arXiv:2603.01222v1 [cs.IT] 1 Mar 2026

大规模无线网络中的通信高效量子联邦学习

Shaba Shaon、Christopher G. Brinton，*IEEE*高级会员，以及Dinh C. Nguyen

摘要—**量子联邦学习（QFL）将量子计算强大的数据处理能力与联邦学习（FL）的隐私保护特性相结合。然而，在大规模无线网络中，优化和速率对于充分发挥QFL的潜力至关重要；在动态信道条件和波动的功率资源下，各设备争夺有限带宽，因此优化和速率有助于实现有效的模型共享与聚合。本文研究了多信道QFL框架中的一种新型和速率最大化问题，该框架专为基于非正交多址接入（NOMA）的大规模无线网络设计。我们通过联合考虑量子设备的信道选择和发射功率，构建了一个和速率最大化问题。所构建的问题是一个非凸混合整数非线性规划（MINLP）难题；即使给定信道选择参数，该问题仍然是非确定性多项式时间（NP）困难的。问题的复杂性促使我们设计一种有效的迭代优化方法，利用先进的量子近似优化算法（QAOA）求得高质量的近似解。此外，我们首次从理论上探讨了全设备参与下的量子联邦学习收敛性质，严格分析了包含非凸损失函数、多样化数据分布以及量子采样噪声影响的真实场景。大量仿真结果表明，我们提出的多信道、基于NOMA的QFL框架能够提升模型训练效果和收敛表现，在准确率和损失方面均优于传统算法。此外，我们以量子计算为核心的联合优化方法在确保快速收敛的同时，使和速率提升超过**100% **，显著优于当前最先进的方法。**

关键词—**量子联邦学习、大规模无线网络、和速率最大化、时延。**

一、引言

量子联邦学习（QFL）是一种创新方法，它将量子计算与联邦学习（FL）相结合，使量子赋能边缘设备（例如量子处理器或智能量子传感器）能够进行去中心化模型训练 [1], [2]。在这一框架中，量子赋能边缘设备（例如量子处理器）与中心基站（BS）处的量子云服务器协作，共享量子机器学习（QML）模型（例如变分量子算法（VQAs））的更新 [2], ，并在保护隐私和数据安全的同时共同改进全局模型。量子边缘设备采用量子计算，能够带来计算效率和算力方面的显著优势，利用量子电路加速学习并提升模型性能 [3]。尤其是在含噪声中等规模量子（NISQ）时代，QML 通常——ically使用一种混合模型进行训练，该模型将量子电路（如参数化量子电路（PQC））与经典优化技术（如梯度下降）相结合。该方法利用量子处理进行特征编码，并利用经典优化来调整电路参数。我们注意到，含噪声中等规模量子固有的量子机器学习模型，即变分量子算法的参数，在模型训练后仍以经典形式而非量子态存在 [1],，从而使其能够通过无线网络 进行共享和传输，以便进一步聚合。

在 6G 等未来无线网络中，由于数据交换量巨大且功率资源有限，大规模量子边缘设备的量子联邦学习（QFL）将面临通信瓶颈 [4]。因此，和速率最大化（例如优化模型传输速率）对于确保量子赋能边缘网络中的学习高效且可扩展至关重要。在大规模量子联邦学习系统中，通信会成为关键瓶颈，因为量子边缘设备生成的参数化量子电路（PQC）参数必须频繁传输至服务器[1]。这些参数的规模随电路深度和量子比特数增加 [3],，会产生巨大的上行链路需求，而该需求受限于带宽 [1], 、信道衰落 [5], 以及量子边缘硬件较低的发射功率 [6]。具体而言，在基于多信道非正交多址接入的系统中，相对于大量参与设备，可用信道数量有限，往往导致设备间对无线资源的竞争加剧，从而进一步恶化通信瓶颈 [7]。此外，量子电路的噪声敏感性会导致参数质量波动 [8],，使稳定的全局更新更加依赖及时可靠的无线链路。尽管量子联邦学习中传输的信息是经典信息[1],，但底层量子硬件的行为会导致不同于经典联邦学习的通信动态 [4]。测量变异性和量子噪声会使参数化量子电路（PQC）参数发生更剧烈的变化 [9],，因此需要更精确且一致的更新传输以维持收敛。再加上新兴量子处理器严格的功率和硬件限制 [6],，这些因素带来了独特的通信压力，凸显了量子联邦学习系统中无线资源优化的重要性。此外，随着量子边缘设备数量不断增加，在大规模系统中训练和优化模型也变得日益复杂。传统优化技术（如凸优化）往往难以应对这类大规模环境中网络编排和资源分配日益增加的复杂性 [10]。这些局限给无线链路上的和速率最大化带来挑战，最终降低联邦学习中的通信效率。幸运的是，

Shaba Shaon 和 Dinh C. Nguyen 任职于美国阿拉巴马大学亨茨维尔分校电子与计算机工程系，地址：美国阿拉巴马州亨茨维尔，邮编 35899；电子邮件：ss0670@uah.edu、dinh.nguyen@uah.edu。

Christopher G. Brinton就职于普渡大学埃尔莫尔电气与计算机工程学院，西拉斐特，印第安纳州，47907，美国，电子邮箱：cgb@purdue.edu。

量子近似优化算法（QAOA）最近成为加速无线量子联邦学习优化的一种有前景的解决方案 [11], [12]。QAOA基于量子启发技术，尤其适用于解决大规模无线网络等复杂系统中出现的优化问题。QAOA可用于高效搜索量子联邦学习系统中的最优资源分配策略，这对于管理面向未来智能无线网络的大规模量子联邦学习网络中的通信带宽、能源使用和计算资源至关重要。

A. 相关工作

1) QFL：近年来，已有多项研究探索了各种量子联邦学习方案 [13]–[21]，展现了该领域中多样化的方法和技术路线。在 [13],中，作者提出了一种动态量子联邦学习框架，利用量子神经网络提升严峻通信条件下协同无人机监控的效率和鲁棒性。在 [14]中，研究人员开发了一种具有深度可控量子神经网络架构的纠缠可伸缩量子联邦学习系统，通过采用叠加编码参数通信和干扰抑制技术，针对多样化的物联网（IoT）环境进行了优化。研究 [15], [16],[21] 致力于通过创新的量子联邦学习方法提升异构环境中的数据隐私、模型准确率和鲁棒性，解决非独立同分布数据和本地模型漂移等挑战。在 [17]中，研究人员通过实现去中心化环形拓扑，并引入量子权重进行完全基于量子的训练，开发了一种量子联邦学习模型，以应对深度学习中的安全性和数据隐私挑战。作者[18] 提出了一种将量子机器学习与5G车联网相结合的量子联邦学习算法，利用量子最小门控单元进行训练，从而提升计算效率、隐私保护水平和语音情感识别准确率。文献 [19], [20] 采用量子联邦学习方法提升网络安全性和效率，其中前者改善了消费网络中的入侵检测，后者增强了智能电网的稳定性。

2）面向通信高效联邦学习/量子联邦学习的优化与和速率最大化：已有多项研究探索了通信高效联邦学习的优化策略，包括逐次凸近似（SCA）和李雅普诺夫优化。基于梯度的技术（如一阶和二阶方法）利用函数导数求解凸问题和非凸问题。例如，文献 [22], [23] 利用逐次凸近似，为无线网络中的联邦学习优化能效资源分配。尽管加速一阶方法、近端方法和随机梯度方法等改进技术提高了效率，但由于设备数量庞大，在优化变量数量较多的大规模联邦学习系统中，这些方法仍显不足。近期研究 [24] 提出了一种基于量子的序贯训练方法，以降低量子联邦学习中的通信成本。然而，尚无研究使用量子近似优化算法进行优化，也未探究量子近似优化算法在大规模量子联邦学习网络中的优势。

在联邦学习/量子联邦学习优化中，和速率最大化至关重要，因为它决定了模型共享的传输速率，直接影响通信效率和聚合速度。例如， [25]中的作者采用基于拉格朗日对偶的方法，研究了经典联邦学习中的和速率最大化问题。 [26], [27]也通过启发式优化等经典技术，考虑了联邦学习的通信效率。然而，该系统模型规模较小，且尚未研究受量子启发的和速率最大化方法。在大规模无线网络中，联邦学习优化面临的挑战更加严峻：设备需要争用有限的带宽，同时确保更新及时可靠。联邦学习设备通信链路上的信道条件动态变化，功率资源也不断波动，这进一步增加了优化过程的复杂性。有效地最大化和速率对于优化通信效率至关重要，例如降低模型共享时延和提升量子联邦学习收敛速率。

3）量子近似优化算法在无线优化中的应用：量子近似优化算法主要用于构建并求解混合整数二进制规划问题 [28]。文献 [29] 将量子近似优化算法应用于智能电网的能源分配，以改善负载均衡、能效和计算速度。文献 [30] 提出了面向6G移动边缘计算的、基于量子近似优化算法的服务部署优化框架，以降低服务成本和时延；文献 [31]则利用这种基于量子的方法，在多径无线环境中优化可重构智能表面。尽管已有这些研究，但在量子联邦学习背景下应用量子近似优化算法仍未得到探究。

B. 主要贡献

鉴于现有文献中的这些局限性，本文首次提出了一种面向大规模无线网络、具有理论保证的优化 QFL 框架。我们的贡献概述如下：

- 我们提出了一种面向基于非正交多址接入的大规模无线网络的新型多信道量子联邦学习框架 II。我们针对该网络建立了和速率最大化问题，以满足此类网络对资源效率和可扩展性的关键需求。为此，我们联合优化设备的信道选择和发射功率 IV。

- 和速率最大化问题的构建带来了复杂的计算挑战，需要采用精密的问题求解方法。经过分析，我们确定该问题是非凸的，属于混合整数非线性规划（MINLP）问题。此外，即使信道选择参数已指定，该问题仍然是 NP 难的。因此，我们开发并实现了一种量子中心化优化方法，结合 QAOA 和块坐标下降（BCD）技术，以有效解决这一问题。我们将原问题划分为两个子问题，并使用 BCD 迭代求解，从而找到原问题的最优解。我们将每个经典的和速率最大化子问题转换为二次无约束二进制优化（QUBO）形式，随后进一步将其重构为哈密顿量表达式，以便与

量子处理器 V 兼容。

- 我们对全设备参与情况下的量子联邦学习进行了严格的收敛性分析，并考虑了非凸损失函数以及量子设备间不同的数据分布（独立同分布和非独立同分布）等实际场景 III。此外，我们还研究了量子采样噪声对量子联邦学习收敛性的影响。据我们所知，这是首次对量子联邦学习的收敛性质进行理论探究。

- 我们进行了大量数值实验，以评估所提出的基于多信道 NOMA 的量子联邦学习框架在大规模无线网络中的模型训练、收敛行为和和速率性能。仿真结果表明，在模型损失和准确率收敛方面，我们提出的量子联邦学习框架优于最先进算法。此外，与最先进方案相比，我们的量子中心化优化方法使和速率提升了 100% 以上 VI。

II. 系统模型

本节介绍我们所研究的基于 NOMA 的多信道大规模无线网络中的量子联邦学习系统模型。首先，我们定义分析过程中经常用到的几个基本术语。

定义 ***1.*** 含噪中等规模量子处理器： 含噪中等规模量子处理器是一种量子计算设备，其量子比特数量有限，且会受到噪声和错误的影响。

定义 ***2.*** 参数化量子电路： 参数化量子电路是一种量子电路，其量子门操作由可调参数定义。通常通过经典反馈回路对其进行优化，并将其用于解决量子计算中的问题，例如变分量子算法。

定义 ***3.*** 变分量子算法： 变分量子算法是一种同时利用量子计算和经典计算的混合方法。该算法使用参数化量子电路生成量子态，并利用经典优化器迭代更新电路参数，以最小化代价函数。图 1 展示了变分量子算法与量子计算及随机梯度下降的结合，用于在量子设备上进行本地模型训练。

定义 ***5.*** 散粒噪声： 散粒噪声是指由于用于估算量子计算结果的测量次数有限而产生的统计波动。

定义 ***7.***退相干： 退相干是指量子系统与周围环境相互作用，导致叠加态和量子纠缠等量子属性退化，进而造成量子信息损失的现象。

定义 ***8***。 量子门保真度误差： 量子门保真度误差用于量化量子门的实际运行与其理想无错误行为之间的偏差，是衡量量子门实现过程中不完善之处或错误的一项指标。

A. 量子联邦学习模型

在本研究中，我们考虑一种量子联邦学习系统，其中中心化服务器负责协调由 N 分布式量子设备组成的网络。假设每台量子设备都采用变分量子算法，例如量子神经网络（QNN），通过在经典损失函数优化方法（如随机梯度下降（SGD））的辅助下执行参数化量子电路，来构建量子机器学习模型。在每次全局迭代中，设备利用各自的私有数据集，通过量子随机梯度下降进行本地训练，持续 T_n次本地迭代。本地训练完成后，设备将经典模型参数发送回中央服务器进行聚合。随后，服务器采用联邦平均更新全局模型。我们将全部设备组成的集合定义为 $\mathcal{N} = \{1, 2, \ldots, N\}$，而 $\mathcal{K} = \{1, 2, \ldots, K\}$ 表示全局迭代集合。对于每台设备 $n \in \mathcal{N}$，其本地迭代集合记为 $\mathcal{T}_n = \{1, 2, \ldots, T_n\}$。相应地，我们将所有设备的本地迭代索引集合定义为 $\mathbb{T} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N\}$。我们假设参数化量子电路在 q 个量子比特上施加酉变换 $U(\boldsymbol{w_n})$。参数化量子电路由向量 $\boldsymbol{w_n} = (\boldsymbol{\theta}_n^1, \boldsymbol{\theta}_n^2, \ldots\ldots, \boldsymbol{\theta}_n^P)$ 参数化，其中P 表示与参数化量子电路相关的参数数量，且每个参数 $\boldsymbol{\theta}_n^p \in \mathbb{R}$ 满足 $p = 1$、2、.....、 P。参数化量子电路的一种常见架构通常称为拟设，它将酉变换规定为:

$$U(\boldsymbol{\theta}_n) = \prod_{p=1}^{P} U_p(\boldsymbol{\theta}_n^p) V_p, \tag{1}$$

其中， $\boldsymbol{\theta}_n^p$ 表示相应酉算子 $U_p(\boldsymbol{\theta}_n^p)$ 的操作。该参数化量子门可表示为

$$U_p(\boldsymbol{\theta}_n^p) = e^{(-i\frac{\boldsymbol{\theta}_n^p}{2} G_p)}, \tag{2}$$

其中， $G_p \in \{I, X, Y, Z\}^{\otimes q}$ 是泡利字符串生成元。需要说明的是，式（1）中的酉算子 V_p 是固定的，与参数 $\boldsymbol{\theta}$ 无关。该参数化量子电路作用于 q 个量子比特，所有量子比特均从基态 $|0\rangle$ 开始。这将得到与设备 n 对应的如下参数化量子态

$$|\Psi_n\rangle = U(\boldsymbol{\theta}_n)|0\rangle. \tag{3}$$

与该量子态相关联的纯态密度矩阵可表示为

$$\Psi_n(\boldsymbol{\theta}_n) = |\Psi_n(\boldsymbol{\theta}_n)\rangle\langle\Psi_n(\boldsymbol{\theta}_n)|. \tag{4}$$

量子神经网络的目标是利用参数化量子电路生成一个能够最优执行指定机器学习任务的量子态。更具体地说，量子神经网络旨在确定参数向量 $\boldsymbol{\theta}_n \in \mathbb{R}^P$ ，以求解以下优化问题

$$\min_{\boldsymbol{\theta}_n \in \mathbb{R}^P} \left\{ f_n(\boldsymbol{\theta}_n) := \mathcal{C}(\langle Z\rangle_{|\Psi_n(\boldsymbol{\theta}_n)\rangle}, y) \right\}, \tag{5}$$

其中， $f_n(\boldsymbol{\theta}_n)$ 表示设备n所对应的通用任务相关损失函数（例如交叉熵或均方误差）。损失 $\mathcal{C}(\cdot,\ y)$ 用于比较参数化量子电路预测的输出（由期望值$\langle Z\rangle_{|\Psi_n(\boldsymbol{\theta}_n)\rangle}$ 给出）与训练样本的真实标签 y。参数化量子电路生成输出量子态 $|\Psi_n(\boldsymbol{\theta}_n)\rangle$，模型的预测结果则通过测量作用于该量子态的可观测量 Z 的期望值获得，即

$$\langle Z\rangle_{|\Psi_n(\boldsymbol{\theta}_n)\rangle} = \langle 0|U^\dagger(\boldsymbol{\theta}_n) Z U(\boldsymbol{\theta}_n)|0\rangle = \mathrm{Tr}(Z\Psi_n(\boldsymbol{\theta}_n)), \tag{6}$$

其中， $\langle 0|$ 和 $U^\dagger(\boldsymbol{\theta}_n)$分别是 $|0\rangle$和 $U(\boldsymbol{\theta}_n)$的厄米共轭。此外，$\mathrm{Tr}(Z\Psi_n(\boldsymbol{\theta}_n))$是可观测量 Z 与密度

矩阵 $\Psi_n(\boldsymbol{\theta}_n)$的乘积的迹。损失函数$f_n(\boldsymbol{\theta}_n)$的最小值表示为

$$f_n{}^* = \min_{\boldsymbol{\theta}_n \in \mathbb{R}^P} f_n(\boldsymbol{\theta}_n). \tag{7}$$

表示可观测量的厄米矩阵 Z 可表示为其特征分解的形式:

$$Z = \sum_{y=1}^{N_z} h_y \Pi_y, \tag{8}$$

其中， $\{h_y\}_{y=1}^{N_z}$ 是 Z 的 $N_z \leq 2^n$ 个互异特征值， $\{\Pi_y\}_{y=1}^{N_z}$ 是投影到相应特征子空间上的投影算子。利用特征分解，设备 n 上可观测量的期望值可表示为

$$\langle Z \rangle_{|\Psi_n(\boldsymbol{\theta}_n)\rangle} = \sum_{y=1}^{N_z} h_y \operatorname{Tr}(\Pi_y \Psi_n(\boldsymbol{\theta}_n)). \tag{9}$$

将此表达式代入式 (5) 中的一般损失定义，损失函数可写为

$$f_n(\boldsymbol{\theta}_n) = \mathcal{C}(\sum_{y=1}^{N_z} h_y \operatorname{Tr}(\Pi_y \Psi_n(\boldsymbol{\theta}_n)), y). \tag{10}$$

对于量子神经网络，通常使用随机梯度下降（ SGD）迭代优化参数化量子电路的参数向量 $\boldsymbol{\theta}_n$ 。因此，可通过实现一种混合量子—经典优化方法来求解式 (5) 中的优化问题。具体而言，对参数化量子电路的输出态 $|\Psi_n(\boldsymbol{\theta}_n)\rangle$ 测量可观测量 Z 共 H 次，得到一组样本 $\{Z_h\}_{h=1}^{H}$，其中 $Z_h \in \{Z_1, Z_2, ..., Z_H\}$。利用这些样本，可计算期望值的估计值:

$$\langle \hat{Z} \rangle_{|\Psi_n(\boldsymbol{\theta}_n)\rangle} = \frac{1}{H} \sum_{h=1}^{H} Z_h. \tag{11}$$

ˆ我们指出，本文中使用的 (·) 符号表示经验估计值。随后，将式 (11) 中的估计期望传入经典损失函数 $\mathcal{C}(\cdot, y)$，以评估损失并计算用于随机梯度下降（SGD）更新的随机梯度。本文重点研究采用 SGD 优化技术实现 QNN 的标准方法。我们将设备 n 的初始参数向量记为 $\boldsymbol{\theta}_n^0 \in \mathbb{R}^M$，参数向量 $\boldsymbol{\theta}_n$ 的更新规则表示为

$$\boldsymbol{\theta}_n^{t+1} = \boldsymbol{\theta}_n^t - \eta_t \hat{g}_n^t, \tag{12}$$

其中， $t = 1,2, ..., T_n$， T_n 表示设备 n 在一个全局轮次中执行的本地随机梯度下降迭代总次数， $\eta_t > 0$ 表示第t次迭代时的学习率，ˆg_n^t 表示第 t次迭代时损失函数 $f_n(\boldsymbol{\theta}_n)$的随机梯度估计值。在此背景下，ˆ$g_n^t$ 表示为

$$\hat{g}_n^t = \nabla f_n(\boldsymbol{\theta}_n)\Big|_{\boldsymbol{\theta}_n = \boldsymbol{\theta}_n^t} = \begin{bmatrix} \frac{\partial f_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n^1} \\ \vdots \\ \frac{\partial f_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n^P} \end{bmatrix}_{\boldsymbol{\theta}_n = \boldsymbol{\theta}_n^t}. \tag{13}$$

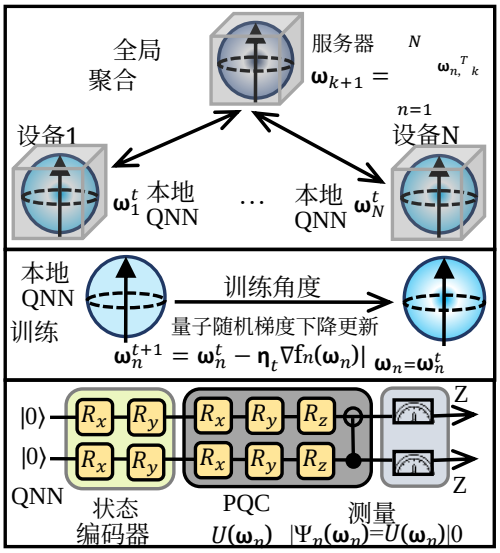


图 1：所提出的 QFL 框架，其中分布式量子设备与中央服务器协作训练共享的机器学习模型。每台设备使用状态编码器将实际数据编码为量子形式，通过包含可训练角度参数的参数化量子电路（PQC）进行处理，然后利用测量结果，采用标准随机梯度下降（SGD）方法更新本地模型参数，随后将更新后的本地模型发送至服务器进行聚合。

B. 网络架构

在这项工作中，我们研究一种多信道量子联邦学习框架，其中量子设备通过多个专用信道与基站通信。具体而言，我们考虑 N个量子设备共享 C 个信道，其中设备数量通常大于信道数量，即$N > C$。设备数量多于信道数量，可通过将有限信道动态分配给更多设备来优化资源利用率，从而提升可扩展性和效率。此外，这种配置也反映了现实世界的约束：可用通信资源通常少于参与设备。我们分别用 $\mathcal{N} = \{1, 2, \ldots, N\}$ 和 $\mathcal{C} = \{1, 2, \ldots, C\}$表示 N 个设备和 C 个信道的集合。在这项工作中，我们假设每个全局轮次都由所有设备参与；借助多信道非正交多址上行架构，所有 N 个量子设备均可通过信道复用同时传输，并通过信号与干扰加噪声比（SINR）表达式显式表征由此产生的同信道干扰。在后续各小节中，我们将详细介绍本研究采用的信道和信号建模方法。

C. 信道模型

我们QFL框架中的所有通信链路均为纯经典无线信道，并采用路径损耗、阴影效应、瑞利衰落和同信道干扰进行建模。量子技术仅用于：(i) 量子边缘设备使用基于PQC的QNN模型进行本地模型训练；(ii) 在基站执行基于QAOA的优化过程。这种设计完全适用于我们的场景，因为量子设备只需与基站交换经典PQC参数 [4]，这意味着无需量子态、纠缠分发或量子通信链路。在我们的系统模型中，量子设备之间的通信链路

以及基站容易受到同信道干扰。当另一台设备同时使用同一通信信道向基站传输模型参数时，就会产生这种干扰。这种重叠可能降低连接质量，影响合法链路的传输效率和可靠性。在我们的框架中，设备 n 具有双重作用：既是与基站建立合法链路的一方，也可能成为干扰源，对任何其他共享同一信道的设备m 所传输的信号造成不利影响 c。假设该系统中的所有信道都具有大尺度和小尺度损伤分量。路径损耗和阴影效应属于大尺度分量，长时间内保持不变；瑞利衰落则代表小尺度分量，其特点是按时隙衰落。在此块衰落模型中，假设信道特性在一定时段内保持不变。为描述瑞利衰落的时间相关性，我们采用Jake 模型 [32]。我们将 b^{th} 和$(b+1)^{\text{th}}$ 时隙内各信道实例的信道系数分别表示为 $g^{(b)}$ 和 $g^{(b+1)}$，则有

$$g^{(b+1)} = \epsilon g^{(b)} + \sqrt{1-\epsilon^2}\delta, \tag{14}$$

其中， $\epsilon(0 \leq \epsilon < 1)$ 表示相关因子，用于刻画连续两个时隙中信道实现之间的相关性； δ 和 $g^{(0)}$ 均为服从分布$\mathcal{CN}(0,\ 1)$ 的复高斯随机变量。此外，分别将设备 n 到基站在信道 c上产生的干扰链路的路径损耗和阴影效应分量记为$\chi_{n,\ m,\ c}$ 和 $\beta_{n,\ m,\ c}$。此外，设备 n 在时隙 b 通过信道 c 到基站所产生的相应瑞利衰落分量表示为 $g_n^{(b)}{}_{,\ m,\ c}$。我们注意到，这些分量会降低设备 m向基站发送的信号质量，其中 $m \in \mathcal{N}$， $m \neq n$ 和 m 使用与 n 相同的信道 c 向基站传输模型参数。因此，设备 n 在时隙 b对设备 n 造成的干扰链路对应的信道增益写作

$$h_{n,m,c}^{(b)} = \chi_{n,m,c}\beta_{n,m,c}|g_{n,m,c}^{(b)}|. \tag{15}$$

D. 帧结构与信号模型

在我们的框架中，鉴于量子设备进行本地模型训练具有异步性，即任意量子设备都可能在某个全局轮次内的任意时刻完成本地模型训练，随后将本地模型参数传输至服务器，因此假设每个全局轮次 k 被划分为 b 个时隙，以便进行信道建模。在所考虑的网络中，对于任何从设备 n 出发、以基站为终点的链路所采用的帧结构，每帧都由预处理阶段和传输阶段组成。在预处理阶段，每台设备都会收集必要的信息，以便为后续传输选择信道和发射功率。作出选择后，每台设备会在传输阶段将更新后的本地模型参数发送至基站。然而，值得一提的是，设备在时隙 b仍可能使用相同的信道传输模型参数，从而可能造成同信道干扰。这是因为正交导频主要用于确保对信道状态信息进行准确估计，但无法消除两个或更多设备使用相同信道同时传输时发生信号重叠的可能性。

我们使用二元变量 $\zeta_n^{(b)}{}_{,c}$ 来表示设备n 是否在时隙 b 期间于信道 c 上传输；若传输则为 $\zeta_n^{(b)}{}_{,c} = 1$，否则为 $\zeta_n^{(b)}{}_{,c} = 0$。每个设备在一个时隙内最多允许选择一个信道，从而使其信道选择之和不超过一。因此，我们有

$$\sum_{c=1}^{C} \zeta_{n,c}^{(b)} \leq 1, \forall n \in \mathcal{N}. \tag{16}$$

我们指出，（16）中的约束在前沿文献 [33]中被广泛采用。我们将设备 n 在时隙 b 期间的发射功率记为 $p_n^{(b)}$，则信道 c上相应的接收信干噪比（信号与干扰加噪声比）可表示为

$$\gamma_{n,c}^{(b)} = \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} p_n^{(b)}}{\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} p_m^{(b)}}, \tag{17}$$

其中 $h_n^{(b)}{}_{,c}$ 表示设备 n在时隙b期间于信道 c 上到基站的合法链路的信道增益， σ^2 为加性高斯白噪声（AWGN）功率。因此，若我们将带宽记为 B，则设备 n 在时隙 b期间于信道 c上可实现的传输速率计算为

$$r_{n,c}^{(b)} = B\log_2(1+\gamma_{n,c}^{(b)}). \tag{18}$$

在本文中，“无线”特指一种面向 6G 的经典上行链路模型，其中量子设备通过多个专用信道与基站通信。该系统采用多信道非正交多址结构，这意味着多个设备可以同时通过同一信道传输 [7], ，从而产生同信道干扰，决定可实现的信号与干扰加噪声比和数据速率。如前所述，所有交换的信息均为经典的参数化量子电路模型参数[4], ，而该框架的量子特性仅体现在本地基于参数化量子电路的量子神经网络训练，以及基站侧基于量子近似优化算法的优化中。

III. QFL收敛行为分析

本节将全面分析所提出的QFL框架的收敛性，考察在全设备参与的场景下，非凸损失函数、数据分布多样性以及量子采样噪声等现实因素的影响。我们的研究结果表明，收敛率受迭代总次数、设备总数、每次全局迭代中参与的设备子集，以及参数化量子电路的测量次数影响。

本小节详细介绍支持全设备参与的量子联邦学习算法，阐述高效本地训练和全局聚合的核心机制。算法1展示了全设备参与模式下量子联邦学习的实施步骤：服务器首先向所有参与设备广播初始全局模型。每个具备量子计算能力的设备 $n \in \mathcal{N}$接收全局模型，并开始使用

量子随机梯度下降在其私有数据集上进行训练，每个全局轮次执行 T_n次迭代。完成本地迭代后，设备根据学习到的参数更新模型。随后，设备将经典模型参数发送回服务器。服务器使用联邦平均算法聚合收到的参数，从而更新并优化全局模型，以便在下一轮分发。此过程在K 个全局轮次中重复，直至模型收敛。算法1采用全设备参与模式，因为服务器在每个全局轮次都会聚合所有设备更新后的参数。

算法1 全设备参与的量子联邦学习算法

输入： 学习率 η_k， $\boldsymbol{\theta}_0$ 作为由所有设备 ，$\mathbb{T}$ 作为所有设备的本地迭代索引集合
对于 $k = 1, 2, ..., K$ **执行：**
  基站调用算法 3以执行联合资源
  分配，从而得到最优信道分配和
  功率级别： $\zeta^{b*}, x^{(b)*}$
  对所有设备 $n \in \mathcal{N}$ **并行执行：**
    对 $t = 1, 2, ... T_n$ **，其中** $\mathcal{T}_n \in \mathbb{T}$，**执行：**
      $\boldsymbol{\theta}_{n,k}^{t+1} = \boldsymbol{\theta}_{n,k}^{t} - \eta_k \tilde{g}_{n,k}^{t}$
    每个设备 $n \in \mathcal{N}$ 发送 $\boldsymbol{\theta}_{n,k}^{T_n}$ 发送至服务器
    结束并行循环
  服务器计算
    $\bar{\boldsymbol{\theta}}_{k+1} = \frac{1}{N} \sum_{n \in \mathcal{N}} [\boldsymbol{\theta}_{n,k}^{T_n}]$
  服务器向所有设备广播 $\bar{\boldsymbol{\theta}}_{k+1}$
  结束循环
结束循环
输出： $\bar{\boldsymbol{\theta}}_K = \frac{1}{N} \sum_{n \in \mathcal{N}} \boldsymbol{\theta}_n^{T_n,\ K-1}$

A. 符号与定义

我们关注以下优化问题：$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \triangleq \sum_{n=1}^{N} f_n(\boldsymbol{\theta})$，其中 $f(\boldsymbol{\theta})$ 为全局目标函数。在此联邦框架中，假设每个设备 n 都在数据集 $\mathcal{S}_n$ 上训练其本地模型，该数据集包含 S_n 个数据点，这些数据点以独立同分布（i.i.d.）的方式从本地分布 $\mathcal{D}_n$ 中采样。我们注意到，由于本地数据集来自不同分布，因此在分析量子联邦学习的收敛性时，我们会仔细考虑这些分布的异质性。我们定义 $g_n = \frac{1}{|\mathcal{S}_n|} \nabla f_n(\boldsymbol{\theta}) \triangleq \frac{1}{|\mathcal{S}_n|} \nabla f(\boldsymbol{\theta}; \mathcal{S}_n)$，其中 $f(\boldsymbol{\theta}; \mathcal{S}_n)$ 表示全梯度。此外，我们将随机梯度记为 $\tilde{g_n} \triangleq \frac{1}{B} \nabla f(\boldsymbol{\theta}; \xi_n)$，其中 $\xi_n \subseteq \mathcal{S}_n$ 是均匀采样的小批量，大小为 $|\xi_n| = B$。在第 k^{th} 个全局轮次的第 t 次局部迭代中，于设备 n 的局部解 $\boldsymbol{\theta}_{n,\ k}^{t}$ 处计算得到的相应量，对于全梯度记为 $g_{n,\ k}^{t}$ ，对于随机梯度记为 $\tilde{g}_{n,\ k}^{t}$ 。我们还定义以下符号：
$\boldsymbol{\theta}_k^t = [\boldsymbol{\theta}_{1,k}^t, \boldsymbol{\theta}_{2,k}^t, \ldots, \boldsymbol{\theta}_{N,k}^t]$,

$$\xi_k^t = [\xi_{1,k}^t, \xi_{2,k}^t, \ldots, \xi_{N,k}^t], \tag{19}$$

用于表示设备在第 k^{th} 个全局轮次的本地迭代 t 期间对应的局部解集合和采样小批量。以下符号将有助于分析 QFL 框架的收敛性： $\bar{\boldsymbol{\theta}}_k^t \triangleq \frac{1}{N} \sum_{n \in \mathcal{N}} \boldsymbol{\theta}_{n,\ k}^t$，$\tilde{g}_k^t \triangleq \frac{1}{N} \sum_{n \in \mathcal{N}} \tilde{g}_{n,\ k}^t$，$g_k^t \triangleq \frac{1}{N} \sum_{n \in \mathcal{N}} g_{n,\ k}^t$. 因此，设备 n 的本地随机梯度下降更新如下：$\boldsymbol{\theta}_{n\ ,\ k}^{t+1} = \boldsymbol{\theta}_{n,\ k}^t - \eta_k \tilde{g}_{n,\ k}^t$ 显然

$$\bar{\boldsymbol{\theta}}_k^{t+1} = \bar{\boldsymbol{\theta}}_k^t - \eta_k \tilde{g}_k^t. \tag{20}$$

需要说明的是， $\mathbb{E}\tilde{g}_k^t = g_k^t$，其中 $\mathbb{E}$ 表示函数的期望。在后续分析中，我们假设 λ 是局部目标函数之间梯度多样性的上界，即 $\frac{\sum_{n=1}^{N} ||g_{n\ k}^t||_2^2}{||\sum_{n=1}^{N} g_n^t}_{\ ,\ ,\ k}||_2^2 \leq \lambda$. 在下一小节中，我们将阐述收敛性分析所依据的基本假设。

B. 假设

为说明我们在收敛分析中所采用假设的背景，首先概述采用这些假设的基本直觉。 L –光滑性条件确保局部目标函数的梯度不会变化得过于剧烈，从而稳定异构设备 [34]上的更新。Polyak-Lojasiewicz（ PL）条件是强凸性的一个现实替代方案。在分析变分量子算法时，人们广泛采用这一假设；此类算法中的损失景观虽为非凸 [35], ，但在实践中通常满足 PL 性质。最后，有界方差假设考虑了小批量采样所引入的随机性，而这种随机性是基于参数化量子电路的训练 [34]固有的。总之，这些假设为在量子联邦学习等框架中建立收敛保证提供了标准且有充分依据的基础。

假设1 (光滑性与下界有界性[34]). 与设备 n相关联的局部目标函数 $f_n(.)$ 对 $1 \leq n \leq N$ 可微，且为 $L-$光滑的，即 $||\nabla f_n(\mathbf{u}) - \nabla f_n(\mathbf{v})|| \leq L||\mathbf{u} - \mathbf{v}||, \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d$。

假设2 (μ-Polyak-Lojasiewicz（PL）[35]). 全局目标函数 $f(.)$ 是可微的，并以常数 μ满足*Polyak-Lojasiewicz*（*PL*）条件，即$\frac{1}{2}||\nabla f(\boldsymbol{\theta})||_2^2 \geq \mu(f(\boldsymbol{\theta}) - f(\boldsymbol{\theta}^*))$成立 $\forall \boldsymbol{\theta} \in \mathbb{R}^d$，其中 $\boldsymbol{\theta}^*$为全局目标函数的最优解。

假设3 (有界局部方差 [34]). 对于每个局部数据集 S_n，$n = 1, 2, ..., N$，我们可以采样一个独立小批量 $\xi_n \subseteq \mathcal{S}_n$ ，且 $|\xi_n| = B$ ，并计算无偏随机梯度 $\tilde{g_n} = \frac{1}{B} \nabla f(\boldsymbol{\theta}; \xi_n)$，$\mathbb{E}[\tilde{g_n}] = g_n = \frac{1}{|\mathcal{S}_n|} \nabla f(\boldsymbol{\theta}; \mathcal{S}_n)$，其方差有界为 $\mathbb{E}[||\tilde{g_n} - g_n||^2] \leq C_1 ||g_n||^2 + \frac{\sigma^2}{B}.$，其中 C_1为非负常数且与小批量大小成反比， σ 为控制方差界的另一个常数。

根据式 (48) 中的更新规则以及目标函数满足 L-光滑性的假设，我们得到以下不等式：
$f\ \bar{\boldsymbol{\theta}}_k^{t+1} - f\ \bar{\boldsymbol{\theta}}_k^t \leq -\eta_k \langle \nabla f\ \bar{\boldsymbol{\theta}}_k^t\ \ \tilde{g}_k^t \rangle + (\ )(\ )(\ )$ 现在，$\frac{\eta_k^2 L}{2} ||\tilde{g}_k^t||^2$.
对不等式两边取期望，得到
$\mathbb{E}[f\ \bar{\boldsymbol{\theta}}_k^{t+1} - f\ \bar{\boldsymbol{\theta}}_k^t\ ] \leq (\ )(\ )(\ )$ 对所 $-\eta_k \mathbb{E}[\langle \nabla f\ \bar{\boldsymbol{\theta}}_k^t\ \ \tilde{g}_k^t \rangle] + \frac{\eta_k^2 L}{2} \mathbb{E}[||\tilde{g}_k^t||^2]$
有本地迭代和全局迭代取平均后，我们得到

$$\begin{aligned} \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \mathbb{E}[f(\bar{\boldsymbol{\theta}}_k^{t+1}) - f(\bar{\boldsymbol{\theta}}_k^t)] \\ \leq \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} (-\eta_k \mathbb{E}[\langle \nabla f(\bar{\boldsymbol{\theta}}_k^t), \tilde{g}_k^t \rangle]) \\ + \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \frac{\eta_k^2 L}{2} \mathbb{E}[||\tilde{g}_k^t||^2]. \end{aligned} \tag{21}$$

接下来，我们系统地确定式 (53) 右侧各项的界。具体而言，我们利用引理 1 确定该式第一项的界。随后，利用引理 3 推导第二项的界。此外，引理 2 关注的是源自引理 1 分析的一项，尤其是引理 1 中列出的最后一项；引理 2 为其提供了界，从而进一步加深我们对整个方程动态特性的理解。

C. 收敛速率

本节分析全设备参与场景下量子联邦学习算法的收敛率。接下来，我们将给出若干用于推导主要结果的引理。

在给出引理 1 之前，我们注意到，假设1，即 $L-$ 光滑性，确保局部目标函数的梯度不会突然变化，这对于界定随机梯度与全梯度之间的相互作用至关重要。下面引理中所建立的不等式直接得益于这一光滑性性质。

引理 1. 设假设*1*成立，则随机梯度与全梯度的期望内积有如下界：

$$
\begin{aligned}
&-\eta_k \mathbb{E}\left[\langle \nabla f(\bar{\boldsymbol{\theta}}_k^t), \tilde{g}_k^t \rangle\right] \leq -\frac{\eta_k}{2}||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||^2 \\
&-\frac{\eta_k}{2}||\sum_{n=1}^{N} \nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2 + \frac{\eta_k L^2}{2}\sum_{n=1}^{N}||\bar{\boldsymbol{\theta}}_k^t - \boldsymbol{\theta}_{n,k}^t||^2.
\end{aligned} \tag{22}
$$

证明. 参见附录 A。

引理 2 依赖于假设3，该假设对小批量采样产生的随机梯度的方差进行了界定。这一条件对于控制局部迭代偏离其平均模型的程度至关重要，由此可得到下面给出的界。

引理 2. 在假设*3*成立的前提下，$\boldsymbol{\theta}_{n,\ k}^t$ 的发散的期望上界给出如下

$$
\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{1}{T_n}\sum_{t=1}^{T_n}\left[\mathbb{E}||\bar{\boldsymbol{\theta}}_k^t - \boldsymbol{\theta}_{n,k}^t||\right] \\
&\leq \frac{(2C_1 + T_n(T_n+1))}{KT}\eta_k^2\frac{N+1}{N}\frac{1}{K}\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{1}{T_n} \\
&\times \sum_{t=1}^{T_n}||g_{n,k}^t||^2 + \frac{\eta_k^2(N+1)(T_n+1)\sigma^2}{NB} \\
&\leq \frac{\lambda\eta_K^2(2C_1 + T_n(T_n+1))}{KT}\frac{N+1}{N}\frac{1}{K}\sum_{k=1}^{K}\sum_{n=1}^{N} \\
&\times \frac{1}{T_n}\sum_{t=1}^{T_n}||g_{n,k}^t||^2 + \frac{\eta_k^2 K T_n(N+1)(T_n+1)\sigma^2}{NB}.
\end{aligned} \tag{23}
$$

证明。 见附录中的B。

引理3同样依赖于假设3，因为有界方差条件使我们能够对随机梯度的二阶矩进行上界估计。该假设确保随机采样引入的随机性不会在局部更新过程中失控地累积。

引理3. 在假设*3*下， $\mathbb{E}[||\tilde{g}_k^t||^2]$的期望上界表示为

$$
\begin{aligned}
\mathbb{E}\left[||\tilde{g}_k^t||^2\right] &\leq \left(\frac{C_1}{N}+1\right)\left[\sum_{n=1}^{N}||\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\right] + \frac{\sigma^2}{NB} \\
&\leq \lambda\left(\frac{C_1}{N}+1\right)\left[\sum_{n=1}^{N}||\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\right] + \frac{\sigma^2}{NB}.
\end{aligned} \tag{24}
$$

证明。 见附录中的C。

引理4利用量子测量噪声的结构，将梯度估计器的方差与在参数化量子电路上执行的测量次数联系起来。该刻画与变分量子算法中使用的标准统计模型一致。

引理4. 梯度估计方差可被上界界定为

$$
var(\xi_k^t) \leq \frac{1}{N}\sum_{n\in\mathcal{N}}\frac{\nu N_z D Tr(Z^2)}{2H}. \tag{25}
$$

证明。 参见附录 D。

定理 ***1***。 设假设*1*、*2*、*3*成立，则考虑全设备参与时，经过 K 个全局轮次后，全局模型训练收敛率的上界满足

$$
\begin{aligned}
&\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\mathbb{E}||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||^2 \leq \frac{2[f(\bar{\boldsymbol{\theta}}_1^0) - f^*]}{\eta_k K\frac{1}{N}\sum_{n=1}^{N}T_n} \\
&+ \frac{L\eta\sigma^2}{NB} + \frac{2\eta_k^2\sigma^2L^2(\frac{1}{N}\sum_{n=1}^{N}T_n+1)}{B}\left(1+\frac{1}{N}\right) \\
&+ \frac{1}{N}\sum_{n\in\mathcal{N}}\frac{\nu N_z D Tr(Z^2)}{2H}.
\end{aligned} \tag{26}
$$

证明。 参见附录 E。

注 1。 定理 *1* 在全设备参与情形下得到的收敛上界表明，量子联邦学习算法的收敛性显著受迭代总次数、设备总数以及参数化量子电路上的测量次数影响。

注2. 由(26) 可以看出，在全设备参与场景下增加设备数量 N 能够提升量子联邦学习的收敛性能。此外，提高参数化量子电路的量子测量次数也有助于实现更快的量子联邦学习收敛。然而，该方程的最后一项包含梯度估计方差的求和。具体而言，这反映了与量子采样噪声相关的噪声方差。因此，随着更多设备参与量子联邦学习，来自所有设备的累积噪声会在每个全局轮次中加剧。因此，同时增加

适度增加设备数量和测量次数被证明对提升收敛性能更为有效。我们将在第六节中进行大量仿真，以进一步研究这一现象。

D. 复杂度分析

现在，我们研究量子训练过程的复杂度分析。假设假设1成立，即局部损失函数 $f_n(\cdot)$ 是 $L-$ 光滑的，也就是说，它具有利普希茨连续梯度；在此条件下，我们得到以下不等式：

$$\nabla f_n\ \boldsymbol{\theta}'\ ^T\ \boldsymbol{\theta}-\boldsymbol{w}'\ +\frac{L}{2}||\boldsymbol{\theta}-\boldsymbol{w}'||^2 \qquad f_n\ \boldsymbol{\theta}\ \leq f_n\ \boldsymbol{w}'\ + \qquad \boldsymbol{\theta}\ \boldsymbol{w}'\ \in\ \Theta$$

( ) ( ) ( ) ( )，对所有，。此外，假设2成立，该假设施加了非凸优化中常用的 μ-PL 条件，以确保函数值下降；存在常数 $\mu>0$ 和 $\mu\leq L$，使得不等式 $||\nabla f_n(\boldsymbol{\theta})||^2\geq 2\mu(f_n(\boldsymbol{\theta})-f_n{}^*)$ 对所有 $\boldsymbol{\theta}\in\Theta$ 均成立。实践中，常数 μ 取决于量子比特数 n 和门参数数量 P，因为规模较大的电路通常具有较小的梯度范数。这种相关性反映了量子电路的复杂度如何影响其梯度的行为：参数更多、规模更大的系统往往表现出更细微且更弱的梯度幅值。下面的定理给出了经过 T_n 次迭代后局部量子训练输出 $\boldsymbol{\theta}^{T_n}$ 的最优性界。该定理的证明源自随机梯度下降（SGD）的经典收敛性分析，详见（[36], 定理6）。在假设 L-光滑且满足 μ-PL 条件的情况下，对于任意给定的初始点 $\boldsymbol{\theta}^0$，任意固定学习率 $\eta_k=\mu\leq\frac{1}{L}$ 下，以下界均成立：$\mathbb{E}[f_n(\boldsymbol{\theta}^{T_n})]-L^*\leq(1-\eta\mu)^{T_n}([f_n(\boldsymbol{\theta}^0)]-L^*)+\frac{1}{2}[\frac{\eta LV}{\mu}]$，其中 $V=\frac{\nu N_z D\mathcal{T}r}{(z^2)\,2H}$ 是梯度估计方差的上界，期望是对测量输出的分布取的。此外，给定某个目标误差水平 $\delta>0$，对于学习率 $\eta=\eta^{\text{shot-noise}}\leq\min\{\frac{1}{L},\frac{\delta\mu}{LV}\}$，迭代次数可表示为 $-$ —

$$T_n^{\text{shot-noise}}=\mathcal{O}\left(\log\frac{1}{\delta}+\frac{V}{\delta\mu}\right)\frac{L}{\mu},\tag{27}$$

足以确保误差为 $\mathbb{E}[f_n(\boldsymbol{\theta}^T)-f_n{}^*]=\mathcal{O}(\delta)$。通过将学习率 η 设置得足够小，可以使该误差任意小，从而保证在不超过 $T_n^{\text{shot-noise}}$次迭代内收敛。这种策略性调整通过稳定训练过程降低计算开销，避免过多迭代对系统资源造成负担。

IV. 和速率最大化问题建模

A. 问题陈述

在本研究中，我们旨在通过联合优化每个时隙内量子设备的信道选择和发射功率，最大化所考虑网络的和速率。最大化上行链路和速率有助于缩短设备向基站发送本地训练的PQC参数所需的总传输时延，而这通常是多信道NOMA无线量子联邦学习系统的主要瓶颈。通过确保模型上传更快、更可靠，系统能够实现更好的同步，并实现更及时的全局聚合。这反过来又能加快整个量子联邦学习训练过程的收敛速度，并提升其可扩展性。因此，我们将问题表述为问题*1*：

———————

$$\max_{\zeta^{(b)},p^{(b)}}\quad\sum_{n=1}^{N}\sum_{c=1}^{C}r_{n,c}^{(b)}\tag{28a}$$

$$\text{s.t.}\quad 0<p_n^{(b)}\leq P^{\max},\quad\forall n\in\mathcal{N},\tag{28b}$$

$$\zeta_{n,c}^{(b)}\in\{0,1\},\quad\forall n\in\mathcal{N},\forall c\in\mathcal{C},\tag{28c}$$

$$\sum_{c=1}^{C}\zeta_{n,c}^{(b)}\leq 1,\quad\forall n\in\mathcal{N},\tag{28d}$$

其中

$\zeta^{(b)}=\{\zeta_{1,1}^{(b)},\zeta_{1,2}^{(b)},\ldots,\zeta_{N,C}^{(b)}\}\in\mathbb{R}_{1\times NC}$ $p^{(b)}=$ 且。$\{p_1^{(b)},p_2^{(b)},\ldots,p_N^{(b)}\}\in\mathbb{R}_{1\times N}$ *1*，

在问题 约束 (28b) 中，设备 n 的发射功率必须保持为正，且不得超过最大限值 $P^{\max}$。约束 (28c) 表示信道选择为二元变量，而约束 (28d) 则源于以下假设：每个设备在一个时隙内最多只能选择一个信道。

B. 所构建问题的性质

问题*1*由于目标函数（28a）而具有非凸性质。此外，由于它包含离散变量和连续变量的组合，因此属于混合整数非线性规划问题（混合整数非线性规划）。此外，即使给定 $\zeta^{(b)}$，由于 $p^{(b)}$ [37] 的耦合，该问题仍是 NP难。由于其性质复杂，问题*1*极难求解，经典优化方法也难以处理。

V. 所提出的量子中心优化方案

在本研究中，我们提出一种以量子为核心的优化策略，利用量子近似优化算法（QAOA）求解问题*1*。首先，我们将经典混合整数非线性规划问题转化为二次无约束二进制优化（QUBO）框架。随后，将 QUBO 公式转换为适用于量子处理器执行的哈密顿量表达式。该 QAOA 方案尤其适用于大规模、多信道、基于非正交多址接入的量子联邦学习系统中遇到的离散高维资源分配问题。其能够高效探索组合解空间，从而实现快速且高质量的近似优化。尽管含噪声中等规模量子设备本身容易受到噪声影响，但通过使用足够数量的测量次数，可以有效缓解这一挑战，确保 QAOA 在真实的含噪量子硬件环境中仍然可靠且实用。为便于计算，我们将问题*1*划分为两个不同的子问题，并对每个子问题应用上述步骤。然后，我们采用块坐标下降（BCD）技术，求解问题*1*的原始形式。

子问题 *1*：

$$\max_{\zeta^{(b)}} \quad \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} \tag{29a}$$

$$\text{s.t.} \quad \zeta_{n,c}^{(b)} \in \{0,1\}, \quad \forall n \in \mathcal{N}, \forall c \in \mathcal{C}, \tag{29b}$$

$$\sum_{c=1}^{C} \zeta_{n,c}^{(b)} \leq 1, \quad \forall n \in \mathcal{N}. \tag{29c}$$

子问题 *2*：

$$\max_{p^{(b)}} \quad \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} \tag{30a}$$

$$\text{s.t.} \quad 0 < p_n^{(b)} \leq P^{\max}, \quad \forall n \in \mathcal{N}. \tag{30b}$$

A. 子问题 *1*

将经典问题转换为 *QUBO* 公式：为了将经典问题转换为 QUBO 公式，我们引入惩罚参数λ_0 和 λ_1。利用这些参数，子问题 *1* 可改写如下：

$$\max_{\zeta^{(b)}} \left[\lambda_0 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} + \lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}\zeta_{n,c}^{(b)} - 1\right)^2\right] \tag{31a}$$

在 (31a) 中，参数 λ_0 确保在目标函数的优化过程中有效考虑量子设备的信道选择，而参数λ_1 则用于施加信道选择约束。值得注意的是，要通过实现量子近似优化算法来求解问题，必须将其表示为特定的二次形式，即

$$\min_{\zeta^{(b)}} {\zeta^{(b)}}^T Q \zeta^{(b)}. \tag{32}$$

根据 (32)，量子近似优化算法通常旨在最小化特定的代价函数。代价函数的构造方式使其最小值对应于该问题的最优解。然而，本研究旨在最大化 QFL 框架的和速率。因此，我们通过对原始函数取负，将最大化问题转换为最小化问题，如下所示：

$$\min_{\zeta^{(b)}} -\left[\lambda_0 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} + \lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}\zeta_{n,c}^{(b)} - 1\right)^2\right] \tag{33a}$$

or,

$$\min_{\zeta^{(b)}} \Bigg[\underbrace{-\lambda_0 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)}}_{\text{(a)}} \underbrace{-\lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}\zeta_{n,c}^{(b)} - 1\right)^2}_{\text{(b)}}\Bigg] \tag{34a}$$

我们现在基于(33a)中的问题表述构建Q矩阵。

构造对应的Q矩阵：二次无约束二进制优化问题的Q矩阵是对相应目标函数的表示。因此，对于子问题 *1*，Q矩阵体现了我们所研究的量子联邦学习网络中用于优化设备信道选择的目标函数。为了构造Q矩阵，我们将式（33a）中的每一项展开如下：

(a):

$$\begin{aligned}
&-\lambda_0 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} \\
&= -\lambda_0 \sum_{n=1}^{N}\sum_{c=1}^{C} B\log_2(1+\gamma_{n,c}^{(b)}) \\
&= -\lambda_0 B \sum_{n=1}^{N}\sum_{c=1}^{C} \log_2\left(1 + \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} p_n^{(b)}}{\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} p_m^{(b)}}\right).
\end{aligned} \tag{35}$$

利用泰勒展开，我们将式（35）写为：

$$-\lambda_0 B \sum_{n=1}^{N}\sum_{c=1}^{C} \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} p_n^{(b)}}{\ln 2(\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} p_m^{(b)})}. \tag{36}$$

此处，采用一阶泰勒近似已足够，因为大规模多信道非正交多址接入上行链路系统通常由于多个设备共享同一信道而产生的同信道干扰，工作在低信号与干扰加噪声比条件下。该工作区间可确保近似保持准确，同时使面向量子近似优化算法的基于二次无约束二进制优化的表述易于处理。此外，该线性化技术已在先前的无线优化工作中被广泛采用（例如， [38]–[40]），以在类似的干扰受限场景下简化对数速率表达式。

(b):

$$\begin{aligned}
&-\lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}\zeta_{n,c}^{(b)} - 1\right)^2 \\
&= -\lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}{\zeta_{n,c}^{(b)}}^2 - 2\zeta_{n,c}^{(b)} + 1\right).
\end{aligned} \tag{37}$$

给定二元变量$\zeta_{n,c}^{b}$，其中 N 表示设备数量， C 为信道数量，矩阵元素确定如下：

由$Q_{(n,c),(n,c)}$表示的对角元素，代表与各个信道选择决策相关的线性项。它们的计算方式为

$$\begin{aligned}
Q_{(n,c),(n,c)} = &-\lambda_0 B \frac{h_{n,c}^{(b)} p_n^{(b)}}{\ln 2(\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} p_m^{(b)})} \\
&+ 2\lambda_1,
\end{aligned} \tag{38}$$

其中，项 $-\lambda_0 B \frac{h_{n}^{(b)}{}_{,c} p_n^{(b)} \ln 2(\sigma^2 + \sum_{m\in\mathcal{N},}}{{}_{m\neq n} \zeta_{m}^{(b)}{}_{,c} h_{m,n,c}^{b} p_m^{(b)})}$ 施加与设备 n相关的信道增益、发射功率及其他噪声项有关的惩罚。此外，项 $2\lambda_1$ 对设备 n 的信道选择约束施加惩罚。

非对角元素记为 $Q_{(n,c),\ (x,y)}$，表示信道选择决策之间的二次相互作用。其计算方式如下：

$$Q_{(n,c),(x,y)} = \begin{cases} -\lambda_1, & \text{if } n = x \quad \text{and} \quad c \neq y, \\ 0, & \text{otherwise}, \end{cases} \tag{39}$$

其中，项 $-\lambda_1$ 是在同一设备 n 被分配到不同信道 c 和 y 时施加的可用性约束惩罚。该惩罚确保每台设备只分配到一个信道。所有其他非对角元素

均设为 0，因为它们代表彼此独立、互不影响的决策。

*QUBO*表述到哈密顿量表达式的转换：为将 QUBO 公式转换为哈密顿量表达式，我们将每个二元变量 $\zeta_{n,\,c}^{(b)} \in \{0,1\}$ 替换为自旋变量 $z_{n,\,c}^{(b)} \in \{-1,1\}$，其中

$$\zeta_{n,c}^{(b)} = \frac{1 - z_{n,c}^{(b)}}{2}. \tag{40}$$

现在需要将 (40) 代入 (34a)。为此，我们将 (40) 代入 (36) 和 (37)。将 (40) 代入 (36)，得到：

$$-\lambda_0 B \sum_{n=1}^{N}\sum_{c=1}^{C} \frac{(1 - z_{n,c}^{(b)}) h_{n,c}^{(b)} p_n^{(b)}}{2\ln 2(\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n}(\frac{1-z_{m,c}^{(b)}}{2}) h_{m,n,c}^{b} p_m^{(b)})}. \tag{41}$$

展开 (41) 中的表达式，可得由常数项和关于 $z_{n,\,c}^{(b)}$ 的线性项构成的组合。同样，将 (40) 代入 (37)，得到：

$$-\lambda_1 \sum_{n=1}^{N}\left(\sum_{c=1}^{C}\left(\frac{1 - z_{n,c}^{(b)}}{2}\right)^2 - 2\left(\frac{1 - z_{n,c}^{(b)}}{2}\right) + 1\right). \tag{42}$$

展开 (42) 中的表达式，可得由常数项、关于 $z_{n,\,c}^{(b)}$ 的线性项，以及关于 $z_{n,\,c}^{(b)} z_{n,\,e}^{(b)}$ 的二次项构成的组合。展开所有项后，接着将其重新整理为哈密顿量形式，其中每个自旋变量 $z_{n,\,c}^{(b)}$ 均由相应的泡利 Z 算符表示。具体而言，每个涉及 $z_{n,\,c}^{(b)}$ 的线性项都映射为一个独立的 $Z_{n,\,c}^{(b)}$ 算符；每个涉及乘积（如 $z_{n,\,c}^{(b)} z_{n,\,e}^{(b)}$）的二次项则表示为两个泡利 Z 算符的乘积 $Z_{n,\,c}^{(b)} Z_{n,\,e}^{(b)}$。因此，哈密顿量可表示为：

$$H_C = \sum_i I_i Z_i + \sum_{i<j} J_{ij} Z_i Z_j, \tag{43}$$

其中，I_i 表示由 $z_n^{(b),c}$得出的每个线性 Z_i 项对应的系数，而 J_{ij} 则是每个二次 $Z_i Z_j$ 项对应的系数，这些系数源自 $z_n^{(b),c}$ 与 $z_n^{(b),e}$之间的二次相互作用。

B. 子问题 *2*

我们注意到，量子近似优化算法专为优化二元变量而设计。然而，在子问题 *2*中，我们要处理发射功率问题，而发射功率必须转换为二元变量，才能与量子近似优化算法兼容。因此，我们将$p_n^{(b)}$转换为二元变量 $x_n^{(b)}$，如下所示：

$$p_n^{(b)} = \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}, \tag{44}$$

其中，q 是将发射功率的十进制值转换为二进制所需的比特数。我们将与发射功率相关的二元变量集合记为 $x^{(b)} = \{x_{1,0}^{(b)}, x_{1,1}^{(b)}, \ldots, x_{1,q-1}^{(b)}, x_{2,0}^{(b)}, \ldots, x_{N,q-1}^{(b)}\}$，将所需比特集合记为 $\iota = \{1, 2, \ldots, q-1\}$。子问题 *2* 现在可表示为：

子问题 *2*（等价形式）：

$$\max_{x^{(b)}} \quad \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} \tag{45a}$$

$$\text{s.t.} \quad x_{n,i}^{(b)} \in \{0,1\}, \quad \forall n \in \mathcal{N}, \forall i \in \iota, \tag{45b}$$

$$\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} > 0, \quad \forall n \in \mathcal{N}, \tag{45c}$$

$$\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} \leq P^{\max}, \quad \forall n \in \mathcal{N}, \tag{45d}$$

where $r_{n,c}^{(b)} = B\log_2(1 + \gamma_{n,c}^{(b)})$, and $\gamma_{n,c}^{(b)} = \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}}{\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i x_{m,i}^{(b)}}$.

T 将经典问题转换为 *QUBO* 公式：现在，我们通过引入惩罚参数 λ_2、 λ_3和λ_4，将经典问题转换为 QUBO 公式。因此，子问题 *2* 的等价形式重新构造如下：

$$\max_{x^{(b)}} \left[\lambda_2 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} + \lambda_3 \sum_{n=1}^{N}\left(0 - \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}\right)^2 + \lambda_4 \sum_{n=1}^{N}\left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} - P^{\max}\right)^2\right].$$

We adjust our maximization problem to a minimization format by negating the original function as follows: $\min_{x^{(b)}} -\left[\lambda_2 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} + \lambda_3 \sum_{n=1}^{N}\left(0 - \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}\right)^2 + \lambda_4 \sum_{n=1}^{N}\left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} - P^{\max}\right)^2\right]$, or,

$$\min_{x^{(b)}} \Bigg[\underbrace{-\lambda_2 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)}}_{\text{(a)}} \underbrace{+\lambda_3 \sum_{n=1}^{N}\left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}\right)^2}_{\text{(b)}} \underbrace{-\lambda_4 \sum_{n=1}^{N}\left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} - P^{\max}\right)^2}_{\text{(c)}}\Bigg]. \tag{46a}$$

现在，我们根据式（46a）中给出的问题表述构建Q矩阵。

相应*Q*矩阵的构建：对于子问题 *2*，Q矩阵表示目标函数，用于优化QFL框架中设备的发射功率。为构建Q矩阵，我们将式（33a）中的每一项展开如下：

(a):

$$-\lambda_2 \sum_{n=1}^{N}\sum_{c=1}^{C} r_{n,c}^{(b)} = -\lambda_2 \sum_{n=1}^{N}\sum_{c=1}^{C} B\log_2(1 + \gamma_{n,c}^{(b)})$$
$$= -\lambda_2 B \sum_{n=1}^{N}\sum_{c=1}^{C} \log_2\left(1 + \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}}{\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i x_{m,i}^{(b)}}\right). \tag{47}$$

利用泰勒展开，我们将式（47）写为：

$$-\lambda_2 B \sum_{n=1}^{N}\sum_{c=1}^{C} \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}}{\ln 2(\sigma^2 + \sum_{m\in\mathcal{N}, m\neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i x_{m,i}^{(b)})}. \tag{48}$$

(b):

$$\lambda_3 \sum_{n=1}^{N} \left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} \right)^2 = \lambda_3 \sum_{n=1}^{N} \sum_{i=0}^{q-1} \left(2^i \right)^2 \left(x_{n,i}^{(b)} \right)^2. \tag{49}$$

(c):

$$\begin{aligned} &- \lambda_4 \sum_{n=1}^{N} \left(\sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} - P^{\max} \right)^2 \\ &= -\lambda_4 \sum_{n=1}^{N} \left(\sum_{i=0}^{q-1} \left(2^i \right)^2 \left(x_{n,i}^{(b)} \right)^2 - 2 \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} P^{\max} \right. \\ &\qquad \left. + \left(P^{\max} \right)^2 \right). \end{aligned} \tag{50}$$

考虑二元变量 $x_{n,\ i}^{b}$，其中 N 表示设备数量，而（$q-1$）表示以二进制表示设备发射功率所需的比特数，矩阵元素确定如下：

我们将对角元素表示为$Q_{(n,i),\ (n,i)}$，其描述了与各个发射功率相关的线性项。它们计算如下

$$\begin{aligned} &Q_{(n,i),(n,i)} \\ &= -\lambda_2 B \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}}{\ln 2 (\sigma^2 + \sum_{m \in \mathcal{N}, m \neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i x_{m,i}^{(b)})}, \\ &\qquad - 2 \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)} P^{\max}, \end{aligned} \tag{51}$$

其中，项 $-\lambda_2 B \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i x_{n,i}^{(b)}}{\ln 2 (\sigma^2 + \sum_{m \in \mathcal{N}, m \neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i x_{m,i}^{(b)})}$ 对设备 n 施加与信道选择、信道增益、发射功率及其他噪声项相关的惩罚。此外，项 $-2 \sum_{i=0}^{q-1} 2^i x_{n}^{(b)}$, $_i P^{\max}$对设备 n的发射功率约束施加惩罚。

非对角元素（记为 $Q_{(n,i),(x,y)}$）表示信道选择决策之间的二次相互作用。其计算方式如下

$$Q_{(n,i),(x,y)} = \begin{cases} (\lambda_3 - \lambda_4) \sum_{i=0}^{q-1} \left(2^i \right)^2, \text{if} & n = x, \\ & i \neq y, \\ 0, & \text{otherwise}, \end{cases}$$

其中，项 $(\lambda_3 - \lambda_4)$ 是针对设备 n 发射功率的约束惩罚项。所有其他非对角元素均设为 0，因为它们表示彼此独立、没有相互作用的决策。

*QUBO*表述到哈密顿量表达式的转换：要将 QUBO 公式转换为哈密顿量表达式，我们将每个二元变量 $x_{n}^{(b)}$, $_i \in \{0,1\}$ 替换为自旋变量 $z_{n}^{(b)}$ $_{,i} \in \{-1,1\}$，其中

$$x_{n,i}^{(b)} = \frac{1 - z_{n,i}^{(b)}}{2}. \tag{52}$$

接着，将 (52) 代入 (46a)。为便于计算，我们将 (48)、(49) 和 (50) 中的 (52) 替换掉。将 (52) 代入 (48)，得到：

$$-\lambda_2 B \sum_{n=1}^{N} \sum_{c=1}^{C} \frac{\zeta_{n,c}^{(b)} h_{n,c}^{(b)} \sum_{i=0}^{q-1} 2^i (1 - z_{n,i}^{(b)})}{2 \ln 2 (\sigma^2 + \sum_{m \in \mathcal{N}, m \neq n} \zeta_{m,c}^{(b)} h_{m,n,c}^{b} \sum_{i=0}^{q-1} 2^i \frac{1 - z_{n,i}^{(b)}}{2})}. \tag{53}$$

展开 (53) 中的表达式，可得到常数项和关于 $z_{n\ i}^{(b)}$、的线性项的组合。同样，将 (52) 代入 (49)，得到：

$$\lambda_3 \sum_{n=1}^{N} \sum_{i=0}^{q-1} \left(2^i \right)^2 \left(\frac{1 - z_{n,i}^{(b)}}{2} \right)^2. \tag{54}$$

展开式 (54) 后，可得到常数项、关于 $z_{n\ i}^{(b)}$, 的线性项，以及关于 $z_{n}^{(b)}$ $_{,i} z_{n}^{(b)}$ $_{,f}$ 的二次项。同样，将式 (52) 代入式 (50)，可得：

$$\begin{aligned} &- \lambda_4 \sum_{n=1}^{N} \left(\sum_{i=0}^{q-1} \left(2^i \right)^2 \left(\frac{1 - z_{n,i}^{(b)}}{2} \right)^2 - 2 \sum_{i=0}^{q-1} 2^i \left(\frac{1 - z_{n,i}^{(b)}}{2} \right) P^{\max} \right. \\ &\qquad \left. + \left(P^{\max} \right)^2 \right). \end{aligned} \tag{55}$$

展开式 (55) 后，可得到常数项、由 $z_{n\ i}^{(b)}$, 表示的线性项，以及涉及 $z_{n}^{(b)}$ $_{,i} z_{n}^{(b)}$ $_{,f}$ 的二次项。因此，将所有项展开后，可将其整理为哈密顿量形式，其中每个自旋变量 $z_{n\ i}^{(b)}$, 都由相应的泡利 Z 算符表示。具体而言，每个涉及 $z_{n\ i}^{(b)}$, 的线性项都映射为单个 $Z_{n\ i}^{(b)}$,算符，而涉及 $z_{n}^{(b)}$ $_{,i} z_{n}^{(b)}$ $_{,f}$ 等乘积的每个二次项，则表示为两个泡利 Z 算符 $Z_{n}^{(b)}$ $_{,c} Z_{n}^{(b)}$ $_{,f}$ 的乘积。因此，哈密顿量表示为：

$$H_C = \sum_{\rho} I_\rho Z_\rho + \sum_{\rho < \phi} J_{\rho\phi} Z_\rho Z_\phi, \tag{56}$$

其中，I_ρ 表示由 $z_{n\ i}^{(b)}$、导出的各线性 Z_ρ 项对应的系数，而 $J_{\rho\phi}$ 表示各二次 $Z_\rho Z_\phi$ 项的系数，这些系数源于 $z_{n\ i}^{(b)}$、与 $z_{n}^{(b)}$ $_{,f}$之间的二次相互作用。

C. 所提算法

基于上述推导，我们提出了一种以量子为核心的优化算法，旨在利用量子近似优化算法解决子问题 *1* 和 子问题 *2*，如算法 (2) 所示。如前所述，我们将每个子问题转换为 QUBO 公式，最终将其转化为哈密顿量表达式。该算法通过梯度下降优化器迭代优化拟设电路。最终解给出经过优化的二元变量矩阵（对于子问题 *1*，为信道选择矩阵 $\zeta^{(b)}$；对于 子问题 *2*，为发射功率矩阵 $x^{(b)}$），以及最小化后的目标值。因此，算法 2 给出了基于量子近似优化算法的量子优化子程序，该子程序作为求解器嵌入算法 3 中。它利用参数化量子电路高效地解决信道选择和功率分配子问题。

此外，我们采用 BCD技术以迭代方式求解原始问题*1*，具体如算法（3）所示。在每次迭代中，算法使用基于QAOA的算法2更新变量 $\zeta^{(b)}$ 和 $x^{(b)}$。该过程持续进行，直至收敛到目标值最小化的最优解。我们还将通过 *VI-E* 中给出的数值结果验证计算复杂度，以评估我们提出的基于*QAOA*的联合优化算法。

算法 2 基于量子近似优化算法的算法，用于求解子-问题*1* 和子问题 *2*1：**输入：** QUBO矩阵 Q、量子比特数量、QAOA层数 p、初始参数（β_0，γ_0）、量子后端及其他网络参数。2：设计一个包含 p 层的QAOA参数化试探电路，并使用参数向量（β_0, γ_0）进行初始化。3：初始化经典优化器，具体而言，初始化用于迭代调整参数的梯度下降优化器。4：**函数** OBJ(β, γ)

- 将参数（β，γ）绑定到试探电路。
- 在量子后端上执行电路，以计算期望值 $\langle\beta, \gamma|H_C|\beta, \gamma\rangle$。

- **返回**测得的期望值，作为优化目标。5：**结束函数**6：优化拟设参数：7：**重复**

- 使用目标函数 OBJ(β, γ) 上的梯度下降优化器更新（β，γ）。

- 跟踪目标函数值以观察收敛情况。8：**直到**收敛。9：**输出：** 优化后的二元变量矩阵（信道选择矩阵 $\zeta^{(b)}$，用于子问题 *1*；或发射功率矩阵 $x^{(b)}$，用于子问题 *2*），以及最小化后的目标值。

算法 3 基于块坐标下降的联合优化算法

1：**输入：** 设置迭代索引 $i = 0$，并提供求解子问题*1* 和子问题 *2* 所需的输入，即相应的二次无约束二进制优化矩阵Q、量子比特数、QAOA层数 p、初始参数（β_0，γ_0）、其他网络参数以及量子后端；2：**重复**设置 $i \leftarrow i + 1$；使用算法 2 求解子问题 *1* 以更新 ζ^b；使用算法 2 求解子问题 *2* 以更新 $x^{(b)}$；3：**直到** 收敛。4：**输出：** 优化后的参数（ζ^{b*}，$x^{(b)*}$）以及最小化后的目标值。

VI. 仿真结果与评估

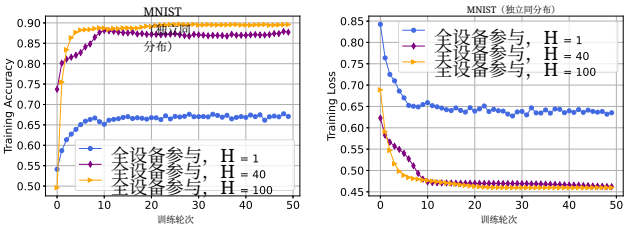
A. *QFL*框架的参数设置

为评估我们提出的QFL框架的性能，所有仿真均在一台服务器上完成，该服务器配置了NVIDIA GeForce RTX 4090图形处理器、64GB内存，并运行Ubuntu 22.04操作系统。我们在用于手写数字识别的标准MNIST数据集上开展实验，该数据集包含60,000个训练样本和10,000个测试样本。它是一个公开且广泛应用的基准数据集，在联邦学习和量子机器学习研究中均有使用，因此适合评估不同数据异质性设置下的收敛行为。在仿真中，我们使用基于*TorchQuan-tum*的量子模型研究QFL的收敛特性。该模型将参数化量子电路与标准损失优化方法随机梯度下降相结合，并采用4个量子比特。模型架构包含随机量子门和可训练旋转层。在实验中，我们探究了联邦设置下全设备参与的情形。训练采用Adam优化器，学习率为0.001。全设备参与时，共有30台设备参与。我们在独立同分布和非独立同分布两种数据分布下测试QFL框架。此外，我们还在MNIST数据集上使用不同数量的量子测量次数（1、40和100）开展实验，以评估这些次数对QFL收敛的影响，尤其考察量子采样噪声如何影响平均损失和准确率等性能指标。

B. 基于量子近似优化算法的和速率最大化参数设置

基于量子近似优化算法的和速率最大化仿真也在同一台服务器上进行，该服务器配备NVIDIA GeForce RTX 4090图形处理器、64GB内存，并运行Ubuntu 22.04操作系统。量子近似优化算法仿真所用的所有无线通信参数均根据第II节中描述的完整系统模型生成。具体而言，无线环境采用距离相关路径损耗、8 dB对数正态阴影衰落以及由Jake 模型描述的时间相关瑞利衰落进行建模，具体形式见公式（14）和（15）。信道分配和干扰场景采用多信道非正交多址接入上行链路结构进行仿真，其中多个量子设备可能同时接入同一信道，从而产生同信道干扰。根据公式（17）和（18）计算每台设备的信号与干扰加噪声比和可达数据速率，以确保理论分析与仿真环境一致。在仿真设置中，我们将系统建模为包含 $C = 4$个信道和 $N = \{50, 200, 500\}$台设备。最大发射功率 $P^{\max}$ 和噪声功率 σ分别设为20 dBm和 -114 dBm [41]。信道相关因子 ϵ取值为 0.6 [41]。拟设由 $p = 2$ 层构成，各层交替使用问题哈密顿量和混合哈密顿量。问题哈密顿量包含由 Q 矩阵元素驱动的受控 RZ 旋转，而混合哈密顿量则使用 RX 门来探索解空间。初始时，所有量子比特均通过哈达玛门制备为均匀叠加态，以确保从均匀探索的起点开始。电路由角度参数化 β 和γ 跨越两层，优化过程通过迭代调整这些参数，以最小化目标函数。仿真在 Python 环境中运行，利用 *PennyLane*进行量子电路模拟和优化，并使用 *NumPy* 进行数值计算，使用*matplotlib* 可视化结果。

C. *QFL*框架性能评估的数值结果

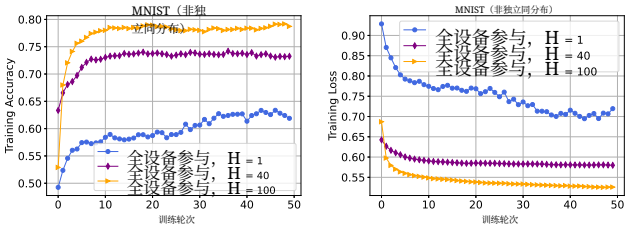


(a) 在全设备参与且量子测量次数不同的情况下，QFL在MNIST数据集（独立同分布）上的训练准确率比较。

(b) 训练损失比较
QFL在MNIST数据集（独立同分布）上的表现
针对全设备参与，且量子数量不同的测量次数。

图2：全设备参与且量子测量次数不同时，量子联邦学习在MNIST数据集（独立同分布）上的训练性能比较。

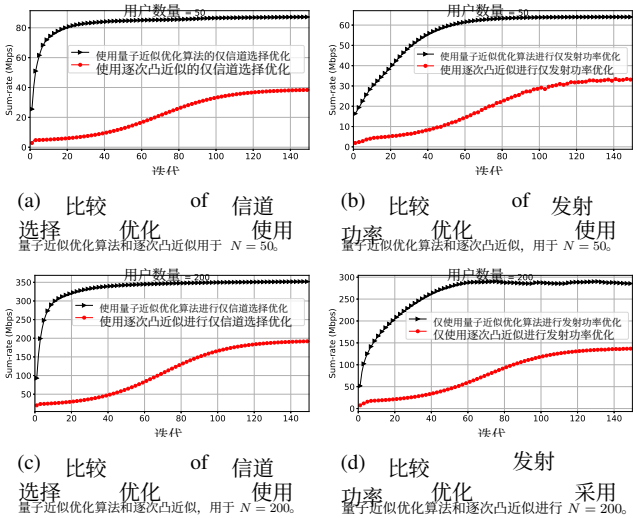
图2展示了在 MNIST 数据集（独立同分布）上进行量子联邦学习训练的表现，表明在以下情况下，随着量子测量次数的增加，准确率得到了提升



（a）量子联邦学习的训练准确率比较（MNIST数据集）（非独立同分布）全体设备参与及不同数量的量子测量次数。

（b）QFL on MNIST 数据集（非独立同分布）下的全设备参与在不同数量的量子测量次数。

图3：全设备参与条件下，量子联邦学习在CIFAR-10和MNIST数据集（非独立同分布）上随量子测量次数变化的训练性能比较。



(a) 比较 of 信道选择 优化 使用量子近似优化算法和逐次凸近似用于 $N = 50$。

(b) 比较 of 发射功率 优化 使用量子近似优化算法和逐次凸近似，用于 $N = 50$。

(c) 比较 of 信道选择 优化 使用量子近似优化算法和逐次凸近似，用于 $N = 200$。

(d) 比较 发射功率 优化 采用量子近似优化算法和逐次凸近似进行 $N = 200$。

图4：针对不同设备数量，使用量子近似优化算法和逐次凸近似进行发射功率与信道选择优化的比较。

全设备参与的场景。图2a展示了在全设备参与且量子测量次数不同（H=1、H=40、 H=100）的情况下，量子联邦学习的训练准确率。从图中可以明显看出，随着量子测量次数的增加，训练准确率显著提升。这一改进归因于通过聚合更多测量结果而降低了量子采样噪声的影响，从而增强了与模型训练相关的量子计算的可靠性和稳定性。从单次测量（H=1）增加到更高次数（H=40），并进一步增加到（H=100），量子联邦学习的准确率持续提升。这一结果表明，增加量子测量次数能够有效降低量子采样噪声引入的不稳定性和错误，从而显著提升量子联邦学习系统的整体训练表现。图2b中量子联邦学习的损失表现与图2a中的准确率表现一致。两幅图均表明，增加量子测量次数（H=1、H=40、 H=100）能够显著提升性能。将测量次数从 H=1 增加到 H=40，再增加到 H=100，能够有效减少量子测量固有的错误和变异性，从而改善损失指标。这一趋势凸显了扩大量子测量次数对于在量子联邦学习设置中实现最佳训练结果的重要作用。

图3展示了非独立同分布设置下、考虑全设备参与及不同量子测量次数时，量子联邦学习在MNIST数据集上的训练性能。其性能趋势与图2中独立同分布设置下观察到的趋势相似。这种一致性表明，无论独立同分布与非独立同分布条件之间的数据分布差异如何，in-的影响

增加量子测量次数保持不变，系统性地增强并提升训练准确率，同时降低两种数据配置下的损失。

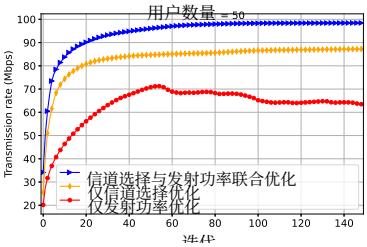
D. 基于量子近似优化算法的和速率最大化评估数值结果

图 4 对比分析了信道选择和发射功率优化技术——量子近似优化算法和逐次凸近似——在 150 次迭代中的性能。仿真

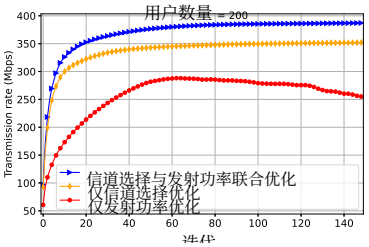
针对这两种优化技术开展了仿真，设备数量从 $N = 50$ 增加到 $N = 200$，以展示各方法在不同网络负载下的可扩展性和有效性。图 4a 展示了信道选择优化中，设备数量为 $N = 200$ 时，量子近似优化算法与逐次凸近似的和速率（单位：Mbps）随迭代次数的变化情况。从图中可以明显看出，无论是在收敛时达到的和速率，还是收敛效率方面，量子近似优化算法都显著优于逐次凸近似。量子近似优化算法在大约第 30th 次迭代时便迅速达到并保持 80 Mbps 以上的和速率，而逐次凸近似则逐渐提升，并在大约第 130^{th}次迭代时稳定在约 40 Mbps。由此可见，在信道选择优化方面，量子近似优化算法优于逐次凸近似，和速率几乎提升了 100%。

图4b展示了和速率（单位：Mbps）随迭代次数的变化，并对比了量子近似优化算法与逐次凸近似技术在 $N = 50$ 设备进行发射功率优化时的性能。图中显示，量子近似优化算法在收敛时达到的峰值和速率以及达到收敛所需的迭代次数方面均优于逐次凸近似。具体而言，量子近似优化算法在大约第 30^{th}次迭代时达到接近60 Mbps的和速率，并保持这一水平；而逐次凸近似的和速率增长较慢，在大约第 130^{th}次迭代时稳定在30 Mbps左右。因此，与信道选择优化类似，量子近似优化算法在发射功率优化方面同样展现出优于逐次凸近似的性能，使和速率提升了近100%。

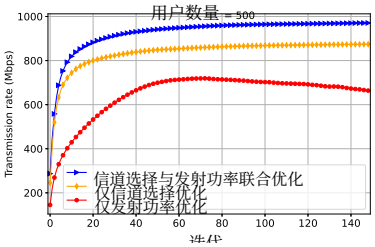
图 4c 比较了 QAOA 与 SCA 技术在为 N =200 台设备优化信道选择时的性能，展示了和速率（单位：Mbps）随迭代次数的变化情况。与设备数量较少时的早期观察结果相似，QAOA 仍显著优于 SCA。在此场景中，QAOA 的和速率比 SCA 高约 75%，并且比 SCA 更早实现收敛。同样，图 4d 比较了 QAOA 与 SCA 技术在为 $N = 200$ 台设备优化发射功率时的性能，展示了和速率（单位：Mbps）随迭代次数的变化情况。从图中可以看出，图 4b 所示的性能趋势在此处同样重现，



(a) 不同优化方案下每个用户和速率的对比分析——各方案下的性能，针对设备数量 $N = 50$。

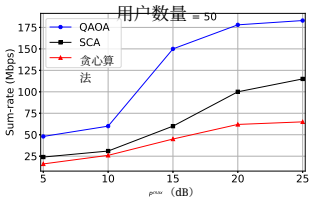


（b）不同优化方案下和速率每个设备的对比分析，针对设备数量 $N = 200$。

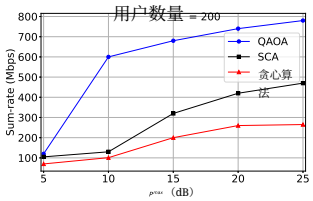


（c）不同优化方案下和速率性能的对比分析，针对设备数量 $N = 500$。

图5：不同优化方案下，设备数量变化时的和速率性能对比分析。

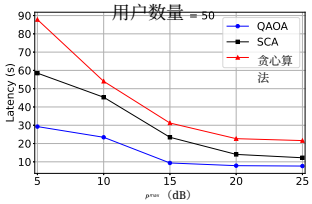


（a）随着设备的最大发射功率增加，在设备数量为 $N = 50$ 时不同方案下和速率性能的比较。

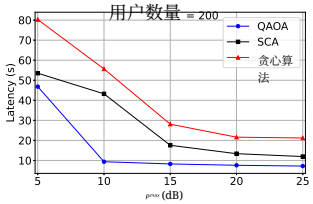


(b) 随着设备数量 $N = 200$ 和设备的最大发射功率增加，不同方案的和速率性能比较。

图6：随着设备的最大发射功率增加，不同方案在不同设备数量下的和速率性能比较。



(a) 比较不同的设备最大发射功率随着设备数量时延性能对比，$N = 50$。



(b) 比较不同方案的性能随着设备的最大发射功率增加，针对设备数量时延在，$N = 200$。

图 7：随着设备最大发射功率增加，不同设备数量下各方案的时延性能比较。

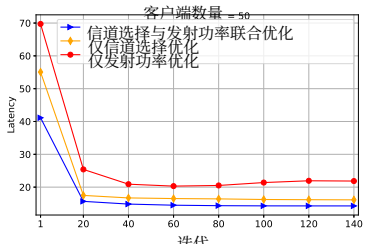
QAOA 继续展现出相较于 SCA 的优势。虽然 QAOA 的和速率几乎提升了 100%，但图中也凸显了其比 SCA 快得多地实现收敛的能力。

图5展示了使用量子近似优化算法（QAOA）针对不同设备数量，对不同优化方案的和速率性能进行的对比分析。图5a展示了对于$N = 50$ 台设备，和速率（单位：Mbps）随迭代次数变化的情况，对比了我们提出的联合优化信道选择与发射功率的方案，以及仅优化其中一个参数的方案。从图中可以清楚看出，我们提出的方案在约第 40th 次迭代后达到稳定的和速率，在最大化和速率及加快收敛方面均显著优于另外两种方案。从数值上看，我们提出的方案所达到的和速率分别比仅优化设备信道选择和发射功率的方案高12.87%和54.87%。图5b和图5c分别将此分析扩展至设备数量为 $N = 200$ 和 $N = 500$的更大规模网络，呈现出相似的趋势。具体而言，图5b中，我们提出的联合优化方案表现优于其他方案，其和速率分别比仅优化信道选择或发射功率的方案高9.99%和52.03%。在图5c中，对于 $N = 500$台设备，我们提出的方案进一步展现出其鲁棒性和效率，其和速率分别比仅优化信道选择和发射功率的方案高11.04%和46.55%。这些结果凸显了我们提出的联合优化方法即使在规模大得多的网络配置中，也具有很强的可扩展性和有效性。

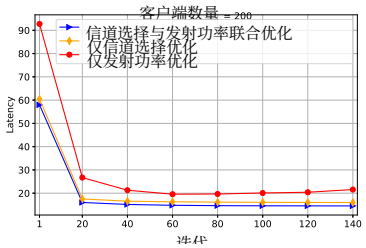
图6展示了设备的最大发射功率（dB）与和速率（Mbps）之间的关系，并比较了量子近似优化算法、逐次凸近似和贪心方案的性能。图6a展示了 $N = 50$ 个设备时的比较结果。图中清楚地表明，随着设备最大发射功率增加，所有优化技术的和速率均有所提升。然而，我们提出的量子优化技术显著优于其他算法，其和速率明显高于逐次凸近似和贪心算法。具体而言，量子近似优化算法的和速率分别比逐次凸近似和贪心算法高59.13%和181.54%。图6b将分析扩展至 $N = 200$个设备，呈现出类似的趋势。在设备数量较多的场景中，量子近似优化算法仍优于其他优化算法，其和速率分别比逐次凸近似和贪心算法高65.96%和194.34%。

同样，如图 7 所示，我们的时延评估也证明了基于量子近似优化算法的优化方法在规模更大的网络中同样有效；随着设备最大发射功率的增加，该网络包含 $N = 50$ 和 $N = 200$ 台设备。例如，在图 7a 的 $N = 50$ 处，与逐次凸近似和贪心方案相比，我们提出的量子近似优化算法方案分别将时延降低了 38.46% 和 63.64%。同样，对于图 7b 中包含 $N = 200$ 台设备的网络，我们的方法将时延降低了 41.67% 和 66.67%。

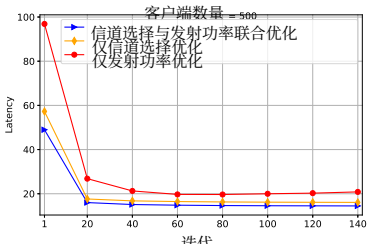
此外，图 8 展示了使用量子近似优化算法对不同设备数量下各优化方案的时延性能进行的对比分析。图 8a 描绘了 $N = 50$ 台设备的时延（单位：秒）随迭代次数的变化，并将我们提出的联合优化信道选择和发射功率的方案



(a) 不同方案的时延对比分析



(b) 时延的对比分析



(c) 不同优化方案在设备数量为 $N = 500$ 时的时延性能对比分析。

性能跨不同优化针对设备数量 $N = 50$ 的方案。

性能跨不同优化设备数量为 $N = 200$ 的方案。

图 8：不同设备数量下各优化方案的时延性能对比分析。

与仅优化其中一个参数的方案进行了比较。显然，在大约第 20th 次迭代后，我们提出的方案能够实现稳定的时延水平，在降低时延和快速收敛方面均显著优于另外两种方案。更具体地说，与仅优化设备信道选择和发射功率的方案相比，我们提出的方案分别可将时延降低 11.53% 和 34.69%。

图 8b 和图 8c 将这一分析分别扩展到由 $N = 200$ 和 $N = 500$ 台设备组成的更大规模网络，表明在这些更大规模下趋势保持一致。从数值上看，在图 8b 中，我们所提出的信道选择与发射功率联合优化方法优于其他方案，其时延分别比仅优化信道选择或仅优化发射功率的方案低 9.07% 和 32.61%。在图 8c 中，对于 $N = 500$ 台设备，我们所提出的方案进一步展现了其鲁棒性与效率，其时延分别比仅关注信道选择和仅关注发射功率的方案低 9.94% 和 30.47%。

E. 计算开销比较

我们还研究并比较了在设备数量不同的情况下，QAOA与基准算法（即SCA算法）应用于联合优化任务时的代码执行时间。表I列出了这些数值，清楚地表明，借助量子计算能力，QAOA在不同设备数量下的执行速度都显著快于SCA。这凸显了量子方法（即QAOA）在优化复杂网络操作方面的高效性。尽管SCA等传统算法广泛用于无线资源分配，QAOA仍能大幅加速，在某些情况下速度提升超过50倍，因此对于需要在实际轮次时长内完成资源协调的大规模量子联邦学习系统而言，它具有很高的实用价值。

表I：QAOA和SCA的执行时间

设备数量	QAOA执行时间（秒）	SCA执行时间（秒）
50	187.33	2245.08
200	730.75	36562.43

此外，表 II 记录了使用量子近似优化算法经过 150 次迭代优化不同参数时的执行时间，设备数量分别为 50、200 和 500。这些数据清晰地展示了执行时间如何随优化任务的复杂度以及所涉及的设备数量而变化。具体而言，同时优化两个参数所需的时间始终长于单参数优化，这反映了其更高的计算需求。然而，执行时间的增加预计会带来更全面的优化结果，凸显了计算开销与优化效果之间的权衡。这项分析对于评估量子近似优化算法在实际应用中的可扩展性至关重要，因为在这些应用中，性能和速度都举足轻重。

表II：不同量子近似优化算法优化策略和设备数量下的执行时间

设备数量	子问题 1（秒）	子问题 2（秒）	联合 (s)
50	46.74	140.33	187.33
200	188.53	540.13	730.75
500	470.79	1365.88	1839.88

VII. 结论

本文研究了一种经过优化的QFL框架，专门面向基于非正交多址接入的大规模无线网络，通过联合优化量子设备的信道选择与发射功率以实现和速率最大化。本文还对我们所提出的QFL框架的收敛行为进行了严格分析。所构建的和速率最大化问题是非凸的混合整数非线性规划难题，即使在信道选择参数已知的情况下仍为NP难。因此，我们提出了一种高效的量子中心迭代优化方法，借助量子近似优化算法与BCD技术来获得最优解。大量数值结果表明，我们基于多信道非正交多址接入的QFL框架同时提升了模型训练与和速率性能。

参考文献

[1] M. Chehimi 等，“经典与量子网络上的量子联邦学习基础”，*IEEE Network*，第38卷，第1期，第124–130页，2023年。[2] Z. Qu 等，“DTQFL：一种用于5G移动网络智能诊断的数字孪生辅助量子联邦学习算法”，*IEEE*生物医学与健康信息学汇刊，2023年。[3] W. Tang 等，“CutQC：使用小型量子计算机进行大型量子电路评估”，载于第*26*届*ACM*编程语言与操作系统架构支持国际会议论文集，2021年，第473–486页。[4] V. K. Quy 等，“从联邦学习到空天地一体化网络的量子联邦学习”，载于*2024*年*IEEE*通信与网络标准会议（*CSCN*）。IEEE，2024年，第402–407页。

[5]Q. Wu 等，“无线供能物联网的频谱与能效：非正交多址接入还是时分多址？”《IEEE车辆技术汇刊》，第67卷，第7期，第6663–6667页，2018年。[6] M. Fellous-Asiani 等，“优化可扩展全栈量子计算机的资源效率”，《PRX Quantum》，第4卷，第4期，040319页，2023年。[7] F. Liu 等，“下行链路多载波非正交多址系统动态功率与信道分配的性能”，《IEEE无线通信汇刊》，第19卷，第3期，第1650–1662页，2019年。[8] S. Endo 等，“混合量子-经典算法与量子误差缓解”，《日本物理学会杂志》，第90卷，第3期，032001页，2021年。[9] S. T. Jose 等，“误差缓解辅助的参数化量子电路优化：收敛性分析”，《IEEE量子工程汇刊》，第3卷，第1–19页，2022年。[10] V. Rana 等，“使用凸优化技术进行无线网络资源分配”，载于2025年普适计算技术国际会议（ICPCT）。IEEE，2025年，第371–376页。[11] B. Gulbahar，“采用递归量子近似优化算法和成本受限均匀采样的多数投票，用于大规模MIMO中的最大似然检测”，《IEEE无线通信汇刊》，2025年。[12]D. Xiang 等，“Choco-Q：用于约束二元优化的基于对易哈密顿量的量子近似优化算法”，载于2025年IEEE高性能计算机体系结构国际研讨会（HPCA）。IEEE，2025年，第275–289页。[13] S. Park 等，“面向无人机自主监视的动态量子联邦学习”，《IEEE车辆技术汇刊》，2025年。[14] ——，“量子纠缠控制的量子联邦学习”，《IEEE物联网汇刊》，2025年。[15] D. Gurung 等，“加权个性化量子联邦学习的性能分析与设计”，《IEEE人工智能汇刊》，2025年。[16] Z. Qu 等，“DAQFL：用于医疗物联网智能诊断的动态聚合量子联邦学习算法”，《IEEE物联网汇刊》，2025年。[17]T. Wang 等，“基于量子网络的量子联邦学习”，载于ICASSP 2024—2024年IEEE声学、语音与信号处理国际会议（ICASSP）。IEEE，2024年，第13 401–13 405页。[18]Z. Qu 等，“QFSM：一种在5G车联网中采用最小门控单元进行语音情感识别的新型量子联邦学习算法”，《IEEE智能车辆汇刊》，2024年。[19] Z. Abou El Houda 等，“一种利用量子联邦学习实现消费者网络高效网络入侵检测的隐私保护框架”，《IEEE消费电子汇刊》，2024年。[20] C. Ren 等，“通过量子联邦学习增强智能电网的动态安全评估”，《IEEE自动化科学与工程汇刊》，2024年。[21] R. Rofougaran 等，“采用差分隐私的联邦量子机器学习”，载于ICASSP 2024—2024年IEEE声学、语音与信号处理国际会议（ICASSP）。IEEE，2024年，第9811–9815页。[22] Z. Yang 等，“无线通信网络上的节能联邦学习”，《IEEE无线通信汇刊》，第20卷，第3期，第1935–1949页，2020年。[23] N. H. Tran 等，“无线网络上的联邦学习：优化模型设计与分析”，载于IEEE INFOCOM 2019—IEEE计算机通信会议。IEEE，2019年，第1387–1395页。[24] B. Hazarika 等，“面向元宇宙赋能车载网络的量子增强联邦学习”，《IEEE通信汇刊》，2024年。[25] S. Hu 等，“OFDMA-F^{2}L：基于OFDMA空中接口的灵活聚合联邦学习”，《IEEE无线通信汇刊》，第23卷，第7期，第6793–6807页，2024年。[26] S. Salgia 等，“一种针对累积遗憾的通信高效自适应联邦学习算法”，《IEEE信号处理汇刊》，第72卷，第735–743页，2024年。[27] X. Sun 等，“具有自适应性的联邦学习中通信高效且保护隐私的聚合”，《IEEE物联网汇刊》，2024年。[28] S. Kasi 等，“基于量子近似优化算法的下一代无线信道编码译码器架构”，载于2024年IEEE量子计算与工程国际会议（QCE），第1卷。IEEE，2024年，第368–379页。[29] M. Manasa 等，“通过人工智能驱动的量子近似优化优化能量管理与负载均衡”，载于*2024*年第*4*届移动网络与无线通信国际会议（*ICMNWC*）。IEEE，2024年，第1–7页。[30] D. Van Huynh 等，“在真实量子硬件上采用以量子为中心的优化实现6G边缘计算的最优服务放置”，*IEEE*通信快报，2024年。[31] E. Colella 等，“通过量子优化可重构智能表面以缓解无线网络中的多径衰落”，*IEEE*多尺度与多物理场计算技术期刊，2024年。[32] T. Kim 等，“对于多天线波束成形系统，频繁的低分辨率反馈是否优于不频繁的高分辨率反馈？”，*IEEE*信号处理汇刊，第59卷，第4期，第1654–1669页，2010年。[33] R. Wang 等，“具有部分信道状态信息的D2D通信中最优的QoS感知信道分配”，*IEEE*无线通信汇刊，第15卷，第11期，第7594–7609页，2016年。[34] X. Li等，“关于FedAvg在非独立同分布数据上的收敛性”，*arXiv*预印本 *arXiv:1907.02189*，2019年。[35] L. Chen 等，“非独立同分布数据上的差分隐私联邦学习：收敛性分析与自适应优化”，*IEEE*知识与数据工程汇刊，第36卷，第9期，第4567–4581页，2024年。[36] A. Ajalloeian 等，“关于有偏梯度下随机梯度下降的收敛性”，*arXiv*预印本 *arXiv:2008.00051*，2020年。[37] Z.-Q. Luo 等，“动态频谱管理：复杂性与对偶性”，*IEEE*信号处理选题期刊，第2卷，第1期，第57–73页，2008年。[38] D.-H. Tran 等，“物联网网络中的无人机中继辅助应急通信：资源分配与轨迹优化”，*IEEE*无线通信汇刊，第21卷，第3期，第1621–1637页，2021年。[39] C. W. Tan 等，“多用户认知无线网络中的频谱管理：最优性与算法”，*IEEE*通信选题期刊，第29卷，第2期，第421–430页，2011年。[40] S. D’Oro 等，“一种用于多载波无线网络能效低复杂度优化的学习方法”，*IEEE*无线通信汇刊，第17卷，第5期，第3226–3241页，2018年。[41] J. Tan 等，“用于D2D网络联合信道选择与功率控制的深度强化学习”，*IEEE*无线通信汇刊，第20卷，第2期，第1363–1378页，2020年。

Shaba Shaon 目前正在美国阿拉巴马大学亨茨维尔分校电气与计算机工程系攻读博士学位。她的研究兴趣包括分布式量子计算、联邦学习和无线网络。

Christopher G. Brinton [IEEE高级会员]目前是美国普渡大学电气与计算机工程系的Elmore副教授。他曾获得四项美国顶级早期职业奖，分别来自美国国家科学基金会（CAREER）、美国海军研究办公室（YIP）、美国国防高级研究计划局（YFA）和美国空军科学研究办公室（YIP）。

Dinh C. Nguyen 是美国阿拉巴马大学亨茨维尔分校电气与计算机工程系的助理教授。他的研究兴趣包括分布式量子计算、联邦学习和无线网络。

附录

本节给出上文所用引理和定理的证明。

A. 引理 *1* 的证明

如前所述，令 $\mathcal{N}=\{1,2,\ldots,N\}$ 表示设备集合，令 $\tilde{g}_k^t = \frac{1}{N}\sum_{n\in\mathcal{N}}\tilde{g}_{n,\ k}^t$ 表示它们在全局轮次 k的局部迭代t 时的局部随机梯度的平均值。我们有

$$
\begin{aligned}
&-\mathbb{E}_{\{\xi_{1,k}^t,\ldots,\xi_{n,k}^t|\boldsymbol{\theta}_{1,k}^t,\ldots,\boldsymbol{\theta}_{N,k}^t\}}\mathbb{E}_{\{1,2,\ldots N\}\in\mathcal{N}}\left[\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\tilde{g}_k^t\rangle\right]\\
&=-\mathbb{E}_{\{\xi_{1,k}^t,\ldots,\xi_{n,k}^t|\boldsymbol{\theta}_{1,k}^t,\ldots,\boldsymbol{\theta}_{N,k}^t\}}\\
&\qquad\mathbb{E}_{\{1,2,\ldots N\}\in\mathcal{N}}\left[\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\frac{1}{N}\sum_{n\in\mathcal{N}}\tilde{g}_{n,k}^t\rangle\right]\\
&\overset{①}{=}-\mathbb{E}_{\{1,2,\ldots N\}\in\mathcal{N}}\\
&\qquad\mathbb{E}_{\{\xi_{1,k}^t,\ldots,\xi_{n,k}^t|\boldsymbol{\theta}_{1,k}^t,\ldots,\boldsymbol{\theta}_{N,k}^t\}}\left[\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\frac{1}{N}\sum_{n\in\mathcal{N}}\tilde{g}_{n,k}^t\rangle\right]\\
&=-\left\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\mathbb{E}_{\mathcal{N}}\left[\frac{1}{N}\sum_{n\in\mathcal{N}}\mathbb{E}_{k,t}\tilde{g}_n\right]\right\rangle
\end{aligned}
\tag{57}
$$

$$
\begin{aligned}
&=-\left\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\mathbb{E}_{\mathcal{N}}\left[\frac{1}{N}\sum_{n\in\mathcal{N}}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)\right]\right\rangle\\
&=-\left\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t),\frac{1}{N}\left[N\sum_{n\in\mathcal{N}}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)\right]\right\rangle\\
&\overset{②}{=}\frac{1}{2}\Bigg[-||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||_2^2-||\sum_{n=0}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2+||\nabla f(\bar{\boldsymbol{\theta}}_k^t)\\
&\qquad\qquad-\sum_{n=0}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2\Bigg]\\
&=\frac{1}{2}\Bigg[-||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||_2^2-||\sum_{n=0}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2\\
&\qquad+||\sum_{n=0}^{N}\left(\nabla f_n(\bar{\boldsymbol{\theta}}_k^t)-\nabla f_n(\boldsymbol{\theta}_{n,k}^t)\right)||_2^2\Bigg]\\
&\overset{③}{\leq}\frac{1}{2}\Bigg[-||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||_2^2-||\sum_{n=0}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2\\
&\qquad+\sum_{n=0}^{N}||\nabla f_n(\bar{\boldsymbol{\theta}}_k^t)-\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2\Bigg]\\
&\overset{④}{\leq}\frac{1}{2}\Bigg[-||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||_2^2-||\sum_{n=0}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||_2^2\\
&\qquad+\sum_{n=0}^{N}L^2||\bar{\boldsymbol{\theta}}_k^t-\boldsymbol{\theta}_{n,k}^t||_2^2\Bigg],
\end{aligned}
\tag{58}
$$

其中 ① 是由于随机变量$\xi_{n,\ k}^t$与 $\mathcal{N}$相互独立， ① 是因为 ② $2\langle a,\ b\rangle=||a||^2+||b||^2-||a-b||^2$， ③ 成立是由于 $||.||_2$的凸性，而 ④ 由假设**1**得出。

B. 引理 *2* 的证明

我们将 $k=i_c$ 表示为最近一次全局通信轮次，因此 $\bar{\boldsymbol{\theta}}^{i_c+1}=\frac{1}{N}\sum_{n\in\mathcal{N}}\boldsymbol{\theta}_n^{i_c+1}$。设备 n 在任意特定迭代 $i>i_c$ 时的局部解，其中假设 i 表示最近一次迭代，涵盖截至当前时刻的所有全局与局部迭代，可写为:

$$
\begin{aligned}
\boldsymbol{\theta}_{n,k}^t=\boldsymbol{\theta}_n^i&=\boldsymbol{\theta}_n^{i-1}-\eta_{i_c}\tilde{g}_n^{i-1}\\
&\overset{①}{=}\boldsymbol{\theta}_n^{i-2}-[\eta_{i_c}\tilde{g}_n^{i-2}+\eta_{i_c}\tilde{g}_n^{i-1}]\\
&=\bar{\boldsymbol{\theta}}^{i_c+1}-\sum_{z=i_c+1}^{i-1}\eta_{i_c}\tilde{g}_n^z,
\end{aligned}
\tag{59}
$$

其中 ②由局部解的更新规则得出。现在，我们按如下方式根据(59)计算迭代 i处的平均虚拟模型:

$$
\bar{\boldsymbol{\theta}}^i=\bar{\boldsymbol{\theta}}^{i_c+1}-\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=i_c+1}^{i-1}\eta_{i_c}\tilde{g}_n^z.
\tag{60}
$$

首先，不失一般性，假设 $i=s_tT_n+r$，其中 s_t 和 r 分别表示全局通信轮次和局部更新的索引。接下来，我们考虑对于 $i_c+1<i\leq i_c+T_n$，$\mathbb{E}_i||\bar{\boldsymbol{\theta}}^i-\boldsymbol{\theta}_n^i||$ 在 $1\leq n\leq N$时不依赖于时间 $i\leq i_c$。因此，对于所有满足 $I=KT_n$的迭代次数 $1\leq i\leq I$，我们可以写出，

$$
\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{1}{T_n}\sum_{t=1}^{T_n}\mathbb{E}||\bar{\boldsymbol{\theta}}_k^t-\boldsymbol{\theta}_{n,k}^t||^2\\
&=\frac{1}{I}\sum_{i=1}^{I}\sum_{n=1}^{N}\mathbb{E}||\bar{\boldsymbol{\theta}}^i-\boldsymbol{\theta}_n^i||^2\\
&=\frac{1}{I}\sum_{s_t=1}^{\frac{I}{T_n}-1}\sum_{n=1}^{N}\sum_{r=1}^{T_n}\mathbb{E}||\bar{\boldsymbol{\theta}}^{s_tE+r}-\boldsymbol{\theta}_n^{s_tE+r}||^2.
\end{aligned}
\tag{61}
$$

我们分三个步骤对项 $\mathbb{E}||\bar{\boldsymbol{\theta}}^i-\boldsymbol{\theta}_l^i||^2$（针对 $i_c+1\leq i=s_tT_n+r\leq i_c+T_n$）进行界定：**(1)** 首先，将该量与随机梯度和全梯度之间的方差联系起来；**(2)** 运用关于无偏估计和独立同分布小批量采样的假设**1**；**(3)** 运用假设**3**对最后几项进行界定。接下来，我们将逐一实施这些步骤。需要注意的是，l与单个设备相关，而 n 用于对所有设备求和。

关于方差:

$$
\begin{aligned}
&\mathbb{E}||\bar{\boldsymbol{\theta}}^{s_tE+r}-\boldsymbol{\theta}_l^{s_tE+r}||^2\\
&=\mathbb{E}||\bar{\boldsymbol{\theta}}^{i_c+1}-\left[\sum_{z=i_c+1}^{i-1}\eta_{i_c}\tilde{g}_l^z\right]-\bar{\boldsymbol{\theta}}^{i_c+1}\\
&\qquad\qquad+\left[\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=i_c+1}^{i-1}\eta_{i_c}\tilde{g}_n^z\right]||^2\\
&\overset{①}{=}\mathbb{E}||\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_l^{s_t+z}-\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_n^{s_t+z}||^2\\
&\overset{②}{\leq}2\left[\mathbb{E}||\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_l^{s_t+z}||^2-\mathbb{E}||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_n^{s_t+z}||^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&\overset{③}{=} 2\Big[\mathbb{E}||\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_l^{s_t+z}-\mathbb{E}\Big[\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_l^{s_t+z}\Big]||^2\\
&-\mathbb{E}||\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_l^{s_t+z}||^2+\mathbb{E}||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_n^{s_t+z}-\mathbb{E}\Big[\frac{1}{N}\\
&\times\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_n^{s_t+z}\Big]||^2\Big]+||\mathbb{E}\Big[\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\tilde{g}_n^{s_t+z}\Big]||^2\\
&\overset{④}{=} 2\mathbb{E}\Bigg(\Big[||\sum_{z=1}^{r}\eta_{i_c}\Big[\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}\Big]||^2\\
&+||\sum_{z=1}^{r}\eta_{i_c}g_l^{s_tT_n+z}||^2\Big]\\
&+||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}\Big[\tilde{g}_n^{s_tT_n+z}-g_n^{s_tT_n+z}\Big]||^2\\
&+||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}g_n^{s_tT_n+z}||^2\Bigg),
\end{aligned}
\tag{62}
$$

其中， ① 成立是因为 $i=s_tT_n+r\leq i_c+T_n$， ② 是由于 $||a-b||^2\leq 2\left(||a||^2+||b||^2\right)$， ③ 源于 $\mathbb{E}[\boldsymbol{\theta}^2]=\mathbb{E}[[\boldsymbol{\theta}-\mathbb{E}[\boldsymbol{\theta}]]^2]+\mathbb{E}[\boldsymbol{\theta}]^2$， ④来自假设**1**。

无偏估计与独立同分布采样

$$
\begin{aligned}
&=2\mathbb{E}\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2||\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}||^2\\
&+\sum_{p\neq q\vee l\neq v}\Big\langle\eta_{i_c}\tilde{g}_l^p-\eta_{i_c}g_l^p,\eta_{i_c}\tilde{g}_v^q-\eta_{i_c}g_v^q\Big\rangle\\
&+||\sum_{z=1}^{r}\eta_{i_c}g_l^{s_tT_n+z}||^2\Big]\\
&+\frac{1}{N^2}\sum_{l\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2||\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}||^2\\
&+\frac{1}{N^2}\sum_{p\neq q\vee l\neq v}\Big\langle\eta_{i_c}\tilde{g}_l^p-\eta_{i_c}g_l^p,\eta_{i_c}\tilde{g}_v^q-\eta_{i_c}g_v^q\Big\rangle\\
&+||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}g_n^{s_tT_n+z}||^2\Bigg)\\
&\overset{⑤}{=}2\mathbb{E}\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2||\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}||^2\\
&+||\sum_{z=1}^{r}\eta_{i_c}g_l^{s_tT_n+z}||^2\Big]+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2||\tilde{g}_n^{s_tT_n+z}\\
&-g_n^{s_tT_n+z}||^2+||\frac{1}{N}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}g_n^{s_tT_n+z}||^2\Bigg)\\
&\overset{⑥}{\leq}2\mathbb{E}\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2||\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}||^2+r\sum_{z=1}^{r}\eta_{i_c}^2\\
&\times||g_l^{s_tT_n+z}||^2\Big]+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}||\tilde{g}_n^{s_tT_n+z}-g_n^{s_tT_n+z}||^2\\
&+\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{s_tT_n+z}^2||g_n^{s_tT_n+z}||^2\\
&=2\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2\mathbb{E}||\tilde{g}_l^{s_tT_n+z}-g_l^{s_tT_n+z}||^2+r\sum_{z=1}^{r}\eta_{i_c}^2\\
&\times\mathbb{E}||g_l^{s_tT_n+z}||^2\Big]+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2\mathbb{E}||\tilde{g}_n^{s_tT_n+z}\\
&-g_n^{s_tT_n+z}||^2+\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2\mathbb{E}||g_n^{s_tT_n+z}||^2\Bigg),
\end{aligned}
\tag{63}
$$

其中， ⑤ 源于独立小批量采样以及无偏估计假设，而 ⑥则由不等式 $||\sum_{i=1}^{m}a_i||^2\leq m\sum_{i=1}^{m}||a_i||^2$得出。

使用假设*3*：接下来，我们利用假设3对（63）中的各项进行界定，具体如下:

$$
\begin{aligned}
&\mathbb{E}||\bar{\boldsymbol{\theta}}_k^t-\boldsymbol{\theta}_{l,k}^t||^2\leq 2\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2\Big[C_1||g_l^{s_tT_n+z}||^2+\frac{\sigma^2}{B}\Big]\\
&+r\sum_{z=1}^{r}\eta_{i_c}^2||g_l^{s_tT_n+z}||^2+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2\Big[C_1||g_n^{s_tT_n+z}||^2\\
&+\frac{\sigma^2}{B}\Big]+\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2||g_n^{s_tT_n+z}||^2\Bigg)\\
&=2\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2C_1||g_l^{s_tT_n+z}||^2+\sum_{z=1}^{r}\eta_{i_c}^2\frac{\sigma^2}{B}\\
&+r\sum_{z=1}^{r}\eta_{i_c}^2||g_l^{s_tT_n+z}||^2\Big]+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2C_1||g_n^{s_tT_n+z}||^2\\
&+\sum_{z=1}^{r}\eta_{i_c}^2\frac{\sigma^2}{NB}+\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2||g_n^{s_tT_n+z}||^2\Bigg).
\end{aligned}
\tag{64}
$$

现在，我们利用式（64）确定 $\sum_{r=1}^{T_n}\sum_{n=1}^{N}[\mathbb{E}||\bar{\boldsymbol{\theta}}_k^t-\boldsymbol{\theta}_{n,k}^t||]$ 的上界如下:

$$
\begin{aligned}
&\sum_{n=1}^{N}\sum_{r=1}^{T_n}\Big[\mathbb{E}||\bar{\boldsymbol{\theta}}^{s_tT_n+z}-\boldsymbol{\theta}_n^{s_tT_n+z}||\Big]\\
&\leq 2\sum_{l=1}^{N}\sum_{r=1}^{T_n}\Bigg(\Big[\sum_{z=1}^{r}\eta_{i_c}^2C_1||g_l^{s_tT_n+z}||^2+\sum_{z=1}^{r}\eta_{i_c}^2\frac{\sigma^2}{B}\\
&+r\sum_{z=1}^{r}\eta_{i_c}^2||g_l^{s_tT_n+z}||^2\Big]+\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2C_1||g_n^{s_tT_n+z}||^2\\
&+\sum_{z=1}^{r}\eta_{i_c}^2\frac{\sigma^2}{NB}+\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}\eta_{i_c}^2||g_n^{s_tT_n+z}||^2\Bigg)\\
&=2\eta_{i_c}^2\Bigg(\Big[C_1\sum_{l=1}^{N}\sum_{r=1}^{T_n}\sum_{z=1}^{r}||g_l^{s_tT_n+z}||^2+\sum_{r=1}^{T_n}\frac{r\sigma^2}{B}\\
&+\sum_{l=1}^{N}\sum_{r=1}^{T_n}\sum_{z=1}^{r}r||g_l^{s_tT_n+z}||^2\Big]+\sum_{r=1}^{T_n}\frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}C_1\\
&\times||g_n^{s_tT_n+z}||^2+\sum_{r=1}^{T_n}\frac{r\sigma^2}{NB}+\sum_{r=1}^{T_n}\frac{r}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{r}||g_n^{s_tT_n+z}||^2\Bigg)
\end{aligned}
$$

$$\overset{①}{\leq} 2\eta_{i_c}^2\Bigg(\Bigg[\sum_{z=1}^{T_n} C_1 \sum_{l=1}^{N} ||g_l^{s_t T_n+z}||^2 + \frac{T_n(T_n+1)\sigma^2}{2B} + \frac{T_n(T_n+1)}{2}\sum_{z=1}^{T_n}\sum_{l=1}^{N}||g_l^{s_t T_n+z}||^2 + \frac{1}{N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{T_n} C_1||g_n^{s_t T_n+z}||^2 + \frac{T_n(T_n+1)\sigma^2}{2NB} + \frac{T_n(T_n+1)}{2N^2}\sum_{n\in\mathcal{N}}\sum_{z=1}^{T_n}||g_n^{s_t T_n+z}||^2$$
$$= \frac{\eta_{i_c}^2(N+1)}{N}\Bigg(\Bigg[(2C_1+T_n(T_n+1))\sum_{z=1}^{T_n}\sum_{n=1}^{N}||g_n^{s_t T_n+z}||^2\Bigg] + \frac{T_n(T_n+1)\sigma^2}{B}\Bigg), \quad (65)$$

其中， ① 是由项 $||g_l||^2$ 为正这一事实得出的。现在，对式（65）中的全局通信轮次求和，得到：

$$\sum_{n=1}^{N}\sum_{s_t=1}^{I/T_n-1}\sum_{r=1}^{T_n}\left[\mathbb{E}||\bar{\boldsymbol{\theta}}^{s_t T_n+z} - \boldsymbol{\theta}_n^{s_t T_n+z}||\right]$$
$$\leq \frac{\eta_{i_c}^2(N+1)}{N}\Bigg(\Bigg[(2C_1 + T_n(T_n+1))\sum_{n=1}^{N}\sum_{s_t=1}^{I/T_n-1}\sum_{z=1}^{T_n}||g_n^{s_t T_n+z}||^2\Bigg] + \frac{I(T_n+1)\sigma^2}{B}\Bigg)$$
$$= \frac{\eta_{i_c}^2(N+1)}{N}\Bigg(\Bigg[(2C_1+T_n(T_n+1))\sum_{n=1}^{N}\sum_{i=1}^{I}||g_n^i||^2\Bigg] + \frac{I(T_n+1)\sigma^2}{B}\Bigg), \quad (66)$$

由此可得

$$\sum_{n=1}^{N}\frac{1}{I}\sum_{i=1}^{I}\left[\mathbb{E}||\bar{\boldsymbol{\theta}}^i - \boldsymbol{\theta}_n^i||\right]$$
$$\leq \frac{(2C_1+T_n(T_n+1))}{I}\frac{\eta_{i_c}^2(N+1)}{N}\sum_{n=1}^{N}\sum_{i=0}^{I-1}||g_n^i||^2 + \frac{\eta_{i_c}^2 I(N+1)(T_n+1)\sigma^2}{NB}$$
$$\overset{①}{\leq} \frac{(2C_1+T_n(T_n+1))}{I}\frac{\lambda\eta_{i_c}^2(N+1)}{N}\sum_{n=1}^{N}\sum_{i=0}^{I-1}||g_n^i||^2 + \frac{\eta_{i_c}^2 I(N+1)(T_n+1)\sigma^2}{NB}, \quad (67)$$

其中， ① 由加权梯度多样性的定义以及主论文中式（30）的上界假设可得。最后，（67）可写为：

$$\frac{1}{K}\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{1}{T_n}\sum_{t=1}^{T_n}\left[\mathbb{E}||\bar{\boldsymbol{\theta}}_k^t - \boldsymbol{\theta}_{n,k}^t||\right]$$
$$\leq \frac{(2C_1+T_n(T_n+1))}{KT_n}\frac{\lambda\eta_{i_c}^2(N+1)}{N}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T_n}||g_{n,k}^t||^2 + \frac{\eta_{i_c}^2 KT_n(N+1)(T_n+1)\sigma^2}{NB}. \quad (68)$$

C. 引理3的证明

We have

$$\mathbb{E}\left[||\tilde{g}_k^t - g_k^t||^2\right] \overset{①}{=} \mathbb{E}\left[||\frac{1}{N}\sum_{n=0}^{N}\tilde{g}_{n,k}^t - g_{n,k}^t||^2\right]$$
$$= \frac{1}{N^2}\mathbb{E}\Bigg[\sum_{n=0}^{N}||(\tilde{g}_{n,k}^t - g_{n,k}^t)||^2 + \sum_{i\neq n}\langle\tilde{g}_{i,k}^t - g_{i,k}^t, \tilde{g}_{n,k}^t - g_{n,k}^t\rangle\Bigg]$$
$$= \frac{1}{N^2}\sum_{n=0}^{N}\mathbb{E}||(\tilde{g}_{n,k}^t - g_{n,k}^t)||^2 + \sum_{i\neq n}\frac{1}{N^2}\mathbb{E}\left[\langle\tilde{g}_{i,k}^t - g_{i,k}^t, \tilde{g}_{n,k}^t - g_{n,k}^t\rangle\right]$$
$$\overset{②}{=} \frac{1}{N^2}\sum_{n=0}^{N}\mathbb{E}||(\tilde{g}_{n,k}^t - g_{n,k}^t)||^2 + \frac{1}{N^2}\sum_{i\neq n}\langle\mathbb{E}\left[\tilde{g}_{i,k}^t - g_{i,k}^t\right], \mathbb{E}\left[\tilde{g}_{n,k}^t - g_{n,k}^t\right]\rangle$$
$$\overset{③}{\leq} \frac{1}{N^2}\sum_{n=0}^{N}\left[C_1||g_{n,k}^t||^2 + C_2^2\right] = \frac{C_1}{N^2}\sum_{n=0}^{N}||g_{n,k}^t||^2 + \frac{C_2^2}{N}, \quad (69)$$

其中，我们在 ① 中使用了 $\tilde{g}_k^t$ 和 g_k^t 的定义，在 ② 中使用了每个设备上的小批量均以独立同分布的方式选取这一事实，而 ③ 可由假设3直接得到。我们注意到，假设3意味着 $\mathbb{E}[\tilde{g}_{n,k}^t] = g_{n,\ k}^t$。因此，我们有

$$\mathbb{E}\left[||\tilde{g}_k^t||^2\right] = \mathbb{E}\left[||\tilde{g}_k^t - \mathbb{E}[\tilde{g}_k^t]||^2\right] + ||\mathbb{E}[\tilde{g}_k^t]||^2$$
$$= \mathbb{E}\left[||\tilde{g}_k^t - g_k^t||^2\right] + ||g_k^t||^2$$
$$\overset{①}{\leq} \frac{C_1}{N^2}\sum_{n=0}^{N}||g_{n,k}^t||^2 + \frac{C_2^2}{N} + ||\frac{1}{N}\sum_{n=0}^{N}g_{n,k}^t||^2$$
$$\overset{②}{\leq} \frac{C_1}{N^2}\sum_{n=0}^{N}||g_{n,k}^t||^2 + \frac{C_2^2}{N} + \frac{1}{N}\sum_{n=0}^{N}||g_{n,k}^t||^2$$
$$= \left(\frac{C_1+N}{N^2}\right)\sum_{n=0}^{N}||g_{n,k}^t||^2 + \frac{C_2^2}{N}, \quad (70)$$

其中， ① 和 ② 由 $||\sum_{i=1}^{m}a_i||^2 \leq m\sum_{i=1}^{m}||a_i||^2$这一事实可得，其中 $a_i \in \mathbb{R}^n$。利用加权梯度多样性的上界 λ,

$$\mathbb{E}\left[||\tilde{g}_k^t||^2\right] \leq \lambda\left(\frac{C_1+N}{N^2}\right)||\sum_{n=0}^{N}g_{n,k}^t||^2 + \frac{C_2^2}{N}, \quad (71)$$

即可得到所述上界。

D. 引理4的证明

为证明引理4，我们固定与全局与局部迭代相关的指标 k 和 t，从而暂时将它们从记号中去掉。令$X_{n,d,\pm} = \langle Z\rangle_{|\Psi_n(\boldsymbol{\theta}_n \pm \frac{\pi}{2}e_d)\rangle} - \langle Z\rangle_{|\Psi_n(\boldsymbol{\theta}_n \pm \frac{\pi}{2}e_d)\rangle}$ 表示可观测量 Z 在量子态 $|\Psi_n(\boldsymbol{\theta}_n \pm \frac{\pi}{2}e_d)\rangle$ 下的估计期望与真实期望之差，其中d^{th}参数被相移 $\pm\frac{\pi}{2}$。在后续分析中，我们使用记号$|\Psi_{n,d,\pm}\rangle = |\Psi_n(\boldsymbol{\theta}_n \pm \frac{\pi}{2}e_d)\rangle$

以求简洁。式(83)中的梯度估计方差写为

$$\begin{aligned}\text{var}(\xi_n) &= \mathbb{E}\Bigg[\sum_{d=1}^{D}\Bigg(\frac{1}{2}(\langle\hat{Z}\rangle_{|\Psi_{n,d,+}\rangle} - \langle\hat{Z}\rangle_{|\Psi_{n,d,-}\rangle}) \\ &\quad - \frac{1}{2}(\langle Z\rangle_{|\Psi_{n,d,+}\rangle} - \langle Z\rangle_{|\Psi_{n,d,-}\rangle})\Bigg)^2\Bigg] \\ &= \sum_{d=1}^{D}\frac{1}{4}\mathbb{E}\Bigg[\Big(X_{n,d,+} - X_{n,d,-}\Big)^2\Bigg] \\ &= \sum_{d=1}^{D}\frac{1}{4}\Big(\mathbb{E}[X_{n,d,+}^2] - \mathbb{E}[X_{n,d,-}^2]\Big), \end{aligned} \tag{72}$$

其中，期望是针对量子态 $|\Psi_n(\boldsymbol{\theta}_n + \frac{\pi}{2}e_d)\rangle$ 和$|\Psi_n(\boldsymbol{\theta}_n - \frac{\pi}{2}e_d)\rangle$的 H 次测量计算的，其中 $d = 1, 2, \ldots, D$。因此，对于 $d = 1, 2, \ldots, D$，随机变量 $X_{n,d,+}$ 和 $X_{n,d,-}$ 相互独立，从而得到等式 (72)。需要注意的是，期望 $\mathbb{E}[X_{n,d,+}^2]$ 等于随机变量 $\langle\hat{Z}\rangle_{|\Psi_{n,d,+}\rangle}$的方差 $\text{var}(\langle\hat{Z}\rangle_{|\Psi_{n,d,+}\rangle})$。令 Y 为定义对可观测量 Z的测量索引的随机变量。因此，$Z = h_Y$表示相应的测量输出。我们用 $W_y = \mathbb{I}\{Y = y\}$ 表示伯努利随机变量，用以确定是否 $Y = y(W_y = 1)$（或否（$W_y = 0$））。我们还指出，量子测量是独立同分布的，因此，根据 $\langle Z\rangle_{|\Psi_{n,d,+}\rangle}$ 的期望定义可知

$$\begin{aligned}\mathbb{E}[X_{n,d,+}^2] &= \frac{1}{H}\text{var}\Big(\sum_{y=1}^{N_z}h_y W_y\Big) \\ &= \frac{1}{H}\mathbb{E}\Bigg[\Big(\sum_{y=1}^{N_z}h_y(W_y - p(y|\boldsymbol{\theta}_n + e_d\frac{\pi}{2}))\Big)^2\Bigg] \\ &\overset{①}{\leq} \frac{1}{H}\Big(\sum_{y=1}^{N_z}h_y^2\Big)\sum_{y=1}^{N_z}\text{var}(W_y) \\ &\overset{②}{=} \frac{1}{H}\Big(\sum_{y=1}^{N_z}h_y^2\Big)\sum_{y=1}^{N_z}v\Big(p(y|\boldsymbol{\theta}_n + e_d\frac{\pi}{2})\Big) \\ &\leq \frac{N_z}{N_y}\Big(\sum_{y=1}^{N_z}h_y^2\Big)v = \frac{N_z\,\text{Tr}(Z^2)}{H}v, \end{aligned} \tag{73}$$

其中，① 由柯西-施瓦茨不等式得出，②是因为伯努利随机变量 W_y 的方差计算如下:

$$\text{var}\Big(W_y\Big) = \mathbb{E}\Big[W_y^2\Big] - \Big(\mathbb{E}\Big[W_y\Big]\Big)^2 = v\Big(p(y|\boldsymbol{\theta}_n + e_d\frac{\pi}{2})\Big), \tag{74}$$

其中，$v(x) = x(1-x)$，对于 $x \in (0, 1)$。最后一个结果由量 v的定义得出。类似地，可以证明以下不等式成立:

$$\mathbb{E}[X_{n,d,-}^2] \leq \frac{N_z\,\text{Tr}(Z^2)}{H}v. \tag{75}$$

由式（73）和式（75），将省略的指标补回后，可写为

$$\text{var}(\xi_{n,k}^t) \leq \frac{\nu N_z DTr(Z^2)}{2H}. \tag{76}$$

对于 N 个 QFL 客户端，我们得到

$$\text{var}(\xi_k^t) \leq \frac{1}{N}\sum_{n\in\mathcal{N}}\frac{\nu N_z DTr(Z^2)}{2H}, \tag{77}$$

从而完成证明。

E. 定理 *1* 的证明

利用引理 1 和引理 2，我们进一步对主论文中的 (34) 式作上界估计如下:

$$\begin{aligned}&\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\mathbb{E}[f(\bar{\boldsymbol{\theta}}_k^{t+1}) - f(\bar{\boldsymbol{\theta}}_k^t)] \\ &\leq \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\Bigg(-\eta_k\mathbb{E}\Big[\langle\nabla f(\bar{\boldsymbol{\theta}}_k^t), \tilde{g}_k^t\rangle\Big]\Bigg) \\ &\qquad + \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\frac{\eta_k^2 L}{2}\mathbb{E}\Big[||\tilde{g}_k^t||^2\Big] \\ &\leq \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\Bigg(-\frac{\eta_k}{2}||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||^2 \\ &\quad - \frac{\eta_k}{2}||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Bigg) + \frac{\lambda\eta_k L^2}{2K\frac{1}{N}\sum_{n=1}^{N}T_n}\frac{N+1}{N} \\ &\quad \times\Bigg(\Big[2C_1 + T_n(T_n+1)\Big]\eta_k^2\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}||^2 \\ &\quad - \frac{\eta_k}{2}||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Bigg) \\ &\quad + \frac{\eta_k L^2}{2K\frac{1}{N}\sum_{n=1}^{N}T_n}\Big(\frac{N+1}{N}\Big)\Big(\frac{KT_n(T_n+1)\eta_k^2\sigma^2}{B}\Big) \\ &\quad + \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\frac{L\eta_k^2}{2}\Bigg(\lambda\Big(\frac{C_1}{N}+1\Big) \\ &\quad \times\Big[||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Big] + \frac{\sigma^2}{NB}\Bigg) \\ &= \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\Bigg(-\frac{\eta_k}{2}||\nabla f(\bar{\boldsymbol{\theta}}_k^t)||^2 \\ &\quad - \frac{\eta_k}{2}||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Bigg) + \frac{\lambda\eta_k L^2}{2K\frac{1}{N}\sum_{n=1}^{N}T_n}\frac{N+1}{N} \\ &\quad \times\Bigg(\lambda\Big[2C_1 + T_n(T_n+1)\Big]\eta_k^2\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}||^2 \\ &\quad - \frac{\eta_k}{2}||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Bigg) + \frac{KT_n(L+1)\eta_k^2\sigma^2}{B} \\ &\quad + \frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\frac{\lambda L\eta_k^2}{2}\lambda\Big(\frac{C_1}{N}+1\Big) \\ &\Big[||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^t)||^2\Big] + \frac{L\eta_k^2}{2}\frac{\sigma^2}{NB}. \end{aligned} \tag{78}$$

由 (78)，我们有

$$\begin{aligned}
&\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\mathbb{E}[f(\bar{\boldsymbol{\theta}}_k^{t+1})-f(\bar{\boldsymbol{\theta}}_k^{t})]\\
&\quad\le -\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\frac{\eta_k}{2}||\nabla f(\bar{\boldsymbol{\theta}}_k^{t})||^2\\
&\quad+\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\Bigg[-\frac{\eta_k}{2}\\
&\quad+\frac{\lambda(N+1)L^2\eta_k^3[2C_1+T_n(T_n+1)]}{2N}\\
&\quad+\frac{\lambda L\eta_k^2}{2}\left(\frac{C_1}{N}+1\right)\Bigg]\left[||\sum_{n=1}^{N}\nabla f_n(\boldsymbol{\theta}_{n,k}^{t})||^2\right]\\
&\quad+\frac{\eta_k^3L^2(T_n+1)\sigma^2}{B}\left(\frac{N+1}{N}\right)+\frac{L\eta_k^2}{2}\frac{\sigma^2}{NB}\\
&\quad\overset{①}{\le} -\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\frac{\eta_k}{2}||\nabla f(\bar{\boldsymbol{\theta}}_k^{t})||^2\\
&\quad+\frac{\eta_k^3L^2(T_n+1)\sigma^2}{B}\left(\frac{N+1}{N}\right)+\frac{L\eta_k^2}{2}\frac{\sigma^2}{NB},
\end{aligned}\tag{79}$$

其中，① 在满足以下条件时成立：

$$\begin{aligned}
&-\frac{\eta_k}{2}+\frac{\lambda(N+1)L^2\eta_k^3[2C_1+T_n(T_n+1)]}{2N}\\
&\qquad+\frac{\lambda L\eta_k^2}{2}\left(\frac{C_1}{N}+1\right)\le 0.
\end{aligned}\tag{80}$$

在任何类型的联邦学习框架中，将局部梯度之和的系数设为零，有助于控制不同客户端更新所带来的方差，从而确保稳定收敛。此条件限制了各个客户端对全局模型的影响，防止振荡或发散。它使更新保持有界，促进模型可靠地收敛至最优解。重新整理 (79)，可得

$$\begin{aligned}
&\frac{1}{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{k=1}^{K}\sum_{t=1}^{T_n}\mathbb{E}||\nabla f(\bar{\boldsymbol{\theta}}_k^{t})||^2\le\frac{2[f(\bar{\boldsymbol{\theta}}_1^{0})-f^*]}{\eta_kK\frac{1}{N}\sum_{n=1}^{N}T_n}\\
&+\frac{L\eta\sigma^2}{NB}+\frac{2\eta_k^2\sigma^2L^2(\frac{1}{N}\sum_{n=1}^{N}T_n+1)}{B}\left(1+\frac{1}{N}\right).
\end{aligned}\tag{81}$$

到目前为止，我们尚未考虑量子客户端局部梯度中的噪声项。然而，我们必须考虑这一项，因为每个量子设备的局部梯度中的噪声项都会影响全局梯度的整体收敛。由于量子联邦学习中的全局梯度是所有设备的局部梯度的聚合，局部梯度估计中的任何噪声或误差也会在全局层面累积。因此，我们求出量子采样噪声导致的每个客户端局部梯度误差方差的上界，并将其加到 (81) 中全局梯度上界上，从而得到正文中的 (39)。

在量子联邦学习中，梯度是通过估计而非显式计算得出的。这种方法利用量子计算近似梯度，从而无需依赖精确的梯度计算即可高效地进行优化。我们的假设 $\mathbb{E}[\tilde{g_n}]=g_n$ 意味着该估计是无偏的。因此，我们可以写作

$$\tilde{g}_{n,k}^{t}=g_{n,k}^{t}+\xi_{n,k}^{t},\tag{82}$$

其中 $\tilde{g}_{n,\,k}^{t}$ 是梯度的随机估计，$g_{n,\,k}^{t}$ 是真实梯度，$\xi_{n,\,k}^{t}$ 是估计过程中引入的误差或噪声，且噪声项满足条件 $\mathbb{E}[\xi_{n,k}^{t}]=0$，并且 $\mathrm{var}(\xi_{n,\,k}^{t})=\mathbb{E}[||\tilde{g}_{n,k}^{t}-g_{n,k}^{t}||]$对所有设备取平均后，我们得到

$$\tilde{g}_k^{t}=g_k^{t}+\xi_k^{t},\tag{83}$$

其中 $\xi_k^{t}=\sum_{n=0}^{N}\xi_{n,k}^{t}$。由于量子联邦学习中的全局梯度是来自所有设备的局部梯度的聚合，局部梯度估计中的任何噪声或错误也会在全局层面累积。如果这些错误显著，它们可能导致聚合后的全局梯度偏离真实下降方向，从而减慢收敛速度或导致次优解。因此，我们使用引理4来求梯度估计方差的上界。因此，(81)中全局梯度平方范数的期望值还将以如下方式额外受到引理4左侧的界定：

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\frac{1}{N}\sum_{n=1}^{N}T_n}\sum_{t=1}^{T_n}\mathbb{E}||\nabla f(\bar{\boldsymbol{\theta}}_k^{t})||^2\le\frac{2[f(\bar{\boldsymbol{\theta}}_1^{0})-f^*]}{\eta_kK\frac{1}{N}\sum_{n=1}^{N}T_n}\\
&+\frac{L\eta\sigma^2}{NB}+\frac{2\eta_k^2\sigma^2L^2(\frac{1}{N}\sum_{n=1}^{N}T_n+1)}{B}\left(\frac{N+1}{N}\right)\\
&+\frac{1}{N}\sum_{n\in\mathcal{N}}\frac{\nu N_zDTr(Z^2)}{2H}.
\end{aligned}\tag{84}$$

在非凸优化中，由于景观的复杂性——充满局部最小值和鞍点——获得全局最小值往往不可行。与其专注于界定连续点之间的距离，另一种方法是界定梯度估计的平方范数。该方法有助于衡量我们距离驻点的接近程度；在驻点处，梯度幅值最小，表明变化极小。通过对平方梯度给出上界，我们可以评估向某一解的收敛情况：该解或许并非全局最优，但在降低损失方面实际上是有效的。