arXiv:2303.03227v2 [quant-ph] 1 Nov 2023

并行混合网络：量子与经典神经网络之间的协同作用

Mohammad Kordzanganeh、Daria Kosichkina 和 Alexey Melnikov*

Terra Quantum AG, Kornhausstrasse 25, 9000 St. Gallen, Switzerland

量子神经网络代表了一种新型机器学习范式，因其潜在优势而近来备受关注。在某些条件下，这类模型能够用截断傅里叶级数近似其数据集的分布。这种拟合的三角函数特性可能导致角度嵌入量子神经网络难以拟合给定数据集中的非谐波特征。此外，神经网络的可解释性仍是一项挑战。在这项工作中，我们提出了一类新型、可解释的混合量子神经网络：它们将数据集的输入并行传递给 1）经典多层感知机和 2）变分量子电路，然后将两者的输出进行线性组合。我们观察到，量子神经网络会基于训练集形成平滑的正弦基底，而经典感知机则填补其中非谐波的空缺。我们通过两个从周期分布中采样、并加入凸起作为噪声的合成数据集，验证了这一观点。训练结果表明，对于带有额外噪声的周期数据集，并行混合网络架构可能改善解的最优性。

I. 引言

近年来，机器学习和量子计算已成为颇具吸引力的研究领域。寻找高效的量子神经网络（QNN），一直是这两种技术交汇领域的研究重点。人们提出了许多关于一种经典计算难以处理的量子机器学习模型潜在内部机制的设想 [1–4], ，但理论和硬件方面的局限性可能会给实现带来挑战。无噪声情况下的贫瘠高原问题 [5]以及维数灾难 [6] ，都是量子神经网络面临的理论挑战。与此同时，硬件限制也体现在行业对精度和量子比特数量的限制上 [7, 8]。因此，当代量子技术在机器学习中的实际应用，应当采用互补的量子—经典架构，即混合量子神经网络（HQNN）：这种架构使用规模相对较小、可实现的量子电路和经典多层感知机（MLP），让二者协同工作。文献 [9–14] 探讨了串行HQNN的适用性和性能，其中多层感知器（MLP）与量子神经网络（QNN）串联，将信息从一个网络传递到另一个网络。串行HQNN可能会在模型的表征能力中引入信息瓶颈，从而限制网络的表达能力。本研究探讨并行HQNN的理论基础：在这种架构中，变分量子电路（VQC）和多层感知器（MLP）并行处理信息。这种方法基于普适性定理，依据来自两个方面：1）多层感知器（MLP）能够生成非调和函数 [15] ；2）量子神经网络（QNN）能够拟合训练数据上的平滑截断傅里叶级数 [16]。本研究受到文献 [17]提出的傅里叶神经算子的启发。

在第二节中，我们回顾多层感知机和变分量子电路的理论基础；在第三节中，我们介绍 PHN 的设计和实验结果。在第三节 B 部分中，我们讨论训练 PHN 时可能出现的组件主导性问题，即变分量子电路或多层感知机中的任一者可能主导训练过程，并提出相应的解决方案。最后，在第四节中，我们总结研究结果并讨论未来方向。

II. 理论基础

在本研究中，我们专注于使用数据集（$\mathbf{x}_i$，y_i）解决监督回归问题，其中 $\mathbf{x}_i \in \mathcal{X}$是特征向量，$y_i \in \mathcal{Y}$ 是标签。我们的目标是发现一个函数 f（$\mathbf{x}$），使其能够近似样本外特征的标签y 。为此，我们创建一个参数为 θ 的机器学习模型，以实现函数 f_θ（$\mathbf{x}$）。我们根据训练样本调整这些参数，以最大化对于给定特征获得正确标签的概率。一般形式的函数 f 是机器学习架构，而其参数的一种具体实现 θ 则是机器学习模型。在接下来的章节中，我们将探讨两种知名架构，并利用这些理论基础为第三节中的 PHN 架构提供依据。

A. 多层感知器

多层感知器构成了一大类成功的机器学习架构。它们是有向图

*通讯作者，电子邮箱：alexey@melnikov.info

请查看已发表版本，其中包含所有最新的增补和修订：Intell. Comput. 2:0028, 2023，DOI：10.34133/icomputing.0028

其节点按一维层排列，每层从前一层接收输入，并将输出传递给下一层。对于全连接多层感知机（FCN），每一层的所有神经元都会向其紧邻前方层中的所有神经元传递信息。图中的每条边都对应一个乘法因子（权重），每个神经元则对应一个加法量（偏置），二者共同构成多层感知机的参数。第一层神经层称为输入层，最后一层称为逻辑值层。参考文献 [18] 对神经网络及其性质进行了全面综述。

参考文献 [19] 表明，多层感知器在渐近意义上是通用逼近器；随着中间神经层中的神经元数量趋于无穷，它们对训练数据的拟合会变得完美。此外， [15, 20]通过一种新颖的图形方法证明了这一点：该方法表明，具有单个中间神经层的全连接网络可以通过拟合矩形波的叠加来逼近任意函数。要将此图用作机器学习架构，我们可以将从样本中选取的数据点的特征编码到输入层上，$\mathbf{x}_i$，然后通过将这些值乘以架构的权重 $\mathbf{w}$ 并加上偏置 $\mathbf{b}$，沿图传播这些值

$$h_i = \sigma(w_{i,j}^{(0)} x_j + b_i), \tag{1}$$

其中 σ 称为激活函数[1]， h_i 表示连续神经层的值，并采用爱因斯坦求和约定。传播过程可以沿着整个图继续进行，直到到达终端节点，即多层感知机的预测结果。

B. 变分量子电路

变分量子电路（VQC）利用变分群旋转，在量子计算机上构建机器学习架构 [4, 21–24]。要构建变分量子电路（VQC），可以先创建一个由多个量子比特组成且处于基态的量子节点。然后，可以对该电路应用一系列变分门和固定量子门。变分门可以包括泡利旋转门，这些门需要相应泡利门的单量子比特时间演化哈密顿量。固定量子门可能包括受控非门（CNOT）和 Hadamard（H）门。我们将变分门分为嵌入门和可训练门，前者用于编码特征 $\mathbf{x}$，后者充当模型参数， θ。在电路末端，我们在某个指定基底（例如泡利基）上测量量子比特，得到 0 或 1 中的一个

[1] 在本研究中，激活函数的具体功能并不重要。为简单起见，我们在所有情况下均采用 sigmoid 函数 $\sigma(t) = \frac{1}{1+e^{-t}}$；只有在明确说明时才例外。

结果。经过多次电路迭代后，我们可以求得得到 0 而非 1 的可能性，并通过取平均值获得该电路的期望值。根据玻恩规则 [25], ，我们可以通过计算测量矩阵的期望值M来求得这一概率，然后将其作为模型的输出：

$$f(\mathbf{x}, \theta) = \langle \psi(\mathbf{x}, \theta) | \, M \, | \psi(\mathbf{x}, \theta) \rangle \,, \tag{2}$$

其中， $|\psi(\mathbf{x}\ \theta)\rangle$ 表示测量前量子电路的状态。通过优化变分量子电路的参数 θ，我们可以提高对标签的近似效果。 [16] 证明了变分量子电路也是通用逼近器，其工作方式是对样本拟合一个截断傅里叶级数：

$$f(x) = \sum_{k=-L}^{L} c_k e^{ikx}, \tag{3}$$

其中，L 是变分量子电路所能表示的最高阶傅里叶项。

在变分量子电路的运行过程中，变分量子门与固定量子门之间的相互作用至关重要。变分门通过一系列变换引导量子系统，根据数据 x 和可训练参数 θ操控量子态。另一方面，固定量子门（如 CNOT 门和哈达玛门）具有可预测的行为，可用于纠缠和操控量子比特。固定门维持系统结构，而变分门则增强适应性，从而使电路能够学习并适应新数据。

III. 结果——并行混合网络

我们将混合量子神经网络（HQNN）的混合接口分为两类：1）串行：经典部分与量子部分直接相互连接；2）并行：经典多层感知机与变分量子电路并行处理相同的信息。本节将深入探讨后一类混合量子神经网络及其所表示的函数。附录 B 对这两类进行了实证比较，但下文将着重探讨后一类。我们将这类网络称为并行混合网络（PHN）。图1展示了 PHN 的总体架构。二者通过带有可训练权重的加权线性相加进行组合。这些权重决定了各网络对最终输出的贡献。此处使用的特定变分量子电路（VQC）是广义数据重上传变分量子电路，其中 K 个量子比特初始化为状态 $|0\rangle^{\otimes K}$。随后，变分层与编码层交替应用。编码层 S将输入特征 $\{x_1, \cdots, x_N\}$编码为幺正变换，

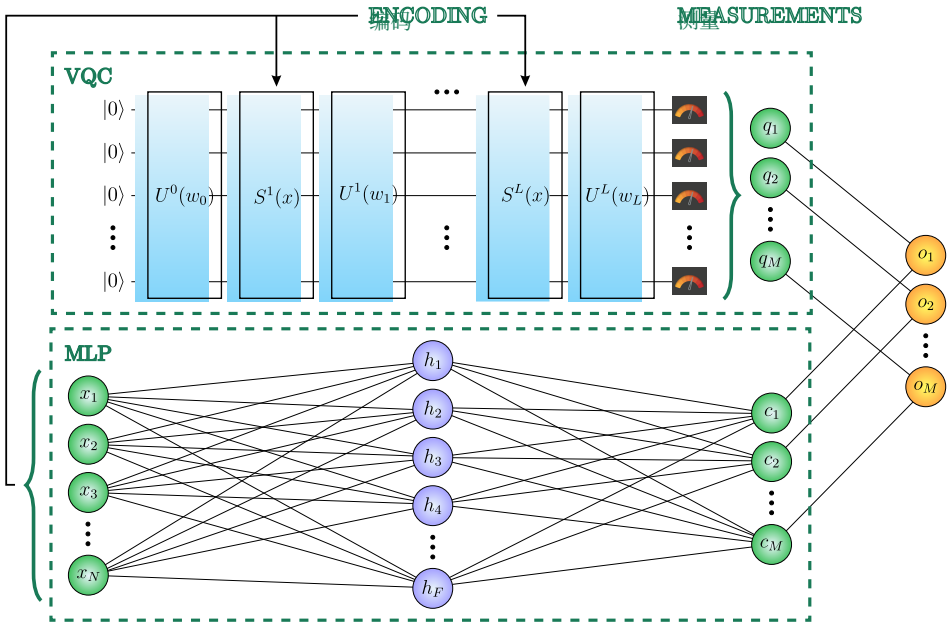

图1：PHN 的总体架构。PHN 接收一个特征输入向量，并将其并行传递给适当架构的角度嵌入变分量子电路（即使用单量子比特泡利门嵌入特征、且不对特征应用任何非线性核的变分量子电路）和一个具有单个隐藏层且规模适当的多层感知机。随后，将变分量子电路的输出与多层感知机的输出进行线性组合，从而生成最终输出向量。

该变换随后作用于量子比特的量子态。变分层 U是幺正算子，它们将 VQC 模型参数封装为可作用于网络量子态的算子。最后，在测量过程中，量子信息坍缩为 M 经典输出；这些输出可通过计算电路相对于测量可观测量的期望值获得。请注意， M（VQC 的经典输出数量）与 K（量子比特数量）有所不同，二者不一定相同，因为我们通常只需测量部分量子比特。

与此同时，全连接多层感知机也接收这些N 特征，并将它们传递至包含 F 个隐藏神经元的单层结构；这一过程通过将特征向量与大小为 $N \times F$ 的权重矩阵相乘来实现。随后，对这些值施加偏置，并使用激活函数进行缩放。激活函数对于向原本为线性的系统引入非线性是必要的。接着，将这些神经元传递至 M 个多层感知机输出神经元，并对其应用各自的偏置和激活函数，记作 $\{c_1, \cdots, c_M\}$。随后，使用一个二对一线性权重层将多层感知机和变分量子电路的输出进行组合，得到 PHN 的输出 $\{o_1, \cdots, o_M\}$。该最终层将变分量子电路的第一个输出与多层感知机的第一个输出组合起来：$o_1 = s_1^c c_1 + s_1^q q_1$；其余 M 个输出也以此类推，其中（$\{s^q\}$，$\{s^c\}$）为可训练参数。

在第二节中，我们看到，只有一个隐藏层的多层感知机为数据集创建了非谐函数拟合，而变分量子电路则创建了截断傅里叶级数，即一个谐函数。因此，将两者结合的网络——可以通过变分量子电路映射平滑的正弦部分，并通过多层感知机填补凸出的部分。这种互补设置有望近似出一个既能在位置空间（多层感知机），又能在共轭动量空间（变分量子电路）拟合数据集的函数。我们可以将这种二重性与参考文献 [17] 中的傅里叶神经算子，或参考文献 [26]中具有良性过拟合的模型相比较。本工作的范围包括并行使用多个变分量子电路（多层感知器）的架构，因为它们总能组合成一个变分量子电路（多层感知机）。

A. 性能

我们从由一个整体单频正弦函数构成的真实值开始，然后向该系统引入高频扰动。具体而言，其函数形式为

$$f = \sin(x) + 0.05\sin(8x) + 0.03\sin(16x) + 0.01\sin(32x),$$

该分布被缩放至 -1 和 1。我们从该分布中选取了 100 个等间距的数据样本用于训练。我们训练了一个简单的 PHN，具体细节见附录 A，其中给出了多层感知机和变分量子电路的确切结构，以尽可能准确地重现真实值。随后，我们训练同一 PHN 架构中的各个组成部分，以观察它们的性能。图 2 展示了训练损失曲线，图 3 展示了各架构对真实值的最佳拟合结果。

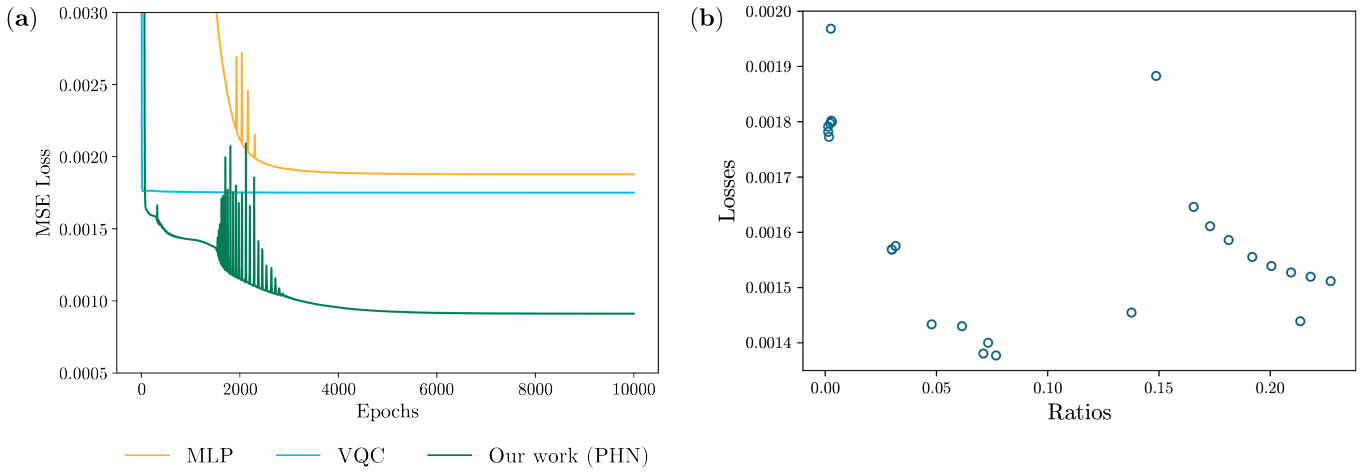

图 2：(a) PHN 的各个独立组件分别训练时的训练损失，以及完整 PHN 的训练损失。(b) 训练 1000 轮次后，最终损失与各自最终比例的散点图。非零比例处可达到最优损失值，而该值两侧的比例则会产生次优损失。请注意，此图仅包含最终损失足够低、可供比较的学习率对应的运行结果。

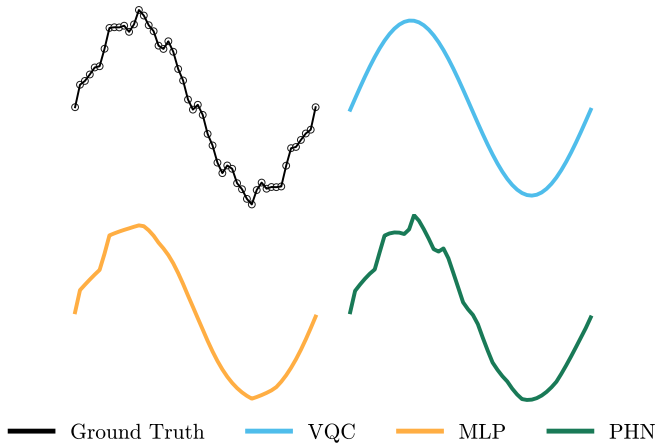

图 3：各架构对真实值的函数拟合。变分量子电路如预期般生成了正弦曲线。多层感知机生成的整体曲线接近正弦曲线，但边缘参差不齐。然而，PHN 的结果最好，成功预测了峰值处的凸起。

PHN 的均方误差训练损失低于其组成部分，这表明加入变分量子电路提升了多层感知机的整体表达能力。此外，通过检查损失曲线，我们发现 PHN 继承了变分量子电路同样快速的下降趋势，同时也具有多层感知机损失曲线中的许多特征，例如尖峰，或训练接近尾声时逐渐趋于平缓的现象。

我们看到，PHN 的表现优于两个单独的组成部分，这意味着变分量子电路和多层感知机都对训练有所贡献，二者都不会变得多余。在第三节 B 中，我们将探讨如何衡量各自的贡献，以及超参数的调整如何改变这种贡献。此外，附录 C 探讨了这种比较的公平性，附录 D 测试了高频率项振幅较大的情况，而附录 E 则考察了 PHN 在该数据集上的泛化能力。

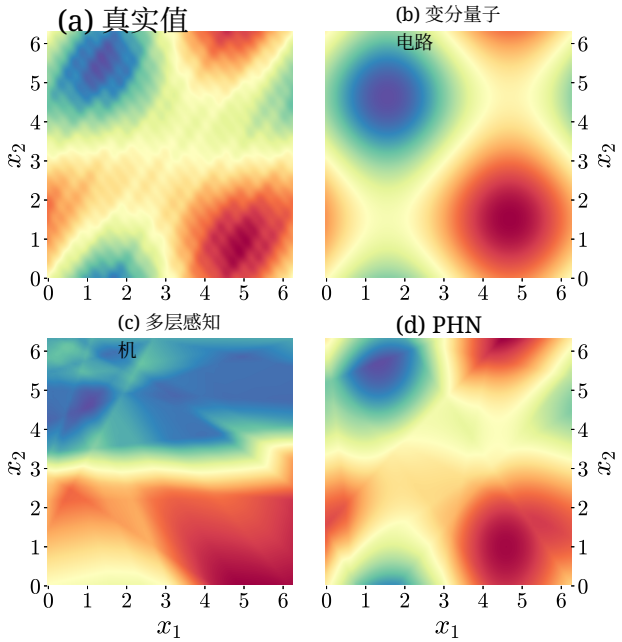

图 4：（a）以等高线图的形式展示了我们二维问题的真实值。图中展示了变分量子电路（b）、多层感知机（c）和 PHN（d）的预测结果。我们看到，变分量子电路的预测平滑且为凸形，而多层感知机则生成锯齿状形状。PHN 利用这两种特性，表示了变分量子电路的调和函数，并加入了必要的凸起。

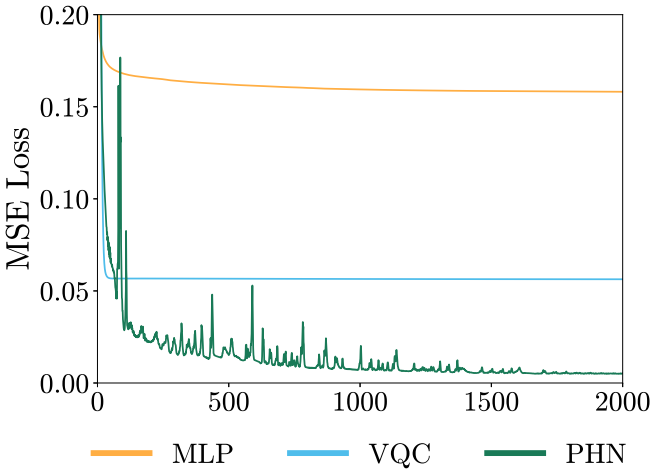

图 5：二维数据集上，图 6(b) 中 PHN 模型及其组成部分（变分量子电路和多层感知机）的损失函数变化。PHN 的训练均方误差损失低得令人印象深刻。

B. PHN 优先性

了解各网络相对贡献的一种方法，是检查组合阶段中分别对应变分量子电路和多层感知机的权重 s_q 和 s_c。本节将探讨在训练 PHN 时，这种贡献会如何变化。

训练 PHN 时，我们必须注意，变分量子电路和多层感知机的训练速率不同。当某个组成架构（变分量子电路或多层感知机）最后一层、位于 PHN 输出层之前的权重变为零时，我们就称该架构相对于另一个架构占据优先地位。等价地，这会使后者网络的输出不再依赖输入特征，也就是说，预测曲线完全由多层感知机或变分量子电路单独构成。这种优先性可能会阻碍 PHN 达到全局最小值，因为它只能发挥其中一个组成部分的能力。

我们使用最终权重的比值 $r = \frac{|s_c|}{|s_q|}$，追踪 PHN 在不同超参数配置下记录到的各类优先性区间。具体而言，我们将变分量子电路的学习率固定为 0.01，然后从 54 个值中为多层感知机选择学习率：$\alpha_c \in \{1.0\text{e}{-7}, 2.0\text{e}{-7}, \cdots, 9.0\text{e}{-7}, 1.0\text{e}{-6}, \cdots, 9.0\text{e}{-2}\}$。随后，我们在固定的初始点上训练 PHN 1000 个训练轮次。最后，我们记录了整个训练过程中比值 r 的变化。比值越大，多层感知机相较于变分量子电路的贡献越大。值得注意的是，即使贡献很小，也可能产生关键影响；只有当某一方的贡献完全消失时，才会出现优先性。

图 2(b) 所示的最终损失对变分量子电路与多层感知机最终权重之比的依赖性，凸显了在 PHN 架构中任一组件主导训练的可能性。结果显示，该比值在 0.1 到 1 之间时处于最优范围，表明变分量子电路与多层感知机的贡献保持平衡有利于取得最佳结果。

此外，显然，当多层感知机完全占据主导地位、即该比值趋近于 0 时，最终损失会变得更差。然而，我们也观察到，调整两个组件的学习率有时可以改善损失。因此，调节变分量子电路与多层感知机的学习率，对于平衡两部分的贡献并防止任一组件主导训练至关重要。

C. 可扩展性与泛化

为展示 PHN 的可扩展性，本节尝试解决一个二维问题，并希望该方法能够扩展到包含大量量子比特的任意复杂问题。为解决图 4(a) 中的问题，我们采用了附录 A 中介绍的一种简单 PHN。用于生成真实值的分布为

$$\begin{aligned} f(x_1, x_2) = &\sin(x_1) + \sin(x_2) + 0.8\sin(x_1 + x_2) \\ &+0.3\sin(x_1 - x_2) + 0.09\sin(8x_1 + 4x_2) \\ &+0.05\sin(16x_1 - 12x_2) + 0.04\sin(12x_1 + 8x_2). \end{aligned}$$

请注意，与一维情形类似，该函数完全随机选取，具有粗略的谐波结构（前四项）和高频噪声（最后三项），并非专门设计来突出 PHN 的优势。从该真实值中等距采样了 100 个点，以构成训练集。随后仅使用变分量子电路、多层感知机和完整的 PHN 对该数据集训练 10,000 个训练轮次。然后在 10,000 个等距采样的数据点上测试训练后的模型，以考察各架构的泛化能力。图 4(b)、(c) 和 (d) 分别展示了变分量子电路、多层感知机和 PHN 的拟合效果，图 5 则展示了它们训练损失的变化过程。变分量子电路生成了对称的正弦模式，而多层感知机则生成锯齿状区域来拟合真实值。然而，PHN 能够结合这两种元素来泛化真实值，从而实现更贴近的拟合。这意味着，对于此类数据集，PHN 的泛化能力可能优于多层感知机或变分量子电路。

IV. 结论

总体而言，我们的研究结果展现了 PHN 作为量子机器学习强大工具的潜力。它是一种混合架构，能够从数据集中提取谐波特征和非谐波特征。借助其独特的架构，PHN 可以学习数据中复杂的模式和关系，而这些模式和关系可能难以通过传统的机器学习算法捕捉。

然而，必须注意的是，PHN 的性能高度取决于超参数的选择。层数、每层的神经元数量

、激活函数和学习率对于决定网络在特定任务上的表现至关重要。因此，超参数调优是成功训练 PHN 的关键步骤。未来研究的一个潜在方向是探索使用自定义学习率调度器，在训练过程中调整学习率。一个学习率调度器可以根据网络在训练集上的性能动态调整学习率，使模型能够更高效地学习并更快收敛。实施学习率调度器或可进一步提升 PHN 在各类任务上的性能。

[1] Yunchao Liu、Srinivasan Arunachalam 和 Kristan Temme。监督式机器学习中严格且稳健的量子加速。 *Nature Physics*, 17(9):1013–1017，2021。[2] Matthias C. Caro、Hsin-Yuan Huang、M. Cerezo、Kunal Sharma、Andrew Sornborger、Lukasz Cincio 和 Patrick J. Coles。少量训练数据下量子机器学习的泛化。 *arXiv:2111.05292*，2021。[3] M. 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。 *Nature Computational Science* 2 9 567 576, 2022 ，( ): –。

[4] Alexey Melnikov、Mohammad Kordzanganeh、Alexander Alodjants 和 Ray-Kuang Lee。量子机器学习：从物理学到软件工程。 *Advances in Physics: X* 8 1 2165452, 2023 ，( ): 。

[5] Jarrod R McClean、Sergio Boixo、Vadim N Smelyanskiy、Ryan Babbush 和 Hartmut Neven。量子神经网络训练景观中的贫瘠高原。*communications* 9 1 4812, 2018 *Nature* ，( ): 。[6] Hsin-Yuan Huang、Michael Broughton、Masoud Mohseni、Ryan Babbush、Sergio Boixo、Hartmut Neven 和 Jarrod R McClean。量子机器学习中的数据力量。 *Nature communications*，12(1):2631，2021。[7]Mohammad Kordzanganeh、Markus Buchberger、Basil Kyriacou、Maxim Povolotskii、Wilhelm Fischer、Andrii Kurkin、Wilfrid Somogyi、Asel Sagingalieva、Markus Pflitsch 和 Alexey Melnikov。使用量子算法和混合算法对模拟及实体量子处理单元进行基准测试。 *Advanced Quantum Technologies*，6:2300043，2023。[8] Samson Wang、Enrico Fontana、Marco Cerezo、Kunal Sharma、Akira Sone、Lukasz Cincio 和 Patrick J Coles。变分量子算法中的噪声诱导的贫瘠高原。 *Nature communications*，12(1):6961，2021。[9]Michael Perelshtein、Asel Sagingalieva、Karan Pinto、Vishal Shete、Alexey Pakhomchik、Artem Melnikov、Florian Neukart、Georg Gesek、Alexey Melnikov 和 Valerii Vinokur。通过混合量子计算实现实用的特定应用优势。 *arXiv:2205.04858*，2022。[10] Asel Sagingalieva、Andrii Kurkin、Artem Melnikov、Daniil Kuhmistrov 等。用于汽车分类的混合量子 ResNet 及其超参数优化。*Quantum Machine Intelligence*，5(2):38，2023。[11] Asel Sagingalieva、Mohammad Kordzanganeh、Nurbolat Kenbayev、Daria Kosichkina、Tatiana Tomashuk 和 Alexey Melnikov。用于药物反应预测的混合量子神经网络。 *Cancers*，15(10):2705，2023。[12] Serge Rainjonneau、Igor Tokarev、Sergei Iudin、Saaketh Rayaprolu、Karan Pinto、Daria Lemtiuzhnikova、Miras Koblan、Egor Barashov、Mo Kordzanganeh、Markus Pflitsch 和 Alexey Melnikov。量子算法应用于地球观测卫星任务规划。 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 7062 7075, 2023 ，: – .

[13] Chao-Han Huck Yang、Jun Qi、Samuel Yen-Chi Chen、Yu Tsao 和 Pin-Yu Chen。BERT 与量子时序卷积学习在异构计算中的文本分类任务上相遇。载于 *ICASSP 2022 - 2022 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* ， ，第 – 页。[14] Jun Qi 和 Javier Tejedor。基于量子神经网络的口语命令识别经典到量子的迁移学习。载于 *ICASSP 2022 - 2022 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* ， ，第 – 页。[15] Michael A Nielsen。《神经网络》，第 25 卷。Determination Press，旧金山，加利福尼亚州，美国，2015 年。[16] Maria Schuld、Ryan Sweke 和 Johannes Jakob Meyer。数据编码对变分量子机器学习模型表达能力的影响。 *Phys. Rev. A*, 103:032430，2021 年。[17]Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。用于参数偏微分方程的傅里叶神经算子。*arXiv:2010.08895*，2020 年。[18] Christopher M Bishop 等。《神经网络》。牛津大学出版社，。[19] G. Cybenko。通过 sigmoid 函数的叠加进行逼近。 *Mathematics of Control Signals and Systems* 2 303 314, 1989 ， ， ，: –。[20] Kurt Hornik、Maxwell Stinchcombe 和 Halbert White。多层前馈网络是通用逼近器。 *Neural Networks*，2(5):359–366，1989 年。[21] Alberto Peruzzo、Jarrod McClean、Peter Shadbolt、Man-Hong Yung、Xiao-Qi Zhou、Peter J. Love、Al´an Aspuru-Guzik 和 Jeremy L. O'Brien。在光子量子处理器上实现变分本征值求解器。 *Nature Communications* 5 4213, 2014 ，: 。

[22] Marcello Benedetti、Erika Lloyd、Stefan Sack 和 Mattia Fiorentini。参数化量子电路作为机器学习模型。 *Quantum Science and Technology*，4:043001，2019 年。[23] Maria Schuld 和 Francesco Petruccione。《量子神经网络》，第 卷。施普林格，2018 年。[24] Jun Qi、Chao-Han Huck Yang、Pin-Yu Chen 和 Min-Hsiu Hsieh。变分量子电路函数回归的理论误差性能分析。*npj Quantum Information*，9(1):4，2023 年。[25] 马克斯·玻恩。关于碰撞过程的量子力学。 *Zeitschrift fur Physik*，37:863–867，1926 年。[26] Evan Peters 和 Maria Schuld。泛化 de-

尽管量子机器学习模型存在过拟合问题。*arXiv:2209.05523*,
2022。

附录 A：确切的实验设置

图 6(a) 和 6(b) 分别展示了用于生成一维和二维结果的 PHN 示例架构。在这两种情况下，q_{out}中的量子测量均在 Z基底下进行；多层感知机分别使用修正线性单元（ReLU）和 Sigmoid 激活函数作为隐藏层和输出层的激活函数，c_{out}。多层感知器采用全连接结构，包含分别具有 256 个和 128 个神经元的层，并设有权重和偏置。多层感知机和变分量子电路的输出分别经 s_c 和 s_q加权后进行线性组合。多层感知机有 769 个参数，变分量子电路有 3 个参数，最终加权层有两个参数。

图 6(b) 描绘了一个简单的二维 PHN，用于展示 PHN 的可扩展性。多层感知机中使用的激活层分别为第一层的 ReLU 和第二层的 Sigmoid。变分量子电路产生一个输出 q_1，该输出通过在 $Z \bigotimes I$ 基底下测量变分量子电路的状态得到；图中省略了恒等测量 I。变分量子电路参数的学习率设为 0.01，其他所有参数的学习率设为 0.001。我们还使用了 Adam 优化器和学习率调度器，每十个训练轮次将所有学习率乘以 $\gamma = 0.99$ 。

附录 B：序贯混合网络中的信息瓶颈

本节聚焦于序贯混合网络中的信息瓶颈，这正是发明 PHN 的主要动机。根据参考文献 [16], ，角度嵌入变分量子电路（例如图 6 中的电路）会生成特征集的截断傅里叶级数，以近似标签。此外，参考文献 [15]表明，多层感知器通过矩形突起在位置空间中拟合函数。在序贯混合网络中，信息流取决于量子过程与经典过程的先后顺序。因此，信息处理能力必然受到变分量子电路或多层感知器处理能力的限制。由于一个过程的输出信息会成为另一个过程的输入，这便构成了信息流中的瓶颈。

例如，在序贯模型中，变分量子电路的输出传递给多层感知机后，多层感知机便会受到变分量子电路所生成的截断傅里叶级数的限制。这可能导致对标签的近似不完整或不精确，因为多层感知机在位置空间中拟合函数的能力可能受到变分量子电路输出质量的制约。类似地，也可以反过来，将多层感知机的输出传递给变分量子电路。在这种情况下，变分量子电路可能难以准确处理多层感知机提供的矩形突起。我们将序贯混合网络中这些信息流与处理能力方面的限制称为信息瓶颈。

相比之下，并行混合网络（PHN）通过允许在量子域和经典域中同时进行信息处理，避开了这些瓶颈。PHN 不再受信息顺序传递的限制，而是能够更高效地利用量子和经典能力。因此，PHN 的处理能力不受单一组件局限的限制，而是由其所有组成部分的累积容量决定。

图 7 展示了在第 IIIA 节的一维数据集上训练时，将变分量子电路依次连接到多层感知机，以及将多层感知机依次连接到变分量子电路所得的结果。

附录 C：与双倍模型的比较

本文将 PHN 与其各个组成部分——多层感知机和变分量子电路进行了比较。一项重要的后续实验是对各个组成部分进行参数化，从而与 PHN 进行更公平的比较。一个直观的实现方法是将每种模型各堆叠两个，即两个变分量子电路并行，或两个多层感知机并行，这就是双倍模型。图 8 展示了在一维数据集上对这种设置的比较，其中将 PHN 与神经元数量翻倍的多层感知机进行比较，并将变分量子电路与另一个规模相同的变分量子电路堆叠。尽管如此，我们仍观察到 PHN 的表现优于这两种模型。

附录 D：具有较大高频项的数据集

为扩展对 PHN 性能的评估，研究人员针对一个以高频分量为主导的函数开展了一组实验。用于该分析的函数表示为 $f(x) = 0.2\sin(x) + \sin(8x)$。该问题采用了两种方法：1）直接将 PHN 应用于数据集，以下简称朴素情形；2）在应用 PHN 之前实施特征缩放。图 9 展示了这两种方法的损失和拟合结果。

在朴素情形下，观察到变分量子电路能够准确识别 $\sin(x)$ 项，但其表达能力局限于频率为 1 的单个傅里叶项。多层感知机准确拟合了最初的峰值，但在这些点之后呈现下降趋势。有趣的是，PHN 提供了最佳拟合，这可能是因为它利用经典分量来适应高频波形，并利用变分量子电路来遵循底层的低频项。这一行为体现了 PHN 作为数据分析起点的潜力，尤其适用于对数据缺乏先验知识的场景。

在采用特征缩放、将输入缩放 8 倍的情况下，变分量子电路和 PHN 的预测效果显著提升，而多层感知机的性能略有下降。这表明，这类预处理方法并不总能提升所有类型网络的表现。

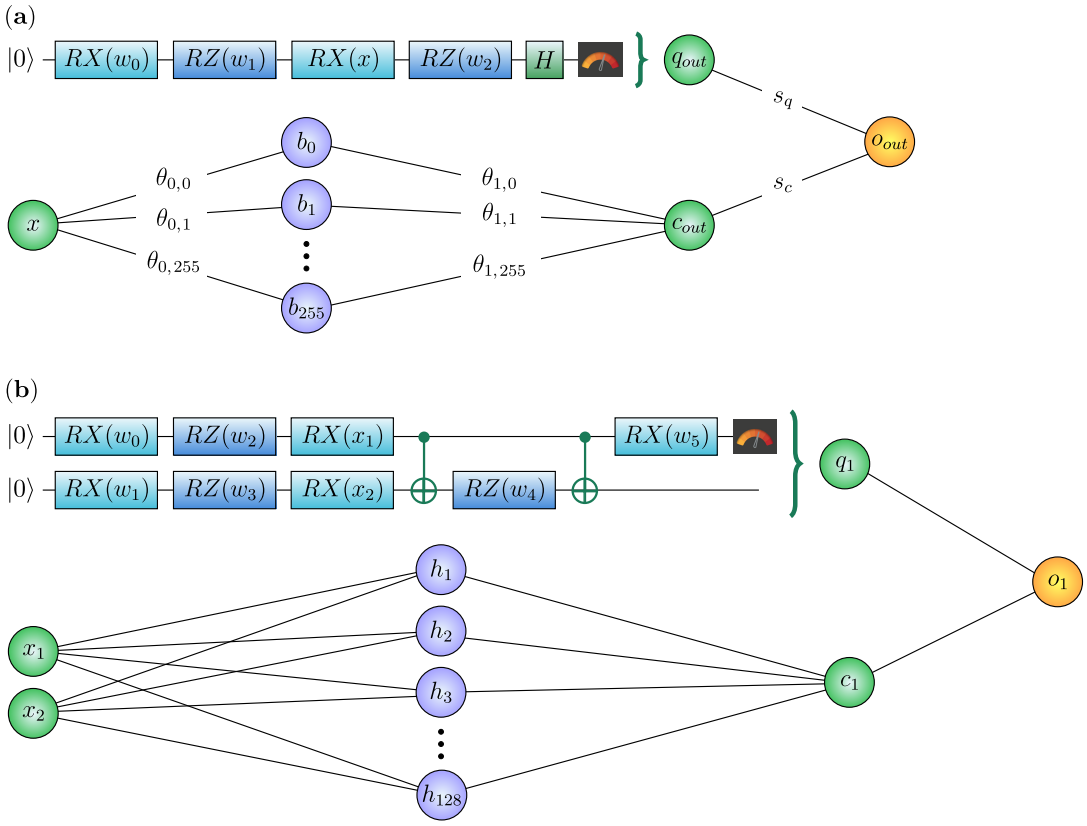

图 6：（a）示意图展示了单量子比特变分量子电路与基本多层感知机架构的协同工作。（b）该架构包括一个双量子比特变分量子电路，仅对第一个量子比特进行测量，以及一个在其唯一隐藏层中含有 128 个神经元的多层感知机。

变分量子电路如今能够识别 $\sin(8x)$ 项，这凸显了特征缩放在释放变分量子电路全部潜力方面的作用。与此同时，PHN 对低频项的拟合效果也更好，优于多层感知机和变分量子电路。

这些结果凸显了 PHN 处理高频成分占主导地位的函数的能力，尤其是在结合特征缩放时。此外，这些结果也表明，在此类情况下采用 PHN 具有潜在价值，其性能优于本研究中考察的其他网络。

附录 E：泛化

为测试 PHN 的泛化能力，我们进行了进一步研究。我们将一维数据集一分为二，仅使用正半部分训练模型，并测试其泛化到未经训练的负半部分的能力。

多层感知机成功拟合了训练数据，但无法预测该范围之外的行为，如图 10 所示。对于周期函数，变分量子分类器（VQC）由于其所用特殊酉矩阵元作为线性映射具有周期性，因此表现出默认的泛化能力。

PHN 同样继承了变分量子电路的外推能力，不过在本例中，并未观察到其泛化能力超越变分量子电路。

三个模型的最终测试损失如下：变分量子电路的损失为 0.008；PHN 的损失为 0.094；多层感知机的损失为 0.133。

这些发现表明，变分量子电路和 PHN 的外推能力更强，尤其是在周期函数方面，同时也凸显了多层感知机在这方面的局限性。此外，研究结果表明，尽管 PHN 继承了变分量子电路的外推能力，但在此特定情况下，它并未在外推方面带来额外优势。

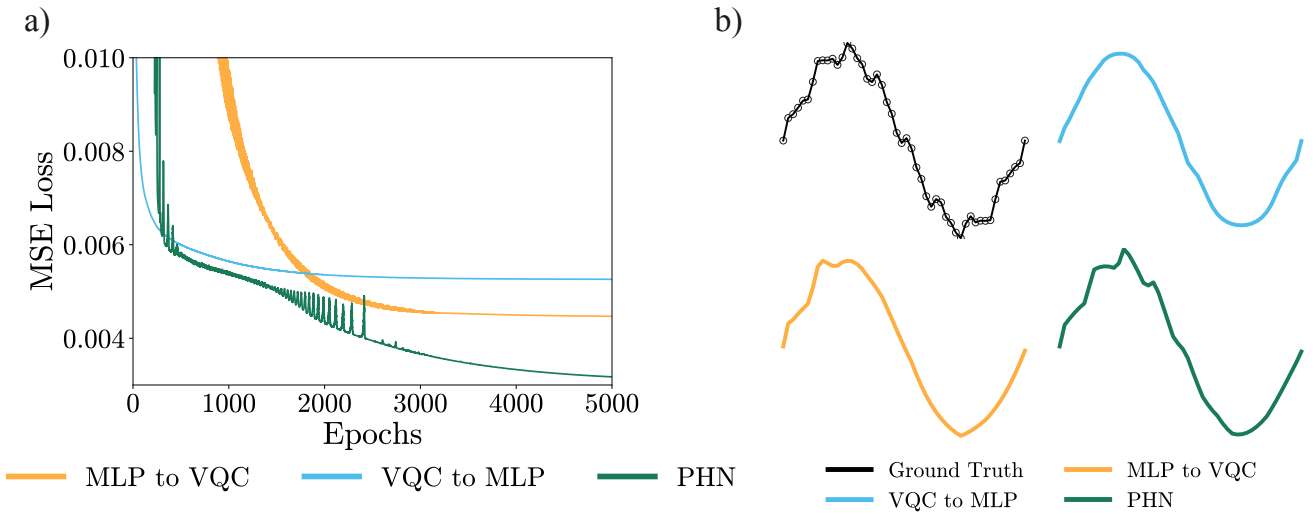

图 7：顺序架构中，在多层感知机到变分量子电路的结构里，多层感知机接收数据集特征，并将其输出传递给变分量子电路。随后，变分量子电路的输出被用作顺序模型的预测结果。变分量子电路到多层感知机模型执行相同的操作，但顺序相反。在这个问题中，PHN 的拟合能力明显优于顺序架构。

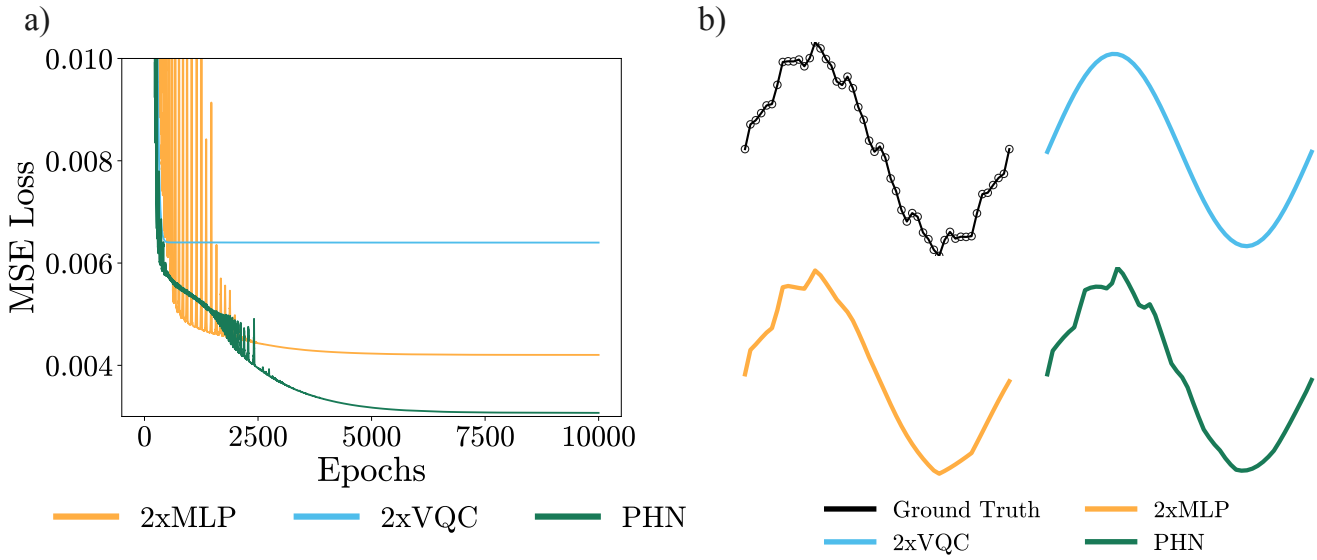

图 8：双 d 网络与 PHN 的结果对比。PHN 仍优于其他架构。

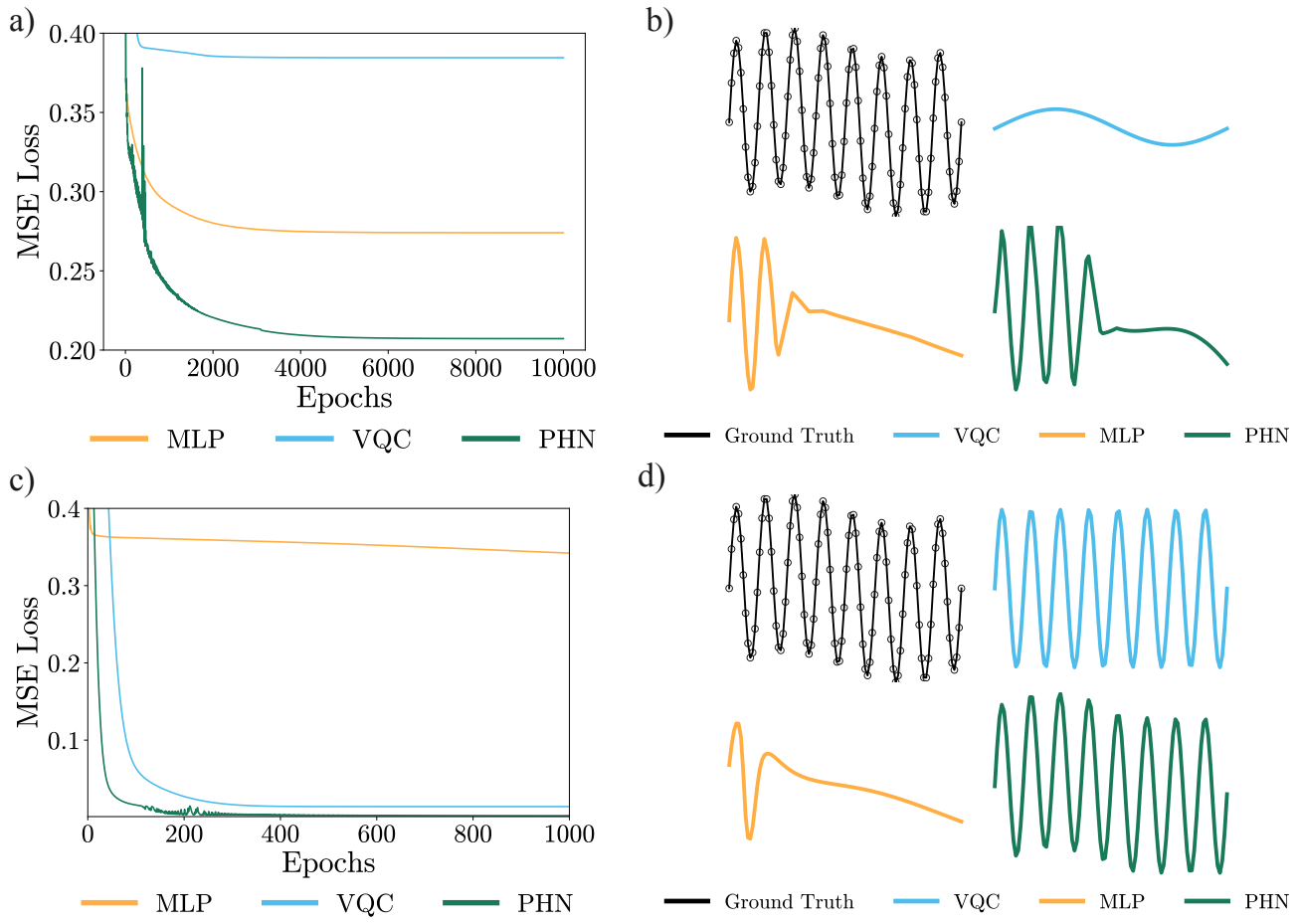

图 9：并行混合网络在具有主导高频分量的函数上的性能示意图。(a) 展示了朴素情形下的损失函数，即直接将 PHN、多层感知机和变分量子电路应用于数据集时，各方法呈现出不同程度的误差幅度。(b) 描绘了朴素情形下的函数拟合，展示了各方法准确拟合函数波形的效果。(c) 展示了缩放情形下的损失函数，即在应用 PHN、多层感知机和变分量子电路之前实施特征缩放，同样突显了不同程度的误差幅度。(d) 展示了缩放情形下的函数拟合，表明缩放后 PHN 和变分量子电路的拟合精度显著提升。此图凸显了在此类情形中采用 PHN 的潜在优势，其性能优于本研究中考察的其他网络。

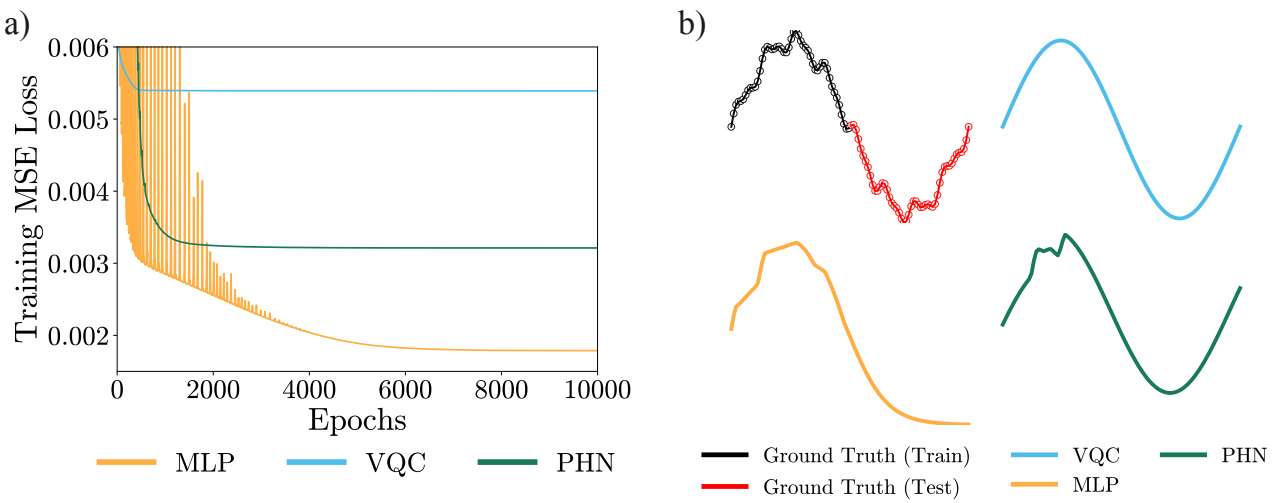

图 10：对变分量子电路、PHN 和多层感知机的外推能力进行比较评估。模型在一维数据集的正半部分上进行训练，随后在未训练的负半部分上进行测试——见（a）。实线表示相应模型学习到的函数，其中多层感知机在未训练区域明显偏离真实函数（点线），表明其外推能力有限。相比之下，变分量子电路和 PHN 展现出更优的外推能力，由于其固有的周期函数处理能力，即使在未训练区域也能很好地拟合该函数。