiani、A. J. Geers 和 F. Aflatouni，面向图像分类的片上光子深度神经网络，Nature **606**，501（2022）。[13] S. Ohno、R. Tang、K. Toprasertpong、S. Takagi 和 M. Takenaka，用于光学神经网络片上推理与训练的硅微环谐振器交叉阵列，ACS Photonics **9**，2614（2022）。[14] S. Bandyopadhyay、A. Sludds、S. Krastanov、R. Hamerly、N. Harris、D. Bunandar、M. Streshinsky、M. Hochberg 和 D. Englund，采用加速训练的单芯片光子深度神经网络，arXiv 电子预印本，arXiv:2208.01623（2022），arXiv:2208.01623 [cs.ET]。[15] N. K. Fontaine、C. R. Doerr、M. A. Mestre、R. R. Ryf、P. J. Winzer、L. L. Buhl、Y. Sun、X. Jiang 和 R. Ringle，利用光子集成电路实现空分复用和全光MIMO解复用，载于 *OFC/NFOEC*（2012）第 1–3 页。[16] A. Annoni、E. Guglielmi、M. Carminati、G. Ferrari、M. Sampietro、D. A. Miller、A. Melloni 和 F. Morichetti，解扰光——自动消除模式间的强混合，Light: Science & Applications **6**，e17110（2017）。[17] D. Melati、A. Alippi、A. Annoni、N. Peserico 和 A. Melloni，用于模式与波分复用传输的集成式全光MIMO解复用器，Opt. Lett. **42**，342（2017）。[18] K. Choutagunta、I. Roberts、D. A. B. Miller 和 J. M. Kahn，调整马赫–曾德尔网格均衡器以用于直接检测模分复用链路，Journal of Lightwave Technology **38**，723（2020）。[19] R. Tanomura、R. Tang、G. Soma、S. Ishimura、T. Tanemura 和 Y. Nakano，基于集成光学酉变换器的全光双偏振MIMO处理器，载于 *2022 European Conference on Optical Communication (ECOC)* 2022 1 4 （）第 – 页。[20] W. R. Clements、P. C. Humphreys、B. J. Metcalf、W. S. Kolthammer 和 I. A. Walmsley，通用多端口干涉仪的优化设计，Optica **3**，1460（2016）。[21] R. Burgwal、W. R. Clements、D. H. Smith、J. C. Gates、W. S. Kolthammer、J. J. Renema 和 I. A. Walmsley，利用不完美光子网络实现随机酉变换，Opt. Express **25**，28236（2017）。[22] S. Bandyopadhyay、R. Hamerly 和 D. Englund，可编程光子学的硬件误差校正，Optica **8**，1247（2021）。[23] D. A. B. Miller，自配置通用线性光学组件 [特邀], Photon. Res. **1**，1（2013）。[24] D. A. B. Miller，设置干涉仪网格——反向局部光干涉法，Opt. Express **25**，29233 2017（）。[25] R. Hamerly、S. Bandyopadhyay 和 D. Englund，渐近容错的可编程光子学，Nature Communications**13**，6831（2022）。[26] R. Hamerly、S. Bandyopadhyay 和 D. Englund，矩形多端口干涉仪的精确自配置，Phys. Rev. Applied **18**，024019（2022）。[27] R. Hamerly、S. Bandyopadhyay 和 D. Englund，大型多端口干涉仪自配置的稳定性，Phys. Rev. Applied **18**，024018（2022）。[28] J.-F. Morizur、L. Nicholls、P. Jian、S. Armstrong、N. Treps、B. Hage、M. Hsu、W. Bowen、J. Janousek 和

H.-A. Bachor，可编程酉空间模式操控，J. Opt. Soc. Am. A **27**，2524（2010）。[29] G. Labroille、B. Denolle、P. Jian、P. Genevaux、N. Treps 和 J.-F. Morizur，基于多平面光转换的高效模式选择性空间模式复用器，Opt. Express **22**，15599（2014）。[30] R. Tang、T. Tanemura 和 Y. Nakano，使用MMI耦合器的集成式可重构酉光学模式转换器，《IEEE光子学技术快报》**29**，971（2017）。[31] R. Tang、T. Tanemura、S. Ghosh、K. Suzuki、K. Tanizawa、K. Ikeda、H. Kawashima 和 Y. Nakano，基于多平面光转换的可重构全光片上MIMO三模式解复用，Opt. Lett. **43**，1798（2018）。[32] R. Tang、T. Tanemura 和 Y. Nakano，使用多端口定向耦合器的稳健可重构光学模式复用/解复用器，载于 *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)* 2017 1 3 ( ) 第 – 页。[33] R. Tanomura、R. Tang、S. Ghosh、T. Tanemura 和 Y. Nakano，使用多端口定向耦合器的稳健集成光学酉变换器，Journal of Lightwave Technology **38**，60（2020）。[34] M. Y. Saygin、I. V. Kondratyev、I. V. Dyakonov、S. A. Mironov、S. S. Straupe 和 S. P. Kulik，用于可编程通用酉变换的稳健架构，Physical Review Letters **124**，010501（2020）。[35] R. Tanomura、R. Tang、T. Umezaki、G. Soma、T. Tanemura 和 Y. Nakano，基于多平面光转换的可扩展稳健光子集成酉变换器，Phys. Rev. Applied **17**，024071（2022）。[36] R. Tanomura、R. Tang、T. Tanemura 和 Y. Nakano，采用紧凑型半整数多模干涉仪的集成式INP光学酉变换器，Opt. Express**29**，43414 2021（）。[37] S. Kuzmin、I. Dyakonov 和 S. Kulik，面向可重构光学干涉仪编程的架构无关算法，Opt. Express **29**，38429（2021）。[38] S. Pai、B. Bartlett、O. Solgaard 和 D. A. B. Miller，通用酉光子器件上的矩阵优化，Phys. Rev. Applied **11**，064044（2019）。[39] R. Tanomura、R. Tang、T. Suganuma、K. Okawa、E. Kato、T. Tanemura 和 Y. Nakano，基于多平面光转换的单片式INP光学酉转换器，Opt. Express **28**，25392（2020）。[40] E. W. Weisstein，单峰，摘自 *Mathworld* – Wolfram 网络资源 https://mathworld.wolfram.com/ Unimodal.html。[41] T. W. Anderson，对称单峰函数在对称凸集上的积分及一些概率不等式，Proceedings of the American Mathematical Society **6**，170（1955）。[42] G. Anescu，单峰优化的启发式快速梯度下降法，Journal of Advances in Mathematics and Computer Science **26**，1（2018）。[43] R. Tang、R. Tanomura、T. Tanemura 和 Y. Nakano，硅光子芯片上的十端口幺正光学处理器，ACS Photonics **8**，2074（2021），https://doi.org/10.1021/acsphotonics.1c00419。[44] R. Tanomura、Y. Taguchi、R. Tang、T. Tanemura 和 Y. Nakano，《基于多平面光转换的光学幺正转换器的模式混合器熵》，载于*(CLEOPR) 2022 Conference on Lasers and Electro-Optics Pacific Rim* 2022 （Optica Publishing Group，）第 CWP13A–02 页。[45] M. Reck、A. Zeilinger、H. J. Bernstein 和 P. Bertani，《任意离散幺正算子的实验实现》，Physical Review Letters **73**，58（1994）。[46] P. Virtanen、R. Gommers、T. E. Oliphant、M. Haberland、T. Reddy、D. Cournapeau、E. Burovski、P. Peterson、W. Weckesser、J. Bright、S. J. van der Walt、M. Brett、J. Wilson、K. J. Millman、N. Mayorov、A. R. J. Nelson、E. Jones、R. Kern、E. Larson、C. J. Carey、 ̇I. Polat、Y. Feng、E. W. Moore、J. VanderPlas、D. Laxalde、J. Perktold、R. Cimrman、I. Henriksen、E. A. Quintero、C. R. Harris、A. M. Archibald、A. H. Ribeiro、F. Pedregosa、P. van Mulbregt 及 SciPy 1.0 Contributors，《SciPy 1.0：Python 科学计算的基础算法》，Nature Methods **17**，261（2020）。[47] R. Fletcher，《 *Practical Methods of Optimization*》，第 2 版（John Wiley & Sons，美国纽约州纽约市，1987 年）。[48] J. Bradbury、R. Frostig、P. Hawkins、M. J. Johnson、C. Leary、D. Maclaurin、G. Necula、A. Paszke、J. VanderPlas、S. Wanderman-Milne 和 Q. Zhang，《JAX：Python+NumPy 程序的可组合变换》（2018 年）。[49] M. Gallagher 和 T. Downs，《使用主成分分析可视化多层感知器网络中的学习过程》，IEEE系统、人和控制论汇刊，B 部分（控制论）**33**，28（2003）。[50] H. Li、Z. Xu、G. Taylor、C. Studer 和 T. Goldstein，《可视化神经网络的损失景观》，arXiv 电子印本，arXiv: 1712 09913 2017 1712 09913 。（），[ ] arXiv：. cs.LG。[51] H. Li、Z. Xu、G. Taylor、C. Studer 和 T. Goldstein，《可视化神经网络的损失景观》，载于 *Advances in Neural Information Processing Systems* 31， ，第 卷，由 S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi 和 R. Garnett 编（Curran Associates, Inc.，2018 年）。[52] T. W. Hughes、M. Minkov、Y. Shi 和 S. Fan，《通过原位反向传播和梯度测量训练光子神经网络》，Optica**5**，864（2018）。[53] M. Jacques、A. Samani、E. El-Fiky、D. Patel、Z. Xing 和 D. V. Plant，《热光移相器设计的优化及 SOI平台上热串扰的缓解》，《光学快报》**27**，10456（2019）。[54] B. V. Gurses、R. Fatemi、A. Khachaturian 和 A. Hajimiri，《硅光子学中的大规模串扰校正热光相移器阵列》，IEEE量子电子学精选主题期刊 **28**，6101009（2022 年）。