

通过聚类和退火优化分布式量子计算的编译过程

Ruilin Zhou^{†*}, Jinglei Cheng^{§*}, Yuhang Gan[†], Junyu Liu[§], Chen Qian[†],

[§]University of Pittsburgh [†]University of California, Santa Cruz

*Authors contributed equally to this research.

摘要—高效地将量子程序映射到分布式量子计算 (DQC) 是一个挑战, 特别是考虑到结构不同的异构量子处理单元 (QPUs)。本文提出了一种综合编译框架, 通过三个关键见解来应对这些挑战: 利用量子电路内的结构模式、使用聚类进行初始量子位放置以及用退火算法调整量子位映射。实验结果证明了我们方法的有效性及其处理复杂异构分布式量子系统的能力。我们的评估显示, 与基线相比, 我们的方法最多可将目标值降低 88.40%。

Index Terms—量子计算, 量子编译, 量子计算机体系结构

I. 介绍

量子硬件的发展在不同的量子平台上迅速且具有开创性, 包括离子阱 [1]、超导 [2]、中性原子 [3]、量子点 [4] 和光子学 [5]。在软件方面, 开发量子算法 [6], [7]、量子纠错 [8] 和编译工具 [9] 取得了显著进展。这些进步共同使我们在实现各种任务中的实用量子优势方面更进一步, 包括量子化学 [10]、量子搜索 [11]、量子机器学习 [12], [13] 和组合优化 [14]。

当我们从近期的中等规模量子计算 (NISQ) 过渡到容错量子计算 (FTQC), 量子计算机的规模已经达到了超过一千个量子比特 [15]。为了进一步扩大量子计算机以应对实际任务, 需要具有多个量子处理单元 (QPUs) 的分布式量子计算 (DQC)。然而, 这种方法引入了若干挑战。其中最关键的一个挑战是映射问题, 即如何在考虑 QPUs 之间连接限制和通信开销的情况下将逻辑量子比特分配给物理量子比特。传统的映射技术 [16]–[18] 是为单个 QPU 设计的, 并未优化以解决由 QPUs 间通信引入的额外开销, 这些通信通常实现为远程门操作和量子隐形传态。因此, 迫切需要专门针对分布式量子系统的新型映射策略。此外, 对于分布式量子系统而言, 映射算法必须具有可扩展性, 因为它们的目的

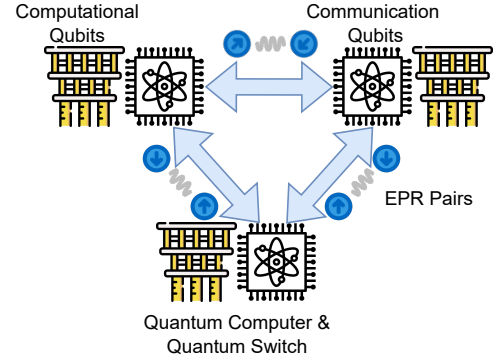


图 1. 我们的 DQC 框架概述。系统架构由具有不同量子位数量和拓扑结构的异构 QPUs 组成, 通过管理远程操作 EPR 对分布的量子开关相互连接。量子开关实现了高效的纠缠资源管理和 QPUs 之间的量子态传输。

标是更大的、更深的并且拥有比 NISQ 时代更多的量子比特的量子程序。

近期的研究 [19]–[23] 提出了多种方法来优化 DQC 的映射和编译过程。例如, [19] 利用了突发通信来减少远程门对整个程序延迟的影响, 而 [21] 探索了量子比特重新分配以最小化不同 QPUs 之间的量子比特移动。然而, 现有的编译方法往往依赖于简化的假设: 它们假定所有 QPUs 的配置相同, 包括一致的量子比特数量、相同的错误率、内部连接的一致性和通信资源的同质性。这些关于同质性的假设无法准确反映分布式量子系统的真实情况, 在这种情况下, 异构性是固有的并且通常是不可避免的。集群内的不同 QPUs 可能在计算和通信量子比特的数量上存在显著差异, 并且具有不同的硬件拓扑结构。当集成来自不同发展代际的 QPUs 时, 这种类型的异构性尤其无法避免, 这在分布式量子计算系统的早期阶段是一种常见情况。

在硬件层面，集群由具有不同规格和能力的 QPU 组成。在程序层面，量子电路也具有非均匀结构，并且在不同区域 [24], [25] 内的门密度各不相同。之前工作的方法并未考虑量子程序内部的门模式。所谓门模式，指的是门倾向于集中在某些量子位上，而不是在整个程序中均匀分布。这些因素表明，仅关注于最小化通信开销的先前 DQC 编译和映射方法可能不适合未来的 DQC 硬件。相反，我们需要能够系统地考虑这些因素——具有不同量子位数量和拓扑结构的异构 QPU、不断演化的量子互连格局以及量子电路内部的结构模式——的编译和映射方法，以提高分布式量子系统的计算效率和可扩展性。

在这项工作中，我们解决了分布式量子计算中有效量子电路映射的先前挑战。该方法包括三个关键步骤：基于门模式将量子电路分割成段、通过时间感知聚类进行初始量子比特分配以及使用模拟退火来优化这一分配。基于退火的优化有效地平衡了局部和远程操作开销之间的权衡，同时考虑了具有不同量子比特数量和拓扑结构的 QPUs 的异构性。通过系统地结合电路模式、我们的交换设计所体现的量子互连能力以及聚类异构性，我们的框架实现了分布式量子系统中逻辑量子比特到物理量子比特的有效映射。

我们的贡献有三点：

- 我们引入了一个感知模式的框架，用于分布式量子电路映射，该框架能够有效地处理具有不同量子位数量和拓扑结构的异构 QPU 集群。
- 我们开发了一种基于量子开关的架构，结合模拟退火优化，高效管理纠缠资源，并平衡局部计算与远程通信开销之间的内在权衡。
- 我们通过全面的方法在分布式量子电路执行中展示了显著改进 (88.40%)，与现有方法相比，实现了更好的资源利用和总体开销的减少。

本文的其余部分组织如下：第 II 节提供了 DQC 的相关背景。第 III 节介绍了我们的动机，并讨论了将量子电路映射到异构 QPU 集群中的挑战。第 IV 节详细描述了我们的方法，包括量子开关架构、模式感知的电路划分以及基于模拟退火的优化。第 V 节通过全面的实验和比较评估了我们的方法。最后，第 VI 节总结了本文。

II. 背景

A. 量子计算基础

量子计算中信息的基本单位是量子位 (qubit)。状态向量用于表示量子位。状态 $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 分别对应经典比特 0 和 1。一个量子位可以存在于这些状态的线性组合中，表达为 $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ ，其中 α 和 β 是复数，满足 $|\alpha|^2 + |\beta|^2 = 1$ 。一个量子比特的状态可以通过单量子比特门（如 Hadamard 门 H ，它可以从基态创建等概率的叠加状态）或 Pauli 门 X, Y 和 Z 等来操控。测量时，量子比特坍缩到状态 $|0\rangle$ 的概率是 $|\alpha|^2$ ，或者坍缩到状态 $|1\rangle$ 的概率是 $|\beta|^2$ 。两个或多个量子比特可以通过双量子比特控制非门（CNOT 门）纠缠在一起。此外，操作三个或更多量子比特的量子门可以分解为单个和两个量子比特门的组合 [26]。大多数现有的量子程序和量子算法使用电路模型，即一系列作用于不同量子比特的门来表示。

B. 分布式量子计算

量子网络和量子互连的最新进展使分布式量子计算成为可能，在这种计算中，大型量子程序可以被分割并分布到不同的量子计算节点上 [27]–[30]。与经典分布式计算类似，分布式量子计算依赖于远程通信。

如果两个量子位处于贝尔态之一，则它们形成一个 EPR 对 [7]。在分布式量子计算中，一对 EPR 中的每个量子位可以位于不同的量子节点以创建远程 EPR 对；而远程通信涉及在两个节点之间建立远程 EPR 对，然后将此 EPR 对作为通信资源来传输量子信息。大多数关于分布式量子计算的先前工作都采用了两种方法之一来执行远程门操作：

- 使用猫纠缠器和猫脱缠器 [31]
- 使用量子隐形传态 [7]

这两种方法都会消耗 EPR 对作为通信资源。在分布式量子计算配置中，每个节点包含两种类型的量子位：通信量子位和计算量子位。计算量子位存储量子态并执行量子门，而通信量子位生成远程 EPR 对然后执行远程门。

III. 动机

A. 异构 QPU

以往 DQC 研究中的一个常见假设是，DQC 将使用同质的全连接 QPU 拓扑结构实现，忽略了更一般的情

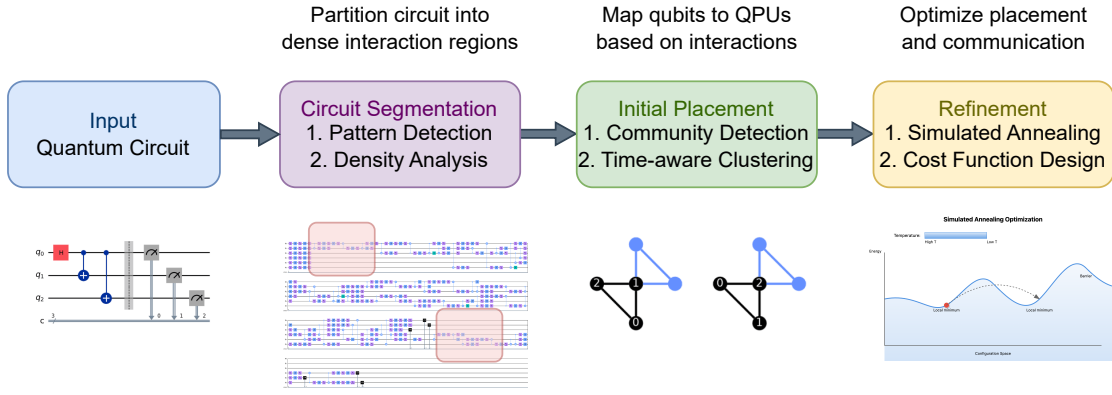


图 2. 我们提出方法的工作流程。主要有三个步骤：电路分割、初始布局和优化。电路分割是识别密集量子门的区域，即电路模式。初始布局是使用社区检测来最小化 QPUs 之间的远程门的可能性，并为进一步的优化提供一个好的起点。优化则是使用模拟退火调整量子比特分配，采用的成本函数考虑了 QPU 间和 QPU 内的成本，在未来不久，与 QPU 间操作相比，QPU 内开销不能被忽视。

况：异构的 QPU 集群。在实际设置中，异构性可以在几个关键方面存在：**不同的 QPU 拓扑结构**：例如今天的 IBM 量子云平台 [32]，它托管具有不同拓扑结构的 QPU。这种变化影响了量子电路如何映射到硬件上，因为某些量子比特交互可能无法直接使用，需要额外开销来实现量子操作。**不同数量的计算和通信量子比特**：一些 QPU 可能有更多量子比特专用于计算，而其他 QPU 则分配更多的量子比特用于通信。这种差异影响了网络中量子算法的分布和执行，并且需要自适应资源分配策略。**受物理距离和环境干扰影响的量子链路质量不一致问题**：QPU 之间的量子链接由于包括距离衰减和环境噪声 [33] 等因素，其保真度会漂移。这导致通信可靠性与性能不稳定，必须在算法设计中予以解决。**来自硬件实现和纠错能力的多种错误率**：不同的 QPU 会因为量子位去相干时间和出厂性质的不同而表现出不同的错误率。这影响了整体计算的准确性，需要专门设计的纠错技术。[34]

B. 更快的量子互连

近期，我们在构建更快更可靠的量子互连方面取得了硬件进步和理论分析方面的改进。一个示例如 [30] 所示，报告了接近 10^5 s^{-1} 的贝尔对速率，保真度约为 0.9999。另一个示例如 [35] 所述，提供了一种使用蒸馏生成高保真贝尔对的方法。从这些进步中，我们可以预见到在不久的将来，生成高质量低开销贝尔对所需的时间将与本地操作相当。因此，在这种情况下，DQC 中仅专注于最小化通信开销或 EPR 对数量的编译方法将

不会是最优的。相反，我们需要方法来系统地探索本地操作开销与远程操作开销之间的权衡，特别是在异构假设下。我们在系统模型中引入了一个量子开关。该开关作为纠缠资源的集中式管理器，协调网络中不同节点之间的 EPR 对生成和分发。这种集中化简化了资源消耗的建模和优化，特别是在远程操作如量子隐形传态和猫纠缠操作上，这些操作各自消耗一个 EPR 对。量子开关抽象还允许我们专注于更高级别的优化决策，同时保持分布式执行所需的纠缠资源的清晰记录。

C. 量子电路中的模式

除了异构 QPU 架构和量子互连的进步所带来的挑战外，某些量子电路中的模式或独特结构也会影响量子电路在 DQC 上的编译。许多量子算法——例如量子傅里叶变换 (QFT)、量子加法器以及用于变分量子本征求解器 (VQE) 的变分电路——具有规则且重复的结构。值得注意的是，即使在大型量子电路中，也有实例表明，在特定时间窗口内只有有限数量的量子位是活跃参与的。在实现量子算法时，某些量子电路被插入了重复的子结构。

识别和利用这些模式使得可以更高效地将量子操作映射到 DQC，从而最小化资源消耗并减少通信开销。探索这些电路模式使编译器能够将特定子电路与具有最适宜拓扑结构和特性的 QPUs 匹配。因此，在设计编译方法时，捕捉这一结构信息至关重要，以便开发出有效平衡异构假设下本地操作与远程操作开销之间权衡的策略。



图 3. 量子电路中门分布模式的可视化。粉红色突出显示的区域表示量子操作集中的区域，展示了重复出现的门模式以及各个量子位上的非均匀门分布。这种自然形成的量子操作聚类表明了通过模式感知编译进行电路优化的机会。

IV. 方法

在本节中，我们介绍了跨多个量子处理单元 (QPUs) 编译分布式量子电路的方法。我们的方法包含三个关键步骤：**电路分割**、**初始布置**和**模拟退火优化**。在电路分割阶段，我们在特定的时间窗口内识别高相互作用的量子比特组，并垂直划分电路以隔离这些交互。基于分割结果，我们编码时间信息并应用时间感知聚类来建立初始量子比特映射，确保频繁近距离未来交互的量子比特位于接近的 QPUs 中。这里，QPUs 的距离由 QPUs 连接的拓扑结构决定。这个假设与之前的工作不同，在之前的工作中，QPUs 是完全连接的。最后，我们使用模拟退火优化量子比特分配，通过最小化本地和远程操作开销来实现，并同时考虑不同的 QPUs 和量子链路的异构特性。

A. 电路段

量子电路的架构经常表现出模式和重复的结构主题。这些模式具有时间片段，在此期间只有特定的量子比特子集积极参与量子操作，或者作为在整个电路执行过程中反复出现的交互模式。这种结构性对分布式量子计算系统具有重要意义，并且常常被之前的工作所忽略，特别是在将计算分布在异构 QPUs 上的时候。在图 3 中展示了这种结构模式的利用，我们显示了两种类型电路的不同模式。我们的目标是将电路分割成其中这些量子比特之间的交互更加密集的时间间隔。为了实现这一目标，我们为每个量子比特计算一个随时间变化的交互密度指标，以量化其在两量子比特门中的活跃程度。在电路的每一层 l 中，我们计算到该层为止涉及量子比特 q 的两量子比特门的累积数量，表示为 $D_q(l)$ 。量子

位 q 在第 l 层的交互密度 $\rho_q(l)$ 定义为 $\rho_q(l) = \frac{D_q(l)}{l}$ 。为了捕捉交互时段并减轻可能的瞬态波动，我们将大小为 w 的滑动窗口应用于交互密度上，计算窗口平均交互密度

$$\bar{\rho}_q(l) = \frac{1}{2w} \sum_{i=l-w}^{l+w} \rho_q(i) \quad (1)$$

。

在每一层 l 中，我们确定集合 S_l ，该集合由具有最高窗口密度 $\bar{\rho}_q(l)$ 的前 k 个量子位组成。为了检测交互模式中的转换，我们使用 Jaccard 指数 $J(S_l, S_{l-1}) = \frac{|S_l \cap S_{l-1}|}{|S_l \cup S_{l-1}|}$ 测量连续集合 S_l 和 S_{l-1} 之间的相似性，并在 $J(S_l, S_{l-1})$ 低于预定义阈值 θ 时，在层 l 启动一个新的片段，其中特定的 $w, k, \text{and } \theta$ 取决于电路的宽度和深度。

B. 初始布置

给定前一步的电路分割，我们通过时间感知聚类确定量子比特在 QPUs 上的初始布局。关键思想是构建一个反映量子比特交互频率并结合时间信息的交互图。时间信息通过赋予近期交互更高的重要性来实现。我们首先构建一个包含所有电路段的整体交互图 $G = (V, E)$ 。每个顶点 $v_i \in V$ 代表一个逻辑量子比特 q_i ，每条边 $e_{ij} \in E$ 连接具有两量子比特门的两个量子比特 q_i 和 q_j 。为了捕获时间信息，我们给边分配了依赖于时间的权重，其中从 q_i 到 q_j 的边的权重 w_{ij} 定义为：

$$w_{ij} = \sum_{s=1}^S \alpha_s \cdot f_{ij}^{(s)} \quad (2)$$

其中 S 是段数总和， $f_{ij}^{(s)}$ 是在 s 段内两个量子比特 q_i 和 q_j 之间的两量子比特门的总数，而 α_s 是段 s 的时间依赖系数，未来的段被分配更大的值。具体来说，我们定义了 $\alpha_s = \exp(-\lambda \cdot s)$ 和 λ 是衰减常数控制器，用于控制未来交互重要性减弱的速率。

借助这一时间感知信息，交互图强调了即将相互作用的量子比特，并确保它们将被放置在接近的 QPU 中，同时也考虑了后期段落中的重要互动。调整超参数将平衡即时和未来的通信需求。为了给每个 QPU 分配量子比特，我们执行聚类以将 G 划分为 P 个不相交子集 $\{V_1, V_2, \dots, V_P\}$ ，目标是最小化总权重边切割

$$C_{twe} = \sum_{\substack{e_{ij} \in E \\ q_i \in V_p, q_j \in V_{p'}, p \neq p'}} w_{ij} \quad (3)$$

，并且用户给出要检测的社区数量 k ，这表示将使用的 QPU。尽管我们通过最小化总加权边缘切割来分配逻辑量子比特给 QPU，但这一初始分配可能会违反 QPU 的容量约束。这些违规问题将在一个细化步骤中得到解决：如果某个 QPU 超出了其容量，我们会调整分配，以贪婪的方式将一些量子比特移动到相邻且具有可用容量的 QPU 上，优先考虑那些在其当前 QPU 内连接较弱的量子比特。这一细化步骤确保始终满足容量约束，并尽可能最小化通信开销。

C. 模拟退火

我们的选择模拟退火算法是基于分布式量子电路优化的内在特性。主要目标是在平衡影响电路性能的各种因素的同时，创建一个复杂的优化格局，在这个格局中，解空间主要由量子比特分配决策决定。模拟退火特别适合此类问题，因为它可以有效地搜索这个复杂的空间。通过允许量子比特在各个段之间移动，我们进一步提高了性能，因为量子比特可以在不同部分的电路中被重新分配以优化交互。在每个段内，我们将量子比特映射到一种方式来增强分布式电路的整体性能。分布式量子电路的性能可能受到多个因素的影响，包括本地开销（例如 QPU 内的门执行和交换操作）、远程开销（例如 QPU 之间的通信以及 QPU 间的量子比特移动）以及段之间量子比特移动的成本。

我们使用一个涉及这些因素的指标来量化性能，并采用模拟退火算法通过最小化以下目标函数来优化量子比特到 QPU 的分配：

$$E = \gamma_1 E_{\text{inter}} + \gamma_2 E_{\text{local}} + \gamma_3 E_{\text{move}}, \quad (4)$$

其中 E_{inter} 表示 QPU 间通信成本， E_{local} 表示 QPU 内操作成本，而 E_{move} 则表示 QPU 之间量子比特移动的成本。

术语 E_{inter} 与 QPU 间的交互成本相关，并惩罚那些将频繁交互的量子比特分配到不同 QPU 上从而导致更高通信开销的配置。每个 QPU 内的操作成本 E_{local} 考虑了由下式定义各个 QPU 内部的开销

$$E_{\text{local}} = \sum_{p=1}^P \sum_{(q_i, q_j) \in G_p} w_{ij} \cdot c_{\text{local}}^{(p)}(q_i, q_j), \quad (5)$$

其中 P 是此量子电路中使用的 QPU 总数， G_p 是分配给 QPU p 的量子比特之间的两量子比特门集，而

$c_{\text{local}}^{(p)}(q_i, q_j)$ 是在 QPU p 上执行从 q_i 到 q_j 的门的成本，考虑了由有限拓扑导致的交换开销。 E_{move} 由

$$E_{\text{move}} = \sum_{q_i \in Q_{\text{move}}} c_{\text{move}}(q_i), \quad (6)$$

定义，表示段之间量子比特的移动，而 $c_{\text{move}}(q_i)$ 是量子比特 q_i 在 QPUs 间移动的成本，这可能取决于 QPUs 之间的距离和量子链路的质量。 $\gamma_1, \gamma_2, \gamma_3$ 是平衡每个项贡献系数的重要程度。在我们的模型中，假设量子比特的移动通过量子隐形传态 [36] 进行，而跨 QPU 操作则由猫纠缠器和猫去纠缠器 [31] 执行。

这些系数将由许多因素进行调整。例如，如果硬件具有更快、更高质量的量子互连，则由于远程开销而导致的性能下降不会很严重。然后，我们可以给 γ_1 分配一个较小的值以减少对最小化跨 QPU 通信成本 E_{inter} 的强调。另一方面，如果某些 QPU 上的本地操作因连接性有限或错误率较高而代价高昂，则需要增加 γ_2 以更重地惩罚内部 QPU 开销 E_{local} 。通过仔细调整这些系数，我们将优化过程针对 DQC 配置的具体特征进行定制。这种灵活性确保了退火过程能够根据其性能的实际影响有效地平衡本地和远程开销之间的权衡。

模拟退火算法迭代搜索一个最小化目标函数 E 的量子比特到 QPU 的分配。从初始分配开始，在每次迭代中，该算法通过进行小改动（交换多个量子比特）或在满足容量约束条件下将一个量子比特移动到不同的 QPU 来处理一个新的分配。对于每次迭代，计算目标函数的变化 $\Delta E = E_{\text{new}} - E_{\text{current}}$ ，并根据以下标准以一定的概率接受新分配：

$$P = \begin{cases} 1, & \text{if } \Delta E \leq 0, \\ \exp\left(-\frac{\Delta E}{T_k}\right), & \text{if } \Delta E > 0, \end{cases} \quad (7)$$

其中 T_k 是第 k 次迭代时的当前温度。温度 T_k 按照冷却计划进行更新。更具体地说，我们使用冷却调度： $T_{k+1} = T_0 / (1 + \alpha k)$ ，其中 T_0 是初始温度， α 是冷却速率。

我们的模拟退火方法利用分割过程中识别的电路模式来指导优化过程。模拟退火方法的一个常见挑战是它们对初始解决方案质量的敏感性——较差的起始点可能导致次优结果和较慢的收敛速度。我们通过使用从前一段得到的高质量布局初始化退火过程来解决这一限制，这为下一阶段提供了完美的初始解决方案。

V. 评估

A. 实验设置

实验评估在一个双插槽的 Intel Xeon Silver 4314 系统上进行, 该系统运行 Ubuntu Linux 22.04.1 LTS。每个插槽包含一个支持超线程的 16 核 CPU, 总共提供 32 个物理核心和 64 个逻辑 CPU。服务器具有 125 GB 的 RAM 和 8 GB 的交换内存。所有基准电路均使用 Qiskit [37] 版本 0.45.0 生成, 并从 [38] 收集。我们评估了三种不同量子比特数量的量子电路类型: 量子傅里叶变换 (QFT)、量子近似优化算法 (QAOA) 和量子算术电路。为了模拟异构 QPU 集群, 我们使用多个虚拟后端 (FakeAuckland、FakeParis、FakeMelbourne 和 FakeBoeblingen) 创建了一个随机拓扑结构。每个模拟的 QPU 的计算量子比特遵循其虚拟后端定义的耦合图。此外, 我们实现了一个量子开关来控制纠缠资源的分配。某个 QPU 的量子开关与其计算量子比特保持全连接。

B. 主要结果

在本节中, 我们展示了我们的方法的有效性和功能性。结果如表 I 所示。我们将采用随机放置找到一组可行的量子比特分配方法作为基线, 并比较了不同层次的方法所实现的目标值减少情况: L1: 使用时间感知聚类 and 电路分割但不进行退火; L2: 采用时间感知聚类和退火, 但电路随机分割且未检测模式; L3: 代表我们的完整方法, 包括通过检测模式来分割电路以及时间感知聚类和退火。主要观察结果是, 对于我们测试的大多数电路而言, 随着我们整合更多层次的方法, 目标值减少情况有所改善, 这验证了整体设计的有效性。另一个值得注意的观察是不同层在各个电路中的相对改进比率存在差异。这种变化可归因于电路本身的特点。量子傅里叶变换 (QFT) 电路表现出独特的电路模式和量子比特局部性, 这一特性使得模式检测更为有利。相比之下, 量子近似优化算法 (QAOA) 电路在整个量子比特集上呈现出重复的模式, 因此基于模式的分割提供的额外优势较少。因此, L2 分割在不同类型的电路中产生了不同程度的改进, 在 QFT 电路中的效果更佳而在 QAOA 电路中则较弱。大多数情况下, 我们的退火方法 (用于 L2 和 L3) 通过优化量子比特映射以有效平衡跨 QPU 和内部 QPU 通信进一步降低了开销。总体而言, 从 L1

到 L3 在不同电路上的逐步改进验证了我们分层方法在提高映射效率方面有效性。

表 I
按层减少目标值。

程序	L1 红色。	L2 红。	L3 红。
qft_100	8.47%	27.84%	34.73%
qft_160	7.05%	7.69%	14.24%
qft_320	7.79%	10.87%	12.42%
qaoa_100	94.19%	92.58%	94.33%
multiplier_75	59.71%	57.18%	66.44%
adder_n118	84.78%	85.04%	88.40%

C. 不同 QPU 间和 QPU 内开销比率的评估

在本小节中, 我们旨在测试我们的方法在不同 Inter-QPU 与 Intra-QPU 开销比率下的有效性。我们将比率从 5:1 变化到 1:1, 这符合量子互连的发展趋势, 即进展导致更快的 Inter-QPU 通信。在此评估中, 我们认为 Inter-QPU 操作和移动都是 Inter-QPU 开销的一部分, 因为它们都依赖于量子互连。我们在加法器 _n118 电路上进行我们的评估, 结果如图 4 所示。从结果观察到, 随着量子互连变得更快 (即 Inter-QPU 开销相对于 Intra-QPU 开销减少), 总能耗降低。具体来说, 不同比率下的 Intra-QPU 开销相对保持不变, 而 Inter-QPU 开销部分则随着更快的互连而减少。这一趋势表明, 当 Inter-QPU 通信变得更加高效且其开销与 Intra-QPU 开销相当时, Inter-QPU 开销对总开销的贡献更小。然而, 这一观察结果符合我们的动机, 在未来的分布式量子计算 (DQC) 系统中, 仅仅最小化 Inter-QPU 开销是不够的。相反, 我们需要考虑影响 DQC 性能的各种因素。鉴于量子算法在不断演变, 不同的量子电路将表现出不同的特征, 这可能会进一步影响这种平衡。我们方法适应不同开销比率的能力表明其在优化各种 DQC 场景下的性能方面具有有效性。

VI. 结论

在本工作中, 我们设计了一个新的分布式量子计算编译框架。我们的框架包括三个主要步骤: 电路模式检测、时间感知聚类和模拟退火, 每个步骤都针对编译过程中的特定挑战。我们工作的主要贡献是考虑了量子电路中固有的模式和特性, 这对 DQC 的性能至关重要。随着量子技术的进步, 我们将不可避免地进入一个异构 QPU 环境, 在这种环境中必须平衡多个因素而不仅仅

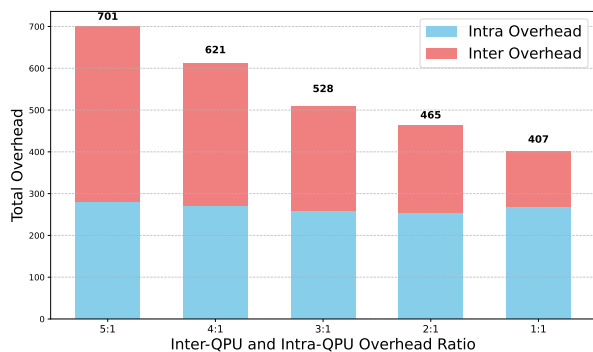


图 4. 不同 QPU 间和 QPU 内开销比率下的总开销

是最小化远程通信开销。我们认为这些挑战迫在眉睫，我们的工作旨在解决这些问题并填补这一空白。我们的实验结果证明了我们方法的有效性（减少了 88.40%），能够处理异构分布式量子系统的复杂性。

参考文献

- [1] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, “Trapped-ion quantum computing: progress and challenges,” *Appl. Phys. Rev.*, 2019.
- [2] J. Clarke and F. K. Wilhelm, “Superconducting quantum bits,” *Nature*, vol. 453, no. 7198, pp. 1031–1042, 2008.
- [3] L. Henriot, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, “Quantum computing with neutral atoms,” *Quantum*, vol. 4, p. 327, 2020.
- [4] D. Loss and D. P. DiVincenzo, “Quantum computation with quantum dots,” *Physical Review A*, vol. 57, no. 1, p. 120, 1998.
- [5] S. Slussarenko and G. J. Pryde, “Photonic quantum information processing: A concise review,” *Applied Physics Reviews*, vol. 6, no. 4, 2019.
- [6] A. Montanaro, “Quantum algorithms: an overview,” *npj Quantum Information*, vol. 2, p. 15023, 2016. [Online]. Available: <https://arxiv.org/abs/1511.04206>
- [7] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.
- [8] D. A. Lidar and T. A. Brun, Eds., *Quantum Error Correction*. Cambridge University Press, 2013. [Online]. Available: <https://www.cambridge.org/core/books/quantum-error-correction/B51E8333050A0F9A67363254DC1EA15A>
- [9] Y. Ge, W. Wenjie, C. Yuheng, P. Kaisen, L. Xudong, Z. Zixiang, W. Yuhan, W. Ruocheng, and Y. Junchi, “Quantum circuit synthesis and compilation optimization: Overview and prospects,” *arXiv preprint arXiv:2407.00736*, 2024. [Online]. Available: <https://arxiv.org/abs/2407.00736>
- [10] I. N. Levine, *Quantum Chemistry*, 7th ed. Pearson, 2014.
- [11] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 1996, pp. 212–219.
- [12] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017. [Online]. Available: <https://arxiv.org/abs/1611.09347>
- [13] M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics*, vol. 56, no. 2, pp. 172–185, 2015. [Online]. Available: <https://arxiv.org/abs/1409.3097>
- [14] E. Farhi, J. Goldstone, and S. Gutmann, “A quantum approximate optimization algorithm,” *arXiv preprint arXiv:1411.4028*, 2014. [Online]. Available: <https://arxiv.org/abs/1411.4028>
- [15] Nature, “Ibm releases first-ever 1,000-qubit quantum chip,” *Nature*, 2023. [Online]. Available: <https://www.nature.com/articles/d41586-023-03854-1>
- [16] G. Li, Y. Ding, and Y. Xie, “Tackling the qubit mapping problem for nisq-era quantum devices,” *Proceedings of the 24th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pp. 1001–1014, 2019. [Online]. Available: <https://arxiv.org/abs/1809.02573>
- [17] A. Zulehner, A. Paler, and R. Wille, “Efficient mapping of quantum circuits to the ibm qx architectures,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 38, no. 7, pp. 1226–1236, 2018. [Online]. Available: <https://arxiv.org/abs/1806.07241>
- [18] Y. Zhang, M. Amolavi, S. Chatterjee, and S. Sankaranarayanan, “Qubit mapping and routing via maxsat,” *arXiv preprint arXiv:2208.13679*, 2022. [Online]. Available: <https://arxiv.org/abs/2208.13679>
- [19] A. Wu, H. Zhang, G. Li, A. Shabani, Y. Xie, and Y. Ding, “Autocomm: A framework for enabling efficient communication in distributed quantum programs,” in *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 2022, pp. 1027–1041.
- [20] Y. Mao, Y. Liu, and Y. Yang, “Qubit allocation for distributed quantum computing,” in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.
- [21] A. Wu, Y. Ding, and A. Li, “Qucomm: Optimizing collective communication for distributed quantum computing,” in *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, 2023, pp. 479–493.
- [22] D. Ferrari, A. S. Cacciapuoti, M. Amoretti, and M. Caleffi, “Compiler Design for Distributed Quantum Computing,” *IEEE Transactions on Quantum Engineering*, vol. 2, pp. 1–20, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9334411/>
- [23] M. Caleffi, M. Amoretti, D. Ferrari, D. Cuomo, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, “Distributed Quantum Computing: a Survey,” Dec. 2022, arXiv:2212.10609 [quant-ph]. [Online]. Available: <http://arxiv.org/abs/2212.10609>
- [24] S. Tan, C. Lang, L. Xiang, S. Wang, X. Jia, Z. Tan, T. Li, J. Yin, Y. Shang, A. Python *et al.*, “Quct: A framework for analyzing quantum circuit by extracting contextual and topological features,” in *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, 2023, pp. 494–508.
- [25] M. Lourens, I. Sinayskiy, D. K. Park, C. Blank, and F. Petruccione, “Hierarchical quantum circuit representations for neural architecture search,” *npj Quantum Information*, vol. 9, no. 1, p. 79, 2023.
- [26] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, “Elementary gates for quantum computation,” *Physical review A*, vol. 52, no. 5, p. 3457, 1995.
- [27] M. Pompili, S. L. Hermans, S. Baier, H. K. Beukers, P. C. Humphreys, R. N. Schouten, R. F. Vermeulen, M. J. Tiggelman, L. dos Santos Martins, B. Dirkse *et al.*, “Realization of a multinode quantum network of remote solid-state qubits,” *Science*, vol. 372, no. 6539, pp. 259–264, 2021.
- [28] S. Hermans, M. Pompili, H. Beukers, S. Baier, J. Borregaard, and R. Hanson, “Qubit teleportation between non-neighbouring nodes in a quantum network,” *Nature*, vol. 605, no. 7911, pp. 663–668, 2022.
- [29] P. Magnard, S. Storz, P. Kurpiers, J. Schär, F. Marxer, J. Lütolf, T. Walter, J.-C. Besse, M. Gabureac, K. Reuer *et al.*, “Microwave quantum link between superconducting circuits housed in spatially separated cryogenic systems,” *Physical Review Letters*, vol. 125, no. 26, p. 260502, 2020.

- [30] Y. Li and J. Thompson, “High-rate and high-fidelity modular interconnects between neutral atom quantum processors,” *arXiv preprint arXiv:2401.04075*, 2024.
- [31] A. Yimsiriwattana and S. J. Lomonaco Jr, “Generalized ghz states and distributed quantum computing,” *arXiv preprint quant-ph/0402148*, 2004.
- [32] IBM, “The ibm quantum development roadmap,” 2023. [Online]. Available: <https://www.ibm.com/quantum/roadmap>
- [33] L. Gyongyosi and S. Imre, “Advances in the quantum internet,” *Communications of the ACM*, vol. 65, no. 8, pp. 52–63, 2022.
- [34] “Suppressing quantum errors by scaling a surface code logical qubit,” *Nature*, vol. 614, no. 7949, pp. 676–681, 2023.
- [35] C. A. Pattison, G. Baranes, J. Ataiades, M. D. Lukin, and H. Zhou, “Fast quantum interconnects via constant-rate entanglement distillation,” *arXiv preprint arXiv:2408.15936*, 2024.
- [36] S. Pirandola, J. Eisert, C. Weedbrook, A. Furusawa, and S. L. Braunstein, “Advances in quantum teleportation,” *Nature photonics*, vol. 9, no. 10, pp. 641–652, 2015.
- [37] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, “Quantum computing with Qiskit,” 2024.
- [38] A. Li, S. Stein, S. Krishnamoorthy, and J. Ang, “Qasmbench: A low-level quantum benchmark suite for nisq evaluation and simulation,” *ACM Transactions on Quantum Computing*, vol. 4, no. 2, pp. 1–26, 2023.