arXiv:2601.01877v1 [quant-ph] 5 Jan 2026

过参数化变分量子电路中的随机矩阵诱导的简单性偏差

Jun Qi[1*]、**Chao-Han Huck Yang**[2]、、**Pin-Yu Chen**[3]、、**Min-Hsiu Hsieh**[4*]

1. 佐治亚理工学院电气与计算机工程学院，美国佐治亚州亚特兰大，邮编 30332 2. 英伟达研究院，美国加利福尼亚州圣克拉拉，邮编 95051 3. IBM 托马斯·J·沃森研究中心，美国纽约州，邮编 10598 4. 鸿海（富士康）量子计算研究中心，中国台湾台北，邮编 114

电子邮件：jqi41@gatech.edu，
min-hsiu.hsieh@foxconn.com* 表示通讯作者

摘要。 过参数化常用于提升变分量子电路（VQC）的表达能力，但更深、参数更多的电路往往可训练性较差，泛化能力也有限。本文从函数类视角为这一现象提供了理论解释。我们证明，表达能力足够强且无结构的变分 ans¨atze 会进入类 Haar 普适类，在该类中，可观测量期望值和参数梯度都会随系统规模增大而呈指数级集中。因此，此类电路所诱导的假设类极有可能坍缩为一小类近常数函数，我们将这一现象称为简单性偏差；贫瘠高原则是其结果，而非根本原因。借助随机矩阵理论和测度集中工具，我们严格刻画了这一普适类，并证明其在有限数据集上会导致一致的假设类坍缩。我们还表明，这种坍缩并非不可避免：张量结构化变分量子电路，包括基于张量网络和张量超网络参数化的电路，均不属于类 Haar 普适类。通过利用有界张量秩或键维数限制可访问的酉系综，这些架构能够避免测度集中，保留局域可观测量的输出变异性，并且即使在过参数化区域中也能保持非退化梯度信号。总之，我们的研究将贫瘠高原、表达能力限制和泛化坍缩统一到一种根植于随机矩阵普适性的结构机制之下，凸显了架构归纳偏置在变分量子算法中的核心作用。

1. 引言

变分量子电路（VQCs）构成了多种量子算法的核心计算模型 [1, 2]，，包括变分量子本征求解器 [3,4]，、量子近似优化算法 [5–7]，以及量子机器学习模型 [8–12]。在这些场景中，增加电路深度和参数数量通常被视为提升表达能力与学习性能的自然途径 [13–16]。然而，

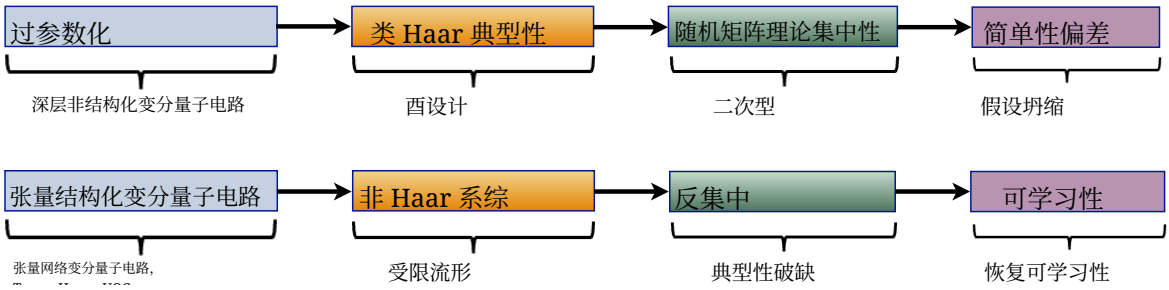


图 1. 变分量子电路中简单性偏差及其缓解方法的示意图。上图：在表达能力足够强且无结构的变分量子电路中，所诱导的酉系综会趋近于某一随机矩阵普适类，进而导致可观测量期望值和梯度发生集中。因此，假设类会坍缩为近常数函数，从而产生简单性偏差和贫瘠高原。下图：张量结构化变分量子电路通过限制张量秩或键维数，约束可访问的幺正流形，从而避免测度集中。这种结构约束保留了输出变异性和信息性梯度，因此即使在过参数化的情况下，也能缓解简单性偏差。

大量实证证据表明，更深且参数更多的变分量子电路往往会面临严重的训练问题，包括梯度消失、平坦损失景观和泛化能力较差 [17–19], ，即使不存在优化噪声和硬件缺陷 [3, 17, 20, 21]。

一项重要的研究脉络将这些失效归因于贫瘠高原 [22–25], ：在这种情况下，典型代价函数的梯度会随系统规模增大而呈指数级消失。尽管这一以梯度为中心的观点影响深远，却仍留下一个更根本的问题：过参数化变分电路实际上表示的是哪一类函数？尤其是，目前尚不清楚观察到的优化困难究竟仅仅源于不利的损失景观，还是反映了电路系综本身固有的、更深层的表征能力限制。

在本研究中，我们通过采用关于过参数化变分量子电路的函数类视角来探讨这一问题。我们不再聚焦于特定的代价函数 [25] 或训练过程 [26], ，而是研究随机参数化下变分电路架构所诱导的假设类。我们证明，在一个广泛且具有实际意义的情形中，通过无结构的过参数化提升表达能力，可能导致函数多样性崩塌。以很高的概率，在给定参数下，电路实现的是一个近似常数函数，且基本不依赖于输入。我们将这一现象称为简单性偏差，以强调它产生于表征层面，而非优化层面。

我们的分析基于这一观察：具有充分表达能力且无结构的变分 ans¨atz 表现出类 Haar 典型性 [27, 28]：其诱导的酉系综能够复现酉群上 Haar 测度的低阶矩。在这一机制下，随机矩阵理论 [29, 30] 和测度集中 [31, 32] 的工具便可以得到应用。我们证明，类 Haar 典型性会同时导致可观测量期望值和参数梯度发生集中，且二者都会随量子比特数呈指数衰减。因此，即使尚未考虑任何学习动力学，此类电路所诱导的假设类也会坍缩为一组狭窄的近常数函数。

更多 i 重要的是，这种表征坍缩并非普遍现象。我们证明 e

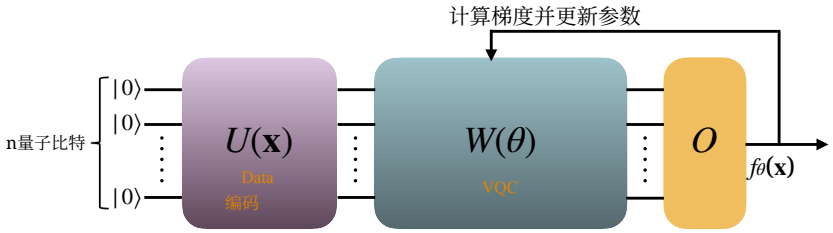


图 2。本文所研究的变分量子电路示意图。变分量子电路由初始化为 $|0\rangle^{\otimes n}$ 的输入量子态开始，接着是数据编码幺正算子 $U(\mathbf{x})$、参数化变分 ans¨atzes $W(\boldsymbol{\theta})$，以及对厄米可观测量O的测量，得到标量输出 $f_{\boldsymbol{\theta}}(\mathbf{x})$。由此产生的函数族 $\{f_{\boldsymbol{\theta}}\}$ 的表达能力和统计性质取决于变分 ans¨atzes $W(\boldsymbol{\theta})$ 的结构。

张量结构变分电路，包括基于张量网络的架构 [26]和张量超网络参数化 [9], ，均不属于类 Haar 普适类。通过施加张量秩或键维数有界等结构约束，这些架构限制了可访问的幺正流形，阻止其收敛至近似酉设计。因此，它们的可观测量期望值呈现反集中现象，并保留非退化梯度信号，从而即使在高度参数化的情形下也能维持函数多样性。

如图 3 所示，本工作的贡献有三点：

(i) 我们对支配过参数化无结构变分量子电路的类 Haar 普适类，给出了严格的随机矩阵理论刻画。(ii) 我们表明，在这一机制下，过参数化会导致假设类在表征层面发生坍缩，这种现象称为简单性偏差；贫瘠高原是其结果，而非根本原因。(iii) 我们证明，张量结构化变分量子电路能够逃离这一普适类，从而提供一种有理论依据的机制，以恢复非平凡假设类和具有信息量的学习信号。

综合来看，这些结果为变分量子算法中若干众所周知的失效模式提供了统一的结构性解释，并表明可学习性不仅由表达能力决定，还取决于所诱导酉系综的几何特性。这一视角凸显了架构归纳偏置在设计可扩展且可训练的变分量子算法中的核心作用。

2. 结果

2.1. 问题设置与假设

图 2 展示了本文分析的通用变分量子电路架构。该电路通过固定酉算符 $U(\mathbf{x})$ 对数据进行编码，应用参数化酉算符 ans¨atze $W(\boldsymbol{\theta})$，并测量所得量子态上的有界可观测量 O ，从经典输入 $\mathbf{x}$ 映射到标量输出 $f_{\boldsymbol{\theta}}(\mathbf{x})$。具体而言，我们考虑如下标准形式的 n 量子比特变分量子电路：

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \left\langle 0^{\otimes n} \middle| U^{\dagger}(\mathbf{x}) W^{\dagger}(\boldsymbol{\theta}) O W(\boldsymbol{\theta}) U(\mathbf{x}) \middle| 0^{\otimes n} \right\rangle, \tag{1}$$

其中，$\boldsymbol{\theta} \in \mathbb{R}^p$ 表示电路参数，O 是满足 $\|O\|_2 \leq 1$ 的厄米可观测量。我们的目标是刻画由 $f_{\boldsymbol{\theta}}$ 诱导的假设类的表达能力和可学习性，尤其是在电路深度和参数化随系统规模扩展的情况下。

定义 1（哈尔随机）。我们首先回顾哈尔随机的标准概念。酉群 $U(2^n)$ 存在唯一的概率测度，该测度在任意酉算子的左乘和右乘下保持不变，称为 *Haar* 测度。从该测度中抽取的酉算子称为哈尔随机，代表希尔伯特空间中的最大统计对称性。

在哈尔随机量子态上计算的有界可观测量的期望值表现出强烈的测度集中：随着希尔伯特空间维数增大，这些量会紧密地集中在其均值附近。在本文中，哈尔随机严格指从 $U(2^n)$ 上的精确 Haar 测度中采样。

相比之下，许多具有物理动机的电路系综并非精确的哈尔随机系综，但能够复现有限阶的 Haar 矩。这些系综通常称为近似酉设计。我们通过假设 1 将过参数化形式化，并用类 Haar 典型性描述这种近似行为。

假设 1（类 Haar 典型性）。变分电路 *ans¨atze* $W(\boldsymbol{\theta})$ 具有足够的深度和表达能力，因此，对于几乎所有参数取值 $\boldsymbol{\theta}$，由此诱导的酉算子分布 $W(\boldsymbol{\theta})$ 在 $U(2^n)$上构成近似酉 2-设计。

在假设 1 中，“几乎所有参数”指从平滑分布（例如均匀分布或高斯分布 [33]）中典型地抽取参数；而过参数化则表示具有足够的表达能力来近似 Haar 低阶矩，而非仅仅意味着参数数量庞大。

这一假设描述了一种已知会出现在多种非结构化电路族中的情形，其中包括具有全连接或最近邻连接的硬件高效 ans¨atze；当电路深度至少随量子比特数线性增长，且参数从平滑分布中独立初始化时，就可能出现这种情形。

假设的适用范围与解释。假设 1 并不要求精确的 Haar 随机性，也不声称所有电路架构都能在有限深度下达到这一情形。相反，它刻画了一个在经验和理论上都具有重要意义的普适类：在该类中，具有足够表达能力的非结构化变分电路在低阶矩上表现出类 Haar 的典型行为。此外，我们的结果仅依赖于足以使有界可观测量呈现测度集中的近似设计行为。因此，只要电路系综在二阶上重现 Haar 统计，无论其具体微观架构如何，这些结论均适用。

此外，尽管类 Haar 典型性是许多已知贫瘠高原结果的基础，但我们关注的是表征，而非算法。假设 1 形式化了由 $f_{\boldsymbol{\theta}}$ 诱导的假设类能否表示具有实际差异的函数，以及无穷小参数扰动是否会引起非消失的函数响应。接下来的理论结果刻画了这种典型性在表征方面的影响，且独立于任何特定的优化过程或噪声模型。

2.2. 理论结果

现在，我们给出本文的主要理论结果。定理1和定理2刻画了表达能力足够强的非结构化变分量子电路所对应的类 Haar 普适类。命题1和命题2证明，张量结构变分量子电路不属于该普适类；定理 3 则形式化地说明了这种结构如何缓解由此产生的塌缩。

随机矩阵诱导的输出集中。我们首先刻画变分拟设变得足够有表达能力、能够近似二阶酉设计的条件；深度随机电路和硬件高效变分拟设中都观察到了这一现象。在假设 1 下，对于固定输入 $\mathbf{x}$和典型参数值 $\boldsymbol{\theta}$，量子态

$$|\psi_{\boldsymbol{\theta},\mathbf{x}}\rangle := W(\boldsymbol{\theta})U(\mathbf{x})|0^{\otimes n}\rangle \tag{2}$$

在 $\mathbb{C}^{2^n}$ 中表现为 Haar 随机态。变分量子电路的输出 $f_{\boldsymbol{\theta}}(\mathbf{x})$ 是一个随机向量的二次型。

定理 1（输出集中）。在假设 *1* 下，对于任意固定输入 $\boldsymbol{x}$,

$$\mathbb{E}\left[f_{\boldsymbol{\theta}}(\boldsymbol{x})\right] = \frac{1}{2^n}\mathrm{Tr}(O), \tag{3}$$

and

$$\mathrm{Var}(f_{\boldsymbol{\theta}}(\boldsymbol{x})) = \mathcal{O}(2^{-n}). \tag{4}$$

此外，对于任意 $\epsilon > 0$,

$$\Pr\left(\left|f_{\boldsymbol{\theta}}(\boldsymbol{x}) - \frac{1}{2^n}\mathrm{Tr(O)}\right| > \epsilon\right) \leq 2\exp\left(-c\epsilon^2 2^n\right), \tag{5}$$

对于某个普适常数 $c > 0$。

定理 1 表明，在类似 Haar 的典型状态下，电路输出偏离其均值的程度会随希尔伯特空间维数呈指数级抑制。因此，对于几乎所有参数设置，过参数化的变分量子电路（VQC）所实现的函数对于任意固定输入都近似为常数。

梯度集中与典型平坦性。接下来我们将展示，同样的随机矩阵机制支配着参数梯度的行为。对于典型的参数化，VQC 输出关于参数 θ_k 的导数可以写为:

$$\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} = \langle \psi_{\boldsymbol{\theta},\mathbf{x}}|G_k|\psi_{\boldsymbol{\theta},\mathbf{x}}\rangle, \tag{6}$$

其中 G_k 是由电路结构决定的有界厄米算子。

定理 2（梯度集中）。在假设 *1* 下，对于任意固定输入 $\boldsymbol{x}$ 和任意电路参数 θ_k,

$$\mathbb{E}\left[\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_k}\right] = 0, \quad \mathrm{Var}\left(\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_k}\right) = \mathcal{O}(2^{-n}). \tag{7}$$

其中，期望和方差是对 $\boldsymbol{\theta}$ 的随机初始化取平均得到的 .

定理 2 表明，在类 Haar 典型性区域，梯度会急剧集中于零附近，其方差随量子比特数增加而呈指数衰减。因此，在初始化时，几乎所有参数方向在局部都不含有效信息，从而形成实际上平坦的损失景观。重要的是，这一现象源于电路系综的表示典型性，而非任何特定的优化策略、代价函数或噪声机制。

简单性偏差与假设类坍缩。定理 1 和定理 2 共同表明，过参数化的非结构化变分量子电路所诱导的有效假设类会发生坍缩。

定义 2（简单性偏差）。如果一个假设类 $\mathcal{F} = \{f_{\boldsymbol{\theta}}\}$ 表现出简单性偏差，则表示无论参数数量如何，随着参数选择，其函数以高概率集中于近常数函数的低复杂度子集 $\mathcal{F}$。

推论 1（有限数据集上的一致假设类坍缩）。设 $\mathcal{D}$ ={$\boldsymbol{x}_1$、.....、$\boldsymbol{x}_m$} 是一个大小为 $m = \mathrm{poly}(n)$ 的有限数据集。在假设 *1* 下，对于电路参数，至少以 $1-\delta$ 的概率，

$$\max_{\boldsymbol{x}_i,\boldsymbol{x}_j\in\mathcal{D}} |f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - f_{\boldsymbol{\theta}}(\boldsymbol{x}_j)| \le \epsilon, \tag{8}$$

其中 $\epsilon = \mathcal{O}(2^{-n/2})$ 以及 $\delta \le m^2 \exp(-c2^n)$。

该推论通过将定理 thm:thm1 应用于每个输入，并对所有输入对取并集界得到。它表明，仅有过参数化并不能保证变分量子电路具有表达能力强的假设类：在类 Haar 状态下，假设类一致地坍缩为近常数函数。

张量结构化变分量子电路与类 *Haar* 典型性的失效。上述确立的简单性偏差，关键依赖于类 Haar 典型性。现在我们证明，张量结构化变分量子电路不属于这一随机矩阵普适类。

张量结构化变分量子电路通过限制张量秩或键维数来施加明确的结构约束，从而将可实现的酉算子限制在$U(2^n)$的一个低维子集中。我们考察两种具有代表性的架构：张量网络变分量子电路和 TensorHyper-VQC（图 3）；二者都限制了纠缠，并采用多项式参数化。

命题 1（近似酉设计的失效）。设 $\mathcal{U}_{\mathrm{ts}} \in U(2^n)$ 表示由固定秩（或键维数）且与 n无关的张量结构化变分量子电路所生成的酉算子系综。那么，对于任意固定的 $t \ge 2$，当 $n \to \infty$时，$\mathcal{U}_{\mathrm{ts}}$ 都不构成近似酉 t-设计。

张量网络结构所隐含的有界算子施密特秩阻碍其收敛到 Haar 矩，而收敛到 Haar 矩需要接近最大纠缠以及在各二分划分上的各向同性。

有界纠缠带来的反集中。类 Haar 典型性的这种失效可以通过纠缠和方差界明确体现。

引理 1 考虑一个张量结构化变分量子电路 $W(\boldsymbol{\theta})$，其张量秩为 $r = \mathcal{O}(1)$。对于任意局域或少体可观测量 $\mathcal{O}$，$\mathcal{O}$ 的支撑集上的约化密度矩阵至多取决于$\mathcal{O}(r^2)$ 个有效自由度。因此，在随机参数初始化下，$\langle O\rangle$的方差有一个下界，该下界为与量子比特总数 n 无关的常数。

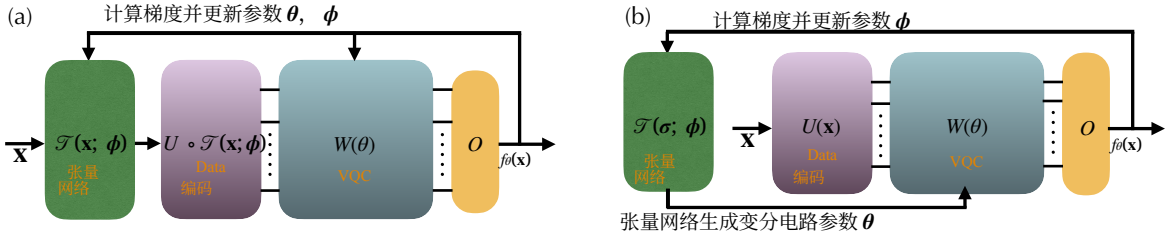


图 3. 张量结构化变分量子电路架构。（a）TN-VQC：张量网络模块将经典输入 $\mathbf{x}$ 转换为低维特征 $\mathcal{T}(\mathbf{x};\phi)$，再通过编码幺正算子$\mathcal{T}(\mathbf{x};\ \phi)$将其进一步转换为量子态 $U\circ\mathcal{T}(\mathbf{x};\phi)$。随后，编码后的量子态由 VQC $W(\theta)$处理，并通过可观测量 O 进行测量，以生成输出 $f_\theta(\mathbf{x})$。编码参数ϕ和电路参数 θ 均通过基于梯度的优化进行更新。（b）TensorHyper-VQC：张量网络充当超网络，通过使用高斯随机向量 σ，直接通过 $\mathcal{T}(\sigma;\phi)$生成变分电路参数 θ。数据编码保持固定，而张量网络在 $W(\theta)$ 的参数之间诱导结构化相关性。在这两种架构中，张量网络结构都会约束可访问的酉系综，打破类 Haar 典型性，并缓解随机矩阵诱导的简单性偏差。

命题 2（反集中）。设 O 为有界的局域或少体可观测量。对于张量秩固定为 r 的张量结构化变分量子电路，存在一个常数 $c(r,O)\ >\ 0$，它与量子比特数 n无关，使得

$$\mathrm{Var}\left(\langle\psi_{\boldsymbol{\theta},\boldsymbol{x}}|O|\psi_{\boldsymbol{\theta},\boldsymbol{x}}\rangle\right)\geq c(r,O). \tag{9}$$

与类 Haar 变分量子电路中方差随 $\mathcal{O}(2^{-n})$ 指数衰减不同，张量结构变分量子电路的方差不会消失。可观测量期望值仍对输入和参数保持敏感，从而避免了表征坍缩。

非平凡假设类与可学习性代理指标。现在我们将反集中的影响形式化。我们在定义 3 中定义非平凡假设类，并在定理 3 和推论 2 中展示张量网络变分量子电路的解决方案。

定义 3（非平凡假设类）。设 $\mathbb{F}=\{f_{\boldsymbol{\theta}}:\ \mathcal{X}\to\mathbb{R}\}$ 为由变分量子电路诱导的假设类。若存在输入$\boldsymbol{x}$、$\boldsymbol{x}'\in\mathcal{X}$ 以及常数 $\triangle$、$\eta>0$，且它们与 n无关，使得 $\mathbb{F}$ 非平凡。

$$\Pr\left(|f_{\boldsymbol{\theta}}(\boldsymbol{x})-f_{\boldsymbol{\theta}}(\boldsymbol{x}')|\geq\triangle\right)\geq\eta. \tag{10}$$

若存在参数索引 k、一个输入$\boldsymbol{x}$以及与 n无关的 $\gamma>0$，使得该类具有非退化的学习信号，

$$\mathrm{Var}\left(\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial\theta_k}\right)\geq\gamma. \tag{11}$$

定理 3（张量结构化变分量子电路缓解简单性偏差）。设 $W(\boldsymbol{\theta})$ 是一个张量秩有界且不随量子比特数 n变化的张量结构化变分量子电路族。假设O 是一个有界局域可观测量，且数据编码映射将输入映射为在 O支撑集上彼此不同的约化态。那么：

(i)（输出反集中）存在一个与 n无关的常数 $c_0 > 0$，使得

$$\mathrm{Var}(f_{\boldsymbol{\theta}}(\boldsymbol{x})) \geq c_0 \tag{12}$$

对于某个$\boldsymbol{x} \in \mathcal{X}$。

(ii)（非退化梯度信号）存在一个参数索引 k 和 $c_1 > 0$，其独立于 n，使得

$$\mathrm{Var}\left(\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_k}\right) \geq c_1. \tag{13}$$

定理 3 表明，张量结构化变分量子电路从根本上改变了过参数化电路的典型行为。通过打破类 Haar 典型性，张量网络结构避免了输出和梯度集中，从而排除了表征坍缩机制。

我们进一步强调，定理 3 的非平凡内容不仅在于张量结构电路无法近似 Haar 随机性，还在于这种结构限制可证明地确保了物理上相关的局域可观测量具有持续的输出变异性和非退化梯度信号，从而即使在过参数化区域也能避免假设类坍缩。

推论 2（在过参数化设置中恢复可学习性）。在定理 *3* 的假设下，张量结构化变分量子电路具有非平凡假设类和非退化学习信号，并且该信号在 $n \to \infty$ 时仍然持续存在。

注 1（关于可学习性）。在本文中，恢复的可学习性指的是避免因类 *Haar* 典型性而导致表征坍缩和梯度坍缩。非零的输出方差和梯度方差确保了初始化时的函数多样性和有信息量的下降方向，但仅凭这些并不能保证训练成功或实现泛化。

统一解释。上述结果可以得到统一解释：贫瘠高原、表达能力坍缩和泛化失败，是同一随机矩阵普适类的不同表现形式，而该普适类支配着过参数化、非结构化的变分量子电路。张量结构化架构通过限制可访问的酉系综，摆脱了这一普适类，从而避免测度集中，并恢复函数和梯度的变异性。

2.3. 数值模拟

为补充理论分析，我们给出一个简化的数值模拟，以展示研究结果所预测的有限尺寸行为。本实验并非旨在进行经验基准测试，而是探究方差坍缩及其张量结构化缓解机制在有限系统规模和有限数据库规模下如何呈现。

我们考虑三类变分量子电路：朴素无结构变分量子电路、张量网络变分量子电路和 TensorHyper-VQC，均以 $n = 12$ 个量子比特和深度 $L = 6$进行演示。对于每种模型，我们都评估其电路输出在大小为m的数据集上的经验方差 $\mathrm{Var}_{\mathbf{x}\sim\mathcal{D}_m}[f_{\boldsymbol{\theta}}(\mathbf{x})]$，其中 $\mathcal{D}_m$ 表示从固定输入分布中抽取的有限样本。数据集大小m从 2^3 变化到 2^8, ，结果取多个随机初始化的平均值。

为将表征效应与任务特定结构分离，我们按如下方式构造合成输入数据集。每个输入样本$\mathbf{x} \in \mathbb{R}^p$ 均独立地从固定的

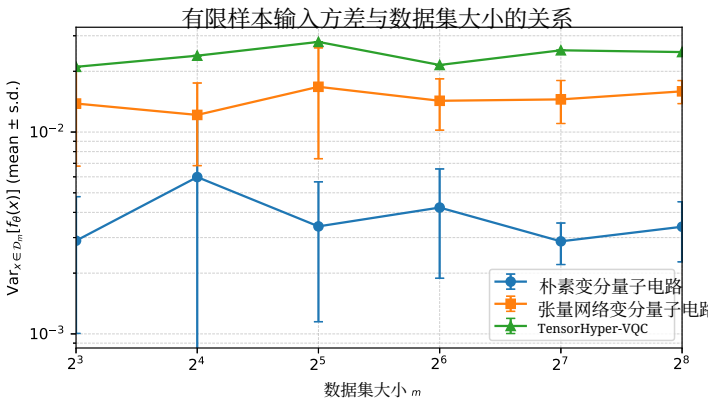


图4. 有限样本输入方差与典型性的失效。在规模为 $m = 2^3, \ldots, 2^8$ 的数据集上，对朴素变分量子电路、张量网络变分量子电路和 TensorHyper-VQC 的模型输出方差 $\mathrm{Var}_{\mathbf{x}\sim\mathcal{D}_m}[f_{\boldsymbol{\theta}}(\mathbf{x})]$ 进行评估。数值取随机初始化的平均值，误差条表示一个标准差。与无结构变分量子电路的高度集中行为不同，张量结构架构在不同数据集规模下表现出稳定且非零的方差。

各向同性分布（标准正态分布）中抽取，并归一化为单位范数。由于目标仅是评估不同输入所引发的电路输出变异性，因此不使用标签。数据集大小从 2^3 到 2^8, 不等，结果取多个随机电路初始化的平均值。误差条表示 10 个随机种子下的一个标准差。

与定理1和定理3一致，如图4所示，朴素变分量子电路的输出值呈现出显著的集中趋势：经验方差保持较小，且对数据集大小的增加基本不敏感，这反映了类 Haar 典型性的出现。相比之下，两种张量结构架构在所有数据集规模下都保持着显著较大且稳定的输出方差。这一行为表明类 Haar 典型性失效，非平凡的函数变异性得以持续存在，正如命题 2和定理 3所预测的那样。

值得注意的是，无结构电路与张量结构电路之间的定性差异在不同数据集规模下始终稳定，这表明观察到的效应并非有限样本波动造成的假象。相反，它为支撑我们理论框架的随机矩阵普适性原理提供了一个有限尺寸示例。

3. 讨论

本研究为变分量子算法中一种持续存在的经验现象提供了理论解释：通过无结构的过度参数化提高电路表达能力，并不一定能改善学习性能，反而可能导致函数多样性坍缩。通过将表达能力足够强的硬件高效变分量子电路建模为类 Haar 普适类中的电路，我们表明，由于测度集中效应，输出和参数梯度都会显著集中。因此，几乎所有参数设置都会产生近常数函数，从而形成强烈的简单性偏差。重要的是，这种坍缩并非仅仅是优化病态性，而是电路系综所诱导的假设类的一种结构性属性。

尽管如此我们的分析是渐近的，既有研究中的大量数值证据表明 s

随着系统规模增大，集中现象可能会迅速出现。由于本文旨在刻画普适类，而非研究有限尺寸性能，因此我们有意聚焦于结构机制，而不是定量阈值。因此，针对有限尺寸交叉效应的系统数值研究留待未来开展。

简单性偏差作为一种表征现象。我们分析的一个关键概念性结论，是区分优化难度与假设类坍缩。现有关于贫瘠高原的大量文献都强调特定代价函数下的梯度消失，突出在损失景观中进行导航的困难。相比之下，我们的结果表明，即使不考虑优化因素，处于类 Haar 典型性区域中的过参数化变分量子电路也具有一种趋向平凡函数的内在归纳偏置。从这一角度看，平坦损失景观和梯度消失，是由高维酉群上的测度集中所支配的更深层表征退化的表现。

这一观点阐明了过参数化在量子模型中的作用。经典深度学习中的过参数化通常通过隐式正则化和特征学习提升性能，而量子电路中的非结构化过参数化则可能产生不利影响，使模型进入一个函数多样性丧失的普适类。因此，应将假设 1 理解为一种描述具有充分表征能力的非结构化电路之极限行为的普适性假设，而不是关于所有架构都能实现有限深度收敛的陈述。

将我们的结果与现有的贫瘠高原理论进行比较至关重要。关于贫瘠高原的既有研究从根本上以梯度为中心：它分析特定代价函数的梯度消失现象，通常基于全局测量或特定优化目标的假设，并从优化难度的角度来界定可训练性。相比之下，我们的分析以函数类为中心。我们证明，在类 Haar 典型性下，由过参数化的无结构变分量子电路诱导的整个假设类，以高概率坍缩为一族范围狭窄的近常数函数，这一结果与所选代价函数或优化动力学无关。

在这种情况下，梯度消失并非可训练性失效的根本原因，而是测度集中所导致的更深层表征退化的结果。据我们所知，现有的贫瘠高原研究既未刻画这种假设类或函数空间坍缩，也未形式化说明仅凭过参数化如何能在学习开始之前就破坏函数多样性。因此，我们的工作通过识别一种不同的表征层面失效模式，并提出一项避免该问题的结构性判据——打破类 Haar 典型性，从而补充并拓展了贫瘠高原理论。

通过张量结构打破类 *Haar* 典型性。我们的第二项主要贡献是指出，张量网络结构是一种有理论依据的机制，可用于摆脱这一普适类。张量结构化变分量子电路，例如张量网络变分量子电路和 TensorHyper-VQC，通过限制张量秩或键维数，明确约束可访问的幺正流形。我们证明，这些约束会阻止电路收敛到近似酉设计，限制纠缠增长，并使可观测量的期望值呈现反集中现象。

从物理学角度看，张量结构电路生成的量子态在希尔伯特空间中并不典型：即使在参数高度密集的情形下，相对于 Haar 随机态，它们仍只占据一个测度为零的子集。从学习理论视角看，这种非典型性有益：它保留了非平凡的函数变异性，并确保在初始化时存在信息性梯度。由此产生的典型性失效，为

结构化 ans¨atze 在实践中为何往往优于更深的非结构化电路，提供了严谨的解释。

变分量子算法的设计原则。分析表明，变分量子算法的一般设计原则是：可学习性不仅受表达能力支配，还取决于所诱导酉系综的几何性质。过于强调表达能力的电路在趋近 Haar 随机行为时，会遭遇表征坍缩。相比之下，经过审慎结构设计的电路能够在表达能力与归纳偏置之间保持有利的平衡。

尽管我们的分析聚焦于基于张量网络的构造，但其核心观点不依赖于具体架构 [34, 35]。任何能够限制电路进入酉设计普适性区域的机制，例如保持局域性的布局、对称性约束，或源自低纠缠物理模型的 ans¨atz，都可以减轻简单性偏差并提升可训练性。

局限性与范围。我们的结果基于假设 1，即足够深且非结构化的变分量子电路会进入类 Haar 典型性区域，其特征是近似酉设计。我们强调，这是一项普适性假设，而非关于特定硬件架构下有限深度收敛的论断。确定特定电路族进入该区域的精确深度阈值仍是一个活跃的研究领域，并且取决于门集、连通性和噪声等因素。

第二个局限性在于，我们的保证具有渐近性质。定理1–3是在 $n \to \infty$ 的极限下表述的，而近期量子设备运行于有限的系统规模下。尽管如此，测度集中界所固有的指数级缩放表明，一旦电路具备足够的表达能力，本文所识别的定性现象（例如输出坍缩、梯度集中和简单性偏差）可能会在相对适中的量子比特数下显现。

最后，我们的分析聚焦于具有有界局部可观测量或少体可观测量的基于期望值的模型，这类模型是变分量子算法和量子机器学习中的标准模型。虽然可以考虑全局可观测量，但它们通常表现出更强的测度集中效应，因此不太可能缓解本文所识别的简单性偏差。将该框架扩展到更一般的测量方案、自适应可观测量或非线性后处理，仍然是一项开放性挑战。

展望。除了解释已有的经验观察外，我们的框架还为未来研究开辟了多个方向。一个自然的延伸方向是，随着张量秩增加，量化典型性程度如何在结构化状态与类 Haar 状态之间进行插值，从而可能在可学习性方面产生类似相变的行为。另一个方向是将噪声和误差缓解纳入随机矩阵分析，以阐明硬件噪声如何与简单性偏差相互作用。

更广泛地说，本文所确立的随机矩阵普适性、纠缠结构与学习动力学之间的联系表明，量子机器学习中的归纳偏置，从根本上说是一个关于系综几何的问题。理解并利用这种几何结构，或许是设计可扩展、可训练且超越含噪中等规模量子时代的变分量子算法的关键。

4. 方法

4.1. 过参数化变分量子电路

当可训练参数的数量随系统规模 m 至少线性增长（通常超线性增长）时，变分量子电路即为过参数化。本文所说的“过参数化”，始终指电路表达能力足以近似酉设计的情形，而非仅指参数数量。在这种情况下，已知缺乏结构的硬件高效型或随机电路 ans¨atze 会表现出类似哈尔测度的行为，从而使输出和梯度都出现测度集中现象。

在实践中，过参数化对应于这样一类电路族：其参数数量和深度增长得足够快，从而能够逼近酉 2-设计，即使可观测量仍具有局域性。更重要的是，此处的过参数化指酉系综的表达能力，而非仅指经典概念中的宽度或深度。我们的理论分析聚焦于渐近区域，其中 $n \to \infty$，同时可观测量局域性保持不变。

4.2. 过参数化变分量子电路的随机矩阵理论视角

随机矩阵理论为刻画大型高维量子系统的典型行为提供了严谨的数学框架。在变分量子电路的背景下，当诱导的酉系综趋近于哈尔随机时，随机矩阵理论用于模拟电路输出和梯度的统计性质。

我们采用随机矩阵理论（RMT）的观点，对过参数化、非结构化变分量子电路出现测度集中现象的情形进行形式化描述。具体而言，当变分 Ansatz 具有足够的表达能力时，例如参数数量众多的深度硬件高效电路或随机电路，典型参数选择所生成的酉算子 $W(\boldsymbol{\theta})$ 的行为近似于从 $U(2^n)$ 上的 Haar 测度中抽取的随机酉算子，或者等价地，近似于低阶矩意义下的近似酉 t 设计。

4.3. 张量结构化变分量子电路

为了超越非结构化的过参数化变分量子电路，我们引入张量结构化变分量子电路，其中电路参数之间的相关性通过低秩张量网络表示加以约束。这些结构明确限制了可访问的酉系综，并阻止其收敛至类 Haar 典型性。具体而言，我们考虑两种代表性架构：TN-VQC [26] 和 TensorHyper-VQC [9]。

在 TN-VQC 架构中，经典张量网络（TN）用于生成依赖于数据的编码特征，而变分电路参数则保持为全局且与输入无关。具体而言，TN 定义了一个映射

$$\mathcal{T}(\mathbf{x};\boldsymbol{\phi}) : \mathcal{X} \to \mathbb{R}^n, \tag{14}$$

其中，$\boldsymbol{\phi}$ 表示张量网络参数，张量网络秩受一个与量子比特数 n无关的常数所限制。随后，输出 $\mathcal{T}(\mathbf{x};\boldsymbol{\phi})$ 用于参数化数据编码幺正算子$U(\mathcal{T}(\mathbf{x};\ \boldsymbol{\phi}))$。由此得到的电路形式为

$$|\psi_{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{x}}\rangle = W(\boldsymbol{\theta})U(\mathcal{T}(\mathbf{x};\boldsymbol{\phi}))|0^{\otimes n}\rangle. \tag{15}$$

在这种设定下，张量网络会在数据编码中引入结构化相关性，且变分参数 $\boldsymbol{\theta}$ 由不同输入共享。可访问态族受到有界张量列车秩的约束，从而限制了任意二分划分下的纠缠增长。

另一方面，TensorHyper-VQC 将上述思想推广为使用张量网络作为超网络，直接根据输入生成变分电路参数。更具体地说，张量网络实现了一个映射

$$\mathcal{T}(\boldsymbol{\sigma};\boldsymbol{\phi}) : \mathcal{X} \to \mathbb{R}^p, \tag{16}$$

其中， $\boldsymbol{\sigma}$ 表示张量网络的高斯随机向量， p 是 W（$\boldsymbol{\theta}$）中的变分参数总数。电路随后定义为

$$|\boldsymbol{\psi}_{\mathbf{x}}\rangle = W(\mathcal{T}(\boldsymbol{\sigma};\boldsymbol{\phi}))U(\mathbf{x})|0^{\otimes n}\rangle. \tag{17}$$

TensorHyper-VQC 的一些关键性质包括：(i) 低秩张量网络结构所带来的强参数相关性；(ii) 有效自由度众多，且仅随 n按多项式增长；(iii) 所有二分划分上的算子施密特秩和纠缠熵均有界。这些性质确保了即使在名义上的过参数化区域，所诱导的酉系综仍远非哈尔随机。

4.4. 施密特秩和算子施密特秩

如命题 1 所述，算子施密特秩刻画了算子的纠缠能力：算子施密特秩较大的酉算子能够在二分划分上生成接近最大值的纠缠。相比之下，施密特秩有界的算子则受限于低纠缠流形。具体而言，在 n 个量子比特上作用的哈尔随机酉算子以压倒性概率具有关于 $|A|$ 呈指数增长的算子施密特秩；而键维数固定的张量网络电路所生成的酉算子，其算子施密特秩则由张量秩的函数所界定，且与系统规模 n无关。

形式上，设 $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ 为与一个 n量子比特系统的二分划分 $A|B$相关联的双分希尔伯特空间。那么，任意纯态 $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ 都可进行施密特分解

$$|\boldsymbol{\psi}\rangle = \sum_{i=1}^{\mathcal{I}} \lambda_i |u_i\rangle_A \otimes |v_i\rangle_B, \tag{18}$$

其中， $\lambda_i > 0$和 $\{|u_i\rangle_A\}$、 $\{|v_i\rangle_B\}$ 均为正交归一集。整数 $\mathcal{I}$ 称为 $|\boldsymbol{\psi}\rangle$ 跨越二分划分 $A|B$ 的施密特秩。它量化了该态中的双分纠缠程度： $\mathcal{I} = 1$ 对应乘积态，而对于高度纠缠的态， $\mathcal{I}$ 达到最大值。更一般地，对于线性算子 $\mathcal{O} \in \mathcal{B}$（$\mathcal{H}_A \otimes \mathcal{H}_B$），其算子施密特分解为

$$O = \sum_{i=1}^{\mathcal{I}_{\mathrm{op}}} A_i \otimes B_i, \tag{19}$$

其中 $\{A_i\} \subset \mathcal{B}(\mathcal{H}_A)$ 和 $\{B_i\} \subset \mathcal{B}(\mathcal{H}_B)$ 是线性无关算子集。这种分解所需的最少项数 $\mathcal{I}_{\mathrm{op}}$ 称为O 在二分划分 $A|B$下的算子施密特秩。

4.5. 定理 *1* 的证明

步骤 *1*：化简为哈尔随机态。假设 1 指出，对于固定的$\mathbf{x}$，状态$|\psi_{\boldsymbol{\theta},\,\mathbf{x}}$是 Haar 典型的（等价地说，其低阶矩与 Haar 测度的低阶矩相匹配；酉 2-设计足以计算均值/方差）。因此，我们可以将其视为

$$|\boldsymbol{\psi}\rangle \sim \text{Haar on the unit sphere in } \mathbb{C}^d,$$

并分析

$$f(\psi) := \langle\boldsymbol{\psi}|O|\boldsymbol{\psi}\rangle. \tag{20}$$

步骤 *2*：计算期望。哈尔随机纯态的一个标准恒等式为

$$\mathrm{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|] = \frac{1}{d}. \tag{21}$$

因此，

$$\mathrm{E}[f(\boldsymbol{\psi})] = \mathrm{E}[\mathrm{Tr}(O|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|)] = \mathrm{Tr}(O\cdot\mathrm{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|]) = \frac{1}{d}\,\mathrm{Tr}\,(O)\,, \tag{22}$$

这证明了式 (3)，若取 $d = 2^n$。

步骤 *3*：计算方差（二阶矩法）。给定交换算子 $\mathcal{F}\in\mathbb{C}^d\otimes\mathbb{C}^d$，利用众所周知的二阶矩公式，我们有

$$\mathrm{E}[|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|\otimes|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|] = \frac{I+\mathcal{F}}{d(d+1)}. \tag{23}$$

此外，我们写出

$$f(\boldsymbol{\psi}) = \mathrm{Tr}(O|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|),\;\; f(\boldsymbol{\psi})^2 = \mathrm{Tr}\,((O\otimes O)(|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|\otimes|\boldsymbol{\psi}\rangle\langle\boldsymbol{\psi}|))\,. \tag{24}$$

取期望后，我们有：

$$\mathrm{E}[f(\boldsymbol{\psi})^2] = \mathrm{Tr}\left((O\otimes O)\frac{I+\mathcal{F}}{d(d+1)}\right) = \frac{1}{d(d+1)}\left(\mathrm{Tr}(O)^2+\mathrm{Tr}(O^2)\right), \tag{25}$$

其中，我们运用了 $\mathrm{Tr}((O\otimes O)I) = \mathrm{Tr}(O)^2$ 以及 $\mathrm{Tr}((O\otimes O)\mathcal{F}) = \mathrm{Tr}(O^2)$。因此，

$$\mathrm{Var}(f(\boldsymbol{\psi})) = \mathrm{E}[f(\boldsymbol{\psi})^2] - \mathrm{E}[f(\boldsymbol{\psi})]^2 = \frac{\mathrm{Tr}(O^2)}{d(d+1)} - \frac{\mathrm{Tr}(O)^2}{d^2(d+1)}. \tag{26}$$

特别地，

$$\mathrm{Var}(f(\boldsymbol{\psi})) \le \frac{\mathrm{Tr}(O^2)}{d(d+1)} \le \frac{d\|O\|_2^2}{d(d+1)} \le \mathcal{O}(d^{-1}) = \mathcal{O}(2^{-n}), \tag{27}$$

这证明了公式 (4)。

步骤 *4*：指数尾界（测度集中）。定义单位球面上的 $g(\boldsymbol{\psi}) := \langle\psi|O|\boldsymbol{\psi}\rangle$ 。可以证明 g 是利普希茨连续的，其常数与 $\|O\|_2$成正比：

$$|g(\boldsymbol{\psi}) - g(\boldsymbol{\theta}) = |\langle\boldsymbol{\psi}|O|\boldsymbol{\psi}\rangle - \langle\boldsymbol{\theta}|O|\boldsymbol{\theta}\rangle| \le 2\|O\|_2\|\boldsymbol{\psi}-\boldsymbol{\theta}\|_2, \tag{28}$$

因此， g 是具有 $L\le 2\|O\|_2$的 L-利普希茨函数。

根据 L´evy 引理 [36] （高维球面上的集中现象），存在一个普适常数 $c_0 > 0$，使得对于所有 $\epsilon > 0$,

$$\Pr\left(|g(\boldsymbol{\psi}) - \mathrm{E}[g(\boldsymbol{\psi})]| > \epsilon\right) \leq 2\exp\left(-c_0 \frac{d\epsilon^2}{L^2}\right) \leq 2\exp(-c\epsilon^2 d), \tag{29}$$

其中 $c := c_0/(4\|O\|_2^2)$。若将 $\|O\|_2$ 视为常数（有界可观测量），即可得到所述界

$$\Pr\left(\left|f_{\boldsymbol{\theta}} - \frac{1}{d}\operatorname{Tr}(O)\right| \geq \epsilon\right) \leq 2\exp(-c\epsilon^2 d) = 2\exp(-c\epsilon^2 2^n), \tag{30}$$

其中我们令 $d = 2^n$，由此证明了式 (5)，并完成证明。值得注意的是，均值和方差只要求满足酉 2-设计假设，而指数尾界则来自球面上利普希茨函数的测度集中。

4.6. 定理 *2* 的证明

步骤 *1*：均值为零。根据假设 1，相关状态是 Haar 典型的（或至少构成一个 2-设计）。对于哈尔随机的 $|\boldsymbol{\psi}\rangle \in \mathbb{C}^d$，我们有

$$\mathrm{E}\left[\langle\boldsymbol{\psi}|A|\boldsymbol{\psi}\rangle\right] = \frac{1}{d}\operatorname{Tr}(A) \tag{31}$$

对于任意固定算子 A。将此应用于 $A = G_k$:

$$\mathrm{E}\left[\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\right] = \frac{1}{d}\operatorname{Tr}(G_k). \tag{32}$$

但 G_k 在酉共轭下等价于一个对易子:

$$G_k \propto i[H_k, O_k] \quad \Rightarrow \quad \operatorname{Tr}(G_k) \propto i\operatorname{Tr}([H_k, O_k]) = 0, \tag{33}$$

因为迹$([A, B]) = $迹$(AB) - $迹$(BA) = 0$。因此，我们得到

$$\mathrm{E}\left[\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\right] = 0. \tag{34}$$

步骤 *2*: 方差 差为 $\mathcal{O}(\frac{1}{d}) = \mathcal{O}(2^{-n})$。对于哈尔随机 $|\boldsymbol{\psi}\rangle$，二阶矩满足 s

$$\mathrm{E}\left[(\langle\boldsymbol{\psi}|A|\boldsymbol{\psi}\rangle)^2\right] = \frac{\operatorname{Tr}(A)^2 + \operatorname{Tr}(A^2)}{d(d+1)}. \tag{35}$$

因此,

$$\operatorname{Var}(\langle\boldsymbol{\psi}|A|\boldsymbol{\psi}\rangle) = \frac{\operatorname{Tr}(A^2)}{d(d+1)} - \frac{\operatorname{Tr}(A)^2}{d^2(d+1)} \leq \frac{\operatorname{Tr}(A^2)}{d(d+1)}. \tag{36}$$

将其与 $A = G_k$ 结合，并利用 $\operatorname{Tr}(G_k^2) \leq d\|G_k\|_2^2$ 和 $d = 2^n$:

$$\operatorname{Var}\left(\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\right) \leq \frac{d\|G_k\|_2^2}{d(d+1)} \leq \frac{\|G_k\|_2^2}{d+1} = \mathcal{O}(d^{-1}) = \mathcal{O}(2^{-n}). \tag{37}$$

4.7. 命题 *1* 的证明

步骤 *1*：归结为 $t = 2$ 的情形。如果某个系综是某个的 --近似酉 --设计，那么它也是 --近似酉 *2*-设计（因为匹配至 阶的 *Haar* 矩，尤其也会匹配至 *2* 阶的 *Haar* 矩）。因此，只需证明对于足够大的 ，不可能是近似酉 *2*-设计。

步骤 *2*：有界的算子施密特秩。固定量子比特 n 的任意二分划分，划分为 $A|B$。考虑任意酉算子 $U \in \mathcal{U}_{\mathrm{ts}}$。对于键维数为 r 且与 n 无关的张量网络结构， U 存在一种张量网络表示，其跨越 $A|B$ 的切割宽度由r控制。因此， U 跨越 $A|B$的算子施密特秩，

$$\mathcal{I}_{\mathrm{op}}(U; A|B) := \min\left\{k : U = \sum_{i=1}^{k} A_i \otimes B_i\right\}, \tag{38}$$

仅由键维数决定的某个函数所界定：

$$\mathcal{I}_{\mathrm{op}}(U; A|B) \le K(r), \tag{39}$$

其中 $K(r) = \mathrm{poly}(r)$（特别是不会随 n增长）。这是张量网络切割界的标准推论：任意切割两侧线性无关项的数量受跨越该切割的键维数乘积所界定。

步骤 *3*：使用 *Choi-Jamiolkowski* 态并将算子施密特秩与纠缠联系起来。对每个酉算子 U，存在

$$|U\rangle := (U \otimes I)|\Phi\rangle, \ \ |\Phi\rangle := 2^{-n/2} \sum_{z\in\{0,1\}^n} |z\rangle|z\rangle, \tag{40}$$

这是双倍系统 $(\mathcal{H}_A \otimes \mathcal{H}_B) \otimes (\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ 上的一个纯态。

一个关键恒等式是， U 跨越 $A|B$ 的算子施密特秩等于 $|U\rangle$ 跨越二分划分 $AA'|BB'$的施密特秩：

$$\mathcal{I}_{\mathrm{op}}(U; A|B) = \mathcal{I}(|U\rangle; AA'|BB'). \tag{41}$$

因此，对于所有 $U \in \mathcal{U}_{\mathrm{ts}}$,

$$\mathcal{I}(|U\rangle; AA'|BB') \le K(r). \tag{42}$$

令 $\rho_{AA'}(U) := \mathrm{Tr}_{BB'}(|U\rangle\langle U|)$为 $|U\rangle$在 AA'上的约化密度矩阵。由于约化态的秩至多为施密特秩，

$$\mathrm{rank}(\rho_{AA'}(U)) \le K(r). \tag{43}$$

对于任意秩为 $\le K$的密度矩阵 ρ ，其纯度满足以下基本界

$$\mathrm{Tr}(\rho^2) \ge \frac{1}{K}, \tag{44}$$

当支撑集维数为 K 时，在最大混合态下取等号。因此，对于每个$U \in \mathcal{U}_{\mathrm{ts}}$,

$$\mathrm{Tr}(\rho_{AA'}(U)^2) \ge \frac{1}{K(r)}. \tag{45}$$

对 $U \sim \mathcal{U}_{\mathrm{ts}}$ 取期望可得

$$\mathrm{E}_{U\sim\mathcal{U}_{\mathrm{ts}}}\left[\mathrm{Tr}(\rho_{AA'}(U)^2)\right] \geq \frac{1}{K(r)}. \tag{46}$$

步骤 *4*：*Haar* 预测（以及2-设计预测）呈指数级小。现在与 Haar 系综进行比较。对于 Haar 随机酉算子 $U \sim$ Haar，Choi 态 $|U\rangle$ 在 $2n$ 量子比特希尔伯特空间上是 Haar 随机的，并且当AA' 不大于 BB'时，其在 AA' 上的约化态几乎是最大混合态。特别地，对于平衡切分 $|A| = |B| = n/2$，有 $\dim(AA') = 2^{2|A|} = 2^n$，标准的 Haar 平均计算给出

$$\mathrm{E}\left[\mathrm{Tr}(\rho_{AA'}(U)^2)\right] = \Theta(2^{-n}), \tag{47}$$

它随 n 呈指数级衰减。

关键在于，函数 $\mathrm{Tr}(\rho_{AA'}(U)^2)$ 是关于 U 和 $U^\dagger$ 的矩阵元的一个次数为 $(2,2)$ 的多项式。因此，任何精确的酉 2-设计都会精确复现 $\mathrm{Tr}(\rho_{AA'}(U)^2)$ 的 Haar 期望，该期望随 ϵ 趋于零。因此，如果 $\mathcal{U}_{\mathrm{ts}}$ 是一个满足 $\epsilon = o(1)$ 的 ϵ-近似酉 2-设计，我们就有

$$\mathrm{E}_{U\sim\mathcal{U}_{\mathrm{ts}}}\left[\mathrm{Tr}(\rho_{AA'}(U)^2)\right] = \Theta(2^{-n}) + o(1), \tag{48}$$

当 $n \to \infty$ 时趋于 0。

步骤 *5*：导出矛盾。但式 (46) 表明，对于张量结构系综，

$$\mathrm{E}_{U\sim\mathcal{U}_{\mathrm{ts}}}\left[\mathrm{Tr}(\rho_{AA'}(U)^2)\right] \geq \frac{1}{K(r)} > 0, \tag{49}$$

这是一个与 n 无关的常数，因为 r 固定，因此 $K(r)$ 也固定。因此，当 n 足够大时，一个正的常数不可能接近 $\Theta(2^{-n})$。这与 $\mathcal{U}_{\mathrm{ts}}$ 匹配 Haar 二阶矩的要求相矛盾（即使是误差趋于零的近似匹配）。因此，当 $n \to \infty$ 时，$\mathcal{U}_{\mathrm{ts}}$ 不是近似酉 2-设计。根据步骤 1，对于任意固定的 $t \geq 2$，它都不可能是近似酉 t-设计。

4.8. 命题 *2* 的证明

步骤 *1*：将问题化归为可观测量支撑集上的局部边缘态。令 S 表示 O 的支撑集，且 $|S| = k = \mathcal{O}(1)$。定义 S 上的约化密度矩阵：

$$\rho_S(\boldsymbol{\theta}, \boldsymbol{x}) := \mathrm{Tr}\left(|\psi_{\boldsymbol{\theta},\mathbf{x}}\rangle\langle\psi_{\boldsymbol{\theta},\mathbf{x}}|\right). \tag{50}$$

于是，可观测量输出为

$$f(\boldsymbol{\theta};\mathbf{x}) := \langle\psi_{\boldsymbol{\theta},\mathbf{x}}|O|\psi_{\boldsymbol{\theta},\mathbf{x}}\rangle = \mathrm{Tr}(O\rho_S(\boldsymbol{\theta},\mathbf{x})), \tag{51}$$

其中 $\|O\|_2$ 有界，$|S|$ 为常数，且 $f(\boldsymbol{\theta}, \mathbf{x})$ 是有界连续函数。

步骤 *2*：张量结构意味着局部参数依赖。由于 $W(\boldsymbol{\theta})$ 具有固定秩 r 的张量结构，电路的诱导态在二分划分之间具有有界关联；对于局域可观测量 O，更重要的是，约化态 $\rho_S(\boldsymbol{\theta}, \mathbf{x})$ 仅取决于有限的“因果邻域”内的参数。

更正式地说，存在一个大小为 $m = m(r, k)$ 的参数子集 $\boldsymbol{\theta}_{\mathrm{loc}} = \mathcal{O}(1)$，以及一个可测映射 g，使得

$$f(\boldsymbol{\theta};\mathbf{x}) = g(\boldsymbol{\theta}_{\mathrm{loc}};\mathbf{x}), \tag{52}$$

其中， f 与其余参数 $\boldsymbol{\theta}_{\text{rest}}$相互独立。这里， m（r， k）与n 相互独立，因为张量秩固定，且可观测量支撑集大小 k 固定。因此，写作 $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{loc}}, \boldsymbol{\theta}_{\text{rest}})$，

$$\text{Var}(f(\boldsymbol{\theta}; \mathbf{x})) = \text{Var}(g(\boldsymbol{\theta}_{\text{loc}}; \mathbf{x})). \tag{53}$$

因此，只需对 $\boldsymbol{\theta}_{\text{loc}}$的方差给出下界，其维度与 n无关。

步骤 *3*：非退化性意味着方差严格为正。根据假设， $\boldsymbol{\theta}$ 的初始化分布具有密度且是非退化的。因此， $\boldsymbol{\theta}_{\text{loc}}$ 的边缘分布在某个开集上也具有密度。

假设在该分布下， $g(\boldsymbol{\theta}_{\text{loc}};\mathbf{x})$并非几乎处处为常数。那么，存在两点$\mathbf{a},\mathbf{b}$ ，它们位于 $\boldsymbol{\theta}_{\text{loc}}$ 的支撑集中，且满足

$$g(\mathbf{a}; \mathbf{x}) \neq g(\mathbf{b}; \mathbf{x}). \tag{54}$$

令 $\triangle := \frac{1}{2}|g(\mathbf{a}; \mathbf{x}) - g(\mathbf{b}; \mathbf{x})| > 0$。由 $g(\cdot; \mathbf{x})$的连续性，存在$\mathbf{a}$和$\mathbf{b}$的开邻域$U_{\mathbf{a}}$、 $U_{\mathbf{b}}$ ，使得

$$g(\boldsymbol{\theta}_{\text{loc}}; \mathbf{x}) \geq g(\mathbf{a}; \mathbf{x}) - \triangle, \quad \forall \boldsymbol{\theta}_{\text{loc}} \in U_{\mathbf{a}}, \tag{55}$$

$$g(\boldsymbol{\theta}_{\text{loc}}; \mathbf{x}) \geq g(\mathbf{b}; \mathbf{x}) - \triangle, \quad \forall \boldsymbol{\theta}_{\text{loc}} \in U_{\mathbf{b}}, \tag{56}$$

特别地，对于任意 $\boldsymbol{\theta}_a \in U_{\mathbf{a}}$ 和 $\boldsymbol{\theta}_{\mathbf{b}} \in U_{\mathbf{b}}$,

$$|g(\boldsymbol{\theta}_{\mathbf{a}}; \mathbf{x}) - g(\boldsymbol{\theta}_{\mathbf{b}}; \mathbf{x})| \geq \triangle. \tag{57}$$

由于 $\boldsymbol{\theta}_{\text{loc}}$ 的边缘分布具有密度，且 $U_{\mathbf{a}}$、 $U_{\mathbf{b}}$ 是具有非零体积的开集，因此这些概率

$$p_{\mathbf{a}} := \Pr(\boldsymbol{\theta}_{\text{loc}} \in U_{\mathbf{a}}) > 0, \;\; p_{\mathbf{b}} := \Pr(\boldsymbol{\theta}_{\text{loc}} \in U_b) > 0 \tag{58}$$

严格为正，且仅取决于边缘律，因此仅取决于 r、 k，而不取决于 n。

现在应用初等方差下界：对于任意随机变量 Z ，

$$\text{Var}(Z) \geq \frac{p_{\mathbf{a}} p_{\mathbf{b}}}{(p_{\mathbf{a}} + p_{\mathbf{b}})^2} \left(\text{E}[Z|U_{\mathbf{a}}] - \text{E}[Z|U_{\mathbf{b}}]\right)^2. \tag{59}$$

取 $Z = g(\boldsymbol{\theta}_{\text{loc}};\mathbf{x})$，上述分离性意味着

$$\text{E}[Z|U_{\mathbf{a}}] - \text{E}[Z|U_{\mathbf{b}}] \geq \triangle, \tag{60}$$

因此

$$\text{Var}(g(\boldsymbol{\theta}_{\text{loc}}; \mathbf{x})) \geq \frac{p_{\mathbf{a}} p_{\mathbf{b}}}{(p_{\mathbf{a}} + \mathbf{b})^2} \triangle^2 = c(r, O) > 0. \tag{61}$$

该常数取决于局部参数化和可观测量类，这由 r 和 O 刻画，但不取决于 n。将其与步骤 2 结合，我们最终得到

$$\text{Var}(\langle \boldsymbol{\psi}_{\boldsymbol{\theta},\mathbf{x}} | O | \boldsymbol{\psi}_{\boldsymbol{\theta},\mathbf{x}} \rangle) = \text{Var}(g(\boldsymbol{\theta}_{\text{loc}}; \mathbf{x})) \geq c(r, O), \tag{62}$$

这证明了命题 2。

5. 致谢

本研究部分由香港研究影响基金（R6010-23）资助。

6. 参考文献

[1] John Preskill, 《含噪中等规模量子时代及其未来的量子计算》, *Quantum*, 第 2 卷, 第 79 页, 2018 年。[2]Kishor Bharti 等, 《含噪中等规模量子算法》, 《现代物理评论》, 第 94 卷, 第 1 期, 第 015004 页, 2022 年。[3] Abhinav Kandala、Antonio Mezzacapo、Kristan Temme、Maika Takita、Markus Brink、Jerry M Chow 和 Jay M Gambetta, 《面向小分子和量子磁体的硬件高效变分量子本征求解器》, 《自然》, 第 549 卷, 第 7671 期, 第 242–246 页, 2017 年。[4] Yu Zhang 等, 《降低电路复杂度的变分量子本征求解器》, *npjQuantum Information*, 第 8 卷, 第 1 期, 第 96 页, 2022 年。[5] Leo Zhou、Sheng-Tao Wang、Soonwon Choi、Hannes Pichler 和 Mikhail D Lukin, 《量子近似优化算法: 近端设备上的性能、机制与实现》, 《物理评论*X*》, 第 10 卷, 第 2 期, 第 021067 页, 2020 年。[6] Zhihui Wang、Stuart Hadfield、Zhang Jiang 和 Eleanor G Rieffel, 《用于最大割问题的量子近似优化算法: 费米子视角》, 《物理评论*A*》, 第 97 卷, 第 2 期, 第 022304 页, 2018 年。[7]Linghua Zhu 等, 《用于在量子计算机上求解组合优化问题的自适应量子近似优化算法》, 《物理评论研究》, 第 4 卷, 第 3 期, 第 033029 页, 2022 年。[8] Marco Cerezo 等, 《变分量子算法》, 《自然综述: 物理》, 第 3 卷, 第 9 期, 第 625–644 页, 2021 年。[9] Jun Qi、Chao-Han Yang、Pin-Yu Chen 和 Min-Hsiu Hsieh, 《TensorHyper-VQC: 由张量列车引导、用于稳健且可扩展的变分量子计算的超网络》, *npjQuantum Information*, 2025 年。[10] Vedran Dunjko, 《量子学习揭示量子系统》, 《科学》, 第 376 卷, 第 6598 期, 第 1154–1155 页, 2022 年。[11] Samuel Yen-Chi Chen、Chih-Min Huang、Chia-Wei Hsing 和 Ying-Jer Kao, 《端到端可训练的经典-量子混合分类器》, *Machine Learning: Science and Technology*, 第 2 卷, 第 4 期, 第 045021 页, 2021 年。[12] Samuel Yen-Chi Chen、Tzu-Chieh Wei、Chao Zhang、Haiwang Yu 和 Shinjae Yoo, 《用于高能物理数据分析的量子卷积神经网络》, 《物理评论研究》, 第 4 卷, 第 1 期, 第 013231 页, 2022 年。[13] Junyu Liu 等, 《迈向大型机器学习模型的可证明高效量子算法》, 《自然·通讯》, 第 15 卷, 第 1 期, 第 434 页, 2024 年。[14] Jacob Biamonte、Peter Wittek、Nicola Pancotti、Patrick Rebentrost、Nathan Wiebe 和 Seth Lloyd, 《量子机器学习》, 《自然》, 第 549 卷, 第 7671 期, 第 195–202 页, 2017 年。[15] Maria Schuld、Ilya Sinayskiy 和 Francesco Petruccione, 《量子机器学习导论》, 《当代物理》, 第 56 卷, 第 2 期, 第 172–185 页, 2015 年。[16] William Huggins、Piyush Patil、Bradley Mitchell、K Birgitta Whaley 和 E Miles Stoudenmire, 《借助张量网络实现量子机器学习》, *QuantumScience andTechnology*, 第 4 卷, 第 2 期, 第 024001 页, 2019 年。[17] Matthias Steffen、David P DiVincenzo、Jerry M Chow、Thomas N Theis 和 Mark B Ketchen, 《量子计算: IBM 视角》, *IBM Journal of Research and Development*, 第 55 卷, 第 5 期, 第 13–1 页, 2011 年。[18] Salonik Resch 和 Ulya R Karpuzcu, 《量子计算机基准测试及量子噪声的影响》, *ACM Computing Surveys*, 第 54 卷, 第 7 期, 第 1–35 页, 2021 年。[19] Yuxuan Du、Min-Hsiu Hsieh、Tongliang Liu、Dacheng Tao 和 Nana Liu, 《量子噪声保护量子分类器免受对抗者攻击》, 《物理评论研究》, 第 3 卷, 第 2 期, 第 023153 页, 2021 年。[20]Marco Cerezo、Guillaume Verdon、Hsin-Yuan Huang、Lukasz Cincio 和 Patrick J Coles,

“量子机器学习中的挑战与机遇”，《自然·计算科学》，第 2 卷，第 9 期，第 567–576 页，2022 年。[21] Abhishek Shukla、Mitali Sisodia 和 Anirban Pathak，“对 IBM 量子处理器所用可直接实现量子门的完整表征”，《物理快报*A*辑》，第 384 卷，第 18 期，第 126387 页，2020 年。[22] Mart´ín Larocca 等，“变分量子计算中的贫瘠高原”，《自然综述：物理》，第 1–16 页，2025 年。[23] Enrique Cervero Mart´ín、Kirill Plekhanov 和 Michael Lubasch，“量子张量网络优化中的贫瘠高原”，*Quantum*，第 7 卷，第 974 页，2023 年。[24]Zo¨e Holmes、Kunal Sharma、Marco Cerezo 和 Patrick J Coles，“将拟设可表达性与梯度幅值及贫瘠高原联系起来”，《*PRX*量子》，第 3 卷，第 1 期，第 010313 页，2022 年。[25]Marco Cerezo、Akira Sone、Tyler Volkoff、Lukasz Cincio 和 Patrick J Coles，“浅层参数化量子电路中依赖代价函数的贫瘠高原”，《自然·通讯》，第 12 卷，第 1 期，第 1791 页，2021 年。[26] Jun Qi、Chao-Han Huck Yang、Pin-Yu Chen 和 Min-Hsiu Hsieh，“基于变分量子电路的函数回归的理论误差性能分析”，*npj QuantumInformation*，第 9 卷，第 1 期，第 4 页，2023 年。[27] Debashis Paul 和 Alexander Aue，“统计学中的随机矩阵理论：综述”，《统计规划与推断杂志》，第 150 卷，第 1–29 页，2014 年。[28] Quynh T Nguyen 等，“等变量子神经网络理论”，《*PRX*量子》，第 5 卷，第 2 期，第 020328 页，2024 年。[29] Terence Tao，随机矩阵理论专题，第 132 卷，美国数学学会，2023 年。[30] Alan Edelman 和 N Raj Rao，“随机矩阵理论”，《数值数学学报》，第 14 卷，第 233–297 页，2005 年。[31] Marshall Hall 和 Nicolaus Tideman，“集中度量”，《美国统计学会杂志》，第 62 卷，第 317 期，第 162–168 页，1967 年。[32] Michel Ledoux，测度集中现象，第 89 号，美国数学学会，2001 年。[33]Kaining Zhang、Min-Hsiu Hsieh、Liu Liu 和 Dacheng Tao，“高斯初始化有助于深层变分量子电路逃离贫瘠高原”，载于神经信息处理系统，2022 年。[34]Yuxuan Du、Tao Huang、Shan You、Min-Hsiu Hsieh 和 Dacheng Tao，“面向变分量子算法的量子电路架构搜索”，*npj Quantum Information*，第 8 卷，第 1 期，第 62 页，2022 年。[35]Shi-Xin Zhang、Chang-Yu Hsieh、Shengyu Zhang 和 Hong Yao，“可微量子架构搜索”，*Quantum Science and Technology*，第 7 卷，第 4 期，第 045023 页，2022 年。[36]Ariel Neufeld 和 Marcel Nutz，“非线性 L´évy 过程及其特征”，《美国数学会会刊》，第 369 卷，第 1 期，第 69–95 页，2017 年。